


Detecting latent topics and trends of digital twins in healthcare: A structural topic model-based systematic review

DIGITAL HEALTH
Volume 9: 1–27
© The Author(s) 2023
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076231203672
journals.sagepub.com/home/dhj



Bo Sheng^{1,2}, Zheyu Wang¹, Yujiao Qiao³, Sheng Quan Xie⁴, Jing Tao¹ 
and Chaoqun Duan¹

Abstract

Objective: Digital twins (DTs) have received widespread attention recently, providing new ideas and possibilities for future healthcare. This review aims to provide a quantitative review to analyze specific study contents, research focus, and trends of DT in healthcare. Simultaneously, this review intends to expand the connotation of “healthcare” into two directions, namely “Disease treatment” and “Health enhancement” to analyze the content within the “DT + healthcare” field thoroughly.

Methods: A data mining method named Structure Topic Modeling (STM) was used as the analytical tool due to its topic analysis ability and versatility. Google Scholar, Web of Science, and China National Knowledge Infrastructure supplied the material papers in this review.

Results: A total of 94 high-quality papers published between 2018 and 2022 were gathered and categorized into eight topics, collectively covering the transformative impact across a broader spectrum in healthcare. Three main findings have emerged: (1) papers published in healthcare predominantly concentrate on technology development (artificial intelligence, Internet of Things, etc.) and application scenarios (personalized, precise, and real-time health service); (2) the popularity of research topics is influenced by various factors, including policies, COVID-19, and emerging technologies; and (3) the preference for topics is diverse, with a general inclination toward the attribute of “Health enhancement.”

Conclusions: This review underscores the significance of real-time capability and accuracy in shaping the future of DT, where algorithms and data transmission methods assume central importance in achieving these goals. Moreover, technological advancements, such as omics and Metaverse, have opened up new possibilities for DT in healthcare. These findings contribute to the existing literature by offering quantitative insights and valuable guidance to keep researchers ahead of the curve.

Keywords

Healthcare, digital twin, structure topic modeling, artificial intelligence, text data mining

Submission date: 28 April 2023; Acceptance date: 8 September 2023

Introduction

Healthcare has undergone tremendous changes after entering the great revolution of Industry 4.0.¹ In the past, healthcare was more patient-oriented, discussing how to cure patients, utilize medical resources, and configure medical equipment. Now, driven by the concept of Healthcare 4.0 and encouraged by a large number of technological devices,² the objects of healthcare are no longer just patients but normal people. With modern technology, healthcare is experiencing a shift toward increasingly

¹School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

²Shanghai Key Laboratory of Intelligent Manufacturing and Robotics, Shanghai University, Shanghai, China

³ShanghaiTech University Center for Innovative Teaching and Learning, ShanghaiTech University, Shanghai, China

⁴School of Electronic and Electrical Engineering, University of Leeds, Leeds, UK

Corresponding authors:

Jing Tao, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China.

Email: taojing1016@shu.edu.cn

Chaoqun Duan, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China.

Email: chaoqun.duan@hotmail.com



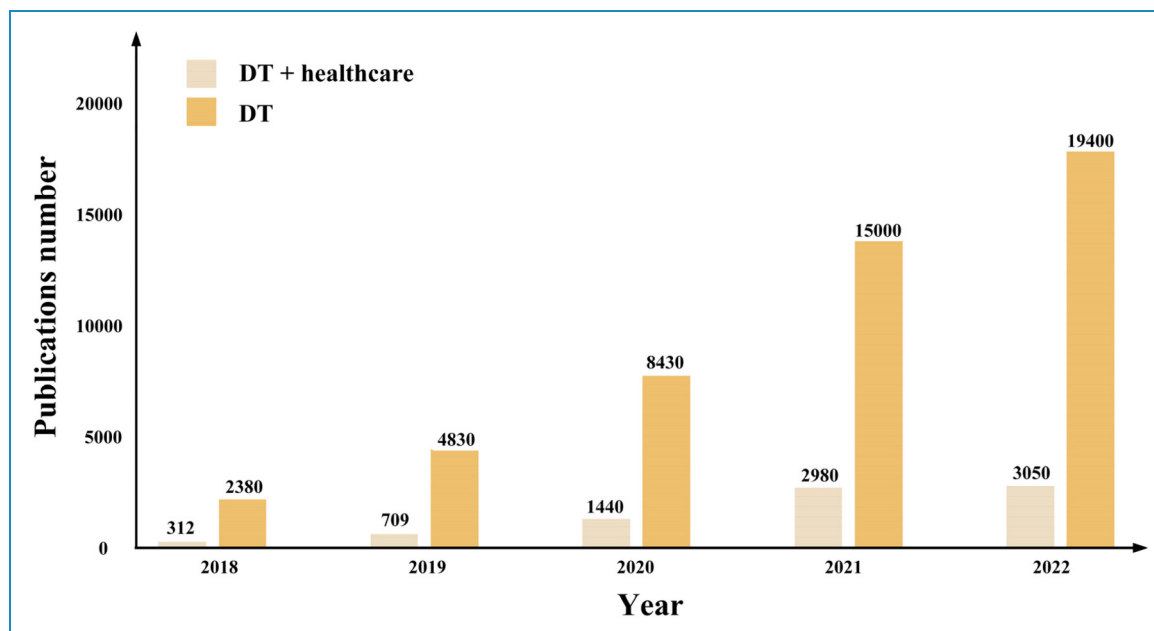


Figure 1. Publication number of “DT” and “DT+healthcare.”

intelligent. A large amount of physiological information can help plan reasonable fitness programs, real-time medical reports, and even predict future diseases.

The emergence of the concept of digital twins (DTs) has introduced new perspectives and possibilities to healthcare. Since it was proposed in 2002 by Professor Grieves,³ DT has expanded the traditional manufacturing concept to whole industries. The world-renowned consulting firm Gartner listed DT as one of the top 10 strategic technology trends for two consecutive years from 2017 to 2018 and predicted that the market size of DT can reach 183 billion dollars in 2031.^{4–6} Besides, the number of papers about DT and healthcare has surged in recent years. Figure 1 shows the changing number of retrieved papers in Google Scholar from 2018 to 2022. The light brown box stands for the retrieved number by the keyword “digital twin+healthcare” and the dark brown box stands for the retrieved number by the keyword “digital twin.” The number on the box stands for the detailed retrieved number. For instance, “1440” and “8430” above the box of 2020 represent the number of results by the search words “digital twin+healthcare” and “digital twin.” It can be found that the number of papers related to healthcare accounts for an increasing proportion of all papers about DT.

DT is a complex concept with no clear and widely accepted definition yet. The basic idea of DT refers to the fusion and mutual mapping of physical and virtual information.⁷ In healthcare, the main advantage of DT lies in its provision of a visual representation. Massive data serves as a bridge that connects the physical world and the virtual model. With the assistance of artificial intelligence (AI), the twin may provide excellent intervention for

humans. Manocha et al.⁸ proposed a DT, and Blockchain combined framework for elderly healthcare. This framework used wearable devices to collect the physical activity data of elderly people. After the transmission and fusion of data, DT can effectively analyze emergencies and accidents involving elderly individuals. Analysis helps mitigate risk factors in their daily lives and offers a tangible solution in the realm of smart healthcare. Alessandra et al.⁹ proposed a DT framework for feedback on dangerous social distance, which can be used to control the spread of COVID-19. The framework was successfully applied in a canteen environment and achieved a good effect. In healthcare, DT is continuously developing, and various physical copies are designed to meet various needs. The above papers excavated the application of DT and illustrated the great potential of DT in healthcare. Compared with traditional treatment and resource allocation methods, DT is more efficient, low cost, and autonomous.

The emergence of DT has accelerated the more detailed division in healthcare, and the applications of DT have driven the publication of review papers. To date, several reviews have summarized the existing research progress of DT in healthcare. Sun et al.¹⁰ combined 22 representative studies in healthcare to analyze the current applications and future possibilities of DT in clinical settings. The research confirmed the significant role of DT and made innovative predictions about its broad applications in real-time disease monitoring, dynamic analysis, and precise treatment. Haleem et al.¹¹ conducted a study exploring the requirements for DT in healthcare. The research comprehensively discussed the concepts, technologies, and applications of DT. They concluded that building digital

replicas through DT can improve care levels, increase data utilization, and more efficient resource allocation. The above reviews summarized the current development of “DT + healthcare” through qualitative methods and predicted the future development of DT.

However, the above reviews have limitations: (1) the discussed DT applications are only focused on the medical field, without exploring other potential applications in health maintenance for normal people; and (2) both papers used qualitative methods to carry out the research, and included papers may not be comprehensive enough to summarize the contents about DT in healthcare fully. To overcome the limitations of previous studies, conducting a quantitative review to analyze the field of “DT + healthcare” is essential. Data-based analysis can offer an objective and compelling overview of research progress. A comprehensive understanding can also be provided, encompassing medical and health improvement aspects.

Specifically, a data mining method called STM (Structure Topic Modeling) was employed to conduct literature analysis. The novelty of STM lies in its ability to automatically generate topics from collected papers based on a fast variant of nonconjugate variational expectation-maximization. The determination of topics is mainly dependent on the content of the papers, making it objective. Additionally, STM can incorporate structured information from papers (e.g. publication date, author gender, and publication venue) and realize a multidimensional analysis. The analysis results can be presented using visualizations, such as graphs and tables, facilitating easier understanding than traditional literature analysis methods.

