

Review

Big Data–Driven Health Portraits for Personalized Management in Noncommunicable Diseases: Scoping Review

Haoyang Du^{1,2}, PhD; Jianing Yu^{2,3}, MD; Dandan Chen^{1,2}, PhD; Jingjie Wu⁴, PhD; Erxu Xue^{1,2}, PhD; Yufeng Zhou^{1,2}, PhD; Xiaohua Pan⁵, PhD; Jing Shao^{2,3}, PhD; Zhihong Ye^{1,6}, Prof Dr Med

¹Sir Run Run Shaw Hospital, Hangzhou, China

²School of Nursing and Institute of Nursing Research, School of Medicine, Zhejiang University, Hangzhou, China

³Department of Nursing, The Fourth Affiliated Hospital of School of Medicine, and International School of Medicine, International Institutes of Medicine, Zhejiang University, Yiwu, China

⁴Ningbo University Affiliated Hospital, Ningbo, China

⁵Binjiang Research Institute of Zhejiang University School, Zhejiang University, Hangzhou, China

⁶The D H Chen School of Universal Health, Zhejiang University, Hangzhou, China

Corresponding Author:

Zhihong Ye, Prof Dr Med

Sir Run Run Shaw Hospital

3# East Qingchun Road, Shangcheng district

Hangzhou, 310020

China

Phone: 86 13606612119

Email: 3192005@zju.edu.cn

Abstract

Background: Health portraits powered by big data integrate diverse health-related data into actionable insights, thereby facilitating precise risk prediction and personalized management of noncommunicable diseases (NCDs). Despite their promise, the adoption and application of health portraits remain fragmented, primarily due to the lack of a standardized conceptual and methodological framework necessary to fully harness their capabilities.

Objective: This study aimed to systematically map and categorize existing research on health portraits in the context of NCD management, evaluate how big data has been used through the lens of the 3V (volume, velocity, and variety) framework, assess the extent of external validation and comprehensiveness, and identify challenges, emerging opportunities, and future research directions in this field.

Methods: A scoping review was conducted following the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines and 6-step framework of Levac et al. A comprehensive search was performed in PubMed, Embase, EBSCO, Ovid, Scopus, Web of Science, and Springer Link, focusing on observational and interventional studies using big data, public databases, electronic health record systems, wearables, and sensors for NCD management from January 2014 to July 2024. Data extraction included study characteristics, modeling approaches, and external validation. Analytical synthesis was conducted using keyword analysis, the 3V framework, and visual tools such as scatter plots, heat maps, and radar charts.

Results: A total of 8707 records were identified, and 89 studies were included for full-text analysis. These studies were categorized into 4 types of health portraits: diagnostic, prognostic, monitoring, and recommender. Evaluation based on the 3V framework showed that only 17.78% of studies met all 3 criteria. In terms of volume, structured data were widely used (64.29%-100% depending on portrait type), while unstructured data usage varied significantly (19.05%-93.33%). Regarding velocity, monitoring and recommender portraits showed high reliance on digital interactive data (over 85%). For variety, only 31.11% of studies incorporated all 3 data attributes (natural, domain, and specific attributes). In terms of comprehensiveness, only 30% of studies reported the external validation, and only 10% met both the external validation and 3V criteria, with recommender portraits outperforming the other types.

Conclusions: This study provides a standardized lens through which to evaluate the development and application of health portraits in NCD management. The findings underscore the need for more robust data integration strategies and emphasize the

importance of artificial intelligence-enabled approaches. Furthermore, enhancing external validation and addressing ethical and privacy considerations are critical for advancing the implementation of personalized health management solutions.

(*J Med Internet Res* 2025;27:e72636) doi: [10.2196/72636](https://doi.org/10.2196/72636)

KEYWORDS

health portraits; big data; non-communicable disease; NCDs; management

Introduction

Noncommunicable diseases (NCDs) account for 77% of the projected global disease burden by 2030, claiming 17 million premature annual deaths and imposing unsustainable pressure on health care systems [1-4]. This threat has accelerated a paradigm shift from reactive treatment to proactive health management [5-7], underscoring the need for integrated care across clinical and community settings [1,8-10] with the emerging emphasis on precision medicine approaches [11,12]. Fueling this transition, digital health technologies—from electronic health records (EHRs) to wearables and remote sensing devices—are generating vast multidimensional data streams, offering unprecedented potential for individualized data-driven interventions and resource optimization [13,14].

Within this data-rich landscape, “health portraits” have emerged as a novel framework to organize and interpret complex health data for precision health management [15]. Originally inspired by Alan Cooper’s “User Personas” concept in 1998 and later adapted to health care by Liu et al [15], Cooper et al [16], and Pietilä et al [17], health portraits refer to an integrated, person-centered representation that synthesizes heterogeneous data, including symptoms, medical history, biochemical tests, and lifestyle factors, into a unified profile of an individual’s health status. Previous studies have shown that health portraits are typically composed of 3 layers of attributes: natural (eg, age and gender), domain (eg, behavior and preferences), and specific attributes (eg, physiological health and disease risk) [15]. Operationally, their construction follows a multistage pipeline: data acquisition, reprocessing, label mapping, and contextualization [18], enabling the translation of raw data into meaningful insights for personalized health management [19,20] in risk stratification, telehealth delivery, and personalized interventions [21-27].

In recent years, a growing number of health portraits have leveraged big data resources, including large-scale EHRs, patients’ self-reported data, and wearable outputs, driven by the defining features of big data: volume, velocity, and variety (the “3V” framework) [28-30]. High-profile examples include the All of Us Research Program in the United States, which integrates genomic, EHR, and wearable data from over 1 million participants to support inclusive biomedical research and personalized health insights. Similarly, IBM Watson Health applies artificial intelligence (AI)-driven analysis of multimodal data (eg, EHRs, genomics) to optimize oncology decision-making, while Oura Ring generates continuous physiological health portraits for disease risk prediction through real-time tracking of heart rate variability, sleep, and activity [31-34]. These initiatives collectively demonstrate the transformative potential of big data-driven health portraits in

advancing real-time, personalized, and context-aware NCD care across clinical and everyday settings [35-37].

While numerous studies have explored the construction and application of health portraits, there has been limited effort to systematically map and categorize their functions or application contexts within NCD management. Some researchers have noted that insufficient processing of data volume, poor integration of diverse data types, and inadequate support for real-time updates continue to constrain the scalability and fidelity of health portraits [38,39]. Likewise, an empirical study found that over 50% of code modules for health analytics failed peer review, largely due to test instability and weak validation [40]. Moreover, the mismatch between datasets and scenarios, such as using inpatient clinical data in community settings, further limits the contextual relevance and broader applicability of health portraits [41].

Despite the growing body of work, no systematic reviews to date have classified the functional types and implementation scenarios of health portraits, assessed their alignment with the big-data “3V” framework (volume, velocity, and variety), or examined their external validation status. To support the theoretical advancement and practical application of big data-driven health portraits, this scoping review aims to (1) categorize health portraits’ functionalities and application scenarios in NCD care, (2) evaluate 3V alignment and external validation status, and (3) propose an implementation roadmap addressing current limitations.

Methods

We performed a scoping review based on the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) statement [42] as shown in [Multimedia Appendix 1](#). We followed the 6-step framework outlined by Levac et al [43], who updated and extended the initial framework developed by Arksey and O’Malley [44].

Search Strategy

We searched PubMed, Embase, EBSCO, Ovid, Scopus, Web of Science, and Springer Link. Search terms included terms describing the concept (health portrait) in combination with population (NCDs) and context (big data). We also included related terms such as “User-centered Design” and “Telemedicine” as search terms because they may be related to the concept or context. The search was restricted to articles published in English or Chinese from January 2014 to July 2024. Furthermore, we scanned reference lists of included publications and published reviews for additional articles. The full search queries can be found in [Multimedia Appendix 2](#).

Selection Criteria

Population is limited to patients diagnosed with NCDs defined by the World Health Organization [45]. Moreover, we also included studies on a broader range of NCDs where relevant, like obesity, mental disorders, and geriatric syndromes, to ensure a comprehensive review of health portrait types and the methodologies used [46,47].

Concept (health portrait) is defined in this study as the models or programs to stratify or classify patients under a certain characteristic or condition based on characteristic information and process data labels related to health status [15], for example, specific social determinants, health risk scores, and lifestyle behaviors.

Context (big data) is defined as the extent to which the study’s modeling process aligns with the “3Vs” framework—volume,

velocity, and variety. Specifically, a study is considered to meet the criteria if it satisfies all of the following: (1) the inclusion of unstructured data; (2) the use of data sources from digital interactive platforms; and (3) the incorporation of patient data covering natural, domain, and specific attributes.

The types of publications included observational studies and interventional studies. Besides, we excluded studies that used only traditional data collection methods (interviews or scales) for data collection. Studies that used public databases, EHR systems, digital application platforms, and wearables and sensors for data collection are included. Detailed explanations about the concept (health portrait), population (NCDs), and context (big data) can be found in Multimedia Appendix 3. The study screening manual is shown in Table 1.

Table 1. Eligibility criteria and their rationale.

Eligibility criteria and variable		Rationale
Inclusion criteria		
Population	Patients diagnosed with NCDs ^a	NCDs are prevalent on a vast scale and impose a significant burden on health care management.
Concept	The health portrait models or programs to stratify or classify the health status of patients into digital labels	It can present a more digital and clearer reference method for personalized health management.
Context	Research data derived from public databases, electronic health care record systems, digital application platforms, and wearables and sensors	Public databases and EHRs ^b are the most common ways for data collection, and digital application platforms and wearables and sensors provide multimodal datasets, and the data transfer speed is fast.
Others	Peer-reviewed	It has greater credibility because the papers have been reviewed by peer experts in the field.
Others	Empirical study design	Empirical studies improve the ability to answer the research questions compared with conceptual commentaries or viewpoints.
Others	Published between January 2014 and July 2024	It was not after 2014 when big data and the Internet of Things (IoT) were relatively widespread and used in more studies.
Others	English or Chinese language	It is about some practical considerations, given the investigators’ language proficiency.
Exclusion criteria		
Population	Research on patients with infectious diseases	The prognosis, influencing factors and profiling requirements of infectious diseases are quite different from those of NCDs.
Concept	Studies not related to health portrait modeling or population stratification	Studies not relative to our research theme or key research questions can hardly help to identify the scope of health portraits.
Context	Research data only derived from traditional questionnaires or qualitative studies	Based on a single cross-sectional study design, the data collection uses traditional scales or interviews to obtain information, which has great limitations in terms of data volume, velocity, and variety.
Others	Editorials, commentaries, opinion articles, and reports	These types of sources are not based on original research or data analysis, which does not ensure a focus on evidence-driven insights.

