Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.18.(1) 2026,pp.Comp 1–23

Available online at www.qu.edu.iq/journalcm
JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS
ISSN:2521-3504(online) ISSN:2074-0204(print)

A Comprehensive Survey on using Segmentation and Density Peaks Clustering (DPC) for Healthcare Data Streams

Ghadeer Qasim Al-jaberi[a,*], Ahmed Al-Shammari[a]

[a] *Department of Computer Science, Faculty of Computer Science and Information Technology, University of Al-Qadisiyah, Al Diwaniyah, Iraq.* Email: *master.student244@qu.edu.iq*, ahmed.alshammari@qu.edu.iq

ARTICLE INFO

Article history:
Received: 26 /11/2025
Rrevised form: 20 /01/2026
Accepted : 02 /02/2026
Available online: 30/03/2026

Keywords: Deep Learning, Machine Learning, Segmentation, Data Clustering, Density Peaks Clustering, Healthcare Data Stream.

ABSTRACT

Healthcare systems have recently undergone a significant digital transformation, driven by the rapid growth of the Internet of Medical Things (IoMT) and smart sensing technologies. The sensing technologies generate continuous, high-speed stream of medical information that require real-time analysis and processing. Healthcare data streams are evolving over time. Recently, medical data segmentation and clustering are considered one of the most important techniques used to enhance IoMT reliability, scalability and to support the online medical decisions. Technically, data segmentation technique is used to divide the medical data into clinically meaningful parts to support machine learning (ML) and deep learning (DL) techniques in providing treatment planning, accurate diagnosis and monitoring. Moreover, the Density Peak Clustering (DPC) is an unsupervised clustering algorithm which is used to facilitate monitoring and near-real-time decision-making, this technique to cluster critical data streams, identify anomalies, and extract patterns. Furthermore, these techniques employ bandwidth optimization by reducing the overhead and transmission delay. To date, several surveys have also been proposed in the literature. However, current challenges such as real-time processing and dynamic maintaining of wide variety of medical data streams, which raise the question of developing intelligent and adaptive analytical systems for use in the medical field. Therefore, we conduct a comprehensive survey on the recent advancements in the segmentation and clustering methods for healthcare data streams. This survey examines healthcare data streams, employing clustering and segmentation techniques to improve diagnostic accuracy and enable early disease prediction.

https://doi.org/10.29304/jqcsm.2026.18.12460

1. Introduction

Recently, healthcare systems have shown a great digital transformation due to the very fast adoption and use of connected medical devices, smart sensors in the Internet of Medical Things (IoMT) [1][2]. This digital transformation has brought about continuous collection of large amounts of real time physiological and clinical data. These data streams are characterized by their wide variety which includes but is not limited to ECGs, EEGs, medical imaging, Electronic Health Records (EHRs) and data generated by mobile health (mHealth) applications [3]. The growth and spread of these data streams present many opportunities for improvement in patient care and bring with them great challenges which should be addressed and managed. Unlike static data sets which are usually easy to handle, these constantly generated data streams are defined by their large volume, diversity and dynamics which in turn complicate real time processing and in-depth analysis [4][5][6]. Healthcare data streams are dynamic,

*Corresponding author: Ghadeer Qasim Al-jaberi

Email addresses: master.student244@qu.edu.iq

Communicated by 'sub etitor'

continuous, huge number of real-time streams of health information making traditional analytical methods inadequate to ensure accurate and reliable data processing. Over the past several years, significant advancements have been made in the extraction of healthcare data streams. This analytical field objective to extract scientific insights from continuously medical data streams. Its goals include improving the quality of healthcare and supporting clinicians in making real-time decisions [7, 8]. Segmentation is a crucial element in analyzing medical data streams.

Segmentation continuous data into relevant time intervals for analysis reduces the computational cost of real-time processing while maintaining essential biological patterns. These technologies enable medical systems and professionals to analyze patient data in a systematic and manageable way to detect even the subtlest variations in electrocardiograms (ECGs) or electroencephalograms (EEGs). Following segmentation, clustering of medical data-particularly density-based algorithms, is crucial for real-time anomaly detection, the identification of significant trends, and the classification of instances. Density Peaks Clustering (DPC) is an effective method for identifying high-density regions within healthcare data streams that may indicate patient issues or high-risk situations. It is also capable of managing chaotic data and does not require prior knowledge of the number of clusters (k). Segmentation and density-based clustering are novel and high-tech strategies for healthcare data stream management and analysis. This method helps uncover clinically significant trends and assist real-time decision-making while scaling and adapting to the ever-changing and complicated medical data environment. Survey papers don't include medical data clustering and segmentation-based clustering algorithms. We overcome these inadequacies by providing a detailed state-of-the-art evaluation. This survey aims to identify the latest research gaps of segmentation and density peak clustering approaches in healthcare data streams. The increasing complexity of medical data necessitates more efficient, flexible, and scalable data analysis approaches.

This paper has the following sections. Section 2 discusses medical data streams management. Section 3 provides the conceptual background and introduces clustering fundamentals. Section 4 presents the datasets used for heart disease segmentation and clustering. Section 5 discusses the main challenges in this survey. Section 6 presents future research trends. Finally, Section 7 concludes the survey.

2. Medical Data Streams Management

The management techniques of medical data streams are the main topics of this section. Effective management allows real-time healthcare data acquisition and integration. Segmentation helps prepare data for clustering and uncover meaningful patterns. Combining these methods allows for competent medical data stream analysis and decision-making.

2.1. Medical Data Stream Management: Overview and Challenges

This section discusses the technical details of proposed medical data stream management. Medical data streams are being produced at an unprecedented rate due to the growth of the Internet of Medical Things (IoMT) and rapid advancements in healthcare technologies [52]. Wearable sensors, hospital information systems, and medical imaging devices give a wealth of data for diagnosis, patient monitoring, and decision-making. The vast number, lightning-fast speed, and persistent volatility of this data make standard storage, analysis, and assessment techniques insufficient. Classifying streams efficiently requires algorithms that use minimal memory and time [1, 2], [9].

Data stream mining provides new challenges compared to static, publicly available data mining. Memory, cognition, response speed, and idea genealogy are impaired [10]. All of these issues demonstrate the need for flexible and scalable healthcare data stream solutions. Continuous clustering offers scalability, flexibility, and analytics. This method may capture critical data in real time for issue identification, patient behavior analysis, and illness trend detection without maintaining prior data [6]. Continuous data stream clustering allows for variable cluster construction and depends less on prior data, making it more efficient and dependable [11]. Modern methods like DPC-AHS simplify cluster centroids [12]. A hybrid technique is more efficient and versatile for real-time monitoring complicated medical data streams. Despite these drawbacks, segmentation algorithms have handled continuous medical data [54].

Online data analysis uses density peak clustering (DPC) because of its versatility with high dimensionality, heterogeneous distributions, non-uniform cluster formation, and real-time methodologies. SWEM revolutionized the game and enabled parallel processing by circumventing memory limits [13]. Amini et al. (2014) [4] study micro-clustering and network clustering, two density-based clustering approaches, to find unusual clusters and decrease streaming noise. DBSTREAM maintains cluster linkages with a common density graph to enhance clustering. This reduces micro-cluster size and memory utilization [14]. More study has focused on evolutionary modifications and real-time cluster upgrades. Gong et al. (2017) [15] presented the EDM stream model, which tracks dynamic clusters effectively using density chains and DP trees. Xu et al. (2017) [16] introduced the Fat Node Leading Tree (FNLT) model to incorporate new data into Martingale theory and locate change locations. They showed that context-adaptive models function in changing situations using fast-variable-concept streams. Li et al. (2019) proposed a density-estimated medical correlation rule system to reduce duplication and increase efficiency [3]. This preprocessing and density estimation approach leverages histograms. Al-Shammari et al. (2019) [17] developed a dynamic clustering methodology using PAA+DBSCAN for initial clustering and Advanced Cluster Maintenance (ACM) for incremental updates to efficiently handle patient data

Yan et al. (2021) [18] reported CDSC-AL, a density-based clustering approach that actively learns from examples, with fewer labeled instances and less idea drift and class overlap. Using quantitative and continuous health data improved results. Chen et al. (2021) presented the single-pass WSWFFP-T2 sliding-window fuzzy recurrent pattern mining method [19]. The DWDP-stream technique [20] used density-peak adaptive clustering and online-offline micro-clustering maintenance to improve clustering. [2] They emphasized the need of scalable stream processing platforms such as Apache Kafka and Spark, along with practical performance and security considerations. The cluster results of this research indicate that direct data clustering algorithms for medical data streams are improving, enabling the real-time clustering of vast, dynamic healthcare data [53]. Table 1 discusses the comparison of the proposed medical data stream management methods.