The review can be trusted to provide a comprehensive perspective to analyze contents, research focus, and trends of DT in healthcare. Specifically, the leading novel contributions of this review are as follows:

1. This review categorized research on DT in healthcare into eight topics, the content covering the principles, technical components, and specific applications. Detailed topic content includes the development of the DT framework, the digitalization process of healthcare, AI and DT fusion, real-time simulation, and personalized health service.
2. This review examines the evolution and transformations of each topic over time by correlating paper publication dates with the eight topics as one covariate. Identifying trends and underlying factors behind these shifts offers valuable insights to guide the future.
3. This review divided papers in healthcare into two main directions: “Health enhancement” and “Disease treatment.” The two directions expand the content in healthcare and provide in-depth insights for present and future research of the “DT + healthcare” field.

Method

This review was conducted according to the PRISMA statement guidelines.¹² The complete survey process consisted of the following three parts: (1) data collection; (2) data preprocessing; and (3) STM analysis. Figure 2 represents a flow chart to demonstrate the study method. Specifically, the data collection section introduces the database and

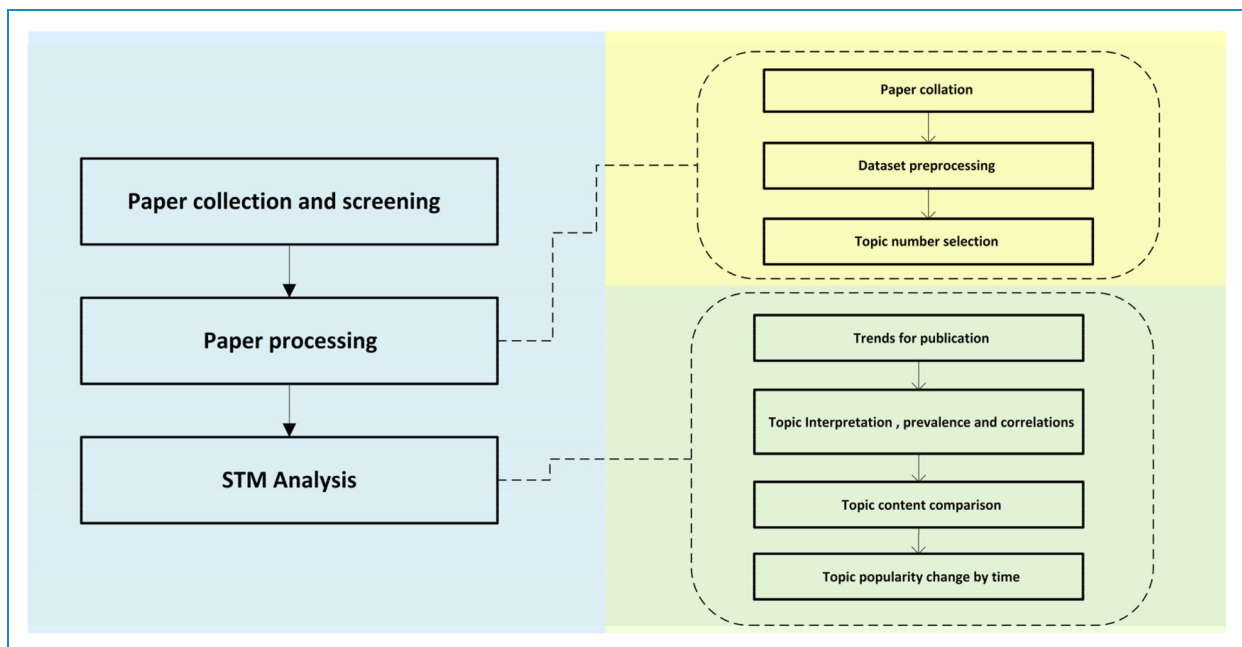


Figure 2. Flow chart for this review.

search methods used in this review. Preprocessing includes two steps, namely manual preprocessing and STM systematic preprocessing. The final STM results will be displayed in Chapter 3.

Paper collection

The collection progress of data is presented in this section. Firstly, this review took three well-known electronic databases as the primary access to obtain data. Google Scholar and Web of Science (WoS) supply most of the papers. The two databases include studies published worldwide, and many previous studies used them as data resources.^{13–15} The China National Knowledge Infrastructure (CNKI) was employed as the Supplementary to Chinese papers. The title, keywords, and abstracts of each paper were chosen as the analysis materials of STM. Usually, the three parts are the first concerns for all readers. The authors also would summarize the core content of the paper in these three parts, which makes the three parts become the most conceptual and semantically parts of the whole paper.^{16–18}

The search keywords were settled as “digital twin/s + healthcare.” All included papers should concern healthcare, and the contents should involve or apply DT. Patents and papers without English abstracts were excluded. Meanwhile, the contents of papers are unrelated to healthcare for humanity and instead focus on production quality were excluded even though their keywords meet the requirements of this review, such as the paper reported by Newrzella et al.¹⁹ The review commenced in 2018 primarily because most general papers in the databases mentioned above were published during that year. Since 2018 can be regarded as the initial year for DT development, it was chosen as the starting point for the review. All included papers were carefully examined by all authors to ensure that each publication appeared only once in the dataset and to exclude papers with highly similar content. Additionally, the authors ensured that the abstract lengths of all included publications did not significantly differ. This manual verification process was implemented to minimize bias risks posed by included papers.

Figure 3 depicts the detailed paper collection and screening progress for this review. Four rounds of screening were conducted. A preliminary search using Google Scholar was conducted to determine whether there were sufficient papers for STM analysis. Approximately 200 papers were identified that met the initial criteria. Meanwhile, WoS and CNKI were utilized for additional searching, resulting in 136 and 136 matching the requirement, respectively. Among them, 193 duplicate papers were excluded. During the second screening round, a total of 279 papers were reviewed by title, resulting in the exclusion of 171 papers. Experts reviewed abstracts and keywords in the third screening round, excluding 24 out of 108 papers.

Subsequently, the references of the 84 papers were reviewed, and 10 papers that came with the references were added to the dataset. Finally, a total of 94 papers were selected as this review’s dataset and used for analysis. All the screening progress was conducted by two experts. To ensure the review contains the latest focus, additional papers published in 2023 have been included. Since the year 2023 has not ended, these papers were not incorporated into the STM analysis. However, they were discussed to supplement the latest research progress and compare it with the results of STM.

Paper collation

After the screening, the 94 papers required a further human process. The purpose of the human process is to extract valid information for further STM preprocessing. Each included paper’s title, keyword, and abstract were sorted into an Excel sheet by the form “Title. Abstract. Keywords.” For a few papers without keywords, the form would be “Title. Abstract.”

Two additional covariates were included to expand the scope of the review, namely “time of publication” and “paper attributes.” Each paper was labeled with these covariates in the Excel sheet. The inclusion of the time variable is essential for tracking the popularity of topics over time, which is a commonly recognized covariate in surveys using the STM methodology. Furthermore, the papers were dichotomized based on their attributes, allowing for a more detailed analysis of the topics based on these attributes.

The attributes considered in this review are based on two research directions derived from the definition of “healthcare.” As previously mentioned, these attributes are labeled as “Health enhancement” and “Disease treatment.” The papers classified under the “Health enhancement” attribute primarily focus on the application of DT in monitoring and predicting changes in the body’s current state.^{20–23} These papers often target healthy individuals, including athletes,^{20,21,24} and discuss the use of DT technology to improve information security,^{25–28} address ethical concerns, reduce social inequality, and enhance overall well-being. In total, 44 out of the 94 papers were distributed with the “Health enhancement” attribute.

On the other hand, the papers with “Disease treatment” attribute focus on patients and have clear directions for specific diseases such as cardiovascular disease or COVID-19.^{29–32} Some papers aim to optimize the allocation of medical resources in hospitals and assist rehabilitation were also labeled with the “Disease treatment” attribute.^{33–38} Papers with this attribute prefer DT to simulate human organs or the effects of different interventions on organs.^{39–42} In short, the “Disease treatment” attribute represents a set of studies that strive to offer effective solutions for curing patients, optimizing medical resources, and

