

Clustering-Based Data Mining for Healthcare Decision Support Systems: A Systematic Review

Francka Sakti Lee¹, Catur Edi Widodo², Johaness Fernandes Andry³

¹Doctoral Program in Information Systems, School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia

²Department of Physics, Faculty of Mathematics and Sciences, Diponegoro University, Semarang, Indonesia

¹³Department of Information Systems, Faculty of Technology and Design, Bunda Mulia University, North Jakarta, Indonesia

Received:

June 30, 2026

Revised:

July 19, 2026

Accepted:

September 10, 2026

Published:

October 24, 2026

Corresponding Author:

Author Name*:

Johaness Fernandes Andry

Email*:

jandry@bundamulia.ac.id

DOI:

10.63158/journalisi.v8i5.1801

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)



Abstract. Clustering can organize heterogeneous healthcare data into clinically or operationally meaningful groups, but evidence of real decision-support use remains uneven. This systematic review examines healthcare domains, data types, clustering methods, validation practices, and implementation gaps in clustering-based Healthcare Decision Support Systems (DSS). Following PRISMA 2020, Scopus records published from 2015 to 2025 were screened using predefined eligibility criteria. Eighty-four records were identified; 36 remained after the open-access filter, 19 full-text reports were assessed, and 10 primary empirical studies were included. The studies covered fraud auditing, ICU profiling, chronic-disease risk and transition analysis, gestational-diabetes support, trauma pathways, radiology text, patient segmentation, and smart-home activity recognition. K-Means was the most frequently reported method, followed by Markov-based approaches, while other clustering techniques appeared less often. Most studies relied on retrospective, internal, or expert-based validation; only one reported prospective evaluation of a deployed web-based clinical DSS. Overall, clustering shows clear analytical value, but evidence for routine clinical deployment and patient-outcome improvement remains limited.

Keywords: Healthcare Decision Support Systems, Clustering, Unsupervised Learning, Systematic Review, PRISMA 2020

1. INTRODUCTION

Healthcare organizations generate heterogeneous data from electronic health records (EHRs), laboratory and imaging systems, physiological monitoring devices, administrative claims, and operational platforms. Healthcare Decision Support Systems (Healthcare DSS) convert these data into information that may assist clinical and managerial decisions. In this review, Healthcare DSS is used as the umbrella term [1], [2]. Clinical DSS refers to patient-oriented applications that assist diagnosis, risk assessment, treatment planning, and monitoring, whereas hospital DSS refers to operational applications such as auditing, resource allocation, workflow management, and service planning [3], [4]. A clustering-based Healthcare DSS is defined here as an application or prototype in which unsupervised grouping is a substantive analytical component and the resulting groups are linked to a clinical or operational decision process [5], [6].

Clustering differs from classification and prediction. Classification assigns observations to predefined classes, while prediction estimates an unknown or future outcome. Clustering instead identifies latent groups based on similarity, density, probability distributions, or graph structure without requiring predefined labels. This is useful in healthcare because labels may be costly to obtain, incomplete, inconsistent, or dependent on expert interpretation [7]–[9]. K-Means remains attractive because of its computational efficiency and transparent centroid representation, although it assumes compact clusters and requires a predefined number of clusters [10]. Density-based approaches such as DBSCAN can capture irregular cluster shapes and noise but are sensitive to density parameters [11]. Deep and graph-based methods can model more complex longitudinal or high-dimensional data, but they increase interpretability, computational, and governance demands [12].

Prior reviews have discussed machine learning in health informatics and data-driven clinical pathways broadly [13]. The remaining problem is that clustering used as an exploratory analytical technique is often discussed together with clustering embedded in an explicit decision-support workflow [14]. This makes it difficult to distinguish technical feasibility from evidence of clinical or operational implementation. The novelty of this review is therefore not the introduction of clustering itself, but a structured synthesis that separates algorithmic use, DSS function, validation maturity,

and implementation evidence across eligible primary empirical studies. The 2015–2025 window was selected to provide a recent eleven-year evidence horizon while retaining early empirical work on smart-home sensing and clustering that remained relevant to Healthcare DSS. The review addresses the following research questions:

- 1) RQ1: In which healthcare domains, data types, and decision-support functions has clustering-based data mining been applied?
- 2) RQ2: Which clustering methods are represented in the eligible studies, and how frequently are they reported under non-exclusive coding?
- 3) RQ3: What benefits, limitations, validation practices, and implementation gaps are reported for clustering-based Healthcare DSS?

2. METHODOLOGY

2.1. Review Design and Reporting Standard

This systematic review was designed as a narrative evidence synthesis and reported in accordance with the PRISMA 2020 statement. The review process covered the rationale and research questions, eligibility criteria, information source, search strategy, study selection, data extraction, quality appraisal, synthesis, and reporting of limitations [15]–[18]. A quantitative meta-analysis was not appropriate because the included studies differed substantially in healthcare domain, dataset structure, clustering method, decision-support function, and validation outcome. The organization of the review was also informed by recent PRISMA-based systematic-review practice, in which search, screening, synthesis, and research-question-based reporting are made explicit to support reproducibility. No protocol was prospectively registered; retrospective registration was not used because it would not represent a genuinely prospective protocol, and this is retained as a methodological limitation [19], [20].