Table 1. Proposed studies on Medical Data Stream Management.

Ref.	Author	Year	Dataset	Proposed Method	Findings	Limitation
[13]	Wee Keong Ng et al.	2009	Synthetic datasets of varying dimensions and distributions.: D10.K4.N100k, D2.K10.N100k, and D4.K5.N100k.	The proposed SWEM method uses (EM) with a time-sliding window and adaptive partitioning and merging.	SWEM showed competitive clustering performance on: D10.K4. N100k: −10.512 to −10.446 D2. K10.N100k: 19.215 to 19.276 D4.K5.N100k: −48.010 to −47.869	The SWEM algorithm efficiently processes data streams under memory and single-pass constraints, adapting to concept drift, though its performance may be limited on non-Gaussian data.
[4]	Hadi Saboohi et al.	2014	This study focuses on datasets from: network intrusion detection, high-dimensional data analysis, environmental monitoring, and medical systems.	comparison of density-based clustering algorithms (Den Stream, D-Stream, rDenStream, MR-Stream,	The study classifying data stream clustering algorithms by operating mode (Online/Offline) and outlier handling,	1. High memory and time consumption (such as rDenStream). 2. Difficulty in parameter selection (MR-Stream).

				and HDD Stream).	compares the performance metrics (Purity, SSQ, and NMI), and third, indicating that Den Stream as the primary reference algorithm.	
[14]	Michael Hahsler et al.	2016	Datasets used include cassini,Noisy Mixture of Gaussians, DS3, DS4, RBF Generator Sensor data, forest cover type data, and KDD CUP'99	DBSTREAM uses online micro-clusters with shared density graphs.	DBSTREAM achieved high clustering quality, with adjusted Rand Index of 1.00 (noisy Gaussian) and 0.88 (Cassini).	The main challenge is scalability as increasing dimensions raise the cost of maintaining the shared density graph.
[15]	Shufeng Gong et al .	2017	Datasets used including: KDDCUP99, Cover Type, and PAMAP2, in addition to a two-dimensional synthetic dataset (SDS) and a real news feed (NADS).	EDM Stream tracking density peak evolution using a Dependency Tree (DP-Tree) for efficient data stream processing.	EDM Stream outperforms competitors by 7–15 times and achieving industry-standard pool quality with 7–23 microsecond update time.	The main challenge is selecting the radius of the cluster cell (r), as reducing it improves clustering quality but increases the computational load, while increasing it speeds up processing but reduces accuracy of the results.
[16]	Ji xu et al	2017	Synthetic datasets: ChameleonDS3, Ex claStar, and MRDS. Real-world datasets: two from the UCI Machine Learning Repository.	The proposed method is DP-Stream based on the FNLT (Thick-Nose Lending Tree) architecture.	The model achieves accurate and efficient clustering, detects varied shapes and densities, monitors concept drift, and supports continuous clustering of streaming with linear time complexity.	Current clustering methods fight to detect clusters with arbitrary shapes and provide continuous results for newly incoming data. DP-Stream improves real-time clustering and complex cluster detection.
[5]	Umesh Kokate et al	2018	1. Real data on household electricity consumption data	Density-micro clustering-based	The D-Stream and DBSTREAM algorithms demonstrated	1.Time constraints: Current algorithms fight to meet real-time

			from UCI. 2. Synthetic data from Stream Generator to test the effects of concept drift and noise (SGC1–SGC4).	methods, such as D-Stream and DB STREAM.	superior performance in terms of accuracy and efficiency, High cluster purity and modified ARI, Faster per-point processing (D-stream).	processing demands. 2.Comprehensiveness: No single algorithm can achieve optimal performance across all evaluation metrics.
[3]	Xiaoffng Li et al	2019	The datasets used are ADNI, MIT-BIH, Obstetric data, and (MRA).	The proposed model extracts correlation rules from medical data stream using density estimation.	Data duplication to ≤2%, (RMSEA) <0.03, high stability and a processing time approximately 3.5 times faster than reference methods.	Traditional methods have high redundancy, RMSEA, and processing time. The Proposed density estimation-based technique addresses these issues.
[17]	Al-Shammari et al	2019	Real-world medical datasets include patient records with metrics: blood flow (x1), blood pressure (x2), and (ECG) (x3).	The proposed approach uses density-based clustering with dynamic maintenance: 1. Initial Clustering via (PAA) with the DBSCAN 2. Dynamic Maintenance via Advanced Cluster Maintenance (ACM).	Accuracy rate: 0.946, AUC: 0.994, Efficient clustering and dynamic tracking of medical data streams.	This paper addresses the limitations of previous approaches, predetermine the number of sets (k), inability to detect sets with arbitrary shapes, and inefficient recalculations for set maintenance.
[19]	Jing Chen et al	2021	The dataset used is synthetic data for medical data streams.	The proposed method is the WSWFFP-T2 algorithm, a weighted sliding window and	High recall and accuracy for (FFPs), reliable and stable for quantitative data streams.	The algorithm is performance evaluation relies on synthetic data, limiting generalizability to real and complex medical data.

				Type-2 fuzzy sets approach for accurate and efficient recurring pattern extraction in data streaming.		
[18]	Xuyang Yan et al	2021	The datasets are: Syn-1, Syn-2, Sea, Gas Sensor, MNIST, CIFAR-10, Poker, Elec, and KDD99.	The proposed method is CDSC-AL framework based on dense clustering, active querying, and sub cluster classification.	Outperformed six out of nine benchmark datasets, average accuracy: 86.68% (Sea), 96.84% (Syn-2), high efficiency in reducing classification costs.	Time and Memory Complexity: Despite the framework's high performance, it faces a computational efficiency challenge due to high time complexity during collection and classification $O(n^2m + knm)$
[20]	Chen et al	2022	The datasets are: 1- Synthetic and real-world: clustering, Flame, R15, 4k2-far, and D31, 2- Biomedical datasets: Haberman, Libras Movement, E. coli, and Iris.	The proposed method is DWDP-Stream, using dynamic weighting and two-step allocation for DPC.	4k2-far: accuracy 1 R15: 0.997 Clustering: 0.996 Iris accuracy: 0.980	The proposed method still requires user-configured parameters: the radius range, decay factor, and outlier threshold. Affecting micro clusters maintenance in the online phase.
[2]	Martin Kostov & kalinka Kaloyanova	2023	The dataset consists of 220,000 medical records in 494 XML files.	The proposed method uses Apache Kafka with Apache Spark for real-time data integration and stream processing.	Time –based batching : 13,872 records per second, average deployment time of 0.072 milliseconds and a 75-millisecond batching interval.	Kafka's throughput is limited by processing delays 75 milliseconds batching clustering. static data processing requires specialized infrastructure.

					Single-file batching: (4,136 records/second).	
[9]	Maniraj S P et al	2023	The dataset are: Healthcare Centre Dataset (unstructured) and Frequency Dataset(structured).	The proposed method is the Nutrition Prescription System (NPS) using Enhanced K-Means Clustering with Map Reduce.	High time efficiency in processing big data, immediate responses to patient inquiries, nutrition-based diagnoses, qualitative improvement in patient healthcare.	This study addresses limitations of current prescription systems focus on medications and lacking real-time responses for malnutrition treatment.
[12]	Ye tian et al	2022	The datasets include synthetic datasets from the Clustering Basic Benchmark, real-world data from the UCI Repository, and image datasets.	The proposed method is DPC-AHS algorithm, using a two-stage center identification strategy with AHS validity index and A-k-NN density estimation.	Outperformed seven contemporary methods, high performance and efficiency, stable and easy to implement.	This study addresses the reliance of earlier DPC algorithms on prior knowledge, expert assistance, and complex iterations, leading to unstable cluster center detection and reduced cluster quality.