^aNCD: noncommunicable disease.

^bEHR: electronic health record.

Study Selection

To increase the consistency of study screening among reviewers, reviewer 1 piloted the study screening manual for database search and study selection based on title and abstract information

available in the databases. Subsequently, reviewer 2 independently cross-checked the study selection of all articles identified in the database search. Both reviewers then discussed results and amended the screening manual before the data charting step. Subsequently, both reviewers 1 and 2

independently piloted the study screening manual for evaluating the eligibility of 10% of all identified full-text reports using a computer-generated random sequence, along with complete data charting for the included articles. Both reviewers then discussed the results and amended the screening manual. Finally, reviewers 1 and 2 independently completed an assessment of the remaining full-text reports for eligibility, along with data charting for all included reports.

Quality Assessment

This review included both observational studies and interventional studies. To reconcile the scope review's mandate for breadth with methodological accountability, we used a dual-tiered screening framework according to some previous research [48,49]: (1) prioritizing journals to leverage rigorous peer-review standards and (2) inclusion of methodologically innovative studies despite their limited overall quality, following thorough internal team deliberations to ensure balanced inclusion. While formal quality appraisal tools were intentionally omitted to preserve conceptual mapping flexibility—an accepted scoping review limitation—we embedded critical safeguards: systematic exclusion of studies with high bias risks (randomization, blinding, and allocation concealment) and study type-specific quality screening aligned with [Multimedia Appendix 4](#). This included external validation thresholds for clinical predictive models, health recommender systems, and interventional studies.

Data Extraction and Synthesis

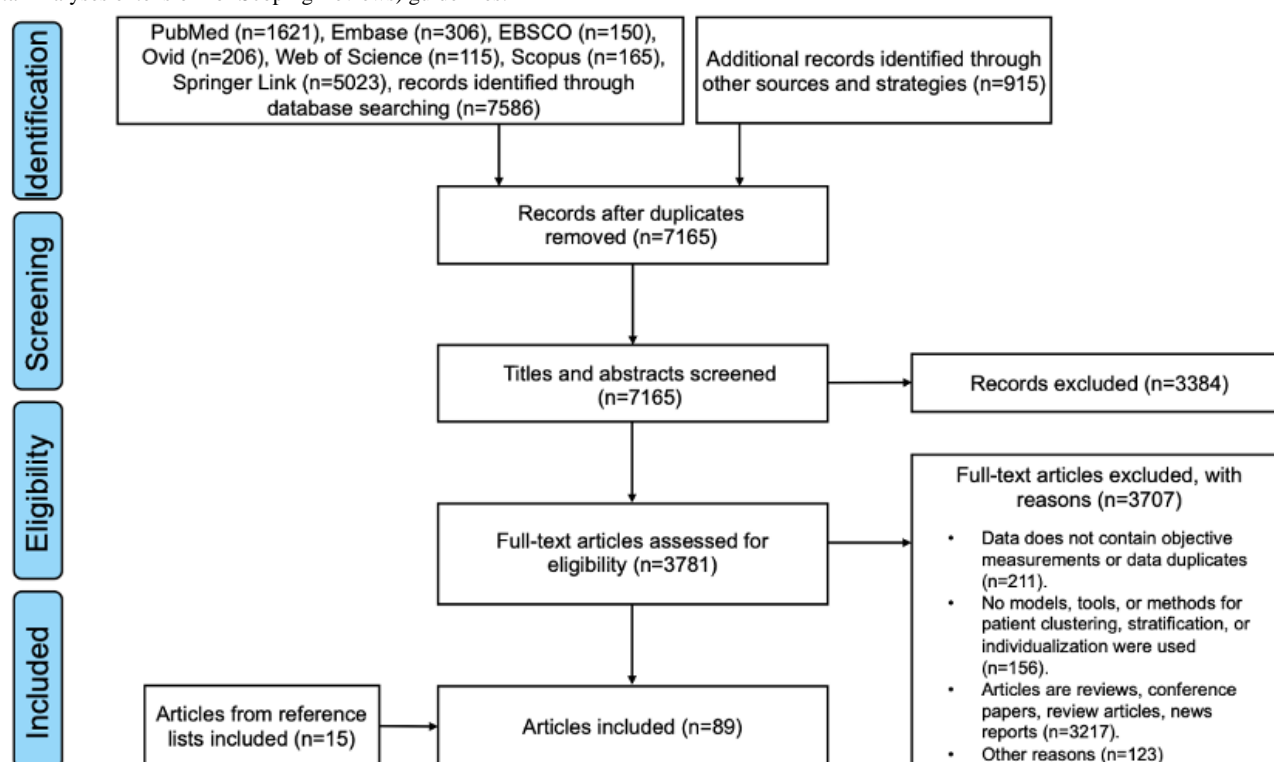
Reviewer 1 extracted basic information from the included articles provided by the first author: published year, country, study design, participants, modeling approaches, external validation status, sample size, and data usage. We group studies

by the scenes, participants, objectives, and modeling approaches of the health portraits, using Python-extracted keywords in the titles and abstracts, and reviewers conduct secondary integration of keywords. An expert panel and the research team determined the 3V framework based on the concept of big data and the comprehensive capability assessment ([Multimedia Appendix 5](#)). All findings are synthesized to identify existing status and knowledge gaps for future research. Key results are summarized using the word cloud analysis to identify different types of health portraits, the heat map to identify the existing status of health portraits that meet the requirements of “3V,” as well as the spider diagrams to perform the comprehensive capability of included records meeting the requirements of both the external validation and “3V.” Any disagreements on study selection and data charting during pilot testing were resolved by consensus, or otherwise with a tiebreaker by reviewer 3 if needed.

Results

A total of 7888 articles were searched in 7 databases for the initial and secondary searches, and 915 articles were manually searched. Of the 8707 records, after the removal of duplicates, 7165 (73.8%) records remained, and we screened titles and abstracts. In this screening, 47.2% (3384/7165) of the records were excluded, and the remaining 52.8% (3781/7165) were assessed for eligibility through full-text review. In addition, the reference lists of these articles were screened, which led to a further 23 articles being identified. After full-text articles were assessed for eligibility, 89 reports were included in this scoping review for data charting and analysis (screening flow diagram in [Figure 1](#)). The basic information of the included articles is described as shown in [Multimedia Appendix 6](#).

Figure 1. Inclusion flowchart. The 4 phases of article selection follow the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines.



Word Frequency Analysis Summarizes the Types and Application Scenarios of Big Data–Driven Health Portraits

The scatter plot (Figure 2) highlights key terms such as “management,” “medical,” “clinical,” “community,” “life,” “home,” and “trajectory,” reflecting the diverse application of health portraits in inpatient and outpatient settings for managing NCDs. Participants in these systems include medical professionals, patients, caregivers, and virtual health coaches, as indicated by terms like “patients,” “experts,” “caregiver,” “coach,” “chatbot,” “nurse,” and “self.” This underscores a collaborative approach spanning hospital care, community support, and patient self-management.

“dataset” indicate that health portrait modeling integrates data from diverse sources, including Internet of Things devices, wearables, digital platforms, EHRs, public datasets, and large language models (LLMs). This rich data foundation enables the extraction of relevant features and the creation of digital labels for comprehensive, individualized health portraits.

Through word frequency analysis and content summarization, this study identified 4 primary types of health portraits—diagnostic, prognostic, monitoring, and recommender—which collectively cover the health management continuum for NCDs. The different types of health portraits and their descriptions are shown in [Table 2](#).

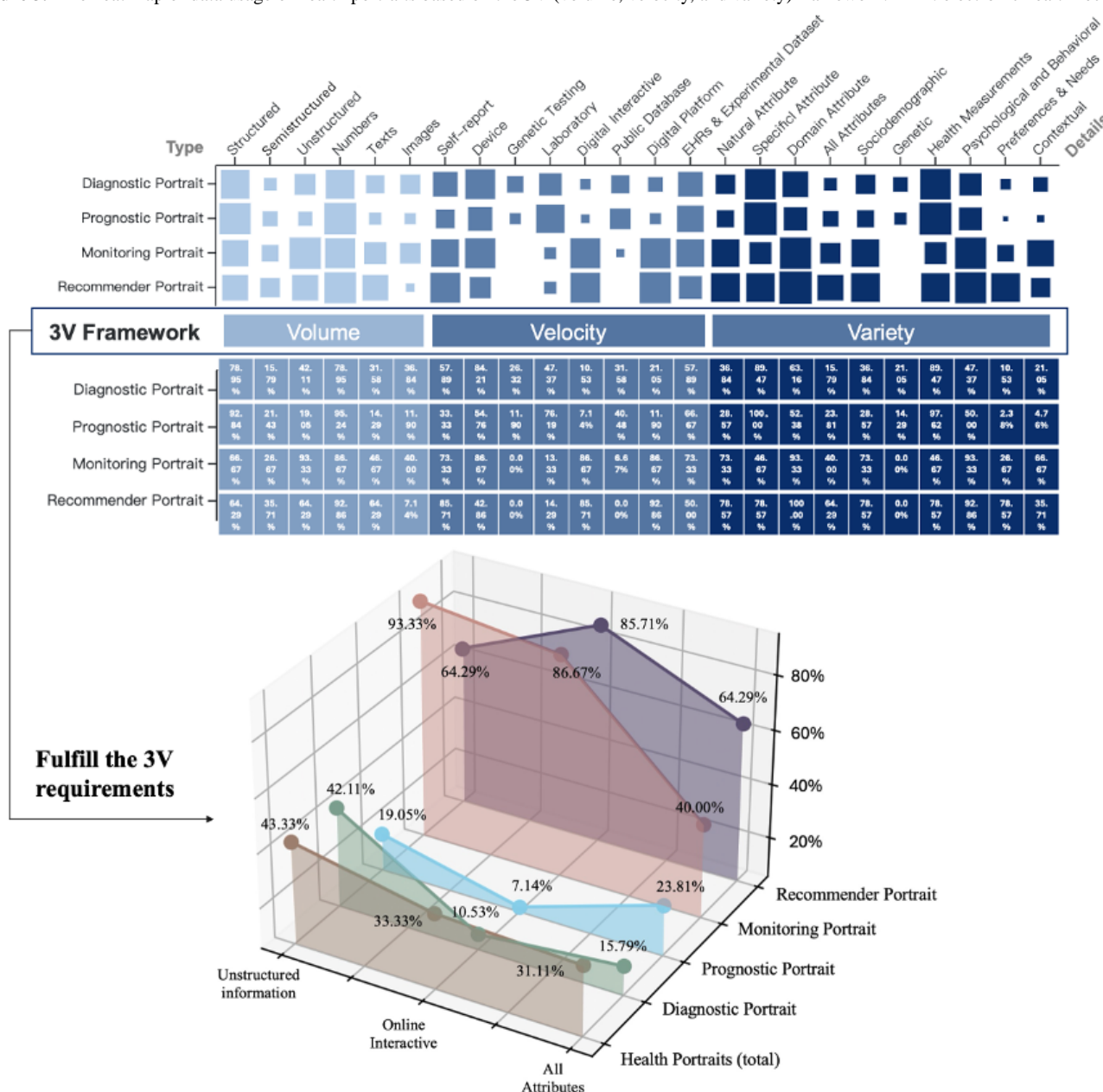
Table 2. Summary of different types of health portraits with management scenes based on full texts.