2.2. Search Strategy and Information Source

Scopus was used as the sole bibliographic database because the review was intended as a focused synthesis across health informatics, medicine, computer science, engineering, and decision-support research. The search covered publications from 1 January 2015 through 31 December 2025 and was last updated on 20 April 2026. The open-access filter was applied at retrieval so that candidate reports could be inspected through a consistent full-text process; it was not treated as a quality criterion. This

decision improves access consistency but may introduce selection bias by excluding relevant subscription-only studies. The search and screening structure follows the transparent reporting principles emphasized by PRISMA 2020 and recent systematic-review examples [18].

TITLE-ABS-KEY ((healthcare OR clinical OR hospital) AND ("decision support system*" OR "clinical decision support system*" OR "healthcare decision support system*") AND ("data mining" OR "machine learning") AND (cluster* OR "unsupervised learning")) AND PUBYEAR > 2015 AND PUBYEAR < 2026 AND (LIMIT-TO (DOCTYPE, "ar")) AND (LIMIT-TO (LANGUAGE, "English"))

2.3. Eligibility Criteria and Study Selection

Studies were eligible when they were: (1) peer-reviewed primary empirical research articles in English published from 2015 to 2025; (2) based on healthcare, clinical, hospital, or health-administrative data; (3) used clustering or unsupervised grouping as a substantive analytical component; (4) described an identifiable clinical or operational decision-support use case; and (5) reported sufficient information on the data, method, and results. Reviews, conceptual overviews, book chapters, studies outside the publication window, non-healthcare studies, and papers in which clustering was incidental or not connected to a discernible DSS function were excluded. These eligibility requirements were translated into a structured set of inclusion and exclusion criteria to support a consistent and reproducible screening process [19]. As summarized in Table 1, each study was assessed according to publication type, publication period and language, healthcare domain, analytical method, decision-support relevance, and reporting completeness.

Table 1. Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Publication type	Peer-reviewed original research article	Review, editorial, book chapter, abstract, thesis, or duplicate
Period and language	2015–2025; English	Outside period or non-English
Domain	Healthcare, clinical, hospital, or health administration	Non-healthcare setting
Analytical method	Clustering/unsupervised grouping is central	Clustering absent or incidental