2.2. Segmentation-Based Medical Data Clustering

This section discusses the technical specifics of a segmentation-based medical data clustering method [50]. Segmentation-based data clustering is essential for medical image analysis tasks such structure identification, anomaly detection, and clinical decision-making in early diagnosis, treatment planning, and sickness monitoring. In the last decade, segmentation methods have improved in accuracy, speed, and clinical relevance, from K-means, Fuzzy C-means, and DBSCAN to hybrid frameworks and advanced deep learning models. K-means and Fuzzy C-means algorithms were combined to improve brain tumor segmentation accuracy and speed up execution by Abdel-Maksoud et al. (2015) [21]. However, utilizing backpropagation-based feature classification, [22] integrated K-means with spatial Fuzzy C-means algorithms to improve segmentation accuracy and processing economy. K-means algorithms reduced false positives in breast cancer screening using classifier-based methods [23], and the Lattice Boltzmann method is a flexible and fast way to define tumor boundaries [24]. Combining clustering and classification greatly improves clinical utility. Meta-optimization approaches like the Sine Cosine algorithm improve segmentation quality and computing efficiency [25]. Deep learning improves medical segmentation. CNNs extracted and classified brain tumor properties better than standard clustering approaches. Researchers are investigating CNN-FCN-RNN-GAN-auto encoder hybrid frameworks for complicated and varied data. Ding et al. (2024) [28] presented S2VNet to effectively segment multi-class data via automatic and interactive slice-to-size propagation [48], [49].

Context-aware models like CCViM [29] and Med-SA [30] improve feature integration, domain adaption, and three-dimensional data processing. Better U-Net models with statistical feature extraction and refined DCNN classification using meta-heuristic methods [31], semi-supervised Cluster Fusion-based cross-teaching [32], unsupervised clustering with dimensionality reduction for cardiac risk assessment [33], and DBSCAN for coronary artery segmentation [34] demonstrate segmentation-based clustering's versatility. These findings show the change from classical clustering to hybrid and deep learning, which have various medical applications and enhance accuracy and computational efficiency [51]. Table 2 discusses the comparison of the proposed segmentation-based medical data clustering methods.

Table 2. Recent Proposed Studies on Segmentation-Based Medical Data Clustering.

Ref.	Author	Year	Dataset	Proposed Method	Findings	Limitations
[21]	Eman et al	2015	The datasets are: 1. DS1 (DICOM): 22 real MRI. 2. DS2 (Brain Web): 152 simulated images. 3. DS3 (BRATS): 81 multi-contrast MRI scans.	The proposed method is KIFCM, a hybrid approach combining K-means and Fuzzy C-means for brain tumor segmentation.	The result showed 100% accuracy on Brain Web and BRATS cohort and 90.5% on DICOM, with reduced iterations and computation time.	FCM is accurate but slow and sensitive to noise, while K-means (KM) is simpler and faster, but less accurate, preprocessing and automated thresholding improve efficiency and reduce human error.
[22]	Malathi & Sinthia	2018	The dataset used is BRATS brain tumor dataset.	The proposed method is: a hybrid methodology combining KISFCM with BPNN.	Execution time: (3.49–6.70 s) Number of iterations: (38–76) BPNN classification, efficiency: 93.28% (improved from 87%)	K-means: inaccurate for malignant tumors, and loses edge information. fuzzy C-means: Slow, unsuitable for noisy images. (KISFCM + BPNN): Improved classification to 93.28%.
[23]	Hassan et al	2021	1. Coimbra Breast Cancer Dataset 2. MIAS Database 3. Wisconsin	Proposed methods: K-means with (SOM) and genetic algorithms and cell nucleus	1-K-Means and (SVMs) or (SOMs): accuracy 95%, 2-Otsu Thresholding: accuracy	K-Means algorithm is sensitive to initial number of clusters and their centers, may produce empty clusters and requires large datasets for high

			Breast Cancer (UCI) Dataset 4. Ultrasound, MRI, and UCSB images 5. Break His Image Set.	segmentation with DWT and SVM.	98.83%, 3-optimized K-Means centers: enhanced early lesion detection sensitivity.	accuracy in Spherical K-Means. It can fragment noise in tumor images have long analysis times with large datasets, and occasionally failure to detect suspicious areas.
[24]	K.K.D. Ramesh et al	2021	Medical images, such as computed tomography (CT) and magnetic resonance imaging (MRI)	In this paper, they proposed the Lattice Boltzmann Method (LBM) an innovative approach based on statistical mechanics	Accuracy and quality 95% , high speed and modeling flexibility.	LBM faces theoretical limits in defining density in medical images and increasing network points raises computational load, limiting its practical efficiency.
[25]	Khrissi, L	2022	Berkeley Image Segmentation dataset: Lena, Baboon, House, 296059, 126007, and Pepper.	This paper proposes an improved clustering method using the sine and cosine (SCA) algorithm for image segmentation.	1-Lower RMSE 2-Higher PSNR 3-Higher separation coefficients 4-Improve cluster coherence and efficiency.	1. No algorithm performed best in all metrics. 2. Low SCA population diversity limits exploration. 3. Suboptimal XB, SC, and CE outcomes despite population cohesiveness. 4. (K-means and FCM) parameters increase randomness and local convergence.
[27]	Yan Xu et al	2024	breast dataset	CNNs combined with Fuzzy K-Means for brain tumors classification.	Accuracy: 98.64% Sensitivity: 100% Specificity: 99%	Deep learning-based approaches require large, annotated datasets and significant computing resources. Making hybrid models more interpretable and accessible in real-world clinical settings also presents a challenge
[28]	Ding et al	2024	Datasets: WORD, BTCV, and AMOS	S2VNet (Slice-to-Volume Network) using	15x faster inference speed ,48.2%	1- Structural: Task-specific design required separate

			(covering CT and MRI patterns).	clustering propagation.	lower memory consumption, Superior performance in automatic and interactive segmentation.	models and iterative training.2- 3D network: Slow inference and difficulty deploying on limited-resource devices 3- Interactivity: Most solutions handle only one class.
[29]	Yun Zhu et al	2025	Kumar, CPM17, ISIC 17, ISIC 18, and Synapse.	CCVIM model (U-shaped architecture combining Context Clustering Module (CC) and the Cross-Scan Module within the Vision Mamba (ViM) framework).	mIoU improvement: 1.17% DSC improvement: 0.71% Best performance in the PQ (Panoptic Quality) on CPM17 dataset.	Fixed scan orientations and contextual clustering layers (CC layers) settings limit performance on complex structures and small lesions, emphasizing the need for adaptive configurations.
[30]	Junde Wu et al	2025	BTCV ,REFUGE2, BraTS2021, TNMIX, and ISIC2019.	The proposed method is the Medical SAM Adapter (Med-SA) using SD-Trans and HyP-Adpt with PEFT.	Dice Score on BTCV: 89.8% Improvement over Swin-UNetr: 2.9% Improvement over original SAM: 34.8% Improvement over Med SAM: 9.4%	Med-SA faces difficulty with ambiguous prompts in overlapping anatomical structures and may require user intervention, needing adaptation for scribbles and text prompts.
[31]	Kusuma et al	2025	Using the BraTS 2020 dataset	The proposed enhanced U-Net, feature extraction and enhanced DCNN with the WSBWO	High accuracy in diagnosing and classifying brain tumors, overcoming long runtime and reliance on expert.	This research addressed previous challenges by using efficient feature extraction (I-GBP and MTH), WSBWO algorithm to improve DCNN training and a fully automated system (Improved U-Net + WSBWO-DCNN) to reduce manual intervention and enhancing diagnostic

						efficiency.
[32]	Zhang et al	2025	LA Segmentation, Left ventricle and Myocardium Segmentation, (KiTS-19).	Cluster Fusion Based Cross Teaching (CFCT) within a semi-supervised learning framework.	CFCT outperformed competing methods in segmenting medical images. Higher Dice Similarity Coefficient (DSC).	Cluster fusion may require high computing resources, and its performance depends on the number and configuration of student models, limiting its applicability in low-resource environments.
[33]	Kaverinskiy et al	2025	Norm dataset (14,863 ECG or HRV records from healthy individuals), Patients' dataset (8,220 ECG or HRV records from heart failure patients).	Unsupervised clustering using UMAP for dimensionality reduction and HDBSCAN for efficient clustering with across dataset strategy.	Norm group: two main clusters, one of which was associated with a reduced risk of heart failure. patient group: three main clusters were identified, one of which was identified as potentially high risk.	Clusters represent preliminary risk classifications for regular screening, with recommendations to integrate cluster analysis into clinical decision support and validate using larger and more varied datasets.