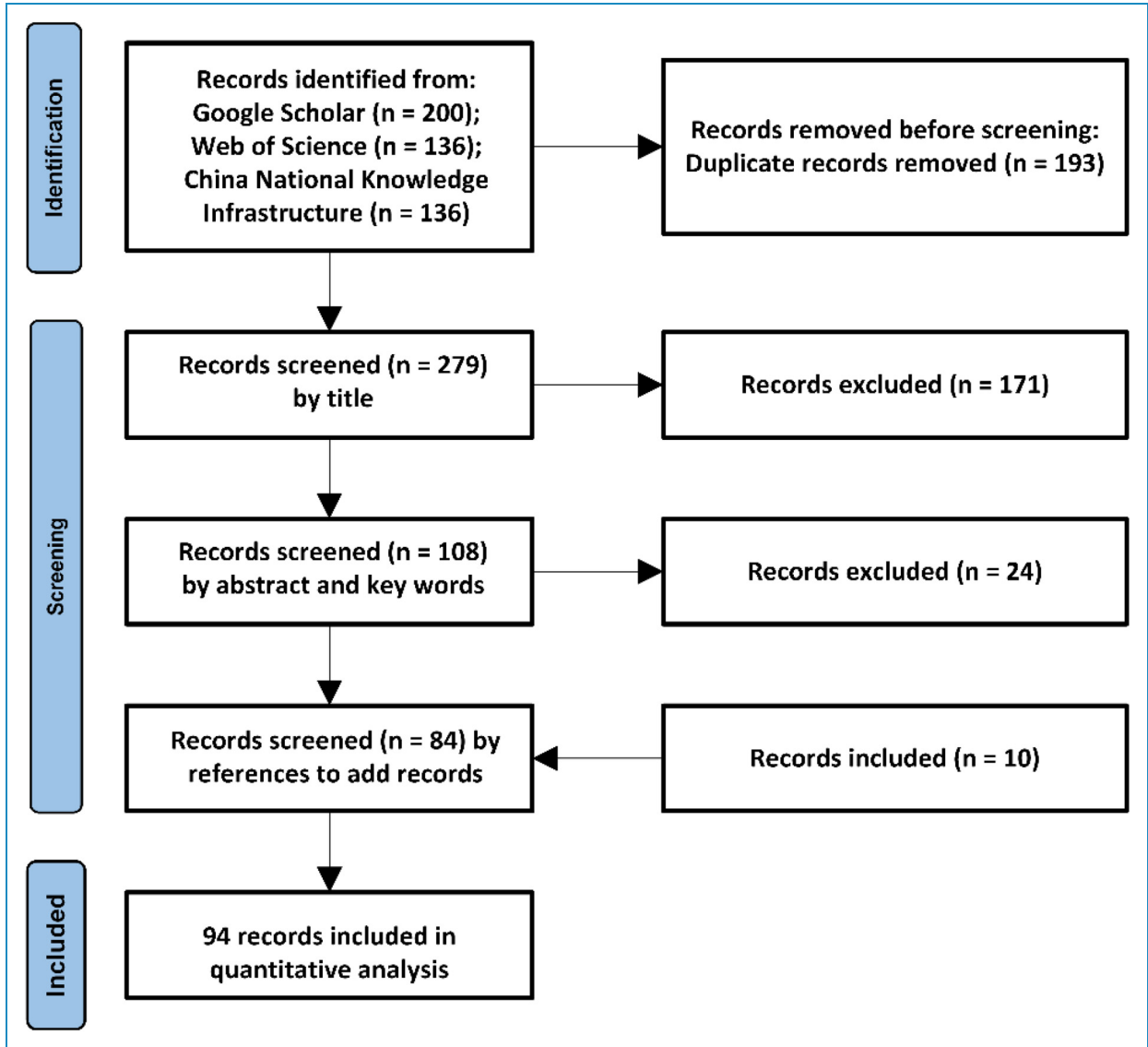


Figure 3. Flow chart for data collection and screening based on the PRISMA flow diagram.

rehabilitating individuals. In total, 50 out of the 94 papers were distributed with the “Disease treatment” attribute.

Dataset preprocessing

As introduced before, a quantitative method named STM was used to make the content analysis. Conceptually, STM is a hierarchical mixed membership model for analyzing topical content within documents. Its main advantage is that various structural information (covariates) can be mixed, such as time and location, helping researchers discover the potential topics, analyze the content of topics, and prevalence changes of topics with kinds of structural information.⁴³

Figure 4 depicts the graphical illustration of STM. Firstly, some variables need to be defined. The number of

documents can be defined as $d \in \{1, 2, \dots, D\}$; $k \in \{1, 2, \dots, K\}$ denotes the defined topic numbers; the words within documents can be defined as $n \in \{1, 2, \dots, N\}$; $v \in \{1, \dots, V\}$ represents the index of all unique observation words.

STM can be divided into three sub-models: the topic prevalence model, topic content model, and core language model.⁴⁴ X and Y in Figure 4 represent the topic prevalence matrix and the topical content covariates matrix, respectively. θ is a $D \times K$ probability matrix of all documents for defined topics. For any document, any row θ_d of θ satisfies the following distribution:

$$\theta_d \sim \text{LogisticNormal}_{K-1}(\gamma'X'_d, \Sigma) \quad (1)$$

where γ is a normal distribution coefficient matrix for the topic prevalence model with half Cauchy parameters. For

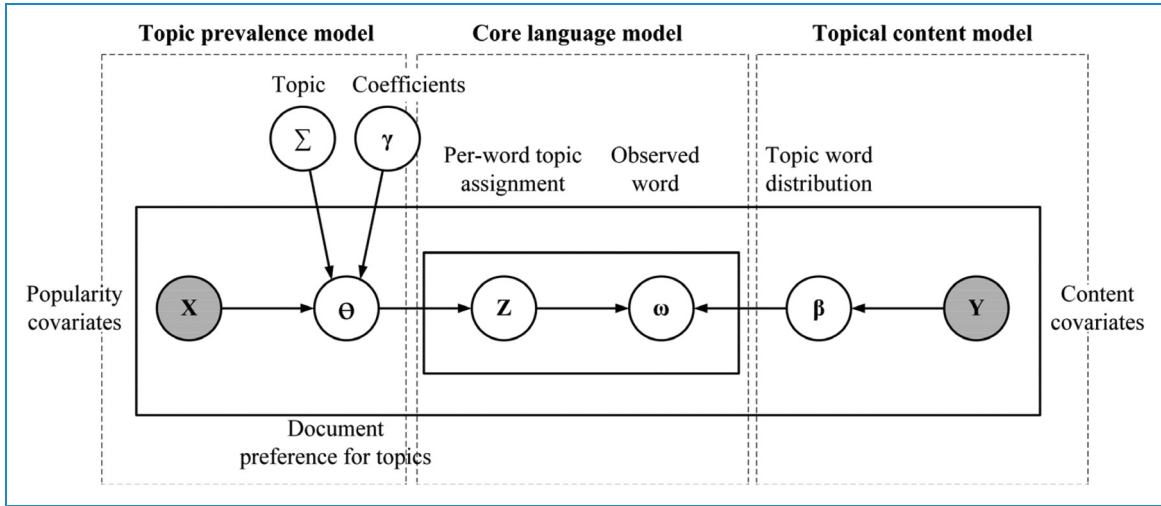


Figure 4. Graphical illustration of STM.⁴⁴⁻⁴⁶

θ_d , it means any document's preference for all topics, and the sum of all values of θ_d is 1. β is a $K \times V$ multinomial logistic topic word matrix with covariates. For each topic, β_k is a vector that combines a set of observed words and follows the below relationship:

$$\beta_k \propto \exp(m_v + \kappa^{(topic)} + \kappa^{(cov)} + \kappa^{(int)}) \quad (2)$$

where m_v is the marginal log-transformed rate of term v . In addition, it can be seen from (2) that the topic content considers the covariates and incorporates the specific deviation of covariates κ . The topic content vector for topic k is influenced by three parts: topic, covariate, and topic-covariate interactions.

Obtaining a preference proportion to all topics, for any document d , all stemmed words in the document will be assigned a topic label according to the proportion of θ_d , which is $z_{d,n}$. The observed words $\omega_{d,n}$ in β_k are paired with the labels of $z_{d,n}$ to verify whether the observed words can belong to the corresponding topics. After multiple iterations, ω will gradually converge and become increasingly accurate, representing the final topic model.

Figure 5 concludes the main contents STM can achieve. STM counts the high-frequency words for all papers in the dataset, which are classified into various topics through algorithms. Different high-frequency words can appear in a paper, and the paper can correspond to several topics. By performing topic statistics on all papers, the proportions of each topic can be obtained. Covariates are attached to each paper as a label, which can be used to divide all papers. Based on the proportions of each topic and the covariates, the popularity of each topic is finally inferred. This review utilized the titles, abstracts, and keywords of the included papers as the primary analysis materials. In cases where keywords were missing, only the titles and abstracts were included for analysis. Additionally, the publication

time of each paper and the attributes defined in dataset preprocessing were assigned as covariate labels to all materials.

STM yields a range of results. The primary focus lies in identifying multiple sets of topic words, referred to as topics. Each topic corresponds to multiple papers within the STM dataset, which serve as the sources for that particular topic. By centering on topics, the proportion of each topic in the entire corpus and the interconnections between topics can be discovered. Incorporating the "time" covariate enables an understanding of how the popularity of each topic changes over time. Similarly, incorporating the "attribute" covariate allows us to comprehend the degree of preference of each topic toward two attributes and the specific details of those preferences. Most of the results are automatically generated by STM and presented in graphical form. Part of the results expressed using parameters will be explained in the corresponding results.

STM has a wide range of application scenarios, and many studies have applied this method in various fields. Chen et al.¹⁸ analyzed 3963 articles published in the "Computers & Education" journal within 40 years, aiming to reveal the journal's research hotspots and trends. The findings provide guidance for future education research directions. Bas Hofstra et al.⁴⁷ employed the STM to figure out the topic terms and extract concepts to explain the diversity-innovation paradox in the science field for underrepresented groups. The above applications have demonstrated the versatility and potential of STM for paper research. Compared with traditional qualitative reviews, the STM reviews can more objectively analyze the synthetic results of all included papers and provide richer analysis angles to cut into the fields.