Type	Description	Management scene	Example
Diagnostic portrait	They are derived from clinical symptoms, signs, and laboratory test results, creating the labels related to patient diagnosis and treatment, and they can help doctors make accurate diagnoses in the early stages of the disease so that timely treatment measures can be taken [53].	Preclinical diagnosis and treatment prediction and preventive risk factor management based on the hospital environment and the medical team	Data were obtained from 9 independent datasets. Using predictors such as demographics, medical history, medication use, neuropsychological assessments, and multimodal neuroimaging, an artificial intelligence model based on a transformer architecture was developed for accurate etiological diagnosis and risk labeling in patients with dementia [54].
Prognostic portrait	They are derived from database information or relevant datasets, primarily consisting of a series of baseline characteristics of the patient, such as age, gender, stage of disease, and biomarkers, creating labels that reflect the health risks in the future [55].	Identification of high-risk patients with poor prognosis and management of risk factors for disease progression based on the in-hospital environment and the medical team	Data were obtained from the Big Data Center at Taipei Veterans General Hospital (VGH). Using predictors such as estimated glomerular filtration rate, hemoglobin, urine protein-to-creatinine ratio, insulin use, β -blocker use, renin-angiotensin system inhibitor use, and hypertension, a machine learning-based predictive model was developed to generate end-stage renal disease risk labels for patients with chronic kidney disease among sepsis survivors [56].
Monitoring portrait	They are derived from health care project processes and outcomes through self-reporting, mobile sensors, or wearables, with key labels derived from various performance metrics and descriptive summaries of patients [57,58].	Collaborative health management based on community and home scenarios, with the purpose of remote monitoring and management efficacy, with the participation of the health management team and patients	Data were obtained from obese adolescents participating in a weight management program. By analyzing interactions with Tess, an artificial intelligence-based behavioral coaching chatbot, a model was developed using natural language processing and machine learning to monitor and support weight management and prediabetes symptoms through digital health labeling [59].
Recommender portrait	They are derived from interactive information on digital platforms, focusing on using intelligent algorithms to recommend personalized information, resources, or interventions relevant to the user's specific health needs [60].	Self-management based on community and home scenarios, centered on knowledge empowerment, and patient focus	Data were obtained from caregivers of patients with dementia recruited through social media advertisements. A knowledge graph-based dementia care intelligent recommendation system was developed using knowledge graph and intelligent recommendation system technologies to provide personalized care plan recommendations for patients with dementia and their caregivers [61].

Data Usage of Big Data–Driven Health Portraits Based on the 3V Framework

Data usage is shown in Figure 3. Overall, only 17.78% of the included studies fulfill the requirements of big data based on the 3V framework. In terms of volume, structured data is highly used across the 4 types, with the lowest rate still reaching 64.29% in recommender portraits. However, the usage of unstructured data (43.33%) shows substantial variation: monitoring portraits lead with 93.33%, while prognostic portraits are at the low end with just 19.05%. Diagnostic portraits, focused similarly on clinical prediction models, exhibit a moderate usage of unstructured data at 42.11%. Regarding velocity, the reliance on digital interactive data (33.33%) is polarized. Diagnostic and prognostic portraits show lower usage

rates, at 15.79% and 7.14%, respectively, whereas monitoring and recommender portraits both exceed 85%. For variety, only 31.11% of health portraits encompass all 3 attributes. However, recommender portraits demonstrate the highest performance, with 64.29% of studies encompassing all 3 attributes, among which domain attribute coverage is at 100%. Further, the domain attribute of recommender portraits primarily comes from preference-demand information, which constitutes 78.57% of its data—significantly higher than that in other portrait types. In contrast, prognostic portraits show the most limitations in the domain attribute. Although they excel in specialized medical data, reaching 100% in specific attribute coverage, they exhibit substantial gaps in preference-demand and contextual information, with usage rates of only 2.38% and 4.76%, respectively.

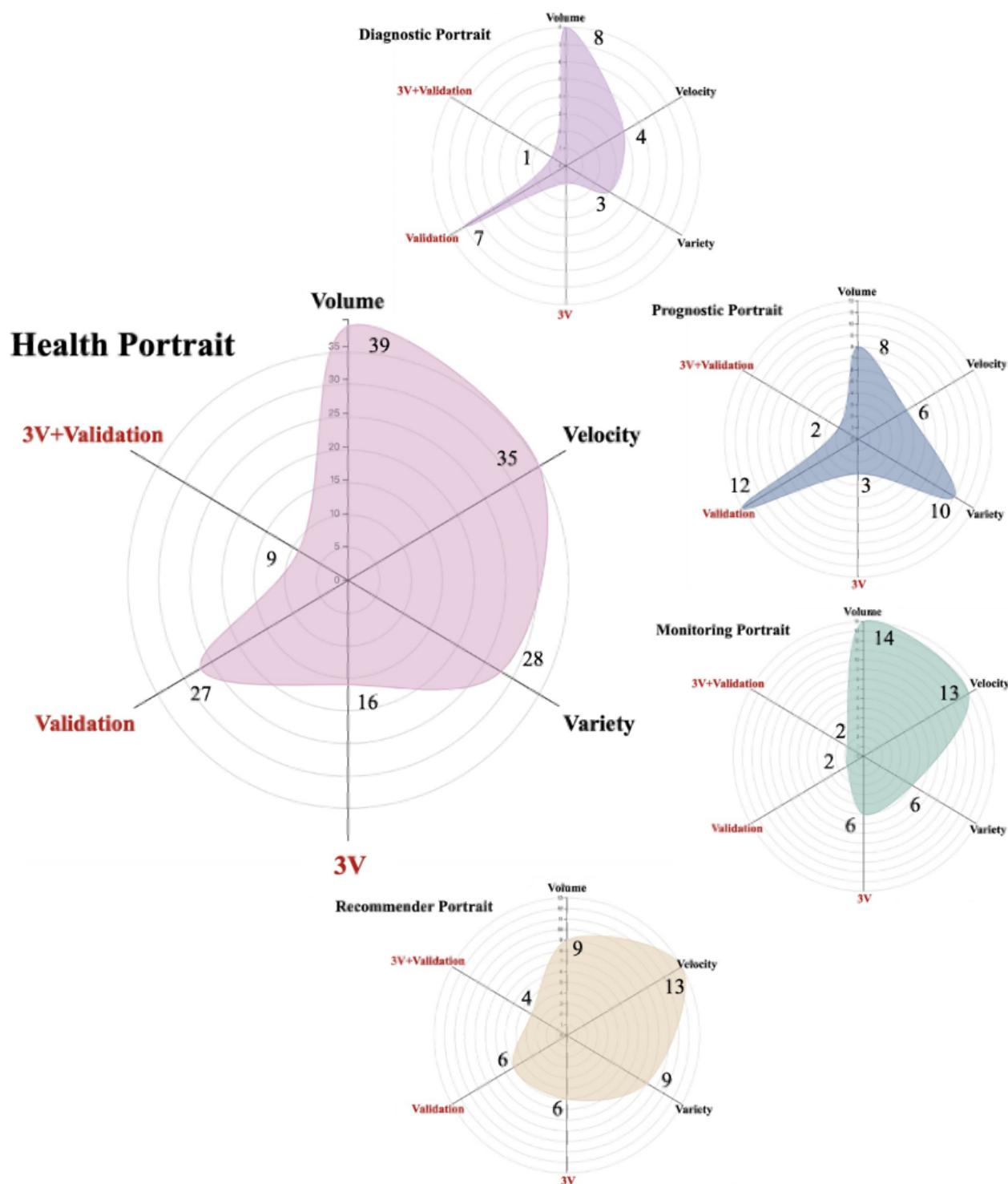
Figure 3. The heat map of data usage of health portraits based on the 3V (volume, velocity, and variety) framework. EHR: electronic health record.

Comprehensive Capability Assessment of Big Data–Driven Health Portraits Based on the 3V Framework and External Validation

A comprehensive capability assessment was conducted in Figure 4. Overall, the external validation of the big data–driven health portraits is suboptimal, with only 27 studies (30.34%) meeting the required standards. Assessed together with each study's

performance in meeting the 3V criteria of big data, only 9 (10.11%) studies meet these standards. Among the 4 types, recommender portraits showed relatively better results, with 4 (28.57%) studies meeting the comprehensive criteria. Finally, we conducted a comprehensive scoping review that synthesizes evidence from diverse literature on big data–driven health portraits, defining the comprehensive scope in Multimedia Appendix 7.

Figure 4. Comprehensive capability assessment radar charts that meet the needs of volume, velocity, and variety (3V)–based big data and external validation.