3. Conceptual Background and Clustering

Research on medical data stream analysis emphasizes how crucial it is to have a solid conceptual framework for pattern recognition and unsupervised analysis in order to comprehend intricate data patterns. This section discusses clustering fundamentals to find hidden patterns in huge, heterogeneous datasets. Our project will expand density peak clustering to manage medical data streams' scalability and temporal dynamics.

3.1 Data Clustering: Fundamental Concepts

Clustering is essential for unsupervised data. Sorts items into categories. Clustering medical data streams may reveal patterns in huge, complex data [35], [55]. Scalable clustering algorithms that enhance spatial and temporal efficiency are needed for enormous medical data streams [36]. The intellectual bias, speed, and volume of these currents are obvious. Figure 1 shows how several popular clustering algorithms build clusters [37]: Dividing methods: Following these stages, data is meticulously categorized into (k) non-overlapping groups. Students learn effective computing essentials like K-Means and K-Medoids. Clusters that aren't circular or irregular may be hard to show [38]. Hierarchical Methods: These methods create a tree-like network of clusters to analyze how data is connected.

Although beneficial for comprehending hierarchical connections, their memory and processing needs may be too high for streaming or big datasets [39]. This concept defines clusters as high-density locations close to one other and separated by low-density areas. DBSCAN and OPTICS can locate clusters with any structure and detect outliers, which is critical for real-time patient health monitoring medical data streams [40]. Recently, hybrid clustering frameworks, which integrate the best of various approaches, have been demonstrated to be ideal for organizing medical data streams. New, cutting-edge methods combine data segmentation with Density Peak Clustering (DPC) to increase scalability, accuracy, and flexibility in changing environments [41], [42]. Integration is the foundation of this architecture to increase clustering performance in real-time medical data streams.

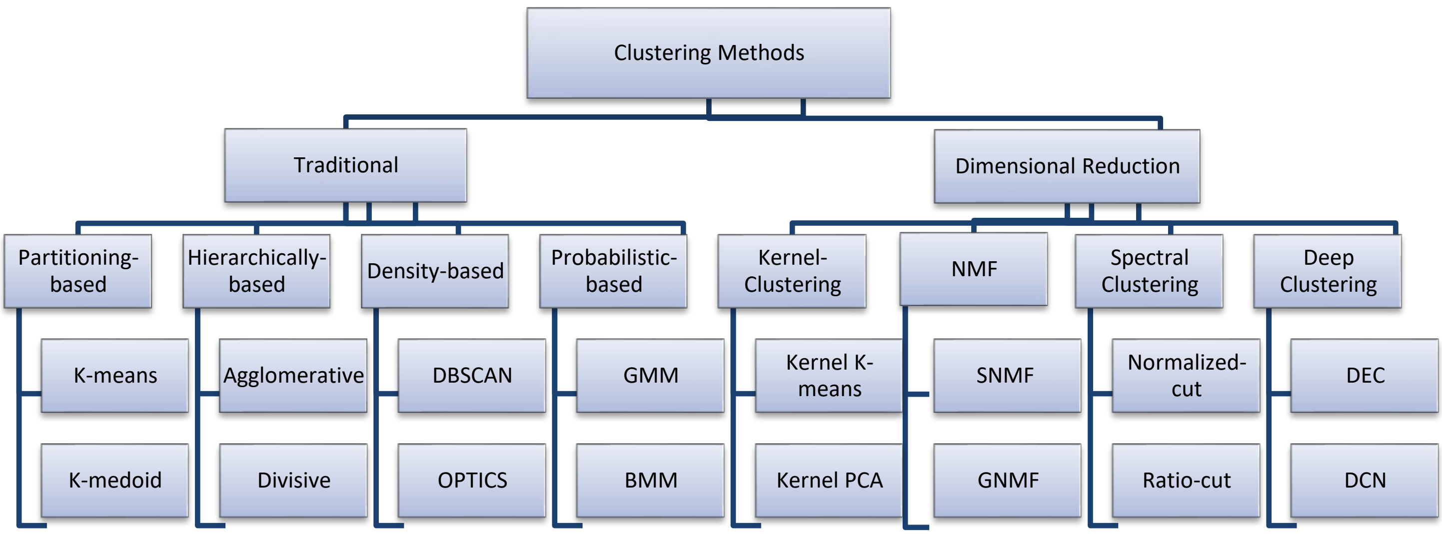

Figure 1. Clustering Algorithms Are Classified in to Several Primary Categories Based On Their Fundamental Methodology for Defining a Cluster [43].

3.2 Density Peak Clustering (DPC)

The Density Peaks Clustering (DPC) method [41] is a traditional density-based clustering approach that efficiently identifies cluster centers and segmentation data using local density (ρ_i) and minimum distance (δ_i) . Cluster centers are identified as data points with a high local density and a large distance from any point with higher-density [44]. In density-based clustering, DPC models clusters a high-density region in the feature space separated by low-density areas, which enables it to model complex (non-spherical) cluster structure.

1. Core Principles and Metrics

The DPC algorithm operates by calculating the local density (ρ_i) for each data point x_i , which is determined by the number of neighbors that fall within a cutoff distance (d_c) from x_i [44]. This(ρ_i) can be calculated using a kernel function (like the Gaussian kernel) or a simpler cutoff kernel (indicator function) [44], [45]:

where d(x_i,x_j) is the distance between points x_i and x_j, and $x(z)$is the indicator function ($x(z)$= 1 if $(z) < 0$, and 0 otherwise).

$$(\rho i) = \sum\nolimits_{j:xi\epsilon x} x\left(d\left(x_i x_j\right)\right) - d_c \;\; or \;\; (\rho i) = \sum\nolimits_{j:xi\epsilon x} exp\left(-\left(\frac{d(x_i\, x_j)}{d_c}\right)^2\right) \qquad (1)$$

The second metric, minimum distance(δ_i), measures the distance from point x_i to the nearest data point x_j that has a higher local density. Formally [46], [47]:

$$\delta_i = \min_{j:\rho_j > \rho_i} \left(d\left(x_i x_j\right)\right) \qquad (2)$$

For the point with the highest local density(ρ_{max}), its (δ_i) is set to the maximum distance from any other point in the dataset.

2. Selection of Cluster Centers and Assignment

The core mechanism for cluster center identification is the decision graph, which is a scatter plot of (ρ_i) $(y - axis)$ versus δi $(x - axis)$ for all data points. Points that are designated as cluster centers appear in the upper-right corner of this graph, characterized by both significantly large (ρ_i) and (δ_i) values.

Once the centers are manually or automatically selected from the decision graph, the remaining data points are assigned to the same cluster as their nearest neighbor with a higher local density. Points with low $(\rho_i)$and high (δ_i) that are not selected as centers are typically considered outliers.

