

Topic Modeling and Keyword Co-occurrence Analysis of Cardiovascular Digital Health Equity Research

Zhiqiang Wang

Yangtze University

Xiaoya Li

Yangtze University

Zhiwen Zhang

fwzzw01@163.com

Fuwai Central China Cardiovascular Hospital

Muwei Li

Fuwai Central China Cardiovascular Hospital

Research Article

Keywords: Cardiovascular diseases, digital health equity, Latent Dirichlet Allocation topic modeling, keyword co-occurrence, literature analysis, social determinants of health, telemedicine

Posted Date: July 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9805442/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Objective

To identify the thematic structure and main research priorities in cardiovascular digital health equity and to guide future research, intervention design, and policy development.

Methods

Relevant articles published between 1995 and 2025 were retrieved from Web of Science and PubMed. Titles and abstracts were extracted and processed through text cleaning, stop-word removal, and lemmatization. A keyword co-occurrence network was constructed, and Latent Dirichlet Allocation topic modeling was used to identify themes. The optimal number of topics was determined by coherence scores, perplexity, and manual semantic review, followed by higher-order thematic synthesis.

Results

Publications on cardiovascular digital health equity showed a sustained upward trend, with especially rapid growth after 2020. The keyword co-occurrence network placed stroke, telehealth, telemedicine, and disparity at the center, highlighting the close intersection of digital technology, disease management, and equity concerns. Eleven latent topics were identified, including digital hypertension management, differences in digital health use across populations, social determinants of health and risk prediction, telemedicine accessibility during the pandemic, rural telestroke care, remote electrocardiographic monitoring and digital diagnostics, and digital literacy and community engagement. Technology integration and implementation strategies to promote equitable access represented the largest topic share (25.37%). These topics were further synthesized into four higher-order clusters: structural inequality and identification of vulnerable populations; digital access and conditions for equitable use; equity-oriented implementation, governance, and system integration; and cardiovascular digital clinical applications and intelligent support.

Conclusion

Cardiovascular digital health equity research has developed a multidimensional and cross-cutting knowledge structure. Its focus has moved beyond technology application alone toward the identification of social disparities, optimization of digital access, and equity-oriented governance, with important implications for targeted interventions and equitable implementation in cardiovascular digital health.

1. Introduction

Cardiovascular disease (CVD) remains a leading source of mortality and disease burden worldwide and a major cause of premature death^[1, 2]. In recent years, with the rapid development of technologies such as mobile health (mHealth), telemedicine, wearable devices, remote patient monitoring, and artificial intelligence-driven data analytics, digital health has been widely applied to CVD prevention, risk factor management, and long-term follow-up, showing substantial potential to expand service coverage, improve the efficiency of continuous monitoring, and support personalized management^[3–5]. However, the expansion of digital technologies does not necessarily translate into equitable benefits across populations. While digital health may improve the accessibility and efficiency of health care, it may also generate new inequalities because of differences in device availability, internet access, digital literacy, platform usability, and institutional design, thereby affecting its real-world effectiveness and public health value in cardiovascular management^[6–10].

Digital health equity is reflected not only in whether services are available, but also in whether individuals can access, use, and continuously benefit from them. Digital interventions that depend on smartphone ownership or stable internet connectivity may create hidden barriers at the recruitment stage, making older adults, individuals with low educational attainment, and low-income populations more likely to be excluded^[11]. People with disabilities also face additional barriers in accessing, understanding, and operating electronic health services, which further limits their effective use of digital tools and services^[12]. At the same time, limited health literacy and eHealth literacy may weaken the ability to obtain, understand, and apply digital health information and services, and may reduce the effectiveness of digital interventions^[13–15]. Older adults, socioeconomically disadvantaged groups, and residents of remote areas are particularly vulnerable to the digital divide; limited device access, weak internet infrastructure, and insufficient digital skills often prevent them from fully benefiting from digital health interventions^[16–18]. This issue is especially important in the CVD field, where prevention and management rely heavily on long-term monitoring, continuous follow-up, and patient self-management. If equity is not adequately incorporated into the design, implementation, and governance of digital tools, digital transformation may fail to reduce existing health disparities and may instead reproduce or even worsen them^[6–10].

Although cardiovascular digital health equity has gradually become an international research priority, the evidence remains scattered across clinical medicine, public health, nursing, health policy, and medical informatics. Existing reviews have mainly focused on specific intervention models or populations, such as telemedicine for CVD and hypertension management, digital hypertension interventions for populations experiencing health disparities, and digital strategies to promote women's cardiovascular health^[5, 19, 20]. In addition, prior studies have proposed action pathways to advance digital health equity in cardiovascular care from the perspectives of digital access, digital literacy, and multilevel interventions^[21]. Other work has examined the potential and limitations of digital technologies in reducing cardiovascular health disparities from the perspectives of global health equity, the digital divide in underrepresented communities, and artificial intelligence-enabled digital health practice in China^{[22–}

^{24]}. Empirical studies in China further show that, although the coverage of cardiovascular telemedicine services has continued to expand, their regional distribution does not fully match levels of economic development or disease burden, suggesting that growth in digital health resources alone does not necessarily improve equity^[25]. These studies provide an important basis for understanding specific issues, but overall they remain dominated by narrative reviews, commentaries, or studies limited to specific countries and settings, making it difficult to systematically reveal the underlying thematic structure, core knowledge units, and evolutionary pathways of this field at an overall level. Although previous bibliometric studies have helped describe publication trends and hotspot distributions, without support from natural language processing and topic modeling they often cannot extract deeper semantic structure or knowledge-evolution patterns from large-scale unstructured text^[26–29].

Therefore, a systematic analysis of cardiovascular digital health equity research using methods suitable for large-scale text data is needed. Latent Dirichlet Allocation (LDA) is a classic unsupervised topic-modeling method that identifies latent thematic structure in a corpus based on term co-occurrence probabilities and provides a reproducible framework for semantic synthesis across large literature sets. Accordingly, this study systematically retrieved literature on cardiovascular digital health equity and combined keyword co-occurrence analysis with LDA topic modeling to identify the core research themes, major areas of focus, and overall knowledge structure of the field, thereby informing future research agendas, policy development, and the design of equity-oriented digital health interventions.

2. Materials and Methods

2.1 Search Strategy

This study systematically searched the Web of Science Core Collection and PubMed for literature related to cardiovascular digital health equity. The search covered publications from January 1, 1995, to December 31, 2025, and was completed on April 9, 2026. The research team conducted three iterative rounds of searching around three core concepts—cardiovascular disease (CVD), digital health, and equity/disparities—and finalized the search strategy after discussion. The full search strategy is provided in Supplementary Table S1. No language or study design restrictions were applied. All records were imported into EndNote 20 for management and deduplication.

2.2 Search Process

After database searching and import into EndNote 20, duplicate records were identified and removed based on title, author, publication year, journal, volume and issue, and digital object identifier (DOI). After deduplication, the research team screened the records according to the predefined study topic. Screening was primarily based on titles and abstracts to determine whether the articles addressed the core concepts of CVD, digital health, and health equity/disparities/accessibility. Articles were excluded if they were clearly unrelated to the study topic, including those in which the study population was not related to CVD, the content did not involve digital health technologies or services, equity- or disparity-

related issues were not addressed, or valid title or abstract information was unavailable for subsequent text analysis. For records with limited information but possible relevance, the research team further assessed eligibility based on the title, abstract, and bibliographic information available in the databases.