Discussion

Principal Results

Our findings indicate that big data–driven health portraits primarily cover the full spectrum of NCD management, categorized into 4 types, including diagnostic, prognostic, monitoring, and recommender portraits, each tailored to specific management scenarios. Despite the diversity of data sources for health portraits, a key limitation is the underuse of unstructured and interactive data, which restricts the potential

of big data–driven approaches. A quantitative evaluation based on the “3V” framework further demonstrates weak comprehensive capabilities of big data–driven health portraits, largely due to insufficient use of diverse data types. Moreover, the lack of external validation metrics hampers the assessment of the generalizability of health portraits, and there is considerable room for improvement in validation efforts. This review emphasizes the need for continued exploration to enhance the adaptability and effectiveness of big data–driven health portraits in NCD management. The following sections

address the key challenges, opportunities, and future roadmap, paving the way for innovative approaches tailored to specific management contexts.

The results showed that 4 primary types of health portraits span the care continuum from hospital-based to home-based settings. Specifically, diagnostic and prognostic portraits are predominantly deployed in clinical settings, where they leverage structured medical metrics and experimental data aligned with traditional health care models [62]. In contrast, monitoring and recommender portraits are increasingly applied in community and home-based contexts, supporting patient self-management, familial support, and interactions with health coaches [63,64]. Each type faces distinct data challenges. Hospital-based portraits benefit from high-quality data but lack continuity, while home-based systems provide continuous input via wearables and mobile devices, often with variable data quality [65]. Monitoring and recommender portraits, in particular, require diverse, high-quality training datasets and rigorous validation to ensure label accuracy and system reliability. Importantly, these types are not mutually exclusive—monitoring and recommender portraits offer real-time responsiveness, while diagnostic and prognostic portraits enhance predictive accuracy. Their integration presents opportunities for more comprehensive, adaptive health management. Emerging technologies, especially the integration of LLMs and knowledge graphs (KGs), are advancing this vision [14,66-70]. LLMs possess robust capabilities for processing unstructured and interactive data, such as chatbot dialogues [71], to extract real-time behavioral embeddings, while KGs structure multimodal biomarkers and social determinants into semantic networks and enable knowledge reasoning based on multimodal information. It must be recommended that this synergy will improve temporal resolution and contextual precision. Such architectures transcend traditional single-dimension health portraits, enabling adaptive health management across biological, behavioral, and environmental tiers [66,67,71]. Numerous experimental studies have shown that integrating LLMs with user historical interaction and behavioral information extracted from KGs can effectively perform relevant predictions and information recommendations. This method has exhibited superior effectiveness and reliability compared to traditional machine learning models across multiple public datasets [72-74]. However, significant challenges persist. Privacy and security concerns are particularly pressing, especially regarding the potential for data breaches during personal data interactions [71]. Addressing these concerns necessitates multifaceted approaches, including blockchain technology, end-to-end encryption, and advanced privacy protocols [75].

This study also reveals that while big data-driven health portraits demonstrate moderate performance across individual dimensions of the 3V framework, their overall effectiveness diminishes sharply when evaluated across all 3 dimensions simultaneously (16 (17.98%) studies). This suggests that health portraits are still in the early stages of development, where the comprehensiveness of data usage remains underdeveloped, which may constrain big data-driven health portraits in patent design or practical application. Upon closer inspection, there are considerable disparities in data use across the 4 types of

health portraits. Diagnostic and prognostic portraits show significant differences in the use of unstructured data (volume). For instance, diagnostic portraits depend heavily on unstructured data like high-dimensional imaging, which encapsulates valuable predictive information [76-78], aligning with the findings of Esteva et al [79]. However, the prognostic model mostly used data from public databases, which provided structured self-reported data, though research has reported that integrating unstructured clinical text with structured data can improve model performance [80]. The possible reason for the problem is that, unlike the relatively clear diagnostic conditions and the relatively low requirements for data continuity, the prediction of prognostic outcomes depends on a large number of cohort data before and after treatment, and it is relatively difficult to obtain data from a single study in the real world, especially under the limited patient adherence [81-84]. Our findings also show that monitoring and recommender portraits excel in data volume, transmission, and variety, which may underscore the potential of wearable and contactless devices in health data capture [85-88] as a cost-effective way for all types of portraits to bridge the gaps in acquisition capability of unstructured and interactive data [80]. Moreover, from a portrait integration perspective, as previously discussed, the integration of KGs and LLMs offers a robust framework for achieving multimodal data fusion and health portrait unification [66,67,71]. The maturation of hardware devices supports cross-modal data interoperability, while the development of time-series fusion algorithms transforms conventional 2D labeling systems into dynamic 3D architectures [89,90]. By incorporating temporal granularity, our approach refines population-level subgroup features into individualized trajectory-aware markers, capturing health progression patterns from clinical to community settings. This evolution facilitates the organic unification of the 4 portrait types, thereby enabling precision health management plans tailored to individual developmental trajectories.

Our findings indicate that only 10.11% (9 studies) of studies meeting the “3V” criteria have undergone the external validation. This highlights a significant gap between theoretical potential and practical implementation, underscoring the need for more external validations in the real world. Unlike Youssef et al [91], who cautioned against the reliability of single models validated on limited datasets, our study emphasizes the importance of integrating external validation within a comprehensive big data framework. External validation is essential for assessing model generalizability and performance, supported by prior research [92-94]. We also found that diagnostic and prognostic models emphasize externally validated results [94]. One possible reason for this is the strong association between external validation and study quality evaluation recommended by *The BMJ*'s guidelines according to the “Prediction Model Risk of Bias Assessment Tool” (PROBAST) [95]. Another possible reason is that the reliability of the prediction results in the real world is the core of the value of model research, aiming to promote results-oriented resource allocation according to accurate risk levels to reduce costs and increase efficiency. Otherwise, models that lack external validation are at high risk of failure in real-world practice. Even though genotype metrics are considered a key component of precision medicine, with advantages such as high

standardization, stability, and potential for causal inference, large-scale retrospective studies of predictive models based on genotype metrics have often failed [96]. This may be attributed to population heterogeneity and gene-environment interactions, which can diminish model performance in diverse populations in the real world [97]. Additionally, evaluations of 31 predictive models related to COVID-19 revealed that most studies had a high risk of bias [98]. Consequently, few are used or disseminated in clinical practice, thus failing to promote practical applications [99]. Monitoring and recommender portraits performed relatively poorly in external validation, probably because the guidance of these 2 types of models for health management was more process-oriented, such as patient intervention adherence or facilitating patients' resource usage in limited resources, so they focused more on data transmission speed and update frequency for the accuracy of external validation results in real-time monitoring [65,100]. Recent research has highlighted that intelligent predictive models extending beyond radiology and bioinformatics possess unique characteristics, which may need new evaluation tools such as PROBAST-AI, TRIPOD-AI (Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence) guidelines, and the criteria for health care conversation powered by AI [101,102]. Thus, future studies are recommended to use appropriate standards to complete external validation in real-world data to clarify the reliability of monitoring results and recommendations.

Strengths and Limitations

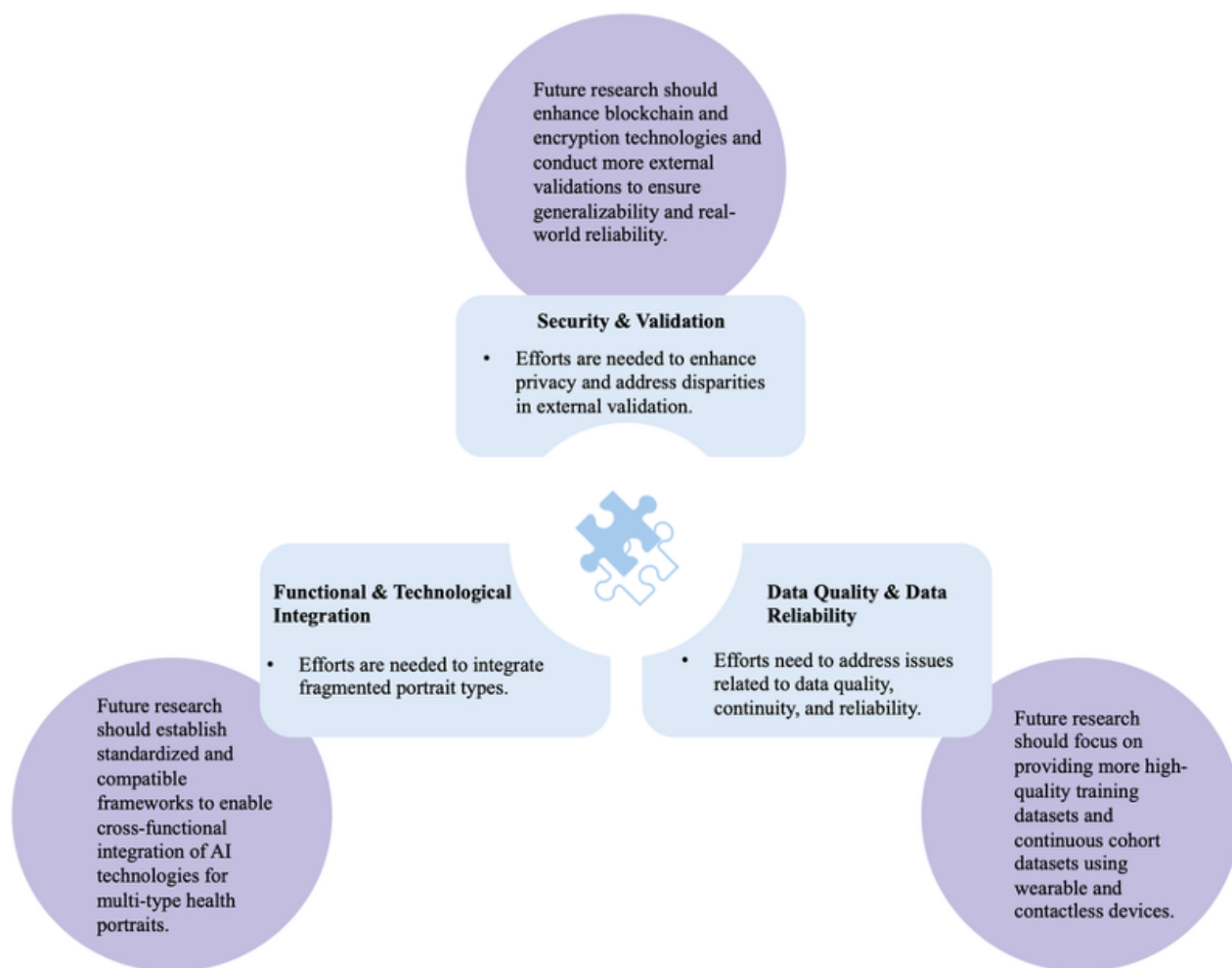
To the best of our knowledge, this study is the first to conduct a scoping review of health portraits from a big data perspective, addressing a critical research gap and laying a solid theoretical foundation for future investigations. One of the key contributions is the introduction of the 3V framework, which operationalizes big data principles to quantify data usage, offering a novel perspective for evaluating health portrait development. Another contribution is that the structured analysis delineates the current state of the field and provides actionable strategies for theoretical advancements and practical applications. Furthermore, compared to traditional health portraits, this scope of big data-driven

health portraits provides us with an opportunity to leverage multimodal data while grasping personal dynamic health status adaptively. Despite these contributions, several limitations must be acknowledged. First, restricting the review to English and Chinese studies may introduce language bias, potentially excluding valuable insights from other languages. Second, using a single binary classification metric for external validation limits the evaluation scope, making it difficult to pinpoint deficiencies in studies that do not meet benchmarks. Third, due to limitations in time and resources, this study lacks a more detailed analysis of health portrait features incorporating disease classifications, which will aid in developing a more comprehensive health portrait labeling system.