3. Advantages

A key advantage of DPC is its ability to identify clusters with arbitrary shapes and is less sensitive to density variations compared to traditional methods. Furthermore, the algorithm is conceptually simple, non-iterative, and requires the tuning of only one parameter (d_c), making it appealing for use in diverse applications [41]. Table 3 discusses the comparison of the proposed DPC methods.

Table 3. Proposed Studies on Density Peak Clustering Methods.

Ref.	Author	Year	Dataset	Proposed Method	Findings	Limitation
[41]	Alex Rodriguez &Alessandro Laio	2014	synthetic point distributions, Olivetti Face Database, wheat seeds	Clustering by fast search and find of density peaks	Wheat Dataset: 97% Synthetic Data: < 1% error Olivetti Faces Dataset: 68%	The algorithm is highly sensitive to the cut-off parameter, relies on manual criteria (ρ, δ, γ) for cluster centers in sparse data, and small sample size negatively affects the estimation of local density and overall accuracy.
[58]	Zekang Bian et al	2020	Real, synthetic datasets: Iris, Segment, Waveform, Wine, Coil20, MNIST, MSRA, USPS, Palm and Medical dataset: diabetes, heart	FDPC: fuzzy extension of DPC, using fuzzy peaks and S-norm operators.	Iris: (ARI = 0.9038, NMI = 0.8846) and MNIST: (ARI = 0.5298, NMI = 0.6282), FDPC outperforms or competes with K-medoid, AP, DPC, PS_DPC, and FN_DPC.	Traditional DPC algorithms suffer from kernel density dependence, ambiguity points assignment, and difficulty determining cluster numbers, which FDPC addresses using fuzzy assignment and sudden drops detecting.

[59]	Yizhang WANG et al	2020	Synthetic datasets: Jain, Circuits (with noise), path-based semi-kernels. Real datasets: Pima, E.coli, Sonar, Iris, Seeds, Wine.	Mc-DPC: a hierarchical multi-core clustering algorithm based on DPC algorithm.	Synthetic datasets: Perfect performance (1.00) on FM, ARI, RI, NMI E.coli dataset: FM = 0.83, NMI = 0.71 Outperformed DPC,AP, DBSCAN.	Mc-DPC addresses DPC limitations detecting multiple or low-density centers but requires more parameters: ρ, δ, pct, and k.
[57]	Yi Lv et al	2020	Synthetic datasets, clustering datasets, and compound datasets.	DPC-SNNACC an improved DPC combining shared nearest neighbors and adaptive clustering.	F-Measure = 0.97843	The DPC-SNNACC: algorithm still needs to specify one of its parameters, the number of neighbors K, in advance, which requires either prior knowledge or an empirical procedure to set the optimal value, which is a methodological challenge.
[60]	Mohamed Abbasa et al	2021	Benchmark repositories: UCI, MNIST, and KEEL. Medical and biological datasets: Appendicitis, Arcene, Breast Cancer, E.coli, WDBC, and Yeast.	Den Mune: a density peaks-based clustering algorithm using the concept of mutual nearest neighbors (MNN).	Strong performance on low- and high-dimensional datasets, outperforming many state-of-the-art clustering algorithms.	Den Mune has slower propagation and a wider distribution of clustering seeds, clustering 50% of the DS7 data points after 1000 iterations, compared to 80% for CSharp, with time complexity is $O(N^2K)$.
[61]	Daichi Amagata & Takahiro Hara	2021	Synthetic datasets: S1, S2, S3, S4, Syn, and real datasets :Airline, Household, PAMAP2, and Sensor.	In this paper, three versions of the DPC algorithm are proposed: EX-DPC, Approx-DPC, and S-	Ex-DPC: up to 145.9 times faster than CFSFDP-A, Approximation algorithms: 10 times speedup over LSH-DDP.	Original DPC algorithm has $O(n^2)$ complexity, Ex-DPC suffers from parallelization difficulties due to the

				Approx-DPC.	Approx-DPC: Rand index up to 0.999 on syn data.	incremental update of the kd-tree, and LSH-DDP lack load balance, limiting efficiency.
[62]	Junyi Guan et al.	2021	Illustrative dataset: E1 Classic dataset: Jain.	In this paper, Fast Hierarchical Clustering of Local Density Peaks via an Association Degree Transfer Method (FHC-LDP) is proposed.	Outperforms traditional clustering algorithms and other DPC variants. Improved execution speed: O(nlogn) compared to the high complexity $O(n^2)$ of the original algorithm.	The research aims to address the fundamental limitation of the DPC algorithm, most notably non-adjacent links that affect the quality of clustering, and the high time complexity $O(n^2)$ with large-scale data.
[46]	Yizhang Wanga et al	2022	Synthetic datasets: Spiral, R15, and others. Real-world datasets (UCI): Iris, Sonar, Wine, and E. coli Medical/biological datasets(UCI): Thyroid and Breast.	PLDPC: Pseudo-Label-Guided Density Peak Clustering using co-occurrence, and mutual information optimization.	Vortex and Twenty datasets: ARI = 1.0, NMI = 1.0 DS5 dataset: ARI = 0.995, NMI = 0.9931.	PLDPC algorithm addresses the challenge of manual parameter assignment in clustering, which is a tedious and time-consuming process, especially for datasets without ground-truths.
[47]	Lifeng Yin et al.	2022	Synthetic datasets, and UCI real datasets: Iris, Wine, Seed, Voel, WDBC, and a medical dataset: Wisconsin Diagnostic Breast Cancer.	F-DPC algorithm: multi-density clustering using KNN partitioning, automatic cut-off and cluster center selection, followed by DPC and cluster merging.	Unbalance synthetic dataset: ARI ≈ 0.994, AMI ≈ 0.982. Iris dataset: ACC ≈ 0.898, FMI ≈ 0.824.	The time complexity of the algorithm is a limitation, as the execution time increases as the data size grows, although the rate of increase remains within average levels compared to other studied algorithms.

[42]	Limin Guo et al.	2024	Synthetic and real datasets, including classic benchmark datasets.	DPC-MS: hybrid clustering combines DPC and Mean-Shift algorithm to identify local maximum density points	High clustering accuracy, Outperformed many state-of-the-art algorithms Overcoming the limitations of the DPC algorithm in processing manifold data.	This study focuses on addressing the main challenges in the original DPC algorithm: missing clusters, sensitivity to truncation distance (d), and cascading misclassification from incorrect initial labels.
[63]	Shibo et al	2024	1. Synthetic datasets: Jain, clustering, Spiral, Path-based, Unbalance, R15, Longquare_1, D31, Noise_1, Noise_2. 2. Real-world datasets: Seeds, Iris, E.coli, Landsat, Zoo, Wine, Ionosphere, Balance Scale. 3. Medical datasets: WDBC.	The SM-DPC algorithm enhances clustering using KNN and Gaussian kernel for density, SNN for two-step mapping, and merging multiple clusters to reduce center selection bias.	Jain dataset: AMI = 1.0, ARI = 1.0, and FMI = 1.0.	The traditional DPC algorithm faces key challenges including poor performance on data with unequal densities, ignoring the spatial structure of sample points, and high sensitivity to errors in assigning non-central points, which reduces clustering accuracy.
[64]	Junyi Guan et al.	2024	Synthetic datasets: Agg, Jain, R15, Flame, S3. Real-world datasets: Iris, Wine, MNIST, and USPS. Medical and biological datasets: Parkinson's, E.coli	R-MDPC: method relies on assigning points within relevant regions, satellite peak attenuator, and kNN for large-scale data	Jain dataset: F1-score = 1.00,DGCI = 0.87.	The research addresses the limitations of the traditional DPC: poor point allocation, inefficient decision scheme from overlapping peaks, and high computational complexity $O(n^2)$, aiming to reduce it to $O(n\log n + nt + nk + nn_p)$.