This review used the R language to realize STM because it is an open-source programming language with simple

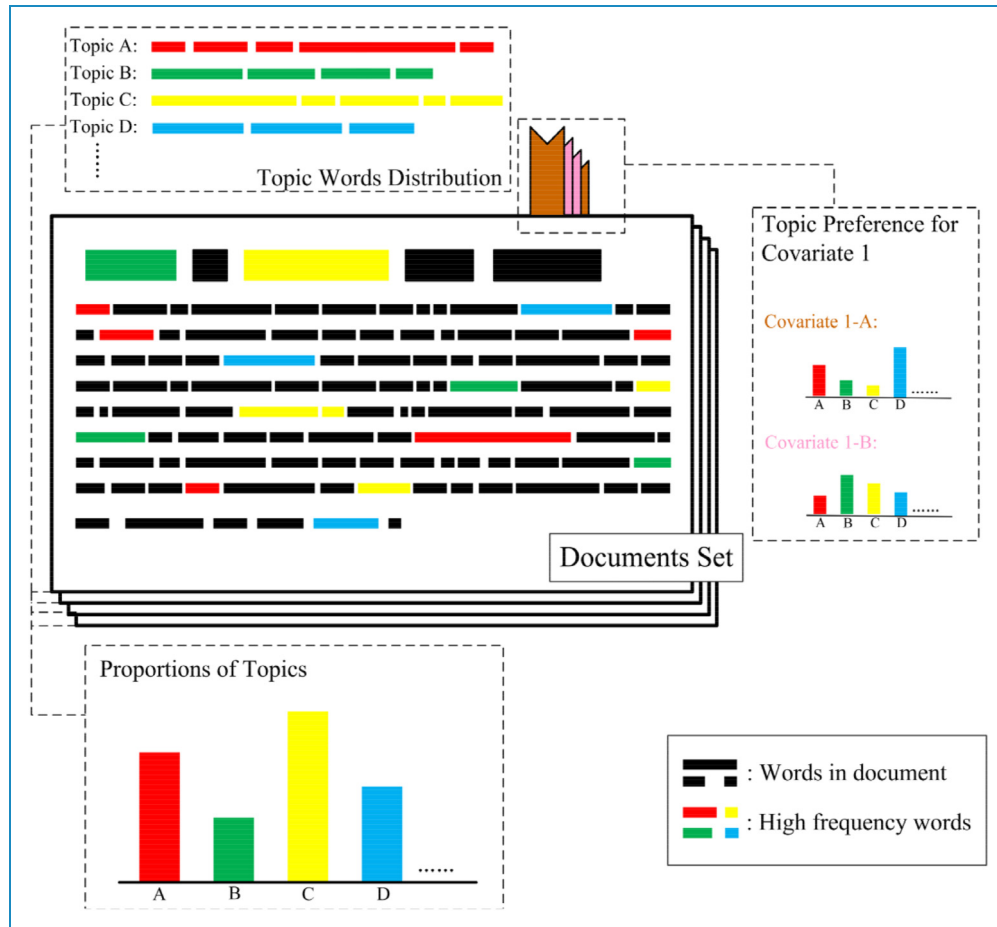


Figure 5. The basic content of STM.

syntax and structure.⁴⁸ In the R language environment, STM is realized by calling the STM package. The package has many built-in functions, which can complete operations such as building STM, drawing images, and outputting data. The review used `prepDocuments` and `textPreprocessor` functions to convert the Excel data into a form that STM can process and prevent information from affecting STM results. The specific processing contents include the following: (1) normalize all the letters and transform them into low cases; (2) removal of stop words (e.g. is, to, are, of), numbers, and punctuations; and (3) extract stemming words.⁴³

Topic number selection

Choosing the appropriate number of topics for an STM-based review is crucial. Topic semantic coherence, exclusivity, held-out likelihood, and the residual are the main indexes for deciding topics in STM. Semantic coherence describes the coherent expression of words under the topics in STM. High semantic coherence indicates that the words used to label each topic appear frequently in

the paper and indicate a strong connection between the topic and the content of the paper. While exclusivity describes the specificity of a topic, the topic words with high exclusivity of a topic will seldom appear in other papers that do not reflect this topic. The Held-out likelihood is to retain a part of words in a group of papers, train the model, and use potential variables to evaluate the probability of retaining parts.⁴³ The residual is an index used to indicate whether the model is reasonable. In an ideal model, the residual equals 1.

Figure 6(a) displays the diagnostic indexes of several topics obtained using the `searchk` function. On the other hand, Figure 6(b) combines the semantic coherence and exclusivity of different numbers of topics into a single figure. Both figures provide evidence for the selection of the optimal number of topics. The first judgment for choosing the number of topics is semantic coherence and exclusivity. In Figure 6(b), the direction pointed by the arrow indicates the ideal number of topics. That is, it has both high semantic consistency and exclusivity simultaneously. Based on Figure 6(a) and (b), eight topics are the first choice.

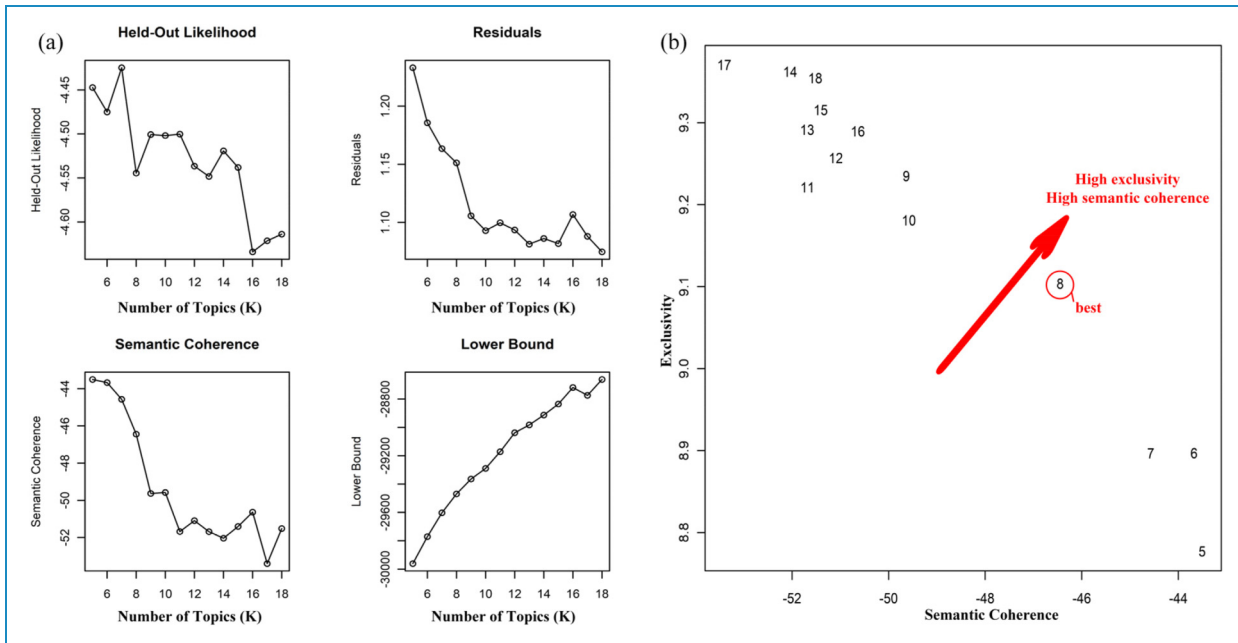


Figure 6. Evidence figure for selecting the appropriate topic numbers: (a) diagnostic values by the number of topics; (b). comparison of the coherence and exclusivity of five ~ 18 topics.

Table 1. Labels for seven, eight, and nine topics.

Seven topics	Eight topics	Nine topics
Conception of DT	Conception of DT	Conception of DT
Healthcare digitalization	Healthcare digitalization	Healthcare digitalization
Precision medicine	Precision medicine	Precision medicine
AI Disease diagnosis	Disease AI diagnosis	AI Disease diagnosis
Health management	Health management	Health management
Model development and application	Model simulation	Model simulation
Personalized healthcare	Personalized healthcare	Personalized healthcare
	Framework development	Framework development
		Smart health service system

Other possible choices, like seven or nine topics, were also tested. The experts gave the names of seven, eight, and nine topics, which are indicated in Table 1. If dividing

the dataset into seven topics, compared with eight topics, the topic “Model simulation” and topic “Framework development” are merged into one topic. There is a clear difference between the two topics. “Model simulation” focuses on the DT’s application, while “Framework development” focuses on the framework construction and future assumptions. When dividing the dataset into nine topics, the papers under the newly added topic “Smart health service system” are frame-type papers. This topic is separated from the “Framework development” topic and bears similarities to the “Health management” topic. However, as an independent topic, it lacks distinct characteristics that set it apart from others. Therefore, seven and nine were excluded, and eight was the most appropriate choice.

Results

Trends of the publications

To explore specific trends for publications in the field “DT + healthcare” and analyze possible reasons behind the trends, Figure 7 depicts the publication time of the papers in this review’s dataset between 2018 and 2022 recorded by quarters. The number of publications generally presents a zigzag upward trend. Two papers were recorded in 2018, but the number increased to 35 in 2022. There were three peaks in published papers, with nine papers published in the fourth quarter of 2020, 11 papers published in the second quarter of 2021, and 14 papers published in the second quarter of 2022. The dotted line presents a trend prediction of the folded line, which indicates an increasing trend with time.

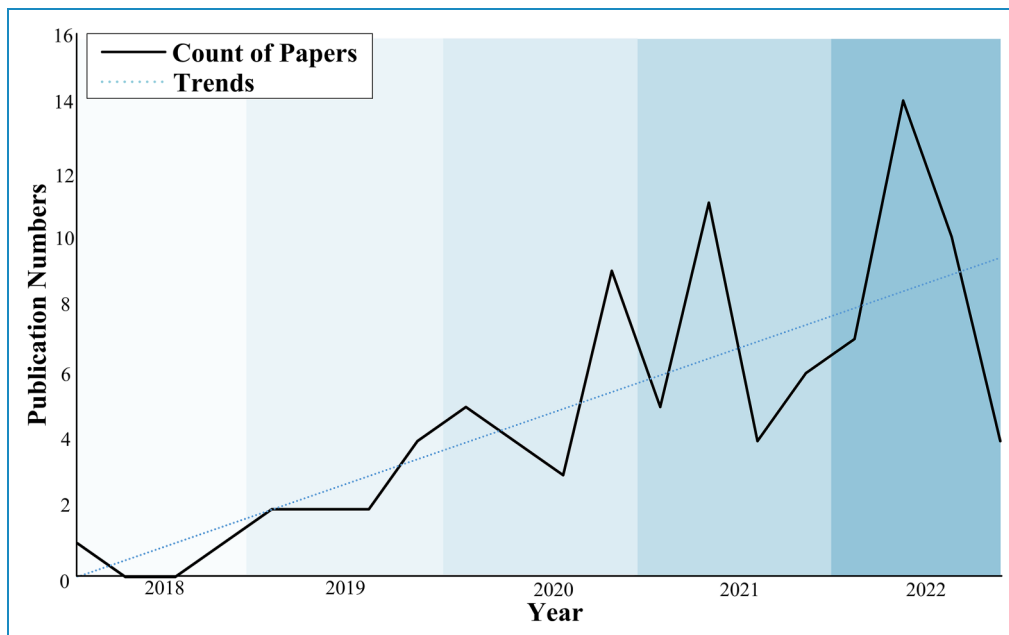


Figure 7. Publication trends record from 2018 to 2022 by quarters.