For the final set of included studies, titles, abstracts, authors, publication years, source journals, and other bibliographic information were extracted to build the raw dataset for subsequent analyses. Combined titles and abstracts were used as the text analysis unit for text preprocessing, keyword co-occurrence analysis, and Latent Dirichlet Allocation (LDA) topic modeling. Publication year was used to describe the temporal distribution of the research, and source journals were used to identify the major channels of knowledge dissemination in this field. The study screening and inclusion process is shown in Fig. 1.

2.3 Data Preprocessing

We extracted the titles and abstracts of all retrieved studies. Texts were preprocessed using standard methods consistent with previous studies^[30–32]. To improve the interpretability and stability of topic modeling, the title and abstract of each included article were combined into a single text unit, and text preprocessing was performed in Python. Specifically, all text was converted to lowercase; punctuation, numbers, special characters, and other noninformative content were removed; and common English stop words were excluded. On this basis, a domain-specific stop-word list was further developed to remove high-frequency terms common in study design and scientific writing but uninformative for topic identification, such as objective, method, result, and conclusion. Lemmatization was then performed using scispaCy. Multiword phrases with consistent semantic meaning, such as machine learning and diabetic retinopathy, were retained as single terms for subsequent analysis. The preprocessed corpus was then used for keyword co-occurrence analysis and LDA topic modeling.

2.4 Data Analysis

2.4.1 Keyword Co-occurrence Network Construction

To reveal the structural relationships among core concepts in cardiovascular digital health equity research, a keyword co-occurrence network was constructed based on the preprocessed titles and abstracts of the included studies. After text cleaning, stop-word removal, lemmatization, and phrase identification, high-frequency terms in the corpus were extracted as network nodes. To reduce noise from low-frequency terms, only terms meeting a prespecified frequency threshold were included in the network analysis. Node size was weighted by the frequency of each term in the full corpus to represent its level of attention in the overall research field. If two terms co-occurred in the title or abstract of the same article, an edge was created between them, and the edge weight was defined as the number of articles in which the two terms co-occurred. To highlight stable and interpretable core relationships, only term pairs with a co-occurrence frequency of at least 30 were retained to construct the keyword co-occurrence network. The network data were then imported into Gephi 0.10.1 for visualization and structural examination.

2.4.2 LDA Topic Modeling

Latent Dirichlet Allocation (LDA), a classic unsupervised topic-modeling method proposed by Blei et al. in 2003, has been widely used to identify latent themes in biomedical literature^[33]. LDA assumes that each document consists of multiple latent topics and that each topic is characterized by a set of terms with different probability distributions. By estimating the document–topic distribution and topic–word distribution, LDA can identify relatively stable semantic structures from large-scale text corpora.

In this study, the title and abstract of each article were combined into a single analysis unit. After text cleaning, stop-word removal, and lemmatization, the LDA model was developed using the Gensim package in Python. Because LDA represents text using a bag-of-words model and does not account for word order or syntactic structure, topic identification mainly depends on term co-occurrence patterns across documents.

To determine the optimal number of topics, topic coherence and perplexity were calculated for models with 2 to 30 topics, followed by a combined evaluation with manual semantic review[34] (Table 1). Higher topic coherence generally indicates better semantic consistency within a topic, whereas perplexity was used as an auxiliary measure of model fit to the corpus. Based on these quantitative metrics, manual assessment was further performed from three perspectives: semantic cohesion within topics, distinctiveness between topics, and alignment with the study focus. After comprehensive comparison, 11 topics were selected as the optimal model to represent the latent thematic structure of cardiovascular digital health equity research (Fig. 2).

During topic interpretation, the highest-probability keywords for each topic were extracted, and topics were named and summarized in conjunction with the content of representative articles to reveal the main research priorities and knowledge structure of the field.

Table 1
Topic coherence and perplexity of the LDA models
with different numbers of topics.

Number of Topics	Coherence	Perplexity
2	0.322974	159.322228
3	0.357295	155.6443
4	0.36456	152.14319
5	0.359026	152.381283
6	0.385943	151.162891
7	0.382845	151.103026
8	0.372052	152.186705
9	0.373573	151.526561
10	0.382287	152.120303
11	0.402633	152.484231
12	0.385556	153.476255
13	0.385887	154.020119
14	0.399813	154.425825
15	0.399112	154.826068
16	0.400129	155.275811
17	0.391394	155.97334
18	0.380069	156.8843
19	0.391906	156.927606
20	0.398559	158.128534
21	0.391325	159.311138
22	0.382504	159.26532
23	0.380233	159.818823
24	0.384814	160.160288
25	0.388391	160.786267
26	0.386906	161.473503
27	0.375362	162.27949

Number of Topics	Coherence	Perplexity
28	0.389811	161.455108
29	0.372388	163.256048
30	0.368563	163.84232

2.4.3 Higher-Order Thematic Synthesis

To further improve the abstraction and interpretability of the thematic structure derived from the first-level LDA results, a higher-order thematic synthesis was conducted for the 11 identified latent topics. Two cardiovascular experts independently classified the topics based on their high-frequency keywords, the content of representative articles, and semantic consistency, and then grouped first-level topics with similar meanings and aligned research orientations into higher-order thematic clusters. To assess classification agreement, Cohen's kappa was used to evaluate the initial judgments of the two experts. The results showed high agreement between the two experts (Cohen's kappa = 0.884). Topics with discrepant classifications were resolved through further discussion to reach a final consensus, resulting in four higher-order thematic clusters. This step constituted a post hoc interpretive synthesis of the model output and was intended to strengthen the structural presentation and overall interpretability of the first-level topic results.

2.4.4 Robustness and Interpretability Controls

To improve the reliability of the findings, multiple robustness and interpretability controls were applied during topic modeling and interpretation. First, topic number selection was based not only on topic coherence and perplexity, but also on manual comparison of several candidate models to avoid excessive topic merging when too few topics were specified and topic fragmentation when too many topics were specified. Second, topic naming was based on both high-frequency keywords and the content of representative articles, rather than on isolated high-frequency terms alone. Third, independent expert judgment and agreement assessment were incorporated into the higher-order synthesis to reduce the influence of subjective interpretation by a single researcher. Together, these procedures strengthened the stability, logical consistency, and interpretive credibility of the identified themes.

2.4.5 Visualization

To present the findings clearly, multiple visualization methods were used to display publication characteristics and thematic structure. Publication trends and journal distributions were plotted using GraphPad Prism 9. The keyword co-occurrence network was laid out and visualized in Gephi 0.10.1. Word clouds of high-frequency terms for each topic were generated using the WordCloud package in Python; for each topic, the top 20 terms with the highest probabilities were displayed, and font size reflected the probability weight of each term within that topic. The relationships between first-level topics and higher-order thematic clusters were illustrated using a Sankey diagram to show the hierarchical structure of the themes.

3. Results

3.1 Publication Trends

As shown in Fig. 3A–B, publications on cardiovascular digital health equity showed a marked upward trend overall. From 1995 to 2010, the field was in an early stage, with annual output remaining at single digits to just over 10 articles, indicating relatively limited attention. From 2011 to 2019, publication output entered a period of steady growth, increasing from 11 to 52 articles per year. After 2020, the field expanded rapidly: annual output reached 144 articles in 2022, 148 in 2023, 238 in 2024, and 398 in 2025. Overall, this field has progressed from early exploration to rapid development, reflecting strong academic activity and substantial research potential.

Figure 3C shows a clustered distribution of journals publishing research on cardiovascular digital health equity. *Telemedicine and e-Health* ranked first with 52 articles, followed by *Journal of Medical Internet Research* with 41 articles, *JMIR mHealth and uHealth* with 22 articles, *JMIR Research Protocols* with 20 articles, and *Cureus* with 20 articles. Overall, the most productive journals were concentrated in telemedicine, digital health, and internet medicine, while also extending to cardiovascular specialty and public health journals, such as *Journal of the American Heart Association*, *Stroke*, and *BMC Public Health*. These findings indicate that this topic is strongly interdisciplinary and has developed a knowledge dissemination pattern centered on digital health journals and jointly supported by cardiovascular medicine and public health journals.