[56]	Hongbo Wang et al	2025	2D industrial datasets: Clustering, Flame, R15, D31. Multidimensional datasets: Iris, WDBC, Ionosphere, Zoo, PASCAL VOC2007. Medical dataset: Breast Cancer Wisconsin Security datasets: NSL-KDD, IOT-23, CIC-IDS-2017.	The DPC-MDNN algorithm: improves the performance of DPC by using flexible spectral distance, nearest natural neighbors, and representative points.	ACC:13.22%, ARI:17.64%, AMI:12.58%, F1: 44.26%, High intrusion detection efficiency on both the NSL-KDD and IOT-23 clusters.	DPC-MDNN addresses the limitations of the original DPC related to density estimation and the domino effect, and reduces the time complexity from $O(N^2)$ to O(N log N) using a KD tree while maintaining high accuracy and operational efficiency.
[45]	Shihu Liu et al	2025	Synthetic and Real-world, and Medical data such as WDBC (Wisconsin Breast Cancer Diagnosis) and Dermatology.	The INSDPC algorithm relies on an interactive neighbor similarity and a two-step assignment strategy.	Five synthetic datasets: Flam, DIM512: ACC=1.00 WDBC dataset: ACC= 0.991, Iris dataset: ACC= 0.980, and the Dermatology medical dataset: ACC= of 0.959	The overall complexity is approximately $O(n^2)$, which is computationally challenging compared to other algorithms that may have lower time complexity (more efficient) when dealing with very large data sets.
[66]	Wei Xingqiong &Li Kang	2025	1. Synthetic datasets such as: clustering, complex, R15, four-line, smile, and circle. 2. Real-world datasets from the UCI repository such as: German, DIM1024, Landsat, and SPAM base. 3. Biomedical datasets such as: liver and thyroid.	DPC-DG: An adaptive peak density clustering method using a Delaunay graph, with automatic parameter tuning and iterative cluster merging via the VIASCKDE index	Smile and Circle datasets: ACC=1.	The algorithm is performance may decline in high-dimensional clusters due to Delaunay triangulation producing less accurate adjacency, leading to potential misclassification.
[68]	Chunhua Ren et al	2025	1.Synthetic Datasets: Jain, Flame, Path based, Clustering, Spiral, Compound,	The WMKNNDPC algorithm relies on weighted mutual	1. WDBC: ARI= 0.8122 2. Parkinson: ARI= 0.7042	The original DPC algorithm failed to identify clustering

			Smile, and D31. 2.Real Datasets: Olivetti Face, Seeds, Libras, WDBC, Parkinson, Glass, Wine, SCADI, E.coli, Dermatology, and Banknote.	neighbors with two-stage point assignment.	3. Dermatology: ARI= 0.9090	centers for low-density clusters, single assignment causes false clustering and fixed value for K, which is unable to account for the local distribution of the data.
[67]	Maixuan Peng et al	2025	Real (UCI) datasets: Sonar, Heart and others. Synthetic datasets: Path based and Jain.	This paper, proposed to perform density clustering in dynamically learned subspaces	Four synthetic datasets (Jain, Spiral, Moon, and CMC): ACC=1.0000 Average classification score on all datasets: 1.5	The WDSC algorithm addresses limitations in k-clustering of irregular clusters, the poor adaptability of density clustering and feature assessment in subspace models.
[65]	Yangming Liu et al	2025	Synthetic Datasets: Clustering, Spiral, R15, Path based, Jain, D31, Zelink1, S1, DS577, T4.8K. UCI Real Datasets: Wine, Ionosphere, Wireless, E. coli, Yeast, Balance, Sat image. Bioinformatics Datasets: Thyroid, Pima, and Breast.	CPDD-ID relies on detecting multidimensional density peaks, partitioning the data to form sub clusters, and then merging them using common nearest neighbor similarity.	Spiral, Jain, and Zelink1: ACC= 1.0 Clustering, R15, Path based, S1, and DS577: ACC= 0.99, D31: ACC= 0.96, T4.8K: ACC= 0.89	Provides solutions for the difficulty of dealing with data with irregular shapes and heterogeneous density. High overlap between clusters and fuzzy boundaries between them.

4.Datasets for Heart Disease Segmentation and Clustering

To evaluate the put forth segmentation-based density peak clustering framework in the healthcare data stream setting we assembled a large set of diverse datasets which include that of cardiovascular diseases. These datasets which range from many patient demographics, physiological measures, and clinical results were very purposeful in our collection. Also it is the diversity and the large scale of these sets which makes them very appropriate for the study of scalable and adaptive clustering methods as they apply to real time healthcare data.

Table 4. An Inventory of Heart Disease Datasets Used in Segmentation and Clustering of Medical data streams.

Datasets	Data Type	Medical Focus	No. of Records / Patients	Availability/ Source
MIT-BIH Arrhythmia Database	ECG signals (2 channels, 30-min segments)	Cardiac arrhythmia detection	48 records from 47 subjects	PhysioNet:https://physionet.org/content/mitdb/1.0.0/
PTB-XL ECG	12-lead ECG, ~10 seconds per record	Cardiac disease diagnosis (infarction, conduction delay, etc.)	21837 records from 18885 patients	Nature / PhysioNet: https://physionet.org/content/ptb-xl/1.0.3/
MIMIC-III Clinical Database	ICU clinical data (ECG, labs, notes, diagnostics)	Critical care and cardiac disease monitoring	40,000 patients	PhysioNet: https://physionet.org/content/mimiciii/1.4/
MIMIC-IV Extended Cardiac Disease	Clinical + ECG data for patients with cardiac diagnoses	Multiple cardiac diseases (20+ categories)	4,761 patient records with a primary diagnosis of one of 20 cardiac diseases	PhysioNet: https://physionet.org/content/mimic-iv-ext-cardiac disease/-1.0.0/
Cardiovascular Disease	Tabular (Static) with demographic and medical features	Heart disease classification	70,000 records	Kaggle:https://www.kaggle.com/datasets/sulianova/cardiovascular-disease-dataset
Heart Failure Prediction	Multivariate and clinical/medical data	Heart failure prediction	918 records with 12 attributes	Kaggle: https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction
UCI Heart Disease Data	Medical and clinical data set	Presence/absence of heart disease	920 records	Kaggle:https://www.kaggle.com/datasets/redwankarimsony/heart-disease-data

Table 4 reports the primary features of each used dataset which includes disease type, data format, and source. This info is critical for the evaluation of the put forth clustering framework. We see that the range of these data sets from very different sources and the volume in which they are presented allows us to test the framework’s performance with large scale, diverse health care data streams, do also see how well it does at performing effective segmentation, and in what role it plays in early diagnosis and personalized healthcare solutions.

5. Challenges

Medical data stream management research has shown several methodological and technical challenges. The healthcare professional requires sophisticated, adaptable analytical tools for real-time processing and diverse data. Existing data stream clustering algorithms include D-stream, Den-stream, and DB-stream suffer from the processing latency and memory utilization, which need precise and fast decision-making in real-time healthcare systems. Numerous medical data analysis methods emphasize static data distributions. The findings suggest that organizations must innovate; previous research have stressed the necessity for real-time medical data response systems. Earlier studies indicating K-Means and Fuzzy C-Means are not suitable for healthcare data stream management. Performance indicators reveal that current algorithms cannot manage the growing volume of medical data, highlighting the need for innovative techniques. The adaptability of hybrid approaches like DPC (density-based segmentation and clustering) shows they may work well on complex datasets. However, significant limits apply:

1. Several segmentation and clustering techniques are unsuitable for real-time healthcare applications due to their high memory requirements, lengthy processing times, and quadratic computational complexity.
2. These findings suggest that modifying parameters such as the cut-off radius, number of neighbors, decay factors, and segmentation thresholds may yield unforeseen outcomes and reduce the generalizability when applied to medical data.
3. Despite advances, some systems still struggle with noise, instability, intensity swings, inconsistent patterns, and idea drift in real-world medical data.

Advanced healthcare systems need increased computing capabilities, flexible, less participative frameworks, and frequent, accurate assessments. DPC-MDNN and DWDP-stream hybrid models improve medical data processing accuracy, efficiency, and conceptual soundness. This work streamlines healthcare analytics hybrid model construction by increasing cardiac data interpretation and usage. If the findings suggest the necessity for further investigation, especially regarding the integration of density-based clustering algorithms with sophisticated deep learning methods, intelligent medical systems could become increasingly adaptable and responsive. Therefore, increased assistance would be attainable.