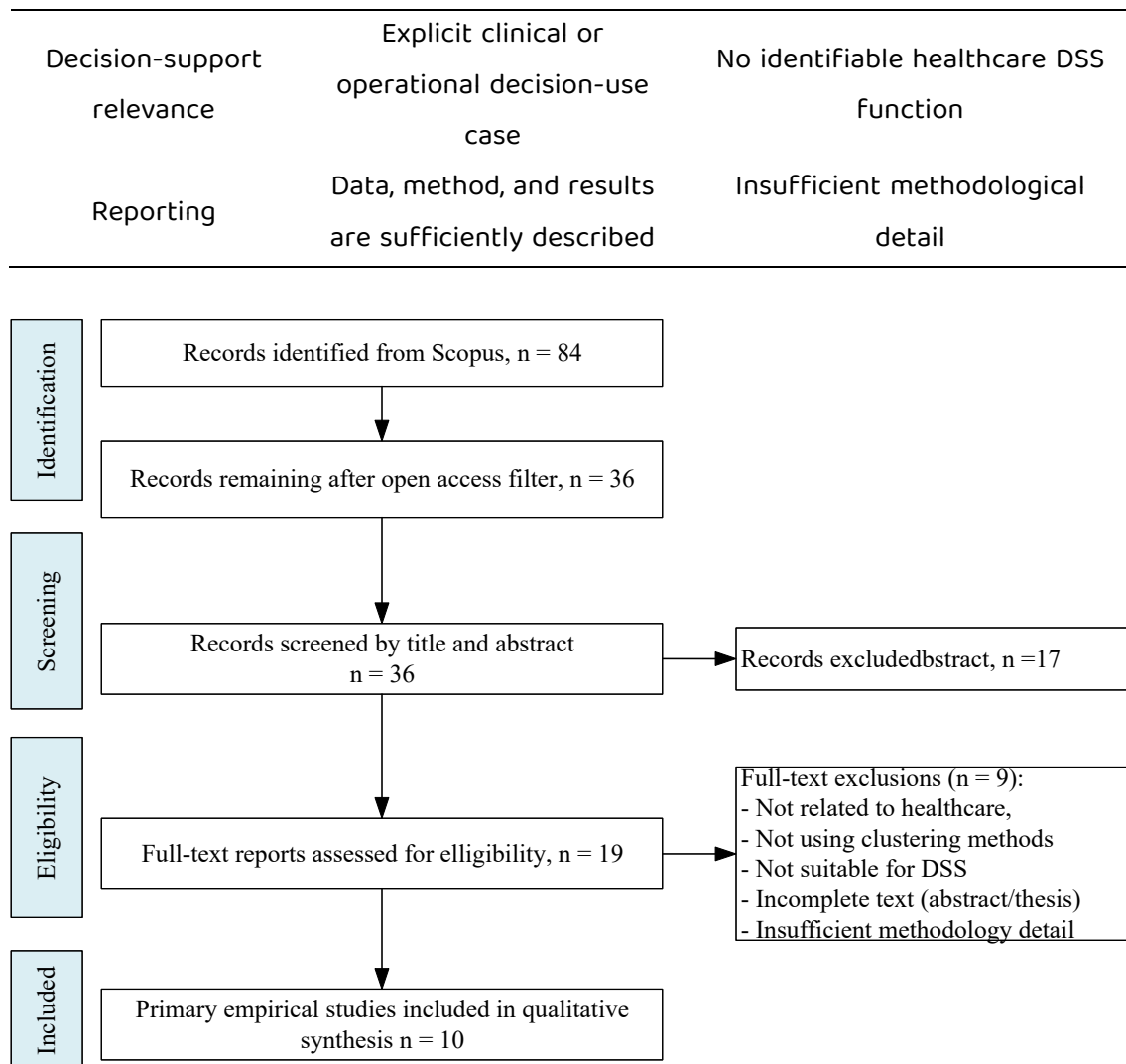


Figure 1. PRISMA Flowchart for the Selection of Healthcare DSS Clustering Studies

Figure 1 summarizes the complete study-selection pathway. During identification, the Scopus search returned 84 records. Applying the open-access filter reduced the set to 36 records that could be screened consistently. During screening, titles and abstracts of these 36 records were reviewed and 17 were excluded because they did not satisfy the review scope, leaving 19 reports for full-text eligibility assessment. At the eligibility stage, nine full-text reports were excluded: six were secondary, conceptual, or otherwise non-empirical; one fell outside the 2015-2025 publication window; one was a book chapter; and one did not provide a sufficiently explicit clustering-based DSS use case. The final included set therefore comprised 10 primary empirical studies for qualitative synthesis. This progression from identification, screening, eligibility, and inclusion is shown explicitly to make the selection process auditable and reproducible.

Screening and extraction were not conducted through a fully independent dual-review process; disagreements were resolved through author discussion, which may introduce selection and interpretation bias [21].

2.4. Data Extraction, Algorithm Coding and Quality Appraisal

For each eligible study, data were extracted on author/year, setting, healthcare domain, dataset type, clustering method, DSS function, reported benefit, limitation, validation type, and quality score. Algorithm frequencies were coded non-exclusively: a study contributed once to every named clustering category that formed part of its reported method. Consequently, the sum of algorithm occurrences can exceed the number of included studies. Reporting quality was appraised using five criteria, not nineteen: C1 domain/problem clarity; C2 data-source and sample transparency; C3 clustering-method reproducibility; C4 explicit connection between clustering results and a DSS function; and C5 validation or implementation evidence. Each criterion was scored 0 (not reported), 1 (partly reported), or 2 (clearly reported), for a maximum score of 10. Scores of 8–10 were classified as high quality, 5–7 as moderate quality, and 0–4 as low quality. This appraisal evaluates reporting completeness and decision-support relevance; it does not establish clinical efficacy.

3. RESULTS AND DISCUSSION

3.1. RQ1: Healthcare Domains, Data Types, and Decision-Support Functions

RQ1 examined where clustering-based data mining has been applied and how the resulting groups support healthcare decisions. The ten included studies covered a broad range of healthcare contexts, including administrative fraud auditing, intensive-care profiling, chronic-disease risk and transition analysis, gestational-diabetes management, traumatic-brain-injury pathway prediction, radiology-report organization, patient segmentation, and smart-home activity recognition for elderly safety. This diversity shows that clustering has been used not only to organize complex clinical data, but also to support operational decisions such as fraud screening, patient prioritization, pathway analysis, and monitoring. The underlying data were equally varied, ranging from administrative and insurance claims to longitudinal electronic health records, clinical monitoring data, registry records, free-text radiology reports, and ambient sensor streams. Despite this heterogeneity, a common feature across the

studies was that the clustering output was connected to a recognizable clinical or operational decision-support function rather than being reported solely as an exploratory grouping exercise. This distinction is important because it indicates that the practical value of clustering depends not only on how well observations are grouped, but also on whether those groups can be interpreted and linked to an actual healthcare decision. Table 2 summarizes the characteristics of the ten included studies, including their healthcare domain, data type, and clustering approach.