3.2 Keyword Co-occurrence Network Analysis

The keyword co-occurrence network, consisting of 100 nodes and 2,042 edges, revealed the core conceptual structure of cardiovascular digital health equity research (Fig. 4). Stroke was located at the center of the network and was closely linked to implementation, telehealth, disparity, telemedicine, and prevention, indicating that this field primarily focuses on the application of digital technologies, implementation, and equity issues in cardiovascular disease contexts, particularly stroke. Strong links were observed for stroke–telestroke, implementation–prevention, stroke–telemedicine, and stroke–telehealth, suggesting that telemedicine, preventive management, and clinical care have become major research lines. In addition, nodes such as disparity, inequity, access to care, and health equity were closely linked to core disease- and technology-related terms, indicating that the field has shifted from a sole emphasis on technological effectiveness to broader concerns about access to resources, differences in service coverage, and unequal benefits across vulnerable populations. Other highly connected nodes, including women, older adults, underserved populations, rural areas, and insurance, further suggest that gender, age, geography, and social determinants have been incorporated into the analytical framework. Overall, the field shows an integrated research pattern at the intersection of technology application, disease management, and health equity.

3.3 LDA Topic Analysis

LDA topic modeling identified 11 latent topics in cardiovascular digital health equity research (Table 2 and Fig. 5), indicating that the field has developed a multidimensional thematic structure centered on digital interventions, disease management, health equity, and implementation governance. Among these, Topic 9, technology integration and implementation strategies to promote equitable access, accounted for the largest proportion (25.37%), suggesting that research related to equitable access, barriers to technology use, and implementation responses occupies a substantial share of the existing literature. The remaining topics mainly covered digital hypertension management, differences in digital health use across populations, social determinants and risk prediction, telemedicine services during the pandemic, rural telestroke care, remote cardiac monitoring, and digital literacy and community engagement. Overall, cardiovascular digital health equity research shows an intersecting thematic pattern across technology application, disease prevention and management, and equity-related concerns.

Table 2
Highest-frequency terms for the 11 topics in cardiovascular digital health equity research.

No.	Topic Title	Keywords (Probability)	Proportion
1	Digital self-management and home monitoring interventions for hypertension	intervention (0.0210), participant (0.0170), control (0.0140), self (0.0130), trial (0.0098), improve (0.0093), program (0.0080), management (0.0079), measure (0.0075), community (0.0068), monitoring (0.0063), hypertension (0.0063), home (0.0058), reduce (0.0058), change (0.0059), randomize (0.0057), engagement (0.0038), adherence (0.0033), systolic (0.0033), SMBP (0.0026)	152 (10.93%)
2	Digital health use and health disparities across racial and age groups	compare (0.0120), rate (0.0089), associate (0.0089), age (0.0085), Black (0.0072), difference (0.0065), Hispanic (0.0057), White (0.0053), factor (0.0051), race (0.0047), female (0.0046), demographic (0.0043), disparity (0.0040), sex (0.0039), woman (0.0039), cohort (0.0036), male (0.0034), racial (0.0030), ethnicity (0.0030), hospitalization (0.0029)	126 (9.04%)
3	Barriers and facilitators to the implementation of digital health equity	intervention (0.0140), review (0.0140), identify (0.0130), interview (0.0084), barrier (0.0070), focus (0.0057), analysis (0.0054), theme (0.0051), evidence (0.0050), systematic (0.0043), develop (0.0042), support (0.0038), implementation (0.0035), framework (0.0034), mixed (0.0033), facilitator (0.0032), qualitative (0.0032), effectiveness (0.0030), experience (0.0028), evaluation (0.0028)	91 (6.55%)
4	Social determinants and geographic disparities in cardiovascular risk	risk (0.0240), factor (0.0230), CVD (0.0140), woman (0.0100), cardiovascular (0.0092), income (0.0071), mortality (0.0062), rural (0.0061), prevention (0.0049), social (0.0048), urban (0.0045), county (0.0045), environmental (0.0042), socioeconomic (0.0041), prevalence (0.0041), community (0.0040), lifestyle (0.0039), exposure (0.0034), obesity (0.0034), determinant (0.0026)	70 (5.06%)
5	Telemedicine accessibility and continuity of care during the pandemic	telehealth (0.0210), visit (0.0180), pandemic (0.0140), service (0.0120), COVID-19 (0.0100), telemedicine (0.0092), provider (0.0092), access (0.0088), clinic (0.0085), medical (0.0054), primary (0.0053), consultation (0.0049), home (0.0045), virtual (0.0044), rural (0.0044), follow (0.0042), medication (0.0041), cardiology (0.0039), appointment (0.0035), video (0.0034)	122 (8.77%)
6	Cardiovascular risk prediction models integrating social determinants	model (0.0360), risk (0.0200), predict (0.0150), identify (0.0140), learning (0.0130), machine (0.0120), SDOH (0.0100), factor (0.0093), clinical (0.0088), prediction (0.0084), feature (0.0078), variable (0.0077), machine learning (0.0075), develop (0.0071), predictor (0.0070), algorithm	137 (9.85%)

No.	Topic Title	Keywords (Probability)	Proportion
		(0.0064), social (0.0047), demographic (0.0046), AUC (0.0041), EHR (0.0028)	
7	Remote complication screening and referral for high-risk patients with cardiovascular comorbidities	diabetes (0.0200), screening (0.0180), screen (0.0160), diabetic (0.0097), telemedicine (0.0095), eye (0.0091), diabetic retinopathy (0.0071), retinopathy (0.0069), referral (0.0066), underserved (0.0055), rural (0.0053), image (0.0053), examination (0.0048), retinal (0.0041), ophthalmology (0.0041), complication (0.0036), teleophthalmology (0.0033), detection (0.0032), fundus (0.0027), glaucoma (0.0046)	68 (4.93%)
8	Telestroke networks in rural and underserved areas	stroke (0.0500), hospital (0.0280), telestroke (0.0150), rural (0.0150), acute (0.0150), care (0.0140), telemedicine (0.0099), access (0.0090), network (0.0082), treatment (0.0071), emergency (0.0066), underserved (0.0062), rehabilitation (0.0052), center (0.0051), neurologist (0.0043), facility (0.0041), neurology (0.0039), physician (0.0039), transfer (0.0037), tPA (0.0036)	99 (7.15%)
9	Technology integration and implementation strategies to promote equitable access	technology (0.0110), healthcare (0.0096), address (0.0093), challenge (0.0092), disparity (0.0070), strategy (0.0070), enhance (0.0064), intervention (0.0055), access (0.0053), reduce (0.0051), cardiovascular (0.0047), digital (0.0041), tool (0.0041), barrier (0.0041), equity (0.0033), policy (0.0032), solution (0.0032), integrate (0.0031), opportunity (0.0029), innovation (0.0025)	352 (25.37%)
10	Remote electrocardiography and digital cardiac monitoring technologies	ECG (0.0100), device (0.0091), heart (0.0078), system (0.0077), diagnostic (0.0074), diagnosis (0.0070), remote (0.0070), image (0.0060), cardiac (0.0056), detect (0.0054), medical (0.0050), algorithm (0.0049), underserved (0.0044), imaging (0.0044), assessment (0.0043), echocardiography (0.0034), CHD (0.0033), wearable (0.0032), accuracy (0.0032), monitoring (0.0031)	72 (5.20%)
11	mHealth adoption, digital literacy, and community engagement	participant (0.0200), adult (0.0160), old (0.0130), self (0.0120), technology (0.0095), mHealth (0.0081), factor (0.0080), individual (0.0076), community (0.0072), education (0.0065), survey (0.0063), user (0.0062), app (0.0060), age (0.0060), support (0.0053), mobile (0.0050), literacy (0.0047), information (0.0044), smartphone (0.0040), digital (0.0031)	98 (7.09%)

Abbreviations for Table 2 and Figure 5: AUC, area under the curve; CHD, congenital heart disease; COVID-19, coronavirus disease 2019; CVD, cardiovascular disease; ECG, electrocardiography; EHR, electronic health record; mHealth, mobile health; SMBP, self-measured blood pressure; SDOH, social determinants of health; tPA, tissue plasminogen activator.