Topic interpretation, prevalence, and correlations

This section presents results about topics via STM, including presenting the content of topic words, the preferences of each topic, and relationships between different topics, enabling the analysis and presentation of topics from various perspectives.

Topic interpretation. Topic interpretation mainly explains the topic words that describe each topic. The topic words summarize the core content of each topic and can help researchers understand the main content and direction of the corresponding topic. STM counts the high-frequency words to figure out eight topics, and each topic has several topic words to indicate the contents. FREX (Frequency and Exclusivity), proposed by Jonathan M. Bischof and Edoardo M. Airolidi,⁴⁹ provides a topic word determination method that considers both semantic coherence and exclusivity. This comprehensive method calculates the topic words by taking the harmonic mean of high-frequency words and a given topic word exclusivity distribution.

Figure 8 presents word cloud figures of the eight topics. Each topic's word cloud displays the top 40 words ranked by the FREX criterion, where word size and color correspond to word frequency. Words are depicted in larger sizes and shades of red (plum red, blue, and black) to indicate higher to lower frequencies. A word cloud figure can provide an intuitive representation of the main content involved in the topic. For example, in the word cloud figure of topic 3, words such as "treatment" and "patient"

indicate that the topic is related to the disease. While words such as "precise," "real-time," and "intelligence" point to the requirements for treating the disease. These words determined the label of topic 3.

Similarly, for topic 5, words such as "service," "support," "process," and "provide" point to the object or the action of the object served by the topic. In combination with words such as "information," "management," and "need," it can be inferred that the topic object should be a hospital-like system that serves healthcare services. The word "simulation" in topic 6 is very unique. Words such as "specify," "optimal," "monitor," and "integration" can be used to associate with the word "simulation," which is the main reason for the label of this topic.

Based on the word cloud figures, further refinement of the aforementioned FREX topic words was conducted, identifying the top 15 topic words (word cloud figures emphasize word frequency and ignore semantic exclusivity). Adjustments were made to the further selected topic words to enhance the interpretation of the topics. Some topic words, such as "digital" and "twin" were combined into the phrase "digital twin" because such continuous expression conveys more information than separate words. High-frequency but limited-value words like "study," "review," "can," and "need" were removed. All the adjustments above were discussed with experts to reduce subjective bias. Finally, the core FREX topic words are presented in Table 2.

Topic 1 contains the words like "physical," "virtual," and "relationship." These words often appear when describing the concept of DT and make the experts decide on the

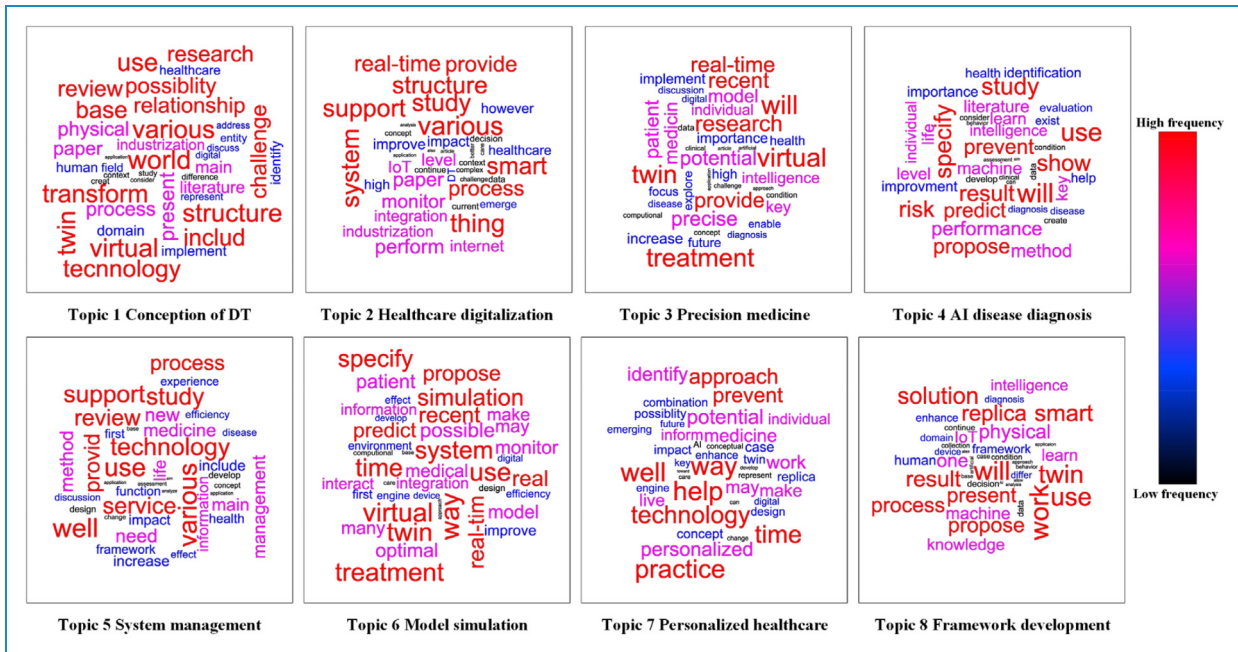


Figure 8. Word cloud figure for the eight topics.

label for topic 1. Topic 2 has many verbs that may describe DT's work in healthcare. The Internet of Things (IoT) appears in a very front position and is the important reason for determining the label. Topic 4 uses "risk" as the first topic word, combined with "disease," "machine," and "learning," explaining that the topic may be related to both disease and AI. "Care," "prevent," "live," and "way" mentioned in topic 7 indicate that the topic may cover a relatively wide range, and the first word "personalized" stands for a tendency or method. The two sides above contribute to the label "Personalized healthcare." The information in topic 8 is scattered and refers to all aspects of DT. With the word "framework," the label is figured.

Topic prevalence. In STM, every paper will be given a weight for each topic. The larger the weight value, the more biased the papers to that topic. By statistically analyzing the frequency of each topic's appearance in papers, the popularity proportion of each topic is found, thus gaining insight into which topic is popular and which is niche. By calculating the average weight of all papers on a topic using the function `colMeans`, the prevalence of the topic among all topics was obtained and presented as a proportion. The total proportion of all topics is equal to 1. Figure 9 depicts the proportions of each topic using a pie chart. The largest segment of the chart represents topic 1, followed by topic 6 and topic 2. Six topics have a proportion exceeding 10%, while topic 4 and topic 5 account for the smallest proportions, 9% and 6%, respectively. The round outside classifies eight topics according to labels,

and three categories are finally gotten: "concept and framework," "related technology," and "application."

Topic correlations. Topic correlations describe connections between topics. Visualizing the relationships between topics helps to reveal the interplay and interdependence among different topics. The weights of each paper enable the calculation of correlation coefficients among various topics through the `topicCorr` function within the STM package. The range of calculated correlation coefficients from the function varies between -1 and 1 . The larger the coefficients, the stronger the correlation between the two topics. However, when the correlation coefficient is less than 0 , it means another case. With the correlation coefficient tending toward -1 , there is a large competition between the targeted topics. For example, a paper covers one topic that competes with another topic and a large probability that the last topic will not appear in this paper. Examining the competition between topics can provide valuable insights into the overall landscape of the paper, and it is interesting to explore why content related to competing topics does not emerge.

Figure 10(a) presents the correlations between topics. The intertwined squares with different abscissas and ordinates represent the relationship between the two topics. If the box is red, it means there is a positive correlation. If the box is blue, it means there is a negative correlation. A diagonal gray box indicates the correlation coefficients between the two same topics. A positive correlation was presented between topic 3 "Personalized Healthcare" and topic 7 "Precision medicine" (correlation coefficient

Table 2. Topic words and labels for topics.

Topic	FREX topic words	Label
1	Field, application, physical, human, relationship, digital- twin, entity, technology, virtual, world	Conception of DT
2	DT, healthcare, IoT, industry, support, complex, challenge, monitor, system, process, decision	Healthcare digitalization
3	Precise, medicine, clinic, computational, treatment, patient, model, enable, challenge, advance	Precision medicine
4	Risk, predict, disease, method, machine, learning, evaluation, life, result, help	AI disease diagnosis
5	Management, service, health, design, function, new, analyze, experiment, framework, technology	System management
6	Medical, model, system, simulation, patient, real-time, interact, environment, efficiency, treatment	Model simulation
7	Personalized, conceptual, emerging, care, prevent, inform, live, way, medicine, potential	Personalized healthcare
8	Knowledge, smart, analysis, propose, device, differ, collect, framework, learn, data	Framework development

equals 0.021). Figure 10(b) expresses the relationship between topics intuitively and visually. The circle's size indicates the topic's popularity, the green connection line indicates a positive correlation, and the red line indicates the competition and exclusion between topics. The strength of the association between topics is different and related to the width of the line. The thicker the line, the stronger the relationship between the two topics. As shown in Figure 10(b), "Personalized healthcare (topic 7)" and "Model simulation (topic 6)" have an obvious competitive relationship (negative correlation), and the corresponding red connection line is very thick.