Table 2. Final included Primary Empirical Studies (n=10)

ID	Study	Healthcare Domain / DSS Use	Data Type	Clustering Method
S1	Massi et al. (2020) [22], Italy	Healthcare administration / fraud audit	Hospital discharge and administrative data	K-Means
S2	Ribeiro et al. (2017) [23], Portugal	Critical care / ICU profiling	INTCare monitoring data	K-Means
S3	Gao et al. (2018) [24], China	Healthcare administration / fraud detection	Large mobile- health insurance claims	Manifold-based activity clustering / outlier detection
S4	Niu et al. (2024) [25], China/UK	Chronic and critical disease / risk prediction	Longitudinal EHRs and medical notes	Deep predictive clustering
S5	Caballero-Ruiz et al. (2017) [26], Spain	Gestational diabetes / treatment support	Home glucose and clinical monitoring data	Expectation- Maximization clustering
S6	Eili et al. (2025) [27], Iran	Trauma / clinical pathway prediction	National trauma registry event logs	Markov-based trace clustering
S7	Hassanpour and Langlotz (2016) [28], USA	Radiology / report organization	1,899,482 Free- text radiology reports	K-Means with distributed MapReduce implementation

ID	Study	Healthcare Domain / DSS Use	Data Type	Clustering Method
S8	Alaeddini et al. (2017) [29], USA	Multiple chronic conditions / transition analysis	Longitudinal Veterans Affairs records	Latent regression Markov clustering
S9	Kaduskar et al. (2025) [30], India	Medical data / patient segmentation	Clinical/medical datasets	Hybrid Fuzzy C- Means and DBSCAN
S10	Nef et al. (2015) [31], Switzerland	Dementia care / activity recognition and elderly safety monitoring	Smart-home ambient data from PIR, temperature, luminance, and humidity sensors	Token-based and room-based clustering

The studies summarized in Table 2 also differed in how directly the resulting clusters were used within a decision-support process. In some cases, clustering formed part of an operational workflow, such as identifying potentially fraudulent records, profiling ICU admissions, prioritizing patients, or supporting treatment recommendations. In other studies, clustering served a more intermediate analytical role by organizing complex data into clinically meaningful groups that could later inform risk assessment, information retrieval, transition analysis, or patient segmentation. This variation suggests that clustering-based healthcare DSS should be viewed along a continuum, from exploratory analytical support to more directly actionable decision support. The strength of the evidence also varied considerably: some studies relied mainly on internal clustering metrics or retrospective evaluation, whereas others incorporated domain-expert assessment or prospective clinical evaluation. Table 3 therefore extends the descriptive characteristics in Table 2 by showing how each clustering approach was translated into a decision-support function, how it was validated, and what limitations remained.

Table 3. DSS Function, Validation, and Main Limitations of Included Studies

ID	DSS Function	Validation Type	Main Limitation
S1	Fraud-screening support for auditors	Internal clustering/robustness analysis on real administrative data	Proof-of-concept; no prospective audit-outcome evaluation
S2	ICU admission profiling and proactive alerts	Internal silhouette-based clustering evaluation	Limited attribute set; no external validation
S3	Fraud identification in mobile healthcare services	Internal model evaluation using real claims data	Dependence on expert interpretation and labeled verification
S4	Personalized disease-risk prediction	Internal clustering and predictive evaluation	High computational demand; no prospective clinical workflow test
S5	Diet/insulin-need recommendations and patient prioritization	Prospective clinical evaluation of deployed web-based CDSS	Single-center evaluation; generalizability requires external validation
S6	Prediction of clinical pathways and outcomes	Real-life registry data with internal predictive validation	Registry-entry dependence; prospective workflow impact not established
S7	Organization of radiology reports for information retrieval/decision support	Domain-expert validation of clusters	Cluster labeling and optimal K depend on expert review
S8	Identification of chronic-condition transition patterns	Large retrospective longitudinal cohort	Assumptions on transition structure; clinical deployment not evaluated
S9	Patient segmentation for precision-medicine support	Experimental clustering validation/case studies	Parameter sensitivity and limited prospective clinical evidence

ID	DSS Function	Validation Type	Main Limitation
S10	Activity-recognition	Experimental evaluation using smart-home	Performance was sensitive to overlapping activities
	support for elderly safety and well-being	ambient sensor data; specificity 96.53%, sensitivity 68.49%, precision 74.41%, F-measure 71.33%	and the presence of visitors; routine clinical deployment was not established

3.2. RQ1 Synthesis: Distribution by Domain, Data Type, and DSS Function

Administrative fraud/audit applications and chronic-disease or longitudinal-risk analysis each appeared in two studies. ICU profiling, gestational-diabetes support, trauma pathways, radiology text, patient segmentation, and dementia/activity recognition each appeared once. Administrative/discharge data, monitoring/clinical data, and longitudinal EHR/records were the most recurrent data types. The DSS functions ranged from fraud detection and risk support to ICU alerts, treatment recommendations, pathway prediction, document organization, patient segmentation, and safety monitoring. Table 4 summarizes these distributions without treating the categories as mutually exclusive.

Table 4. Distribution of Final Studies by Domain, Data Type, and DSS Function

Dimension	Distribution
Healthcare domain	Administration/fraud (2); chronic disease/risk (2); ICU (1); gestational diabetes (1); trauma (1); radiology (1); patient segmentation (1); dementia/activity recognition (1)
Data type	Administrative/discharge (2); monitoring/clinical (2); longitudinal EHR/records (2); trauma registry (1); clinical text (1); medical/clinical datasets (1); smart-home/IoT sensor data (1)
DSS function	Fraud detection/audit (2); risk/transition support (2); ICU alerts (1); treatment recommendation (1); pathway prediction (1); document organization (1); segmentation (1); activity-recognition/safety monitoring (1)

3.3. RQ2: Clustering Methods and Frequency Patterns

RQ2 examined which clustering methods were represented in the final corpus. K-Means was the most frequently reported method, appearing in three studies (S1, S2, and S7). Markov-based clustering appeared in two studies (S6 and S8). DBSCAN, Fuzzy C-Means, deep clustering, Expectation-Maximization, trace clustering, manifold/outlier clustering, and the token-/room-based method reported by Nef et al. (S10) each appeared in one study. Because methods were coded non-exclusively, a study could contribute to more than one category. These frequencies describe methodological prevalence only and should not be interpreted as evidence that one algorithm is clinically superior.