To further reveal the higher-order structural relationships among the LDA-identified topics, we invited two cardiovascular experts (a doctoral researcher in cardiovascular medicine and a professor of cardiovascular medicine) to independently classify the 11 first-level latent topics based on their high-frequency keywords, representative articles, and semantic consistency (Table 3). Agreement analysis showed high consistency in the experts' initial classifications (Cohen's kappa = 0.884). Topics with discrepant classifications were resolved through discussion to reach a final consensus. Ultimately, the 11 first-level topics were synthesized into four higher-order thematic clusters: structural inequality and identification of vulnerable populations; digital access and conditions for equitable use; equity-oriented implementation, governance, and system integration; and cardiovascular digital clinical applications and intelligent support. The first two clusters mainly focused on racial, age, income, urban-rural, and geographic disparities, as well as conditions for digital access such as telemedicine accessibility, digital literacy, and community engagement. The latter two clusters extended to equity-oriented implementation mechanisms, technology integration strategies, and the application of digital technologies in chronic disease management, risk prediction, remote screening, acute care, and digital diagnostics. The Sankey diagram (Figure 6) showed clear aggregation patterns between the 11 first-level topics and the four higher-order thematic clusters, supporting the stability and interpretability of the higher-order synthesis. Overall, cardiovascular digital health equity research has developed a hierarchical thematic structure that progresses from identifying structural inequalities to optimizing digital access and implementation, and further to expanding clinical applications and intelligent support.

Table 3. Higher-order thematic synthesis results of cardiovascular digital health equity research.

Higher-Order Thematic Cluster	Description	Corresponding Topic Numbers	Corresponding Specific Topics
Structural inequality and identification of vulnerable populations	Focuses on the upstream social foundations of digital health inequity and identifies disparities and vulnerabilities in cardiovascular digital health across populations and regions.	Topic 2, Topic 4	Digital health use and health disparities across racial and age groups; Social determinants and geographic disparities in cardiovascular risk
Digital access and conditions for equitable use	Focuses on the real-world conditions that enable individuals and groups to enter the digital health service system, sustain use, and develop the capacity for continued engagement.	Topic 5, Topic 11	Telemedicine accessibility and continuity of care during the pandemic; mHealth adoption, digital literacy, and community engagement
Equity-oriented implementation, governance, and system integration	Focuses on the process through which digital health equity moves from concept to implementation, including barrier identification, enabling mechanisms, resource integration, and strategy optimization.	Topic 3, Topic 9	Barriers and facilitators to the implementation of digital health equity; Technology integration and implementation strategies to promote equitable access
Cardiovascular digital clinical applications and intelligent support	Focuses on the specific applications and intelligent expansion of digital technologies in cardiovascular disease prevention, management, screening, diagnosis, referral, and emergency care.	Topic 1, Topic 6, Topic 7, Topic 8, Topic 10	Digital self-management and home monitoring interventions for hypertension; Cardiovascular risk prediction models integrating social determinants; Remote complication screening and referral for high-risk patients with cardiovascular comorbidities; Telestroke networks in rural and underserved areas; Remote electrocardiography and digital cardiac monitoring technologies

4. Discussion

4.1 Expansion of Digital Health Applications: Chronic Disease Management and Remote Care as Major Research Directions

Based on the LDA results and higher-order thematic synthesis, this study found that topics related to digital chronic disease management, remote screening and diagnosis, and acute care accounted for a substantial share of cardiovascular digital health equity research. Specifically, Topic 1 focused on digital self-management and home monitoring for hypertension, Topic 5 reflected telemedicine accessibility and continuity of care during the pandemic, Topic 8 centered on telestroke networks in rural and underserved areas, and Topic 10 involved remote electrocardiography and digital cardiac diagnostic technologies. These findings suggest that digital technologies have been widely embedded in applications related to CVD prevention, monitoring, management, and continuous care. At the same time, the publication trend showed a marked increase after 2020, indicating sustained growth in research at the intersection of digital health and cardiovascular equity.

This finding is broadly consistent with the existing literature. Previous studies have shown that telemedicine and digital interventions have demonstrated potential in stroke care networks, hypertension management, and maintaining continuity of cardiovascular care during the pandemic^[19,35-39]. Our topic identification results further suggest that these application settings are not only important components of cardiovascular digital health research, but also frequently discussed in the context of equity. In other words, the current literature does not focus solely on the usability of digital technologies themselves, but also on their implementation conditions and service coverage across different disease settings.

It should be noted that our findings primarily reflect the thematic distribution of the literature and cannot directly demonstrate that the field has fully transitioned from one research stage to another. However, in the keyword co-occurrence network, disease- and technology-related terms such as stroke, telehealth, and telemedicine were strongly linked to equity-related terms such as disparity, inequity, and access to care, indicating a close association between technology application and equitable access in the current literature. Together with previous studies, this suggests that although digital cardiovascular care may expand service coverage, barriers to device availability, digital access, and sustained use remain across populations^[38,40]. A more cautious interpretation of our findings is that, within cardiovascular digital health research, technology application and equitable access have developed as parallel and intertwined lines of inquiry. Future studies should therefore pay closer attention to differences across populations in access, use, and sustained benefit.

4.2 Intersecting Drivers of the Digital Divide: Social Determinants, Digital Literacy, and Implementation Barriers

Our findings show that equity-related issues occupy a prominent position in the overall thematic structure. In the keyword co-occurrence network, disparity and inequity were closely linked to core terms such as stroke and telehealth. In the LDA results, Topic 9, technology integration and implementation strategies to promote equitable access, accounted for the largest proportion, while Topics 2, 4, and 11 highlighted major equity-related concerns from the perspectives of utilization differences across racial and age groups, social determinants and geographic disparities, and mHealth adoption and digital literacy, respectively. These findings suggest that digital health equity is not an isolated issue in the

current literature, but is closely intertwined with technology application, disease management, and service implementation.

This finding is consistent with previous framework and review studies. Prior research has shown that the digital divide is reflected not only in differences in device ownership, but also in structural differences across digital access, digital skills, contexts of use, and institutional support^[7,9]. The thematic distribution identified in this study shows similar patterns: Topic 2 highlights racial, age, and gender differences; Topic 4 emphasizes income, rural residence, and socioeconomic factors; and Topic 11 focuses on older adults, mHealth adoption, and digital literacy. Taken together, the current literature tends to examine digital inequity not through a single factor, but within an integrated framework that includes sociodemographic characteristics, resource conditions, and digital capabilities.