Topic preference comparison

One of the key advantages of STM is its ability to estimate the relationship between papers and topics from multiple perspectives. Through the integration of covariates, STM has the capacity to contrast topic preferences across varying covariates. Such capability imparts a heightened level of understanding regarding the topics, enabling a thorough and profound examination of the paper content linked to each topic. The **estimateEffect** function within the STM package calculates the average effect size of each topic within the covariate model. It further furnishes statistical inference regarding these effects by means of confidence intervals and significance tests. The preference of each topic can be represented in Figure 11.

The point of each topic in Figure 11 corresponds to the x-label representing "Estimate," one output of function **estimateEffect**. "Estimate" is derived by computing a weighted average of both the global weight and the topic's impact strength on the covariate within each paper. "Estimate" serves as a means to ascertain the magnitude and direction of the topic's influence on the

covariate. The corresponding line of points represents the confidence interval of the effect size of each topic, which is 95% in this study. The large "Estimate" (larger than 0) means the topic has more preference for "Health enhancement," and the small "Estimate" (less than 0) means the topic has more preference for "Disease treatment." It is observed that topics 1, 2, 5, 7, and 8 prefer the covariate "Health enhancement," while topics 3, 4, and 6 prefer the covariate "Disease treatment."

In addition to examining the relative preferences of topics for the two covariates, the absolute preferences of topics for the two covariates were also calculated using the topic focus index. These results were visualized using a radar figure, as shown in Figure 12.⁵⁰ The radar figure showcases eight topics represented by the eight corners, which are illustrated in Figure 9. These topics are further divided into three distinct areas: "Concept and framework," "Related technology," and "Application." The two graphs extending from the center depict the topic's preference for each covariate, respectively.

From the figure, it can be observed that concerning the covariate "Health enhancement," the topics "Framework development" and "System management" demonstrate a strong absolute preference. In terms of the overall pattern, topics related to "Concept and framework" and "Related technology" show a greater inclination toward this covariate. On the other hand, for the covariate "Disease treatment," the topic "AI disease diagnosis" exhibits a clear absolute preference, whereas the topic "Conception of DT" displays the weakest absolute preference. The overall distribution of "Disease treatment" seems more balanced, with topics related to applications exhibiting a higher preference for this covariate.

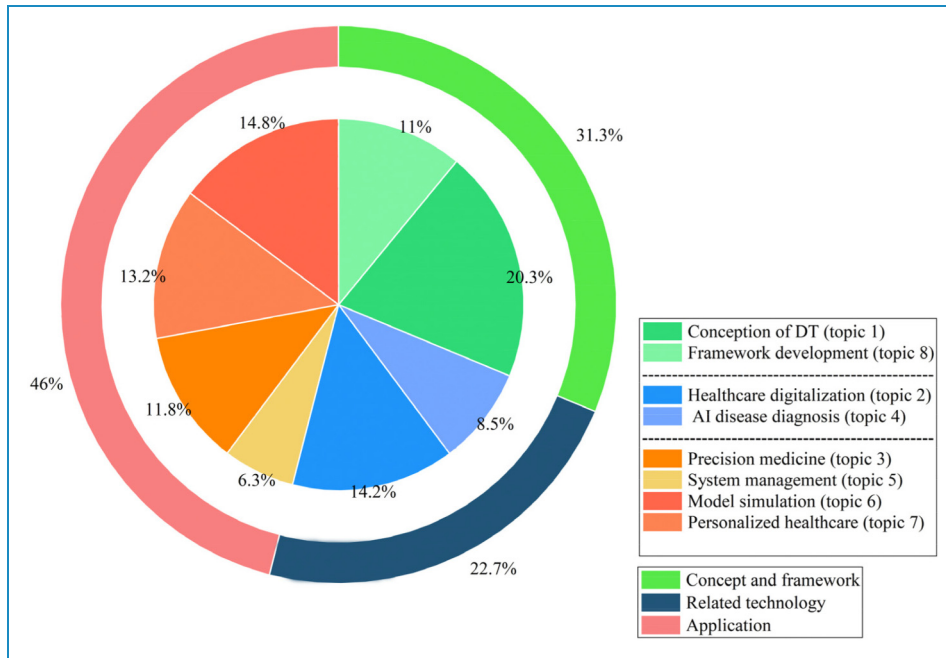


Figure 9. The proportion of topics.

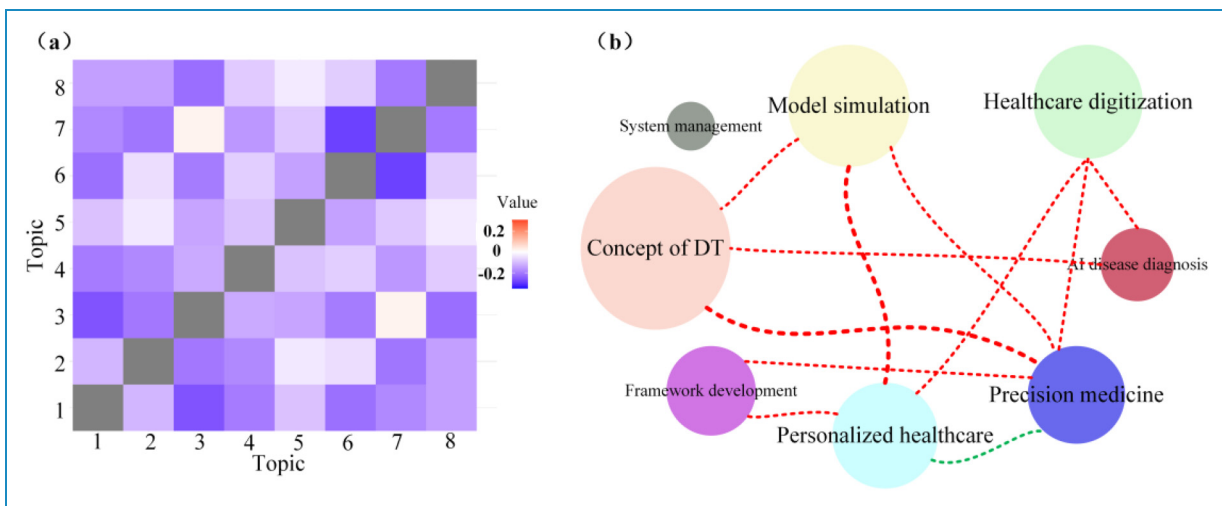


Figure 10. The correlations of topics: (a) Topic hotspot correlations figure; (b) Correlations figure of each topic.

Topic content comparison

The analysis of preference differences for the covariates “Health enhancement” and “Disease treatment” concerning a particular topic content can provide a detailed insight into the specifics of the research preference. Topic content comparison in STM is achieved by categorizing topic words based on their covariates.

Table 3 displays the specific content of the topic comparison. For a particular topic, the topic words were classified into a covariate, and the further away the topic word is

from the dotted line in the table, the more evident its preference for the covariate it was assigned to. The closer the word away from the dotted line in the middle, the weaker the topic word’s preference for the covariate. From another perspective, the words close to the line are ambiguous and require discussion.

As can be seen from Table 3, there are differences in the specific preferences of the two covariates “Disease treatment” and “Health enhancement” for each topic. “Conception of DT (topic 1)” divided the words “digital twin” and “healthcare” into two sides, respectively.