Table 5. Transparent Algorithm-Coding Matrix

Algorithm Category	Included Study IDs	Count
K-Means	S1, S2, S7	3
Markov-based clustering	S6, S8	2
DBSCAN	S9	1
Fuzzy C-Means	S9	1
Deep clustering	S4	1
Expectation-Maximization	S5	1
Trace clustering	S6	1
Manifold/outlier clustering	S3	1
Token-/room-based clustering	S10	1

Frequency is interpreted as descriptive prevalence only, not evidence that one algorithm is clinically superior. Table 5 provides the coding matrix used to generate Figure 2. The persistence of K-Means is understandable: it is computationally efficient, familiar to practitioners, easy to explain through centroids, and comparatively straightforward to integrate into broader workflows. These properties are useful when a DSS must support human interpretation rather than only maximize predictive performance. The trade-off is that K-Means assumes relatively compact clusters, depends on initialization and the selected value of K, and can be affected by scale and outliers [6]. Deep clustering offers richer representation learning for high-dimensional EHR and text data [11], but its computational cost, lower transparency, and more complex governance requirements may impede routine clinical deployment. Density-based and fuzzy approaches can better accommodate noise or overlapping

membership S9, but parameter selection and clinical interpretation remain important implementation issues.

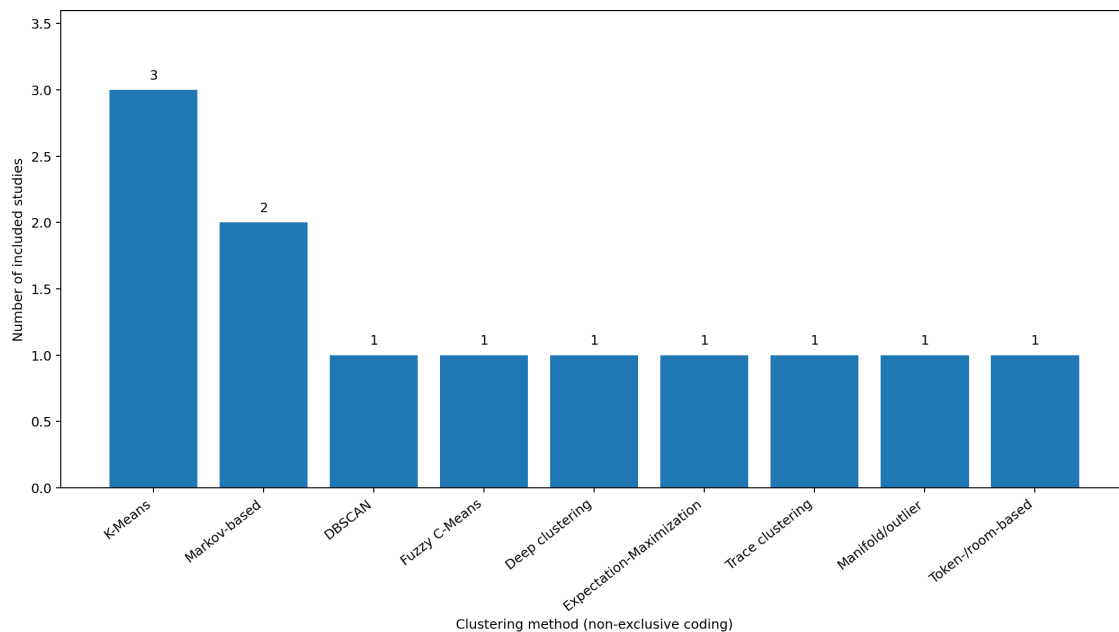


Figure 2. Non-Exclusive Frequency of Clustering Methods in the Final Included Studies

3.4. RQ3: Benefits, Limitations, Validation Practices, and Implementation Gaps

RQ3 focused on the maturity of validation and the gap between analytical clustering and implemented decision support. Seven studies were classified as high quality and three as moderate quality using the five-criterion appraisal; none was rated low. Most studies relied on internal clustering metrics, retrospective evaluation, predictive discrimination, or expert interpretation. One important exception was S5, in which clustering was embedded within a web-based gestational-diabetes clinical decision-support system that underwent prospective clinical evaluation. Even there, the reported benefits centered on safety, workload reduction, patient prioritization, and service utilization rather than definitive downstream patient-outcome improvement. Across the corpus, external multi-institutional validation, independent clinician validation, and prospective workflow evaluation remained limited.