6. Future Research Trends

The segmentation and density peaks clustering (DPC) have become advanced models for medical data streams because they can localize analysis in time through segmentation and discover non-spherical, evolving structures through density peaks clustering. However, the clinical realities of streaming data, heterogeneous sampling rates, resource restrictions, and severe privacy requirements are considered as open research gaps. There are several future research trends that can be explored, and further developed:

1- Hybrid segmentation models: Based on the combination of statistical and advanced models to produce segments that clinicians can interpret and validate.
2- Multi-rate and asynchronous segmentation.
3- Dynamic Density Peak Clustering (DDPC) for evolving clinical phenotypes. Although density peaks clustering (DPC) is a traditional clustering algorithm, it has not been developing to the data streams, because of the challenges of incremental density-updating and concept drift.
4- Multimodal and relational medical streams.

7. Conclusion

This comprehensive survey examines the most recent techniques for segmentation and clustering of medical data streams. The review highlights the pros and cons of previous studies. This study suggests that segmentation and density peak clustering (DPC) might enhance efficiency and processing time in healthcare companies. Increased detection accuracy. Adaptive models are needed to enhance medical data streams, especially for real-time vital sign and cardiology monitoring, for best patient care. Different data sources and formats make technique comparisons tougher, requiring more assessment labor. Segmentation and clustering increase real-time data processing. The suggested models were tested using big cardiac healthcare datasets including MIT-BIH, PTB-XL, and MIMIC-III. This work analyzes Density Peak Clustering (DPC) and its recent expansions, such as DPC-AHS, DPC-MDNN, and DWDP-Stream, and their theoretical basis. Future research should test the models in real healthcare settings to confirm the results. Integrated medical data stream design improves data management.

References

[1] Huang, C., Wang, J., Wang, S., & Zhang, Y. (2023). Internet of medical things: A systematic review. *Neurocomputing*, *557*, 126719.

[2] Kostov, M., & Kaloyanova, K. (2023). Real-time data integration in information systems using stream processing for medical data. *Annual of Sofia University St. Kliment Ohridski. Faculty of Mathematics and Informatics*, *110*, 101-110.

[3] Li, X., Wang, Y., & Li, D. (2019). Medical data stream distribution pattern association rule mining algorithm based on density estimation. *IEEE Access*, *7*, 141319-141329.

[4] Amini, A., Wah, T. Y., & Saboohi, H. (2014). On density-based data streams clustering algorithms: A survey. *Journal of Computer Science and Technology*, *29*(1), 116-141.

[5] Kokate, U., Deshpande, A., Mahalle, P., & Patil, P. (2018). Data stream clustering techniques, applications, and models: comparative analysis and discussion. *Big Data and Cognitive Computing*, *2*(4), 32.

[6] Ghesmoune, M., Lebbah, M., & Azzag, H. (2016). State-of-the-art on clustering data streams. *Big Data Analytics*, *1*(1), 13.

[7] L. Sun, J. Ma, Y. Zhang, and H. Wang. Exploring data mining techniques in medical data streams. In Australasian Database Conference, pages 321–332. Springer, 2016.

[8] S. J. Miah, E. Camilleri, and H. Q. Vu. Big data in healthcare research: a survey study. Journal of Computer Information Systems, 62(3):480–492, 2022.

[9] Saravanakumar, R., Maniraj, S. P., Barshan, A. D., Das, S., Hasan, H., & Alazzam, M. B. (2023, November). Clustering big data for novel health care system. In *AIP Conference Proceedings* (Vol. 2587, No. 1, p. 050006). AIP Publishing LLC.

[10] Nguyen, H. L., Woon, Y. K., & Ng, W. K. (2015). A survey on data stream clustering and classification. *Knowledge and information systems*, *45*(3), 535-569.

[11] Wang, Y., & Jing, B. (2023). A semi-supervised density peaks clustering algorithm. *Statistics and Its Interface*, *16*(3), 363-377.

[12] Zhang, R., Miao, Z., Tian, Y., & Wang, H. (2022). A novel density peaks clustering algorithm based on Hopkins statistic. *Expert Systems with Applications*, *201*, 116892.

[13] Dang, X. H., Lee, V., Ng, W. K., Ciptadi, A., & Ong, K. L. (2009, April). An EM-based algorithm for clustering data streams in sliding windows. In *International conference on database systems for advanced applications* (pp. 230-235). Berlin, Heidelberg: Springer Berlin Heidelberg.

[14] Hahsler, M., & Bolaños, M. (2016). Clustering data streams based on shared density between micro-clusters. *IEEE transactions on knowledge and data engineering*, *28*(6), 1449-1461.

[15] Gong, S., Zhang, Y., & Yu, G. (2017). Clustering stream data by exploring the evolution of density mountain. *Proceedings of the VLDB Endowment*, *11*(4), 393-405.

[16] Xu, J., Wang, G., Li, T., Deng, W., & Gou, G. (2017). Fat node leading tree for data stream clustering with density peaks. *Knowledge-Based Systems*, *120*, 99-117.

[17] Al-Shammari, A., Zhou, R., Naseriparsaa, M., & Liu, C. (2019). An effective density-based clustering and dynamic maintenance framework for evolving medical data streams. *International journal of medical informatics*, *126*, 176-186.

[18] Yan, X., Homaifar, A., Sarkar, M., Girma, A., & Tunstel, E. (2021). A clustering-based framework for classifying data streams. *arXiv preprint arXiv:2106.11823*.

[19] Chen, J., Li, P., Fang, W., Zhou, N., Yin, Y., Zheng, H., ... & Wang, R. (2021). Fuzzy Frequent Pattern Mining Algorithm Based on Weighted Sliding Window and Type-2 Fuzzy Sets over Medical Data Stream. *Wireless Communications and Mobile Computing*, *2021*(1), 6662254.

[20] Chen, D., Du, T., Zhou, J., Wu, Y., & Wang, X. (2022). DWDP-stream: a dynamic weight and density peaks clustering algorithm for data stream. *International Journal of Computational Intelligence Systems*, *15*(1), 96.

[21] Abdel-Maksoud, E., Elmogy, M., & Al-Awadi, R. (2015). Brain tumor segmentation based on a hybrid clustering technique. *Egyptian Informatics Journal*, *16*(1), 71-81.

[22] Malathi, M., & Sinthia, P. (2018). MRI brain tumour segmentation using hybrid clustering and classification by back propagation algorithm. *Asian Pacific Journal of Cancer Prevention: APJCP*, *19*(11), 3257.

[23] Hassan, N. S., Abdulazeez, A. M., Zeebaree, D. Q., & Hasan, D. A. (2021). Medical images breast cancer segmentation based on K-means clustering algorithm: a review. *Asian Journal of Research in Computer Science*, *9*(1), 23-38.

[24] Ramesh, K. K. D., Kumar, G. K., Swapna, K., Datta, D., & Rajest, S. S. (2021). A review of medical image segmentation algorithms. *EAI Endorsed Transactions on Pervasive Health & Technology*, *7*(27).

[25] Khrissi, L., El Akkad, N., Satori, H., & Satori, K. (2022). Clustering method and sine cosine algorithm for image segmentation. Evolutionary Intelligence, 15(1), 669-682.

[26] Mohammed, Y. M., El Garouani, S., & Jellouli, I. (2023). A survey of methods for brain tumor segmentation-based MRI images. Journal of Computational Design and Engineering, 10(1), 266-293.

[27] Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., & Liu, F. (2024). Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. Bioengineering, 11(10), 1034.

[28] Ding, Y., Li, L., Wang, W., & Yang, Y. (2024). Clustering propagation for universal medical image segmentation. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 3357-3369).

[29] Zhu, Y., Zhang, D., Lin, Y., Feng, Y., & Tang, J. (2025). Merging context clustering with visual state space models for medical image segmentation. *IEEE Transactions on Medical Imaging*.