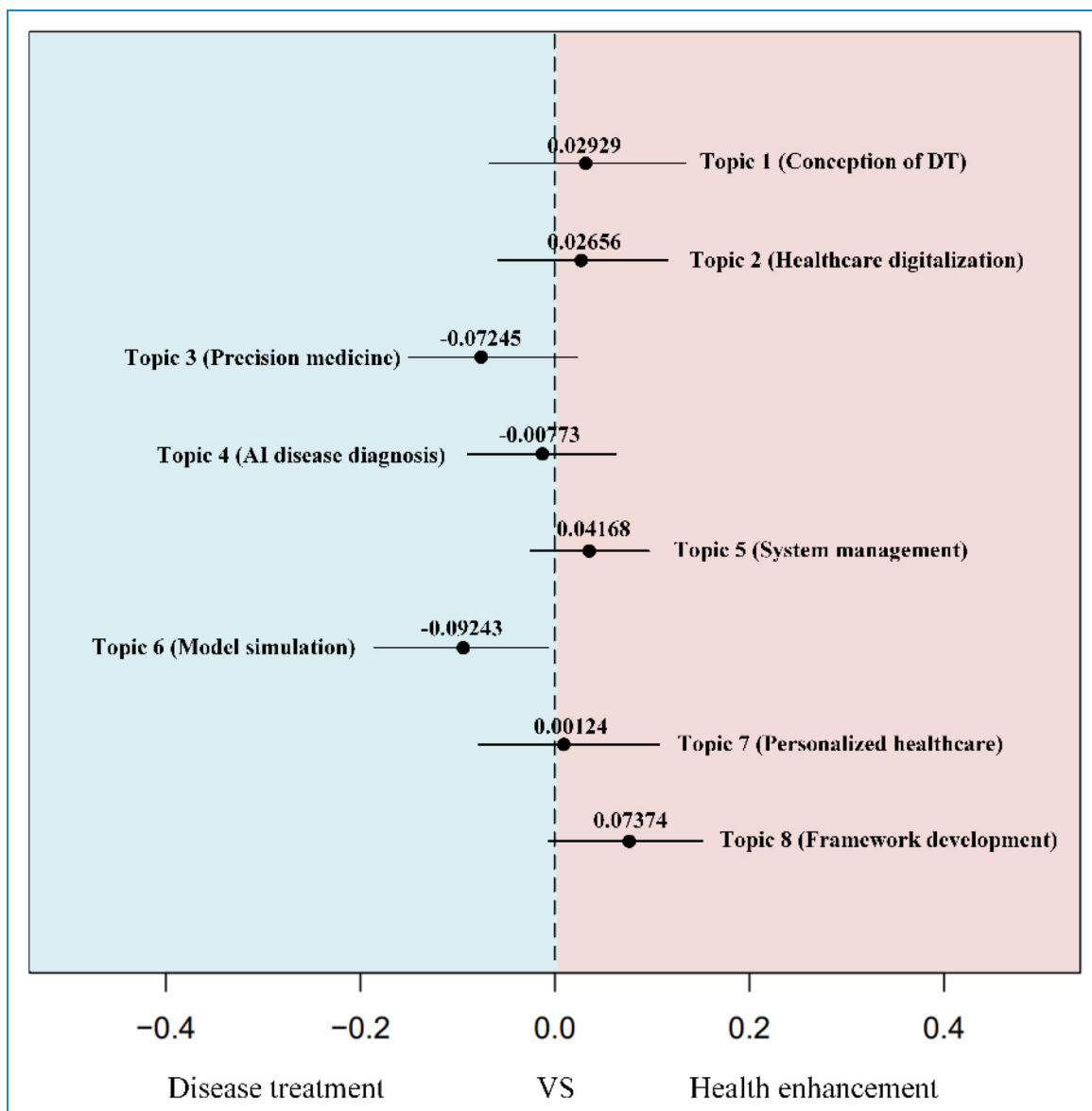


Figure 11. The forest plot for topic preference comparison.

“Healthcare digitalization (topic 2)” brings IoT into “Disease treatment,” while “AI disease diagnosis (topic 4)” splits “machine learning (ML)” into two sides. As for other topics, the two attributes also correspond to different contents, which will be discussed in Chapter 4.

Topic popularity changes over time

STM generates topic prevalence over time by plotting the distribution of topic proportions and provides a clear and effective way to analyze changes in topic relevance. However, due to the limited number of papers in the “DT + healthcare” field, some topics may exhibit trend changes that are difficult to explain during certain periods.

Figure 13 shows the expected topic proportion of the eight topics. The figure uses the linear fit to smooth the curves.⁴³ The time gap for 2018–2022 was divided into 20 parts by quarters, which adds the range of values on the x-axis. The number of papers published in the 4th quarter of 2022 is not fully counted by the review conducted, and Figure 13 only records the trends of the 2018 1st quarter to 2022 3rd quarter. “Conception of DT (topic 1)” shows a high growth percentage in 2020–2021. “Healthcare digitalization (topic 2)” presents a surge in the last two quarters of 2022. “Precision medicine (topic 3)” and “Model simulation (topic 6)” maintain a growth trend. “AI Disease diagnosis (topic 4),” “Health management (topic 5),” and “Framework development (topic 8)” are the three topics

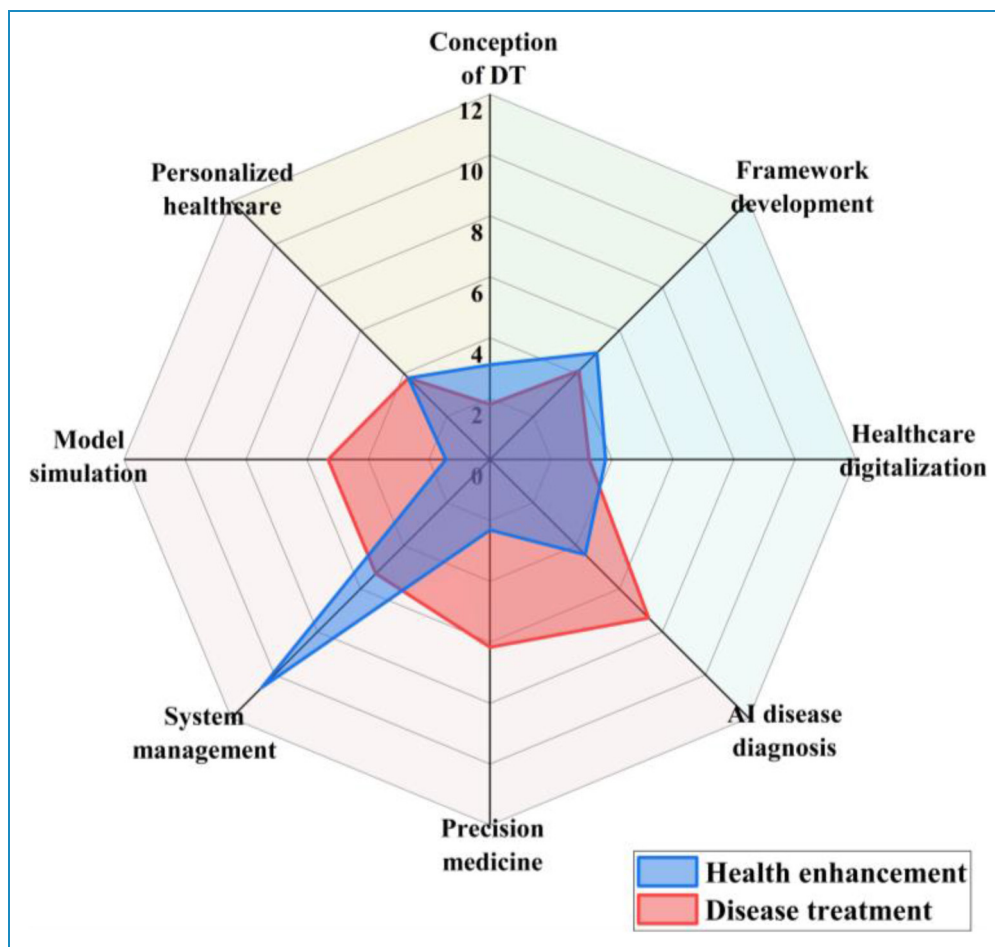


Figure 12. The topic preference of “health enhancement” and “disease treatment” attributes under the radar chart.

Table 3. Topic content comparison.

Topic	Disease treatment	Health Enhancement
1. Conception of DT	Healthcare, application, health, field, use	Physical, technology, human, twin, digital
2. Healthcare digitalization	Support, decision, thing, Internet, DT, industry	Technology, challenging, emerge, system
3. Precision medicine	Medicine, patient, computational, clinic, individual	Precise, use, data, new, health
4. AI disease diagnosis	Develop, study, model, risk, data, learn	Health, machine, framework, predict, disease
5. Health management	Healthcare, efficient, new service	Analysis, technology, medical, design, management
6. Model simulation	System, patient, simulation, treatment, develop	Base, device, collect, medical, model, data
7. Personalized healthcare	Personalized, emerging, conceptual, framework	Individual, technology, prevent, design, help, Health
8. Framework development	Diagnosis, disease, behavior, AI, data	Support, patient, system, learn, knowledge, Decision

Note: The comparative topic words in this table are extracted from the content comparison figure in STM. It should be noted that the frequency determines the preference for such topic words.

with the lowest proportion, and the trend is relatively stable. “Personalized healthcare (topic 7)” has extensive undulation changes, falling from a high percentage in the whole year of 2018 and rising sharply in the last two quarters of 2022.

Figure 14 presents the integrated and summarized popularity trends of all topics obtained through local polynomial regression fitting. The overall topic variations are depicted in a straightforward graphical form, facilitating the

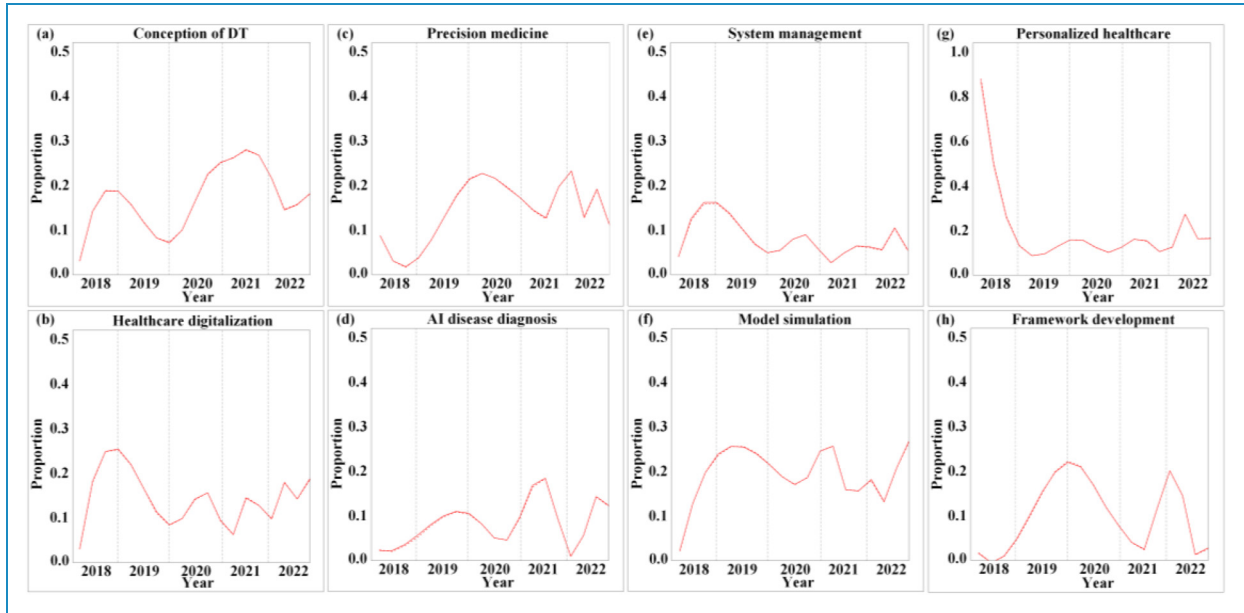


Figure 13. Excerpted topic proportion by quarters.