Table 6. Qualitative Appraisal of the Final Included Studies

Study	C1	C2	C3	C4	C5	Total
S1 / Massi et al.	2	2	2	2	1	9 (High)

Study	C1	C2	C3	C4	C5	Total
S2 / Ribeiro et al.	2	1	2	1	1	7 (Moderate)
S3 / Gao et al.	2	2	2	2	1	9 (High)
S4 / Niu et al.	2	2	2	1	1	8 (High)
S5 / Caballero-Ruiz et al.	2	1	2	2	2	9 (High)
S6 / Eili et al.	2	2	2	1	1	8 (High)
S7 / Hassanpour & Langlotz	2	2	2	1	1	8 (High)
S8 / Alaeddini et al.	2	2	2	1	0	7 (Moderate)
S9 / Kaduskar et al.	1	1	2	1	1	6 (Moderate)
S10 / Nef et al.	2	2	2	1	2	9 (High)

C1 represents domain/problem clarity, C2 data transparency, C3 method reproducibility, C4 DSS relevance, and C5 validation or implementation evidence. Each criterion was scored from 0 to 2, with total scores of 8–10 classified as High, 5–7 as Moderate, and 0–4 as Low. While the appraisal captures the reporting quality and decision-support relevance of the included studies, validation maturity varied considerably across the corpus and was generally less developed than algorithmic performance. Most studies relied on internal clustering metrics, retrospective evaluation, predictive discrimination, or expert interpretation. S5 represents an important exception because the clustering component was embedded within a web-based gestational-diabetes clinical decision support system that underwent prospective clinical evaluation. However, even in this study, the reported evidence focused primarily on safety, workload reduction, patient prioritization, and service utilization rather than on definitive improvements in downstream patient outcomes. Overall, external multi-institutional validation, independent clinician validation, and prospective workflow evaluation remain limited across the included studies.

3.5. Discussion: Interpreting the Findings Across RQ1 – RQ3

Taken together, RQ1-RQ3 show that clustering has clear analytical value but occupies different positions in the healthcare decision-support chain. In some studies, clusters directly informed an audit list, alert, treatment recommendation, pathway prediction, or safety-monitoring function (S1, S2, S5, S6, and S10). In others, clustering primarily organized complex data into groups that could later support information retrieval, risk modeling, or precision-medicine decisions (S4, S7, S8, and S9). This distinction matters

because an exploratory clustering experiment should not automatically be described as an implemented Healthcare DSS. A more defensible interpretation is a continuum from exploratory grouping, through prototype decision support, to prospective workflow deployment.

The method distribution in RQ2 also helps explain why K-Means remains common. Its computational efficiency, transparent centroid representation, and familiarity make it comparatively easy to integrate into human-interpretable workflows. These strengths are balanced by sensitivity to initialization, scale, outliers, and the selected value of K. Deep, density-based, fuzzy, and graph-oriented approaches can represent more complex structures, but they also raise additional challenges in parameter selection, explainability, computation, and governance. For real DSS use, cluster separation alone is insufficient: clinicians and operational users need understandable profiles, meaningful labels, uncertainty information, and a clear connection between clusters and actions.

The validation findings in RQ3 suggest a practical hierarchy for future work. Internal validation establishes whether clusters are mathematically coherent; external validation tests reproducibility in other populations or institutions; clinician validation determines whether groups are interpretable and actionable; and prospective deployment evaluation assesses whether a system improves workflow, decision quality, safety, or patient outcomes under real operating conditions. Most of the included evidence remains concentrated in the earlier stages of this hierarchy. Future clustering-based Healthcare DSS research should therefore report not only algorithmic metrics but also usability, clinician agreement, workflow impact, failure modes, fairness, interoperability, privacy protection, and post-deployment monitoring [32].

4. CONCLUSION

This systematic review identified 10 primary empirical studies published from 2015 to 2025 that linked clustering-based data mining to an identifiable healthcare decision-support function. The evidence covered administrative fraud detection, ICU profiling, chronic-disease risk and transition analysis, gestational-diabetes treatment support, traumatic-brain-injury pathways, radiology-report organization, patient segmentation,

and smart-home activity recognition, with K-Means reported most frequently and Markov-based approaches appearing next most often. Across the three research questions, the central finding is that clustering is useful for structuring heterogeneous healthcare data and supporting exploratory or prototype decisions, but the evidence for routine clinical deployment remains substantially weaker than the evidence for analytical feasibility. Most studies relied on retrospective, internal, or expert-based validation, and only one included prospective evaluation of a deployed web-based clinical DSS; external multi-institutional validation, independent clinician assessment, and demonstrated patient-outcome improvement were uncommon. These findings indicate that future research should move beyond cluster quality alone toward explainable profiles, reproducibility across institutions, clinician-centered validation, privacy-preserving integration, EHR/HIS interoperability, fairness assessment, and prospective workflow evaluation. The conclusions should also be read in light of this review's own limitations, including reliance on Scopus, use of an open-access filter, heterogeneity that precluded meta-analysis, absence of a prospectively registered protocol, and screening and extraction that were not fully duplicated by independent reviewers.