[30] Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., ... & Jin, Y. (2025). Medical sam adapter: Adapting segment anything model for medical image segmentation. Medical image analysis, 102, 103547.

[31] Kusuma, P. V., & Reddy, S. C. M. (2025). MRI brain tumor segmentation and classification with improved U-Net model. Multimedia Tools and Applications, 84(4), 1671-1696.

[32] Zhang, H., Lu, X., Ma, P., Liu, J., Lian, J., & Ma, Y. (2025). Cluster fusion based cross teaching for semi-supervised medical image segmentation. Neurocomputing, 618, 129147.

[33] kaverinskiy, V., Chaikovsky, I., Mnevets, A., Ryzhenko, T., Bocharov, M., & Malakhov, K. (2025). Scalable Clustering of Complex ECG Health Data: Big Data Clustering Analysis with UMAP and HDBSCAN. Computation, 13(6), 144.

[34] Mardani, K., Maghooli, K., & Farokhi, F. (2025). Segmentation of coronary arteries from X-ray angiographic images using density based spatial clustering of applications with noise (DBSCAN). Biomedical Signal Processing and Control, 101, 107175.

[35] Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on neural networks*, *16*(3), 645-678.

[36] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, *46*(4), 1-37.

[37] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern recognition letters*, *31*(8), 651-666.

[38] MacQueen, J. (1967). Multivariate observations. In *Proceedings ofthe 5th Berkeley Symposium on Mathematical Statisticsand Probability* (Vol. 1, pp. 281-297).

[39] Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, *32*(3), 241-254.

[40] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd* (Vol. 96, No. 34, pp. 226-231).

[41] Rodriguez, A., & Laio, A. (2014). Clustering by fast search and find of density peaks. *science*, *344*(6191), 1492-1496.

[42] Guo, L., Qin, W., Cai, Z., & Su, X. (2024). Hybrid clustering algorithm based on improved density peak clustering. *Applied Sciences*, *14*(2), 715.

[43] Berahmand, K., Saberi-Movahed, F., Sheikhpour, R., Li, Y., & Jalili, M. (2025). A comprehensive survey on spectral clustering with graph structure learning. *arXiv preprint arXiv:2501.13597*.

[44] Xingqiong, W., & Kang, L. (2025). Adaptive density peak clustering based on Delaunay graph. *PLoS One*, *20*(6), e0325161.

[45] Liu, S., He, Y., Yang, X., & Yu, Z. (2025). INSDPC: A density peaks clustering algorithm based on interactive neighbors similarity. *AIMS MATHEMATICS*, *10*(4), 9748-9772.

[46] Wang, Y., Qian, J., Hassan, M., Zhang, X., Zhang, T., Yang, C., ... & Jia, F. (2024). Density peak clustering algorithms: A review on the decade 2014–2023. *Expert Systems with Applications*, *238*, 121860.

[47] Yin, L., Wang, Y., Chen, H., & Deng, W. (2022). An improved density peak clustering algorithm for multi-density data. *Sensors*, *22*(22), 8814.

[48] Huang , M., Zou, J., Zhang, Y., Bhatti, U. A., & Chen, J. (2024). Efficient click-based interactive segmentation for medical image with improved Plain-ViT. *IEEE Journal of Biomedical and Health Informatics*.

[49] Oskouei, A. G., Abdolmaleki, N., Bouyer, A., Arasteh, B., & Shirini, K. (2025). Efficient superpixel-based brain MRI segmentation using multi-scale morphological gradient reconstruction and quantum clustering. *Biomedical Signal Processing and Control*, *100*, 107063 .

[50] Mittal, H., Pandey, A. C., Saraswat, M., Kumar, S., Pal, R., & Modwel, G. (2022). A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*, *81*(24), 35001-35026 .Zhu, Y., Zhang, D., Lin, Y., Feng, Y., & Tang, J. (2025). Merging context clustering with visual state space models for medical image segmentation. *IEEE Transactions on Medical Imaging*.

[51] Huang, C., Wu, Z., Xi, H., & Zhu, J. (2025). kMaXU: Medical image segmentation U-Net with k-means Mask Transformer and contrastive cluster assignment. *Pattern Recognition*, *161*, 111274 .

[52] Baseer, K. K., Sivakumar, K., Veeraiah, D., Chhabra, G., Lakineni, P. K., Pasha, M. J., ... & Harikrishnan, G. (2024). Healthcare diagnostics with an adaptive deep learning model integrated with the Internet of medical Things (IoMT) for predicting heart disease. *Biomedical Signal Processing and Control*, *92*, 105988 .

[53] Aljohani, A. (2024). Optimizing patient stratification in healthcare: A comparative analysis of clustering algorithms for ehr data. *International Journal of Computational Intelligence Systems*, *17*(1), 173.

[54] Wu, Y., Chen, J., Hu, L., Xu, H., Liang, H., & Wu, J. (2025). OmniFuse: A general modality fusion framework for multi-modality learning on low-quality medical data. *Information Fusion*, *117*, 102890.

[55] Kowsar, I., Rabbani, S. B., Akhter, K. F. B., & Samad, M. D. (2023, December). Deep clustering of electronic health records tabular data for clinical interpretation. In *2023 IEEE International Conference on Telecommunications and Photonics (ICTP)* (pp. 01-05). IEEE.

[56] Wang, H., Zhang, J., Shen, Y., Wang, S., Deng, B., & Zhao, W. (2025). Improved density peak clustering with a flexible manifold distance and natural nearest neighbors for network intrusion detection. *Scientific Reports*, *15*(1), 8510.

[57] Lv, Y., Liu, M., & Xiang, Y. (2020). Fast searching density peak clustering algorithm based on shared nearest neighbor and adaptive clustering center. *Symmetry*, *12*(12), 2014.

[58] Bian, Z., Chung, F. L., & Wang, S. (2020). Fuzzy density peaks clustering. *IEEE Transactions on Fuzzy Systems*, *29*(7), 1725-1738.

[59] Wang, Y., Wang, D., Zhang, X., Pang, W., Miao, C., Tan, A. H., & Zhou, Y. (2020). McDPC: Multi-center density peak clustering. *Neural Computing and Applications*, *32*(17), 13465-13478.

[60] Abbas, M., El-Zoghabi, A., & Shoukry, A. (2021). DenMune: Density peak based clustering using mutual nearest neighbors. *Pattern Recognition*, *109*, 107589.

[61] Amagata, D., & Hara, T. (2021, June). Fast density-peaks clustering: multicore-based parallelization approach. In *Proceedings of the 2021 International Conference on Management of Data* (pp. 49-61).

[62] Guan, J., Li, S., He, X., Zhu, J., & Chen, J. (2021). Fast hierarchical clustering of local density peaks via an association degree transfer method. *Neurocomputing*, *455*, 401-418.

[63] Zhou, S., Peng, B., Xu, W., & Ren, L. (2024). SM-DPC: Clustering by fast search and find of density peaks based on SNN with multi-cluster fusion strategy. *IEEE Access*, *12*, 76413-76431.

[64] Guan, J., Li, S., Zhu, J., He, X., & Chen, J. (2024). Fast main density peak clustering within relevant regions via a robust decision graph. *Pattern Recognition*, *152*, 110458.

[65] Liu, Y., Ding, J., Wang, H., & Du, Y. (2025). A Clustering Algorithm Based on the Detection of Density Peaks and the Interaction Degree Between Clusters. *Applied Sciences*, *15*(7), 3612.

[66] Xingqiong, W., & Kang, L. (2025). Adaptive density peak clustering based on Delaunay graph. *PLoS One*, *20*(6), e0325161.

[67] Peng, M., Wu, Y., Lu, Y., Li, M., Zhang, Y., & Cheung, Y. M. (2025, April). Weighted density for the win: Accurate subspace density clustering. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE.

[68] Ren, C., Li, C., Yu, Y., Yang, W., & Guo, R. (2025). Density peak clustering algorithm based on weighted mutual K-nearest neighbors. *Frontiers in Applied Mathematics and Statistics*, *11*, 1598165.