Future Work

This study underscores the challenges of big data-driven health portraits in NCD management in Figure 5, highlighting the possible future solutions that can fully realize the transformative potential. Among them, 2 main steps, according to our results, need to be highlighted: (1) prioritize LLM-driven integration of functionalities, contextual applications, and multisource data, while embedding ethical, privacy, and cost considerations at the design stage; and (2) promote more robust external validations according to appropriate standards. Notably, hyperpersonalization, which integrates multiomics data for tailored therapies, is highly recommended as a part of integrated health portraits [103-105]. With its convincing power in uncovering disease mechanisms, it brings the potential to integrate precision medicine with lifestyle nursing based on big data. Additionally, from a health equity perspective, future research could apply the PROGRESS-Plus (place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital) framework to health portraits, generating equity-based labels from natural attribute indicators to promote more equitable resource allocation and care for vulnerable populations [106]. Finally, building on our findings and previous research [18], we propose an integrated detailed roadmap for future practice as shown in Multimedia Appendix 8.

Figure 5. Challenges and future directions in big data–driven health portraits. A summary of the current challenges and possible future directions of big data–driven health portraits based on discussion and analysis of research results. AI: artificial intelligence.



Conclusions

Big data–driven health portraits offer significant potential to enhance personalized and precise management of NCDs. With the hardware advancements and cross-domain collaborations for holistic portrait design, chances are that the transformation of conventional 2D labeling systems into dynamic 3D

architectures of integrated big data–driven health portraits through AI-driven approaches will reveal new opportunities. However, future research should focus on the privacy-utility tradeoffs and ethical dilemmas. Addressing these gaps will advance the development of personalized health management solutions.

Acknowledgments

This review was supported by Technologies and Applications for Universal Access to Public Services for Shared Prosperity (2022YFF0902000), Double First-Class Construction Specialized Discipline Project at Zhejiang University (HL202406), China Postdoctoral Science Foundation funded project (2024M762877), and the National Social Science Fund of China (20BGL275).

Authors' Contributions

HYD was involved in the conceptualization of the study; data curation was done by HYD (lead) and JNY (equal); formal analysis was done by HYD; funding was acquired by ZHY (lead) and JS (equal); investigation was done by HYD (lead), YF Zhou (supporting); HYD is responsible for methodology; project administration was done by ZHY (lead) and JS (equal); XHP (lead) is responsible for resources; support was provided by EXX; supervision was undertaken by ZHY (lead); JS (equal) and JJW provided support; validation was done by JJW (lead) and XHP (equal); HYD involved in visualization; writing—original draft was done by HYD; writing—review and editing was done by HYD (lead); JNY and DDC provided support.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA-ScR checklist.

[\[DOCX File , 32 KB-Multimedia Appendix 1\]](#)

Multimedia Appendix 2

Database-specific search strategy.

[\[DOCX File , 23 KB-Multimedia Appendix 2\]](#)

Multimedia Appendix 3

The principle and explanation of “Population,” “Concept,” and “Context.”.

[\[DOCX File , 26 KB-Multimedia Appendix 3\]](#)

Multimedia Appendix 4

Criteria for external validation.

[\[DOCX File , 25 KB-Multimedia Appendix 4\]](#)

Multimedia Appendix 5

The 3V (volume, velocity, and variety) framework and the comprehensive capability assessment.

[\[DOCX File , 22 KB-Multimedia Appendix 5\]](#)

Multimedia Appendix 6

Basic information about the included articles. An overview of articles included in the scoping review (N=90).

[\[DOCX File , 165 KB-Multimedia Appendix 6\]](#)

Multimedia Appendix 7

The scope of big data–driven health portraits in the management of noncommunicable diseases.

[\[DOCX File , 284 KB-Multimedia Appendix 7\]](#)

Multimedia Appendix 8

The roadmap of big data–driven health portraits in the management of noncommunicable diseases.

[\[DOCX File , 419 KB-Multimedia Appendix 8\]](#)

References

1. Bhuiyan MA, Galdes N, Cuschieri S, Hu P. A comparative systematic review of risk factors, prevalence, and challenges contributing to non-communicable diseases in South Asia, Africa, and Caribbeans. *J Health Popul Nutr*. 2024;43(1):140. [\[FREE Full text\]](#) [doi: [10.1186/s41043-024-00607-2](#)] [Medline: [39252085](#)]
2. Noncommunicable Diseases Progress Monitor 2022. 1st ed. World Health Organization. URL: <https://www.who.int/publications/i/item/9789240047761> [accessed 2022-04-11]
3. Ben-Shlomo Y, Kuh D. A life course approach to chronic disease epidemiology: conceptual models, empirical challenges and interdisciplinary perspectives. *Int J Epidemiol*. 2002;31(2):285-293. [Medline: [11980781](#)]
4. Shu J, Jin W. Prioritizing non-communicable diseases in the post-pandemic era based on a comprehensive analysis of the GBD 2019 from 1990 to 2019. *Sci Rep*. 2023;13(1):13325. [\[FREE Full text\]](#) [doi: [10.1038/s41598-023-40595-7](#)] [Medline: [37587173](#)]
5. McPake B. The need for cost-effective and affordable responses for the global epidemic of non-communicable diseases. *Lancet Global Health*. 2019;7(10):e1293-e1294. [doi: [10.1016/s2214-109x\(19\)30375-4](#)]
6. Gottfredson LS. The transition to noncommunicable disease: how to reduce its unsustainable global burden by increasing cognitive access to health self-management. *J Intell*. 2021;9(4):61. [\[FREE Full text\]](#) [doi: [10.3390/jintelligence9040061](#)] [Medline: [34940383](#)]
7. Omotayo O, Maduka CP, Muonde M, Olorunsogo TO, Ogugua JO. The rise of non-communicable diseases: a global health review of challenges and prevention strategies. *Int Med Sci Res J*. 2024;4(1):74-88. [doi: [10.51594/imsrj.v4i1.738](#)]
8. Witter S, Zou G, Diaconu K, Senesi RGB, Idriss A, Walley J, et al. Opportunities and challenges for delivering non-communicable disease management and services in fragile and post-conflict settings: perceptions of policy-makers and health providers in Sierra Leone. *Confl Health*. 2020;14(1):3. [\[FREE Full text\]](#) [doi: [10.1186/s13031-019-0248-3](#)] [Medline: [31921333](#)]