Previous studies further support the interpretation of these themes. For example, substantial heterogeneity exists in the measurement of digital literacy among older adults, and the oldest populations are more likely to face difficulties in understanding, operating, and sustaining the use of digital tools^[41]. Digital health literacy is also closely related to chronic disease self-management capacity^[42]. Intervention designs that assume smartphone ownership may exclude some of the populations most in need at the very beginning of program implementation^[11]. Gender, caregiving roles, and socioeconomic position may also jointly shape opportunities to access and use digital health tools^[20]. Taken together with our findings, a more appropriate interpretation is that inequity in cardiovascular digital health is typically presented in the current literature as a multifactorial structure rather than the result of any single barrier. Accordingly, future research and practice should further examine the role of age-friendly design, cultural adaptation, device and connectivity support, and digital literacy improvement in reducing digital exclusion^[9,21].

4.3 Data-Driven Approaches and Equity-Oriented Governance: An Important Direction in Cardiovascular Digital Health Research

The thematic structure identified in this study also suggests that data-driven decision-making and equity-oriented governance have emerged as relatively distinct lines of inquiry in this field. Topic 6, cardiovascular risk prediction models integrating social determinants, directly reflects the combination of social determinants and machine learning-based prediction, whereas Topic 9 captures governance-related issues through technology integration, implementation challenges, and strategic responses. This indicates that the current literature addresses not only the application of digital tools themselves, but also model development, implementation barriers, and institutional responses.

This result is broadly consistent with related review studies. Previous research has shown that incorporating social determinants into machine learning-based prediction models for some CVD outcomes may improve model performance and strengthen the ability to explain risk differences in real-world populations^[43]. Machine learning models based on electronic health records have also shown potential for cardiovascular risk identification, although their generalizability, quality of external

validation, and interpretability still require improvement^[44,45]. In our study, the concurrent prominence of Topics 6 and 9 suggests that the literature is concerned not only with model performance, but also with the equity issues that may arise during technology implementation.

At the same time, our findings suggest that data-driven research is not equivalent to equity-oriented research. In Topic 9, terms such as disparity, barrier, strategy, and access carried substantial weight, indicating that the current literature places considerable emphasis on equitable access, implementation barriers, and response strategies. Together with previous studies, this can be interpreted as follows: if model development and deployment overlook sample representativeness, contextual variation, and institutional constraints, digital technologies may still reproduce or even amplify existing inequities^[7,46,47]. In addition, prior work has shown that structural factors such as racial and ethnic segregation, poverty, and income inequality are closely associated with differences in CVD burden and related health care costs^[48]. Therefore, a more cautious conclusion from the thematic structure identified here is that the current literature has begun to position data-driven approaches and equity-oriented governance side by side as important topics. Future research should further strengthen attention to sample diversity, contextual adaptability, algorithmic transparency, and fairness evaluation^[7,21,46].

5. Limitations and Future Directions

5.1 Limitations

This study has several limitations. First, the literature sources were limited to the Web of Science Core Collection and PubMed. Although both databases are highly representative, relevant studies indexed in Scopus, Embase, and other specialized databases may have been missed, which could affect the completeness of corpus coverage. Second, the text analysis was based mainly on article titles and abstracts rather than full texts. Although titles and abstracts can reasonably reflect study themes, they provide limited information on research context, methodological details, and the substantive dimensions of equity, which may have led to the simplification or underrepresentation of some latent topics. Third, as an unsupervised topic-modeling method based on a bag-of-words framework, LDA relies primarily on term co-occurrence to identify latent semantic structure and cannot fully capture contextual meaning, syntactic dependencies, or deeper relationships among complex concepts. In addition, although topic number selection was based on topic coherence, perplexity, and manual semantic review, a certain degree of methodological dependence and interpretive subjectivity remains inevitable. Furthermore, the higher-order synthesis of identified topics, although supported by independent expert judgment and agreement assessment, is still essentially a post hoc interpretive integration. The results should therefore be understood as a synthesized representation of the current knowledge structure rather than fixed objective boundaries. Finally, this study mainly describes the overall thematic structure and its relative distribution, but does not examine in depth the dynamic evolution of research priorities across

different time periods. Any inference regarding shifts in research focus or thematic evolution should therefore be made with caution.

5.2 Future Directions

Future research can be extended in several directions. First, more international databases and gray literature sources could be incorporated to improve sample coverage and strengthen the robustness of the findings. Second, future analyses could integrate full-text data, author keywords, and citation network information to construct knowledge maps and thematic frameworks based on multisource heterogeneous data. Third, at the methodological level, LDA could be combined with BERTopic, dynamic topic modeling (DTM), or semantic clustering methods based on large language models to improve semantic granularity, temporal sensitivity, and interpretability. Fourth, in terms of result interpretation, more stratified analyses could be conducted across specific populations, disease settings, and national or health system contexts to more precisely identify vulnerable groups, implementation barriers, and policy entry points in cardiovascular digital health equity research. With further advances in data sources, analytical methods, and interpretive frameworks, future studies may more systematically reveal the knowledge evolution of cardiovascular digital health equity research and its priorities for practical implementation.

6. Conclusion

Based on keyword co-occurrence analysis and LDA topic modeling, this study identified 11 latent topics in cardiovascular digital health equity research and further synthesized them into four higher-order thematic clusters. The findings indicate that this field mainly centers on digital chronic disease management and remote care, health equity and social determinants, digital health implementation and data-driven decision-making, and remote screening, diagnosis, and acute care applications. Overall, cardiovascular digital health equity research has developed a multidimensional and cross-cutting thematic structure. Its core concerns include not only the application of digital technologies in disease prevention and management, but also equity in access, use, and sustained benefit across populations. These findings may inform future research agendas, intervention design, and policy development.

Declarations

Ethics approval and consent to participate

Not applicable. This study was based on published literature and bibliographic records retrieved from Web of Science and PubMed and did not involve human participants, human data, human tissue, or animals. Therefore, ethics approval and consent to participate were not required.

Consent for publication

Not applicable. This manuscript does not contain any individual person's data, images, or videos.

Availability of data and materials

The data supporting the findings of this study were derived from bibliographic records retrieved from Web of Science and PubMed. The datasets generated and/or analysed during the current study are not publicly available due to restrictions associated with bibliographic database licensing, but are available from the corresponding author on reasonable request where permitted by the relevant database terms of use.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (82270474, 82470344) and Henan Cardiovascular Disease Center (Central China Subcenter of National Center for Cardiovascular Diseases) (2023-FZX18). The funders had no role in the design of the study; collection, analysis, and interpretation of data; or writing of the manuscript.

Authors' contributions

ZW and XL contributed equally to this work. ZW and ZZ conceived and designed the study. ZW and XL performed the literature search, data collection, and data preprocessing. ZW conducted the keyword co-occurrence analysis, LDA topic modeling, and visualization. XL contributed to data interpretation and manuscript drafting. ZZ and ML supervised the study, critically revised the manuscript, and provided important intellectual input. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Martin SS, Aday AW, Allen NB, et al. 2025 heart disease and stroke statistics: a report of US and global data from the American Heart Association. *Circulation*. 2025;151(8):e41-e660. doi:10.1161/CIR.0000000000001303.
2. Mensah GA, Fuster V, Murray CJL, Roth GA; Global Burden of Cardiovascular Diseases and Risks Collaborators. Global burden of cardiovascular diseases and risks, 1990-2022. *J Am Coll Cardiol*. 2023;82(25):2350-2473. doi:10.1016/j.jacc.2023.11.007.
3. MacKinnon GE, Brittain EL. Mobile health technologies in cardiopulmonary disease. *Chest*. 2020;157(3):654-664. doi:10.1016/j.chest.2019.10.015.
4. Ahmad OM, Ibrahim R, Odunsi DI, et al. Role of mobile health and wearable devices in cardiovascular disease prevention: a comprehensive review. *Cureus*. 2025;17(5):e83673.

doi:10.7759/cureus.83673.