REFERENCES

- [1] M. I. Arias, L. Cadavid, and J. D. Velásquez, "Advancing Healthcare Analytics: A Thematic Review of Machine Learning, Health Informatics, and Real-World Data Applications," *J. Biomed. Inform.*, vol. 171, 2025, doi: 10.1016/j.jbi.2025.104934.
- [2] B. S. Raja and S. Asghar, "Beyond the Horizon: A Meticulous Analysis of Clinical Decision-Making Practices," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, 2020, doi: 10.14569/IJACSA.2020.0110287.
- [3] D. Garrido-Delgado, J. Lillo-Vidal, M. Escudero-Bell, V. Hurtado-Meneses, and M. Villalobos-Cid, "Spatiotemporal Characterisation of Stroke Care Accessibility and Hospital Coverage in the Santiago Metropolitan Region," in *21st International Symposium on Medical Information Processing and Analysis, SIPAIM 2025*, Institute of Electrical and Electronics Engineers Inc., 2025. doi: 10.1109/SIPAIM67325.2025.11283388.
- [4] J. F. Andry and H. Tannady, "Data Management Analysis for Predicting Stroke using RapidMiner," *J. Inf. Syst. Technol. Eng.*, vol. 2, no. 3, pp. 318–322, 2024.

- [5] W. Radhwan and A. Alnahdi, "Towards Medical Ontology Construction using Data Mining: An Approach For Creating A Diabetic Ontology using Clustering," in *6th International Conference on Medical and Health Informatics, ICMHI 2022*, Association For Computing Machinery, 2022, pp. 85–88. doi: 10.1145/3545729.3545747.
- [6] L. Lella, L. Gentile, C. Pristipino, and D. Toni, "Predictive Clustering Learning Algorithms for Stroke Patients Discharge Planning," in *14th International Conference on Health Informatics, HEALTHINF 2021 - Part of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC 2021*, SciTePress, 2021, pp. 296–303.
- [7] K. A. Arasan, E. Ramaraj, and A. Padmapriya, "Clinical Support System for Healthcare Providers Using Big Data Analytics," *J. Eval. Clin. Pract.*, vol. 31, no. 1, 2025, doi: 10.1111/jep.70014.
- [8] R. A. Dinesh, J. Shanmugam, and K. Biswas, "Prediction of Heart Disease," *Leveraging Potential Artif. Intell. Real World Smart Cities Healthc.*, pp. 21–33, 2024, doi: 10.1201/9781032667508-2.
- [9] E. Rozaki *et al.*, "An Unsupervised Learning Approach to Optimising Tumour Therapy Through Clinical Data Mining," *Healthc. Anal.*, vol. 9, p. 100450, 2026, doi: 10.1016/j.health.2026.100450.
- [10] M. Alharbi and S. Ahmad, "Enhancing COVID-19 Detection using CT-Scan Image Analysis and Disease Classification: The DI-QL Approach," *Health Technol. (Berl.)*, vol. 15, no. 2, pp. 477–488, 2025, doi: 10.1007/s12553-025-00952-0.
- [11] J. F. Andry, H. Tannady, G. D. Rembulan, and D. F. Dinata, "Analysis of the Omicron Virus Cases using Data Mining Methods in Rapid Miner Applications," *Microbes Infect. Dis.*, vol. 4, no. 2, pp. 323–334, 2023, doi: 10.21608/MID.2023.194619.1469.
- [12] X. Li, J. Zhang, and F. Safara, "Improving the Accuracy of Diabetes Diagnosis Applications through a Hybrid Feature Selection Algorithm," *Neural Process. Lett.*, vol. 55, no. 1, pp. 153–169, 2023.
- [13] L. Natrayan, M. D. R. Kamal, K. K. Manivannan, and G. Sunil, "Machine Learning and Data Mining Approaches for Infectious Disease Surveillance and Outbreak Management in Healthcare," in *2nd International Conference on Advancements in Smart, Secure and Intelligent Computing, ASSIC 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/ASSIC60049.2024.10507990.