9. Lin Y, Fu H. Challenges in improving non-communicable diseases management and achieving universal health coverage in China. *Lancet Reg Health West Pac*. 2024;44:101007. [FREE Full text] [doi: [10.1016/j.lanwpc.2024.101007](https://doi.org/10.1016/j.lanwpc.2024.101007)] [Medline: [38283261](https://pubmed.ncbi.nlm.nih.gov/38283261/)]
10. Budreviciute A, Damiati S, Sabir DK, Onder K, Schuller-Goetzburg P, Plakys G, et al. Management and prevention strategies for non-communicable diseases (NCDs) and their risk factors. *Front Public Health*. 2020;8:574111. [FREE Full text] [doi: [10.3389/fpubh.2020.574111](https://doi.org/10.3389/fpubh.2020.574111)] [Medline: [33324597](https://pubmed.ncbi.nlm.nih.gov/33324597/)]
11. Canfell OJ, Davidson K, Woods L, Sullivan C, Cocoros NM, Klompas M, et al. Precision public health for non-communicable diseases: an emerging strategic roadmap and multinational use cases. *Front Public Health*. 2022;10:854525. [FREE Full text] [doi: [10.3389/fpubh.2022.854525](https://doi.org/10.3389/fpubh.2022.854525)] [Medline: [35462850](https://pubmed.ncbi.nlm.nih.gov/35462850/)]
12. Alton DM. Population health management: from principal to practice. *J Integr Care*. 2023;31(4):417-423. [doi: [10.1108/jica-05-2023-0033](https://doi.org/10.1108/jica-05-2023-0033)]
13. Busnatu S, Niculescu AG, Bolocan A, Petrescu GED, Păduraru DN, Năstasă I, et al. Clinical applications of artificial intelligence-an updated overview. *J Clin Med*. 2022;11(8):2265. [FREE Full text] [doi: [10.3390/jcm11082265](https://doi.org/10.3390/jcm11082265)] [Medline: [35456357](https://pubmed.ncbi.nlm.nih.gov/35456357/)]
14. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. *Artif Intell Healthcare*. 2020:25-60.
15. Liu L, Zhu X, Li J. Application of health portraits in COPD personalized health management system. *Chin J Med Phys*. 2020;37(7):918-926.
16. Cooper A, Reimann R, Cronin D, Cooper A. *About Face 3: The Essentials of Interaction Design*. 3rd ed. United States. Wiley Pub; 2007.
17. Pietilä AM, Eirola R, Oikarinen K. [Conceptual system of health portrait]. *Hoitotiede*. 1998;10(2):78-86. [Medline: [10437438](https://pubmed.ncbi.nlm.nih.gov/10437438/)]
18. Long H, Zhu D. Research on intelligent portrait of Chinese elderly based on big data and deep learning. 2019. Presented at: IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCLOUD/SocialCom/SustainCom); 2019 December 18; Xiamen, China. [doi: [10.1109/ispa-bdcloud-sustaincom-socialcom48970.2019.00174](https://doi.org/10.1109/ispa-bdcloud-sustaincom-socialcom48970.2019.00174)]
19. Li R. A literature review of user portrait research. *Sci Technol Innov*. 2021;(23):12-19. [doi: [10.15913/j.cnki.kjycx.2021.23.002](https://doi.org/10.15913/j.cnki.kjycx.2021.23.002)]
20. Wen T, Liu X. A review of health portrait researches at home and abroad. *J Med Inform*. 2023;44(3):2-9.
21. Lee SM, Kim SH, Yoon H. Prediction of incident chronic kidney disease in a population with normal renal function and normo-proteinuria. *PLoS One*. 2023;18(5):e0285102. [FREE Full text] [doi: [10.1371/journal.pone.0285102](https://doi.org/10.1371/journal.pone.0285102)] [Medline: [37134104](https://pubmed.ncbi.nlm.nih.gov/37134104/)]
22. Kluckert J, Hötter AM, Da Muten R, Konukoglu E, Donati OF. AI-based automated evaluation of image quality and protocol tailoring in patients undergoing MRI for suspected prostate cancer. *Eur J Radiol*. 2024;177:111581. [FREE Full text] [doi: [10.1016/j.ejrad.2024.111581](https://doi.org/10.1016/j.ejrad.2024.111581)] [Medline: [38925042](https://pubmed.ncbi.nlm.nih.gov/38925042/)]
23. Tavakoli H, Chen W, Sin DD, FitzGerald JM, Sadatsafavi M. Predicting severe chronic obstructive pulmonary disease exacerbations. Developing a population surveillance approach with administrative data. *Ann Am Thorac Soc*. 2020;17(9):1069-1076. [doi: [10.1513/annalsats.202001-070oc](https://doi.org/10.1513/annalsats.202001-070oc)]
24. Surian NU, Batagov A, Wu A, Lai WB, Sun Y, Bee YM, et al. A digital twin model incorporating generalized metabolic fluxes to identify and predict chronic kidney disease in type 2 diabetes mellitus. *NPJ Digit Med*. 2024;7(1):140. [FREE Full text] [doi: [10.1038/s41746-024-01108-6](https://doi.org/10.1038/s41746-024-01108-6)] [Medline: [38789510](https://pubmed.ncbi.nlm.nih.gov/38789510/)]
25. da Rosa Tavares JE, Victória Barbosa JL. Apollo SignSound: an intelligent system applied to ubiquitous healthcare of deaf people. *J Reliable Intell Environ*. 2020;7(2):157-170. [doi: [10.1007/s40860-020-00119-w](https://doi.org/10.1007/s40860-020-00119-w)]
26. Deng Y, Tang Z, Yang Z, Chai Q, Lu W, Cai Y, et al. Comparing the effects of aquatic-based exercise and land-based exercise on balance in older adults: a systematic review and meta-analysis. *Eur Rev Aging Phys Act*. 2024;21(1):13. [FREE Full text] [doi: [10.1186/s11556-024-00349-4](https://doi.org/10.1186/s11556-024-00349-4)] [Medline: [38764039](https://pubmed.ncbi.nlm.nih.gov/38764039/)]
27. Faíl LB, Marinho DA, Marques EA, Costa MJ, Santos CC, Marques MC, et al. Benefits of aquatic exercise in adults with and without chronic disease-a systematic review with meta-analysis. *Scand J Med Sci Sports*. 2022;32(3):465-486. [doi: [10.1111/sms.14112](https://doi.org/10.1111/sms.14112)] [Medline: [34913530](https://pubmed.ncbi.nlm.nih.gov/34913530/)]
28. NANA. Big hopes for big data. *Nat Med*. 2020;26(1):1. [doi: [10.1038/s41591-019-0740-8](https://doi.org/10.1038/s41591-019-0740-8)] [Medline: [31932805](https://pubmed.ncbi.nlm.nih.gov/31932805/)]
29. Li Y, Hu T, Wang J. The application and prospect of "User Profile" based on big data in patient experience improvement. *Inf Manage*. 2020;20(5):776-779.
30. Laney D. 3-D Data Management: Controlling Data Volume, Velocity, and Variety. 2001. URL: <https://xueshu.baidu.com/usercenter/paper/show?paperid=b1d4cfacbab42842b49068c026a31ceb> [accessed 2024-09-19]
31. Dawber TR, Kannel WB. An epidemiologic study of heart disease: the Framingham study. *Nutr Rev*. 1958;16(1):1-4. [doi: [10.1111/j.1753-4887.1958.tb00605.x](https://doi.org/10.1111/j.1753-4887.1958.tb00605.x)]
32. All of Us Research Program Genomics Investigators. Genomic data in the all of us research program. *Nature*. 2024;627(8003):340-346. [FREE Full text] [doi: [10.1038/s41586-023-06957-x](https://doi.org/10.1038/s41586-023-06957-x)] [Medline: [38374255](https://pubmed.ncbi.nlm.nih.gov/38374255/)]
33. Baxi V, Edwards R, Montalto M, Saha S. Digital pathology and artificial intelligence in translational medicine and clinical practice. *Mod Pathol*. 2022;35(1):23-32. [FREE Full text] [doi: [10.1038/s41379-021-00919-2](https://doi.org/10.1038/s41379-021-00919-2)] [Medline: [34611303](https://pubmed.ncbi.nlm.nih.gov/34611303/)]

34. Thigpen N, Patel S, Zhang X. Oura ring as a tool for ovulation detection: validation analysis. *J Med Internet Res*. 2025;27:e60667. [FREE Full text] [doi: [10.2196/60667](https://doi.org/10.2196/60667)] [Medline: [39889300](https://pubmed.ncbi.nlm.nih.gov/39889300/)]
35. Shilo S, Rossman H, Segal E. Axes of a revolution: challenges and promises of big data in healthcare. *Nat Med*. 2020;26(1):29-38. [doi: [10.1038/s41591-019-0727-5](https://doi.org/10.1038/s41591-019-0727-5)] [Medline: [31932803](https://pubmed.ncbi.nlm.nih.gov/31932803/)]
36. Burdick H, Pino E, Gabel-Comeau D, McCoy A, Gu C, Roberts J, et al. Effect of a sepsis prediction algorithm on patient mortality, length of stay and readmission: a prospective multicentre clinical outcomes evaluation of real-world patient data from US hospitals. *BMJ Health Care Inform*. 2020;27(1):e100109. [FREE Full text] [doi: [10.1136/bmjhci-2019-100109](https://doi.org/10.1136/bmjhci-2019-100109)] [Medline: [32354696](https://pubmed.ncbi.nlm.nih.gov/32354696/)]
37. Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, Saraiva A, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*. 2019;572(7767):116-119. [FREE Full text] [doi: [10.1038/s41586-019-1390-1](https://doi.org/10.1038/s41586-019-1390-1)] [Medline: [31367026](https://pubmed.ncbi.nlm.nih.gov/31367026/)]
38. Bena YA, Ibrahim R, Mahmood J, Al-Dhaqm A, Alshammari A, Nasser M, et al. Big data governance challenges arising from data generated by intelligent systems technologies: a systematic literature review. *IEEE Access*. 2025;13:12859-12888. [doi: [10.1109/access.2025.3528941](https://doi.org/10.1109/access.2025.3528941)]
39. Onnela JP. Opportunities and challenges in the collection and analysis of digital phenotyping data. *Neuropsychopharmacology*. 2021;46(1):45-54. [FREE Full text] [doi: [10.1038/s41386-020-0771-3](https://doi.org/10.1038/s41386-020-0771-3)] [Medline: [32679583](https://pubmed.ncbi.nlm.nih.gov/32679583/)]
40. Maipradit R, Wang D, Thongtanunam P, Kula RG, Kamei Y, McIntosh S. Repeated builds during code review: An empirical study of the OpenStack community. 2023. Presented at: 38th IEEE/ACM International Conference on Automated Software Engineering (ASE); September 11-15, 2023; Luxembourg. [doi: [10.1109/ase56229.2023.00030](https://doi.org/10.1109/ase56229.2023.00030)]
41. Qian L, Yi-fei W, Jian-wei S, Wen-ya Y, Liang Z, Xiang G, et al. The theory and practice of community health portraits from the perspective of general practitioners. *Chin J Gen Pract*. 2021;19(11):1797-1801. [doi: [10.16766/j.cnki.issn.1674-4152.002170](https://doi.org/10.16766/j.cnki.issn.1674-4152.002170)]
42. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med*. 2018;169(7):467-473. [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
43. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5:69. [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]
44. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol*. 2005;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
45. Noncommunicable diseases. URL: <https://www.who.int/health-topics/noncommunicable-diseases> [accessed 2024-12-10]
46. Chen X, Giles J, Yao Y, Yip W, Meng Q, Berkman L, et al. The path to healthy ageing in China: a Peking University-Lancet Commission. *Lancet*. 2022;400(10367):1967-2006. [doi: [10.1016/s0140-6736\(22\)01546-x](https://doi.org/10.1016/s0140-6736(22)01546-x)]
47. Zhang J, Shi C, Liang Z, Jin C, Wang L, Zhong Y, et al. Burden of noncommunicable diseases among children and adolescents aged 10-24 years in China, 1990-2019: a population-based study. *Cell Rep Med*. 2023;4(12):101331. [FREE Full text] [doi: [10.1016/j.xcrm.2023.101331](https://doi.org/10.1016/j.xcrm.2023.101331)] [Medline: [38118417](https://pubmed.ncbi.nlm.nih.gov/38118417/)]
48. Ding Y, Hooper J, Bates I, IMPALA Consortium. Common criteria for evaluating cross-disciplinary research in global health: a scoping review. *BMC Glob Public Health*. 2024;2(1):82. [doi: [10.1186/s44263-024-00113-x](https://doi.org/10.1186/s44263-024-00113-x)] [Medline: [39681972](https://pubmed.ncbi.nlm.nih.gov/39681972/)]
49. Gottlieb M, Haas MRC, Daniel M, Chan TM. The scoping review: a flexible, inclusive, and iterative approach to knowledge synthesis. *AEM Educ Train*. 2021;5(3):e10609. [FREE Full text] [doi: [10.1002/aet2.10609](https://doi.org/10.1002/aet2.10609)] [Medline: [34222749](https://pubmed.ncbi.nlm.nih.gov/34222749/)]
50. Tsai CW, Huang HC, Chiang HY, Chung C, Chang S, Chu P, et al. Longitudinal lipid trends and adverse outcomes in patients with CKD: a 13-year observational cohort study. *J Lipid Res*. 2019;60(3):648-660. [doi: [10.1194/jlr.p084590](https://doi.org/10.1194/jlr.p084590)]
51. Wu R. Research on Disease Prediction and Medication Recommendation Based on Electronic Health Record. Southeast University. Jun 2022. URL: <https://link.cnki.net/doi/10.27014/d.cnki.gdnau.2022.002109> [accessed 2024-02-15]
52. Nghiem N, Atkinson J, Nguyen BP, Tran-Duy A, Wilson N. Predicting high health-cost users among people with cardiovascular disease using machine learning and nationwide linked social administrative datasets. *Health Econ Rev*. 2023;13(1):9. [FREE Full text] [doi: [10.1186/s13561-023-00422-1](https://doi.org/10.1186/s13561-023-00422-1)]
53. Rajcáni J, Krobová J, Málková D. Distribution of Lednice (Yaba 1) virus in the chick embryo. *Acta Virol*. Nov 1975;19(6):467-472. [Medline: [1991](https://pubmed.ncbi.nlm.nih.gov/1991/)]
54. Ormerod WE. Ecological effect of control of African trypanosomiasis. *Science*. Feb 27, 1976;191(4229):815-821. [doi: [10.1126/science.2977](https://doi.org/10.1126/science.2977)] [Medline: [2977](https://pubmed.ncbi.nlm.nih.gov/2977/)]
55. Kent P, Cancelliere C, Boyle E, Cassidy JD, Kongsted A. A conceptual framework for prognostic research. *BMC Med Res Methodol*. Jun 29, 2020;20(1):172. [FREE Full text] [doi: [10.1186/s12874-020-01050-7](https://doi.org/10.1186/s12874-020-01050-7)] [Medline: [32600262](https://pubmed.ncbi.nlm.nih.gov/32600262/)]
56. Lee KH, Chu YC, Tsai MT, Tseng W, Lin Y, Ou S, et al. Artificial intelligence for risk prediction of end-stage renal disease in sepsis survivors with chronic kidney disease. *Biomedicines*. Feb 24, 2022;10(3):546. [FREE Full text] [doi: [10.3390/biomedicines10030546](https://doi.org/10.3390/biomedicines10030546)] [Medline: [35327348](https://pubmed.ncbi.nlm.nih.gov/35327348/)]
57. Steenkamer BM, Drewes HW, Heijink R, Baan CA, Struijs JN. Defining population health management: a scoping review of the literature. *Popul Health Manag*. Feb 2017;20(1):74-85. [doi: [10.1089/pop.2015.0149](https://doi.org/10.1089/pop.2015.0149)] [Medline: [27124406](https://pubmed.ncbi.nlm.nih.gov/27124406/)]