5. Jackson TN, Sreedhara M, Bostic M, et al. Telehealth use to address cardiovascular disease and hypertension in the United States: a systematic review and meta-analysis, 2011-2021. *Telemed Rep.* 2023;4:67-86. doi:10.1089/tmr.2023.0011.
6. Funes Hernandez M, Rodriguez F. Health tech equity: opportunities for digital health innovations to improve equity and diversity in cardiovascular care. *Curr Cardiovasc Risk Rep.* 2023;17(1):1-20. doi:10.1007/s12170-022-00711-0.
7. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. *npj Digit Med.* 2022;5:119. doi:10.1038/s41746-022-00663-0.
8. Kim KK, Backonja U, et al. Digital health equity frameworks and key concepts: a scoping review. *J Am Med Inform Assoc.* 2025;32(5):932-944. doi:10.1093/jamia/ocaf017.
9. Wilson S, Tolley C, Mc Ardle R, et al. Recommendations to advance digital health equity: a systematic review of qualitative studies. *npj Digit Med.* 2024;7(1):173. doi:10.1038/s41746-024-01177-7.
10. World Health Organization. Global strategy on digital health 2020-2027. Geneva: World Health Organization; 2025. <https://www.who.int/publications/i/item/9789240116870>. Accessed 21 Apr 2026.
11. Bommakanti KK, Smith LL, Liu L, Do D, Cuevas-Mota J, Collins K, et al. Requiring smartphone ownership for mHealth interventions: who could be left out? *BMC Public Health.* 2020;20:81. doi:10.1186/s12889-019-7892-9.
12. Pettersson L, Johansson S, Demmelmaier I, Gustavsson C. Disability digital divide: survey of accessibility of eHealth services as perceived by people with and without impairment. *BMC Public Health.* 2023;23(1):181. doi:10.1186/s12889-023-15094-z.
13. Sundell E, Wångdahl J, Grauman Å. Health literacy and digital health information-seeking behavior: a cross-sectional study among highly educated Swedes. *BMC Public Health.* 2022;22:2278. doi:10.1186/s12889-022-14751-z.
14. Sun C, Meijer E, Chavannes NH, Dai H, Li X, Wang Y, et al. eHealth literacy in the general population: a cross-sectional study in China. *BMC Public Health.* 2025;25:211. doi:10.1186/s12889-025-21389-0.
15. Barbati C, Maranesi E, Giammarchi C, Lenge M, Bonciani M, Barbi E, et al. Effectiveness of eHealth literacy interventions: a systematic review and meta-analysis of experimental studies. *BMC Public Health.* 2025;25:288. doi:10.1186/s12889-025-21354-x.
16. Cui Y, He Y, Xu X, Zhou L, Nutakor JA, Zhao L. Cultural capital, the digital divide, and the health of older adults: a moderated mediation effect test. *BMC Public Health.* 2024;24:302. doi:10.1186/s12889-024-17831-4.
17. Xu Z, Zhou L, Gu W, Yang Z, Zhang L. Telehealth, medical decisions and new health inequality in China. *BMC Public Health.* 2025;25:793. doi:10.1186/s12889-025-22039-1.
18. Xylander AAP, Cichosz SL, Hejlesen O, Witt Udsen F. Telemedicine as a tool for bridging geographical inequity: insights in geospatial interactions from a study on chronic heart failure patients. *BMC*

- Public Health. 2024;24:2953. doi:10.1186/s12889-024-20438-4.
19. Katz ME, Mszar R, Grimshaw AA, et al. Digital health interventions for hypertension management in US populations experiencing health disparities: a systematic review and meta-analysis. *JAMA Netw Open*. 2024;7(2):e2356070. doi:10.1001/jamanetworkopen.2023.56070.
 20. Azizi Z, Adedinsewo D, Rodriguez F, et al. Leveraging digital health to improve the cardiovascular health of women. *Curr Cardiovasc Risk Rep*. 2023;17(11):205-214. doi:10.1007/s12170-023-00728-z.
 21. Mihan A, Van Spall HGC. Interventions to enhance digital health equity in cardiovascular care. *Nat Med*. 2024;30(3):628-630. doi:10.1038/s41591-024-02815-z.
 22. Ogungbe O, Longenecker CT, Beaton A, et al. Advancing cardiovascular health equity globally through digital technologies. *J Am Heart Assoc*. 2024;13(2):e031237. doi:10.1161/JAHA.123.031237.
 23. Johnson T, Chilazi M, Isakadze N, et al. Bridging the digital divide: applying digital tools to improve cardiovascular health in underrepresented communities. *JACC Adv*. 2023;2(7):100587. doi:10.1016/j.jacadv.2023.100587.
 24. Zhao D, Zhang Y, Wang J, He L, He Y, Ma C. Cardiovascular disease prevention in China: challenges and opportunities in the artificial intelligence-enabled digital health era. *Nat Rev Cardiol*. 2026;23:363-374. doi:10.1038/s41569-025-01222-2.
 25. Liu X, Chen W, Qiu Y, Li X, Liu F, Jiang Z, et al. Improving access to cardiovascular care for 1.4 billion people in China using telehealth. *npj Digit Med*. 2024;7(1):376. doi:10.1038/s41746-024-01381-5.
 26. Li C, Liu Z, Shi R. A bibliometric analysis of 14,822 researches on myocardial reperfusion injury by machine learning. *Int J Environ Res Public Health*. 2021;18(15):8231. doi:10.3390/ijerph18158231.
 27. Shiferaw KB, Wali P, Waltemath D, Zeleke AA. Navigating the AI frontiers in cardiovascular research: a bibliometric exploration and topic modeling. *Front Cardiovasc Med*. 2024;10:1308668. doi:10.3389/fcvm.2023.1308668.
 28. Damar M, Pinto AD, Hosseini B, et al. Trends in health literacy discussions within primary health care research: a topic analysis using machine learning techniques. *Aten Primaria*. 2026;58(3):103441. doi:10.1016/j.aprim.2025.103441.
 29. Ma L, Chen R, Ge W, Rogers P, Lyn-Cook B, Hong H, et al. AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. *Exp Biol Med (Maywood)*. 2025;250:10389. doi:10.3389/ebm.2025.10389.
 30. Cheng X, Cao Q, Liao SS. An overview of literature on COVID-19, MERS and SARS: using text mining and latent Dirichlet allocation. *J Inf Sci*. 2022;48(3):304-320. doi:10.1177/0165551520954674.
 31. Min KB, Song SH, Min JY. Topic modeling of social networking service data on occupational accidents in Korea: latent Dirichlet allocation analysis. *J Med Internet Res*. 2020;22(8):e19222. doi:10.2196/19222.
 32. Chen J, Williams M, Huang Y, Si S. Identifying topics and evolutionary trends of literature on brain metastases using latent Dirichlet allocation. *Front Mol Biosci*. 2022;9:858577.

doi:10.3389/fmolb.2022.858577.

33. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. *J Mach Learn Res.* 2003;3:993-1022. doi:10.1162/jmlr.2003.3.4-5.993.
34. Meaney C, Stukel TA, Austin PC, Moineddin R, Greiver M, Escobar M. Quality indices for topic model selection and evaluation: a literature review and case study. *BMC Med Inform Decis Mak.* 2023;23:132. doi:10.1186/s12911-023-02216-1.
35. Sharrief AZ, Guzik AK, Jones E, Okpala M, Love MF, Ranasinghe TIJ, et al. Telehealth trials to address health equity in stroke survivors. *Stroke.* 2023;54(2):396-406. doi:10.1161/STROKEAHA.122.039566.
36. Ain HQ, Tahir MJ, Ahmed KAHM, Ahmed F, Ali MMI, Elhaj EHS, et al. Tele-stroke: a strategy to improve acute stroke care in low- and middle-income countries. *Ann Med Surg (Lond).* 2024;86(7):3808-3811. doi:10.1097/MS9.0000000000002187.
37. Towett G, Snead RS, Marczika J, Ambalavanan R, Kairichi MM, Malioukis A. Bridging the gap in digital health: a framework for leveraging digital health technologies in cardiovascular diseases, hypertension, and diabetes—a narrative review. *Digit Health.* 2025;11:20552076251406654. doi:10.1177/20552076251406654.
38. Eberly LA, Khatana SAM, Nathan AS, Snider C, Julien HM, Deleener ME, et al. Telemedicine outpatient cardiovascular care during the COVID-19 pandemic: bridging or opening the digital divide? *Circulation.* 2020;142(5):510-512. doi:10.1161/CIRCULATIONAHA.120.048185.
39. Battineni G, Sagaro GG, Chintalapudi N, Amenta F. The benefits of telemedicine in personalized prevention of cardiovascular diseases (CVD): a systematic review. *J Pers Med.* 2021;11(7):658. doi:10.3390/jpm11070658.
40. Hon JJ, Carr-White G. The digital divide in cardiovascular care: who gets left behind? *Eur Heart J Digit Health.* 2026;7(2):ztag011. doi:10.1093/ehjdh/ztag011.
41. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. *J Med Internet Res.* 2021;23(2):e26145. doi:10.2196/26145.
42. Rodríguez Parrado IY, Achury Saldaña DM. Digital health literacy in patients with heart failure in times of pandemic. *Comput Inform Nurs.* 2022;40(11):754-762. doi:10.1097/CIN.0000000000000883.
43. Zhao Y, Wood EP, Mirin N, Cook SH, Chunara R. Social determinants in machine learning cardiovascular disease prediction models: a systematic review. *Am J Prev Med.* 2021;61(4):596-605. doi:10.1016/j.amepre.2021.04.016.
44. Teshale AB, Htun HL, Vered M, Owen AJ, Freak-Poli R. A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction. *J Med Syst.* 2024;48(1):68. doi:10.1007/s10916-024-02087-7.
45. Liu T, Krentz A, Lu L, Curcin V. Machine learning based prediction models for cardiovascular disease risk using electronic health records data: systematic review and meta-analysis. *Eur Heart J Digit Health.* 2025;6(1):7-22. doi:10.1093/ehjdh/ztae080.

46. Nair P, Dai M, Patel DA, Pearlman K, Shah SD. Equity by design: using digital technology to overcome cardiovascular health disparities. *Curr Cardiol Rep.* 2025;27(1):157. doi:10.1007/s11886-025-02319-3.
47. Kim DS, Eltahir AA, Ngo S, Rodriguez F. Bridging the gap: how accounting for social determinants of health can improve digital health equity in cardiovascular medicine. *Curr Atheroscler Rep.* 2025;27(1):9. doi:10.1007/s11883-024-01249-9.
48. Sun F, Yao J, Du S, Qian F, Appleton AA, Tao C, et al. Social determinants, cardiovascular disease, and health care cost: a nationwide study in the United States using machine learning. *J Am Heart Assoc.* 2023;12(5):e027919. doi:10.1161/JAHA.122.027919.

Figures

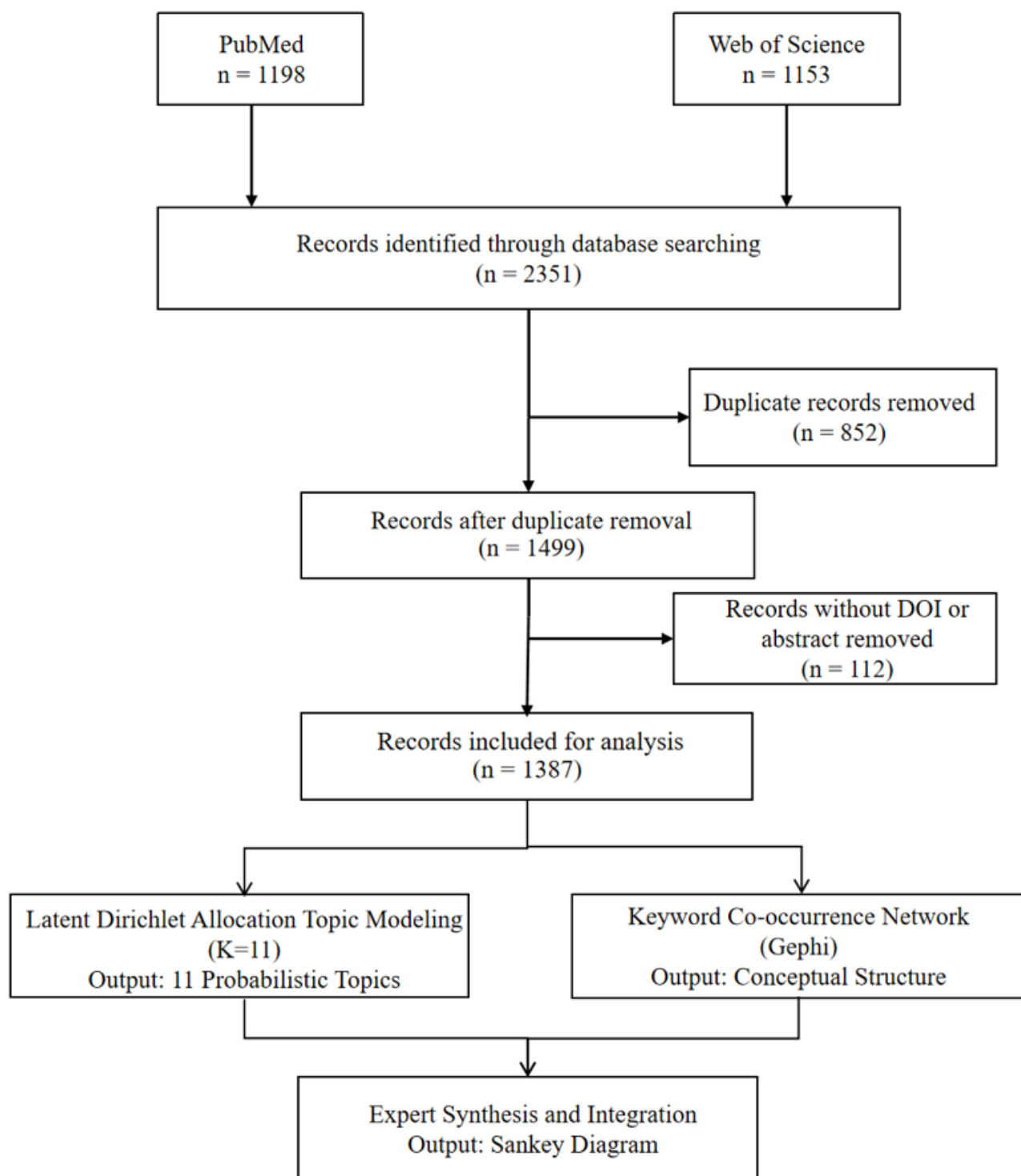


Figure 1

Flowchart of study inclusion.