- [14] S. H. Moz, M. A. Hosen, N. S. Santo, S. S. Kabir, M. N. Adnan, and S. M. Galib, "Precision Cardiodiet: Transforming Cardiac Care with Artificial Intelligence-Driven Dietary Recommendations," *Radioelectron. Comput. Syst.*, vol. 4, no. 108, p. 2026, 2023, doi: 10.32620/reks.2023.4.02.
- [15] J. Schwarzerova *et al.*, "CyberKnife and Data Mining: Exploring Opportunities for Clinical Advancements," in *12th International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO 2025*, Springer Science and Business Media Deutschland GmbH, 2026, pp. 219–229. doi: 10.1007/978-3-032-08452-1_18.
- [16] F. S. Lee, J. F. Andry, Y. Purnomo, K. Christianto, D. V. Cong, and R. E. Tanuwijaya, "Corporate Responsibility and Sustainable Business Practices," in *Green Society, Environmental Strategies and Sustainable Development*, 2025, pp. 109–124. doi: 10.1007/978-3-031-94086-6.
- [17] H. Geremew *et al.*, "Late-stage Diagnosis: The Driving Force Behind High Breast Cancer Mortality in Ethiopia: A Systematic Review and Meta-Analysis," *PLoS One*, vol. 19, pp. 1–10, 2024, doi: 10.1371/journal.pone.0307283.
- [18] L. Muyama, A. Neuraz, and A. Coulet, "Machine Learning Approaches for the Discovery of Clinical Pathways from Patient Data: A Systematic Review," *J. Biomed. Inform.*, vol. 160, no. 104746, 2024, doi: 10.1016/j.jbi.2024.104746.
- [19] X. Chen, G. Cheng, F. L. Wang, X. Tao, H. Xie, and L. Xu, "Machine and Cognitive Intelligence for Human Health: Systematic Review," *Brain Informatics*, vol. 9, no. 5, 2022, doi: 10.1186/s40708-022-00153-9.
- [20] R. Chahar, "Computational Decision Support System in Healthcare: A Review and Analysis," *Int. J. Adv. Technol. Eng. Explor.*, vol. 8, no. 75, 2021, doi: 10.19101/IJATEE.2020.762142.
- [21] L. Souza-pereira, S. Ouhbi, and N. Pombo, "A Process Model for Quality in Use Evaluation of Clinical Decision Support Systems," *J. Biomed. Inform.*, vol. 123, no. May, p. 103917, 2021, doi: 10.1016/j.jbi.2021.103917.
- [22] M. C. Massi, F. Ieva, and E. Lettieri, "Data Mining Application to Healthcare Fraud Detection: A Two-Step Unsupervised Clustering Method for Outlier Detection with Administrative Databases," *BMC Med. Inform. Decis. Mak.*, vol. 20, no. 160, pp. 1–11, 2020, doi: 10.1186/s12911-020-01143-9.
- [23] A. Ribeiro, F. Portela, M. Santos, A. Abelha, J. Machado, and F. Rua, "Patients ' Admissions in Intensive Care Units: A Clustering Overview," *Information*, vol. 8, no. 1, pp. 1–12, 2017, doi: 10.3390/info8010023.

- [24] Y. Gao, C. Sun, R. Li, Q. Li, L. Cui, and B. Gong, "An Efficient Fraud Identification Method Combining Manifold Learning and Outliers Detection in Mobile Healthcare Services," *IEEE Access*, vol. 6, pp. 60059–60068, 2018, doi: 10.1109/ACCESS.2018.2875516.
- [25] S. Niu, J. Ma, Q. Yin, L. Bai, C. Li, and X. Yang, "A Deep Clustering-Based State-Space Model For Improved Disease Risk Prediction In Personalized Healthcare," *Ann. Oper. Res.*, vol. 341, no. 1, pp. 647–672, 2024, doi: 10.1007/s10479-023-05817-1.
- [26] E. Caballero-Ruiz, G. García-Sáez, M. Rigla, M. Villaplana, B. Pons, and M. E. Hernando, "A Web-Based Clinical Decision Support System for Gestational Diabetes: Automatic Diet Prescription and Detection of Insulin Needs," *Int. J. Med. Inform.*, vol. 102, pp. 35–49, 2017, doi: 10.1016/j.ijmedinf.2017.02.014.
- [27] M. Y. Eili, J. Rezaeenour, and M. H. Roozbahani, "Predicting Clinical Pathways of Traumatic Brain Injuries (TBIs) through Process Mining," *npj Digit. Med.*, vol. 8, no. 112, pp. 1–12, 2025, doi: 10.1038/s41746-025-01484-7.
- [28] S. Hassanpour and C. P. Langlotz, "Unsupervised Topic Modeling in a Large Free Text Radiology Report Repository," *J Digit Imaging*, vol. 29, no. 1, pp. 59–62, 2016, doi: 10.1007/s10278-015-9823-3.
- [29] A. Alaeddini, C. A. Jaramillo, S. H. A. Faruqui, and M. J. Pugh, "Mining Major Transitions of Chronic Conditions in Patients with Multiple Chronic Conditions," *Methods Inf. Med.*, vol. 56, no. 5, pp. 391–400, 2019, doi: 10.3414/ME16-01-0135.Mining.
- [30] V. Kaduskar, A. Srivastava, S. O. Husain, M. Gupta, and V. Patil, "Integrating Fuzzy C-Means and DBSCAN: A Hybrid Approach to Medical Data Mining," *Fuzzy Inf. Eng.*, vol. 17, no. 1, pp. 108–119, 2025, doi: 10.26599/FIE.2025.9270055.
- [31] T. Nef *et al.*, "Evaluation of Three State-of-the-Art Classifiers for Recognition of Activities of Daily Living from Smart Home Ambient Data," *Sensors (Switzerland)*, vol. 15, no. 5, pp. 11725–11740, 2015, doi: 10.3390/s150511725.
- [32] M. Li, Y. Lin, H. Chen, and R. R. Aparasu, "An Unsupervised Embedding Harmonization System for Privacy-Preserving Data Mining in Healthcare," *IJSE Trans. Healthc. Syst. Eng.*, vol. 14, no. 1, pp. 1–17, 2024, doi: 10.1080/24725579.2023.2231011.