58. Varady AB, Wood RM. Improving uptake of population health management through scalable analysis of linked electronic health data. *Health Informatics J*. 2024;30(3):14604582241259344. [FREE Full text] [doi: [10.1177/14604582241259344](https://doi.org/10.1177/14604582241259344)] [Medline: [39095387](https://pubmed.ncbi.nlm.nih.gov/39095387/)]
59. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. *Transl Behav Med*. May 16, 2019;9(3):440-447. [doi: [10.1093/tbm/ibz043](https://doi.org/10.1093/tbm/ibz043)] [Medline: [31094445](https://pubmed.ncbi.nlm.nih.gov/31094445/)]
60. Wiesner M, Pfeifer D. Health recommender systems: concepts, requirements, technical basics and challenges. *Int J Environ Res Public Health*. Mar 03, 2014;11(3):2580-2607. [FREE Full text] [doi: [10.3390/ijerph110302580](https://doi.org/10.3390/ijerph110302580)] [Medline: [24595212](https://pubmed.ncbi.nlm.nih.gov/24595212/)]
61. Leng M, Sun Y, Li C, Han S, Wang Z. Usability evaluation of a knowledge graph-based dementia care intelligent recommender system: mixed methods study. *J Med Internet Res*. Sep 26, 2023;25:e45788. [FREE Full text] [doi: [10.2196/45788](https://doi.org/10.2196/45788)] [Medline: [37751241](https://pubmed.ncbi.nlm.nih.gov/37751241/)]
62. Lino C, Barrias S, Chaves R, Adega F, Martins-Lopes P, Fernandes JR. Biosensors as diagnostic tools in clinical applications. *Biochim Biophys Acta Rev Cancer*. May 2022;1877(3):188726. [FREE Full text] [doi: [10.1016/j.bbcan.2022.188726](https://doi.org/10.1016/j.bbcan.2022.188726)] [Medline: [35367530](https://pubmed.ncbi.nlm.nih.gov/35367530/)]
63. Liu D, Gao Y, Su X. Self-care ability and associated factors in community-dwelling older adults living with oral frailty using the COM-B model. *J Oral Rehabil*. Aug 2024;51(8):1530-1541. [doi: [10.1111/joor.13713](https://doi.org/10.1111/joor.13713)] [Medline: [38725254](https://pubmed.ncbi.nlm.nih.gov/38725254/)]
64. Olmedo-Aguirre JO, Reyes-Campos J, Alor-Hernández G, Machorro-Cano I, Rodríguez-Mazahua L, Sánchez-Cervantes JL. Remote healthcare for elderly people using wearables: a review. *Biosensors (Basel)*. Jan 27, 2022;12(2):73. [FREE Full text] [doi: [10.3390/bios12020073](https://doi.org/10.3390/bios12020073)] [Medline: [35200334](https://pubmed.ncbi.nlm.nih.gov/35200334/)]
65. Tan SY, Sumner J, Wang Y, Wenjun Yip A. A systematic review of the impacts of remote patient monitoring (RPM) interventions on safety, adherence, quality-of-life and cost-related outcomes. *NPJ Digit Med*. 2024;7(1):192. [FREE Full text] [doi: [10.1038/s41746-024-01182-w](https://doi.org/10.1038/s41746-024-01182-w)] [Medline: [39025937](https://pubmed.ncbi.nlm.nih.gov/39025937/)]
66. Çiftci N, Sarman A, Yıldız M, Çiftci K. Use of ChatGPT in health: benefits, hazards, and recommendations. *Public Health*. 2024;228:e1-e2. [doi: [10.1016/j.puhe.2023.12.032](https://doi.org/10.1016/j.puhe.2023.12.032)] [Medline: [38346914](https://pubmed.ncbi.nlm.nih.gov/38346914/)]
67. Kharko A, McMillan B, Hagström J, Muli I, Davidge G, Hägglund M, et al. Generative artificial intelligence writing open notes: a mixed methods assessment of the functionality of GPT 3.5 and GPT 4.0. *Digit Health*. 2024;10:20552076241291384. [FREE Full text] [doi: [10.1177/20552076241291384](https://doi.org/10.1177/20552076241291384)] [Medline: [39493632](https://pubmed.ncbi.nlm.nih.gov/39493632/)]
68. Atmakuru A, Shahini A, Chakraborty S, Seoni S, Salvi M, Hafeez-Baig A, et al. Artificial intelligence-based suicide prevention and prediction: a systematic review (2019–2023). *Inf Fusion*. 2025;114:102673. [doi: [10.1016/j.inffus.2024.102673](https://doi.org/10.1016/j.inffus.2024.102673)]
69. Budiharto W, Andreas V, Gunawan AAS. Deep learning-based question answering system for intelligent humanoid robot. *J Big Data*. 2020;7(1):77. [doi: [10.1186/s40537-020-00341-6](https://doi.org/10.1186/s40537-020-00341-6)]
70. Li YH, Li YL, Wei MY, Li GY. Innovation and challenges of artificial intelligence technology in personalized healthcare. *Sci Rep*. 2024;14(1):18994. [FREE Full text] [doi: [10.1038/s41598-024-70073-7](https://doi.org/10.1038/s41598-024-70073-7)] [Medline: [39152194](https://pubmed.ncbi.nlm.nih.gov/39152194/)]
71. Hasal M, Nowaková J, Ahmed Saghair K, Abdulla H, Snášel V, Ogiela L. Chatbots: security, privacy, data protection, and social aspects. *Concur Comput*. 2021;33(19):e6426. [doi: [10.1002/cpe.6426](https://doi.org/10.1002/cpe.6426)]
72. Geng B, Huan Z, Zhang X. Breaking the length barrier: LLM-enhanced CTR prediction in long textual user behaviors. 2024. Presented at: SIGIR '24: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 14-18, 2024; Washington, DC, USA. URL: <https://doi.org/10.1145/3626772.3657974> [doi: [10.1145/3626772.3657974](https://doi.org/10.1145/3626772.3657974)]
73. Yang S, Ma W, Sun P, Ai Q, Liu Y, Cai M, et al. Sequential recommendation with latent relations based on large language model. *arXiv:2403.18348*. 2024. [doi: [10.48550/arXiv.2403.18348](https://doi.org/10.48550/arXiv.2403.18348)]
74. Tan J, Xu S, Hua W, Ge Y, Li Z, Zhang Y. IDGenRec: LLM-recSys alignment with textual ID learning. 2024. Presented at: SIGIR '24: Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval; July 14-18, 2024:355-364; Washington, DC, USA. [doi: [10.1145/3626772.3657821](https://doi.org/10.1145/3626772.3657821)]
75. Yang J, Chen YL, Por LY, Ku CS. A systematic literature review of information security in chatbots. *Appl Sci*. 2023;13(11):6355. [doi: [10.3390/app13116355](https://doi.org/10.3390/app13116355)]
76. Azher ZL, Vaickus LJ, Salas LA, Christensen BC, Levy JJ. Development of biologically interpretable multimodal deep learning model for cancer prognosis prediction. 2022. Presented at: SAC '22: Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing; April 25-29, 2022:636-644; Virtual Event. [doi: [10.1145/3477314.3507032](https://doi.org/10.1145/3477314.3507032)]
77. Lipkova J, Chen RJ, Chen B, Lu MY, Barbieri M, Shao D, et al. Artificial intelligence for multimodal data integration in oncology. *Cancer Cell*. 2022;40(10):1095-1110. [FREE Full text] [doi: [10.1016/j.ccell.2022.09.012](https://doi.org/10.1016/j.ccell.2022.09.012)] [Medline: [36220072](https://pubmed.ncbi.nlm.nih.gov/36220072/)]
78. Lobato-Delgado B, Priego-Torres B, Sanchez-Morillo D. Combining molecular, imaging, and clinical data analysis for predicting cancer prognosis. *Cancers (Basel)*. 2022;14(13):3215. [FREE Full text] [doi: [10.3390/cancers14133215](https://doi.org/10.3390/cancers14133215)] [Medline: [35804988](https://pubmed.ncbi.nlm.nih.gov/35804988/)]
79. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, et al. A guide to deep learning in healthcare. *Nat Med*. 2019;25(1):24-29. [doi: [10.1038/s41591-018-0316-z](https://doi.org/10.1038/s41591-018-0316-z)] [Medline: [30617335](https://pubmed.ncbi.nlm.nih.gov/30617335/)]