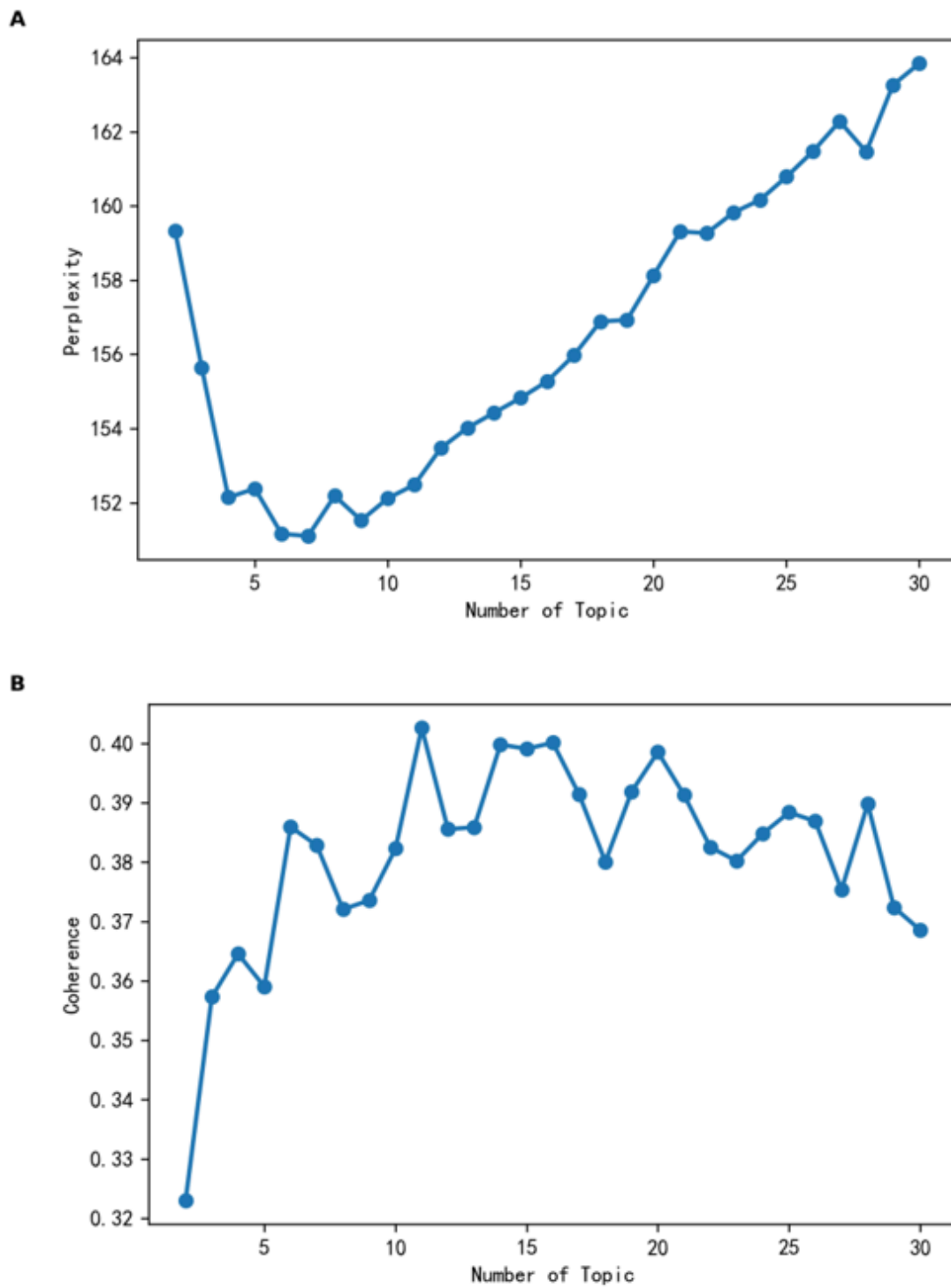


Figure 2

Line plots of topic coherence and perplexity for the LDA models with different numbers of topics.

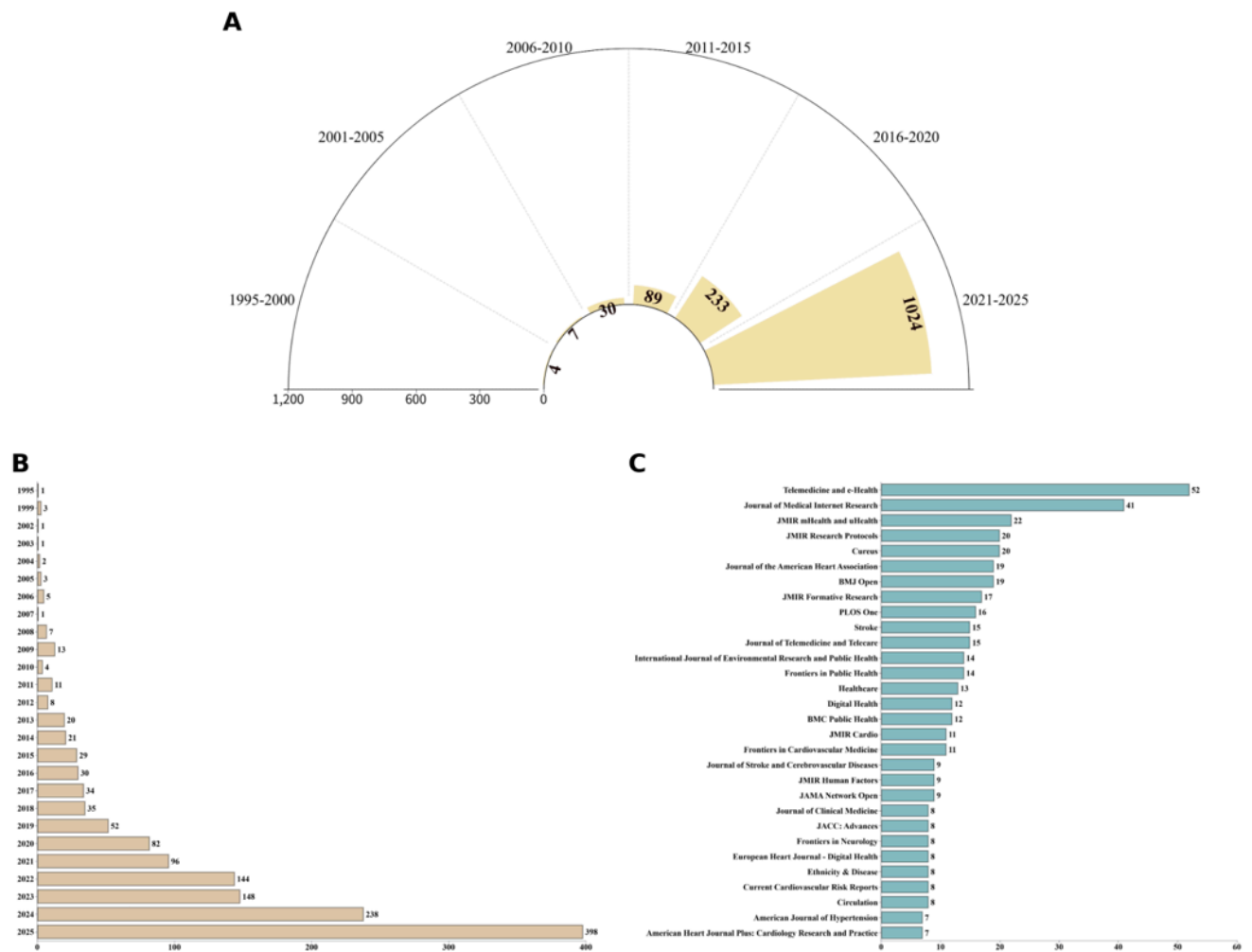


Figure 3

Number of articles by publication year and source journal.

Note: The left side shows the 11 first-level topics identified by LDA (T1–T11), and the right side shows the four higher-order thematic clusters derived from further synthesis (C1–C4). The links indicate the assignment of first-level topics to higher-order thematic clusters, and link width reflects the number of first-level topics contained in each higher-order cluster.