80. Seinen TM, Fridgerisson EA, Ioannou S, Jeannetot D, John LH, Kors JA, et al. Use of unstructured text in prognostic clinical prediction models: a systematic review. *J Am Med Inform Assoc*. 2022;29(7):1292-1302. [[FREE Full text](#)] [doi: [10.1093/jamia/ocac058](#)] [Medline: [35475536](#)]
81. Carini E, Villani L, Pezzullo AM, Gentili A, Barbara A, Ricciardi W, et al. The impact of digital patient portals on health outcomes, system efficiency, and patient attitudes: updated systematic literature review. *J Med Internet Res*. 2021;23(9):e26189. [[FREE Full text](#)] [doi: [10.2196/26189](#)] [Medline: [34494966](#)]
82. Lyles CR, Nelson EC, Frampton S, Dykes PC, Cembali AG, Sarkar U. Using electronic health record portals to improve patient engagement: research priorities and best practices. *Ann Intern Med*. 2020;172(11_Supplement):S123-S129. [doi: [10.7326/m19-0876](#)]
83. Lu L, Zhang J, Xie Y, Gao F, Xu S, Wu X, et al. Wearable health devices in health care: narrative systematic review. *JMIR Mhealth Uhealth*. 2020;8(11):e18907. [[FREE Full text](#)] [doi: [10.2196/18907](#)] [Medline: [33164904](#)]
84. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: scoping review and future directions. *JMIR Mhealth Uhealth*. 2019;7(9):e12861. [[FREE Full text](#)] [doi: [10.2196/12861](#)] [Medline: [31512582](#)]
85. Sun Y, Zhou J, Ji M, Pei L, Wang Z. Development and evaluation of health recommender systems: systematic scoping review and evidence mapping. *J Med Internet Res*. 2023;25:e38184. [[FREE Full text](#)] [doi: [10.2196/38184](#)] [Medline: [36656630](#)]
86. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. *NPJ Digit Med*. 2022;5(1):3. [[FREE Full text](#)] [doi: [10.1038/s41746-021-00548-8](#)] [Medline: [35017634](#)]
87. Valero-Ramon Z, Fernandez-Llatas C, Valdivieso B, Traver V. Dynamic models supporting personalised chronic disease management through healthcare sensors with interactive process mining. *Sensors (Basel)*. 2020;20(18):5330. [[FREE Full text](#)] [doi: [10.3390/s20185330](#)] [Medline: [32957673](#)]
88. Tyler J, Choi SW, Tewari M. Real-time, personalized medicine through wearable sensors and dynamic predictive modeling: a new paradigm for clinical medicine. *Curr Opin Syst Biol*. 2020;20:17-25. [[FREE Full text](#)] [doi: [10.1016/j.coisb.2020.07.001](#)] [Medline: [32984661](#)]
89. Wu X. BEKO: bidirectional enhancement with a knowledge ocean for LLMs and KGs. *Chin J Comput*. 2025;1-18. [[FREE Full text](#)]
90. Fang Y. Design and Implementation of an EHR Temporal Analysis System Enhanced by LLM Knowledge. Beijing, China. Peking University; 2024.
91. Youssef A, Pencina M, Thakur A, Zhu T, Clifton D, Shah NH. External validation of AI models in health should be replaced with recurring local validation. *Nat Med*. 2023;29(11):2686-2687. [doi: [10.1038/s41591-023-02540-z](#)] [Medline: [37853136](#)]
92. Wessler BS, Nelson J, Park JG, McGinnes H, Gulati G, Brazil R, et al. External validations of cardiovascular clinical prediction models: a large-scale review of the literature. *Circ Cardiovasc Qual Outcomes*. 2021;14(8):e007858. [doi: [10.1161/circoutcomes.121.007858](#)]
93. Roy D, Dutta M. A systematic review and research perspective on recommender systems. *J Big Data*. 2022;9(1):59. [doi: [10.1186/s40537-022-00592-5](#)]
94. Ramspek CL, Jager KJ, Dekker FW, Zoccali C, van Diepen M. External validation of prognostic models: what, why, how, when and where? *Clin Kidney J*. 2021;14(1):49-58. [[FREE Full text](#)] [doi: [10.1093/ckj/sfaa188](#)] [Medline: [33564405](#)]
95. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58. [doi: [10.7326/m18-1376](#)]
96. Schetelig J, Baldauf H, Heidenreich F, Massalski C, Frank S, Sauter J, et al. External validation of models for KIR2DS1/KIR3DL1-informed selection of hematopoietic cell donors fails. *Blood*. 2020;135(16):1386-1395. [[FREE Full text](#)] [doi: [10.1182/blood.2019002887](#)] [Medline: [31932846](#)]
97. Herrera-Luis E, Benke K, Volk H, Ladd-Acosta C, Wojcik GL. Gene-environment interactions in human health. *Nat Rev Genet*. 2024;25(11):768-784. [doi: [10.1038/s41576-024-00731-z](#)] [Medline: [38806721](#)]
98. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. *Br Med J*. 2020;369:m1328. [[FREE Full text](#)] [doi: [10.1136/bmj.m1328](#)] [Medline: [32265220](#)]
99. Hangfeng M, Yaping C, Hui H, Yaping Z, Yujin L, Mei Z, et al. Methods and procedures of clinical predictive model. *Chin J Evid Based Med*. 2024;24(2):228-236.
100. Serrano LP, Maita KC, Avila FR, Torres-Guzman RA, Garcia JP, Eldaly AS, et al. Benefits and challenges of remote patient monitoring as perceived by health care practitioners: a systematic review. *Perm J*. 2023;27(4):100-111. [[FREE Full text](#)] [doi: [10.7812/TPP/23.022](#)] [Medline: [37735970](#)]
101. Abbasian M, Khatibi E, Azimi I, Oniani D, Shakeri Hossein Abad Z, Thieme A, et al. Foundation metrics for evaluating effectiveness of healthcare conversations powered by generative AI. *NPJ Digit Med*. 2024;7(1):82. [[FREE Full text](#)] [doi: [10.1038/s41746-024-01074-z](#)] [Medline: [38553625](#)]
102. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based

- on artificial intelligence. *BMJ Open*. 2021;11(7):e048008. [FREE Full text] [doi: [10.1136/bmjopen-2020-048008](https://doi.org/10.1136/bmjopen-2020-048008)] [Medline: [34244270](https://pubmed.ncbi.nlm.nih.gov/34244270/)]
103. Robson M, Im SA, Senkus E, Xu B, Domchek SM, Masuda N, et al. Olaparib for metastatic breast cancer in patients with a germline mutation. *N Engl J Med*. 2017;377(6):523-533. [doi: [10.1056/nejmoa1706450](https://doi.org/10.1056/nejmoa1706450)]
 104. Ference BA, Ginsberg HN, Graham I. Low-density lipoproteins cause atherosclerotic cardiovascular disease. 1. Evidence from genetic, epidemiologic, and clinical studies. A consensus statement from the European Atherosclerosis Society Consensus Panel. *Eur Heart J*. 2017;38(32):2459-2472. [doi: [10.3410/f.727555798.793563055](https://doi.org/10.3410/f.727555798.793563055)]
 105. Mollenhauer B, Locascio JJ, Schulz-Schaeffer W, Sixel-Döring F, Trenkwalder C, Schlossmacher MG. α -Synuclein and tau concentrations in cerebrospinal fluid of patients presenting with parkinsonism: a cohort study. *Lancet Neurol*. 2011;10(3):230-240. [doi: [10.1016/s1474-4422\(11\)70014-x](https://doi.org/10.1016/s1474-4422(11)70014-x)]
 106. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. *J Clin Epidemiol*. 2014;67(1):56-64. [doi: [10.1016/j.jclinepi.2013.08.005](https://doi.org/10.1016/j.jclinepi.2013.08.005)] [Medline: [24189091](https://pubmed.ncbi.nlm.nih.gov/24189091/)]

Abbreviations

3V: volume, velocity, and variety

AI: artificial intelligence

EHR: electronic health record

KG: knowledge graph

LLM: large language model

NCD: noncommunicable disease

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

PROBAST: Prediction Model Risk of Bias Assessment Tool

PROGRESS-Plus: place of residence, race/ethnicity/culture/language, occupation, gender/sex, religion, education, socioeconomic status, and social capital

TRIPOD-AI: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis–Artificial Intelligence

Edited by J Sarvestan; submitted 13.02.25; peer-reviewed by LP Gorrepati, L Zhang, S Mohamed Shaffi; comments to author 01.04.25; revised version received 21.04.25; accepted 24.04.25; published 05.06.25

Please cite as:

Du H, Yu J, Chen D, Wu J, Xue E, Zhou Y, Pan X, Shao J, Ye Z

Big Data–Driven Health Portraits for Personalized Management in Noncommunicable Diseases: Scoping Review

J Med Internet Res 2025;27:e72636

URL: <https://www.jmir.org/2025/1/e72636>

doi: [10.2196/72636](https://doi.org/10.2196/72636)

PMID:

©Haoyang Du, Jianing Yu, Dandan Chen, Jingjie Wu, Erxu Xue, Yufeng Zhou, Xiaohua Pan, Jing Shao, Zhihong Ye. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 05.06.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research (ISSN 1438-8871), is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.