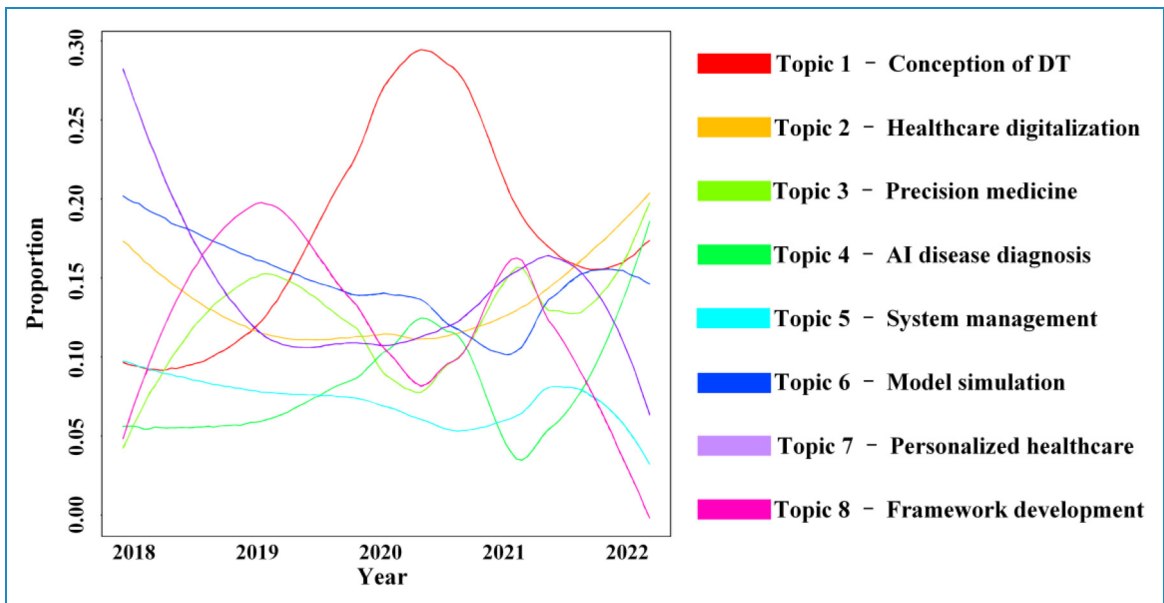


Figure 14. Summary chart of topic popularity trends.

comparison of their respective popularity changes. The overall graph shows that topics 1, 2, 3, and 4 exhibit an upward trend in popularity, with topic 4 indicating the most pronounced increase. Conversely, topics 5, 6, 7, and 8 demonstrate a declining trend in popularity, with topic 8 experiencing the most significant decrease. This graphical representation allows for a clear and coherent assessment of the popularity dynamics among different topics.

Discussion

Chapter 3 has shown the detailed topic information analyzed by STM. In Figure 7, the rising publications presented the broad prospects for the “DT+healthcare” field. The fusion of advanced technologies, such as AI, IoT, Block Chain, and Big Data Analytics, has paved the way for innovative solutions that improve patient outcomes and enhance people’s quality of life.^{51,52} The

emergence of the COVID-19 pandemic further strengthened the public's requests for digital and efficient healthcare, highlighting the importance of remote patient monitoring, virtual consultations, and data-driven decision-making.^{32,53,54} The above factors have made "DT + healthcare" an exciting research area with significant potential to transform healthcare and promote digitalization.

The convergence of cross-technologies in digital twin: Fully penetrating the healthcare industry

In healthcare, DT is utilized to create virtual replicas of physical objects or systems, including organs, medical devices, and even the entire human body. The primary objective is to facilitate real-time interaction and seamless integration between the digital and physical realms. To harness the full potential of DT, it becomes crucial to develop robust structures, which is why numerous studies are dedicated to exploring DT frameworks (topic 8). Song et al.⁵⁵ constructed a five-layer DT healthcare facility management framework. The framework can widely integrate a large amount of heterogeneous data and real-time data queries and achieve better resource allocation. The framework is used in the facility management of Shanghai Tongji Hospital and achieved good results. Pirbhulal et al.⁵⁶ proposed an automated DT framework for the cybersecurity of IoT healthcare systems. The agriculture simulated potential threats and vulnerabilities by generating DT of IoT systems, preventing possible cyberattacks, and improving the resistance of IoT healthcare systems to attacks and overall security. In general, frameworks help better understand, analyze, or predict the behavior of an actual entity. By extracting entity features, DTs can have visibility and expansibility, and research can continuously optimize and integrate new ideas based on the current framework.

Digital twins in healthcare: Technical components. Supported by the existing framework, the composition of DT can be modularized to provide the impetus for construction. Specifically, the main technical components of DT can be divided into three aspects. Firstly, big data is the foundation of the composition of DT and the key to real-time analysis. In terms of healthcare, the source of various physiological data usually comes from commercial health data acquisition devices or sensors in the environment. These devices/sensors should be ergonomically designed to avoid making people uncomfortable and gather diverse data. Existing studies have used smartphones, watches, and other wearable devices to collect parameters like daily food intake, exercise, sleep time, and mood changes for DT development.^{21,57,58} Intrabody microrobots offer another pathway for data collection. Negnab et al.⁵⁹ have developed controllers for magnetic levitation microrobots. These microrobots have the advantage of non-contact movement within the

body, providing significant convenience for drug delivery, imaging, and collecting sensitive data. Besides new equipment, a wider range of heterogeneous data is essential for realizing big data. Motion data, electroencephalogram signals, electromyogram signals, genome data, and microbial information hold the potential to contribute to enhancing the accuracy and effectiveness of DT.⁶⁰

The second aspect of DT-related technology is establishing a channel to communicate reality entities and virtual twins, now known as the IoT technology. IoT significantly contributes to healthcare information digitalization (topic 2). After being applied to healthcare, IoT can be further differentiated according to different purposes. One of the differentiations is the Internet of Medical Things (IoMT), which aims to network various medical equipment to achieve intelligence and automation. By combining with DT, the IoMT can construct real-time internal medical imaging of the human body, offering remote diagnosis and surgery possibilities^{61,62} and solving issues of delayed cancer diagnosis.⁶³ Another differentiation in IoT technology specific to healthcare is the Internet of Healthcare Things (IoHT). Devices connected to the IoHT can be various, such as smart bracelets, smart watches, smart scales, and smart medicine boxes. By recording changes in data, the IoHT can construct DT of humans, helping individuals better manage their health, prevent diseases, and improve their quality of life.^{64,65}

The third technical aspect of DT involves maximizing the potential of collected multimodal data with high quality and minimal latency. One promising approach to achieve the goal is AI (topic 4), which can continuously process data and realize evolution using iterative algorithms. Kaul et al.⁶⁶ analyzed the critical role of AI in healthcare DT through the case of cancer care. AI can easily integrate comprehensive personal DT data and support care suggestions considering treatment effects and patient preferences. Besides, AI can simultaneously analyze a large amount of clinical data and learn from published journals, which help accelerate the update of treatment methods. Existing research has promoted the combination of healthcare DT and AI through various ML algorithms. Ferdousi²³ developed a dynamic learning algorithm selection framework for predicting diseases. The framework can automatically match the best algorithm for related diseases timely. Allen et al.⁶⁷ utilized Variational Autoencoder ML methods, which can reconstruct input data and predict outcomes, making up for any gaps or mistakes in the dataset. Edge computing has the potential to expedite medical data validation and computation. It can improve the time-intensive process of traditional drug development, especially after integrating IoMT and deep learning algorithms.⁶⁸ Distributed computing nodes are crucial in reducing data transmission latency and alleviating server burdens. Additionally, edge computing offers the advantage of avoiding potential data

leakage, server crashes, and data loss issues arising from traditional single-point computing.

In summary, diverse data acquisition channels and efficient, secure data transmission methods play a crucial role. Additionally, advancements in AI models and ML algorithms contribute significantly to the continuous enhancement of DT's technical components. Furthermore, technologies like 3D modeling and cyber-physical systems continually augment the capabilities of DT, leading to its growing complexity. These advancements have revolutionized traditional healthcare, empowering individuals to take greater control over their health. However, paying attention to data security, personal privacy protection, and AI ethics in data acquisition is also essential. As a result, the scientific regulation of technology is anticipated to emerge as a crucial concern for the future of DT in healthcare.

Digital twins in healthcare: The applications. Up to now, DT has found diverse applications in various healthcare purposes. One of the primary and currently achievable applications is DT-based simulation (topic 6). Leveraging its excellent visibility and superior data transportation capabilities, DT simulation can provide real-time, cost-effective insights into the status of complex systems. The generated data by DT simulation can compensate for data lacking of rare diseases and overcome some ethical hurdles.⁶⁹ Operations simulations on DT can significantly enhance surgeons' ability to improve the success rate and efficiency of operations, particularly when they encounter complex and rare conditions.⁷⁰ The implementation of distributed DT healthcare units enables the effective monitoring of the transmission of COVID-19 in medical units with vulnerable populations. By rapidly conducting simulations based on real-time data, the spread of the virus can be disrupted, thereby improving the effectiveness and safety of treatment.⁷¹ DT breaks down the barrier that traditional computer simulation may have huge errors with physical entities. By establishing reliable and intuitive simulation models, DT can collect all-around influencing factors in the physical world and accelerate the accuracy of healthcare, increasing the success rate of treatments.

In other applications, DT can be categorized into two distinct types. The first category employs DT to create personal copies, while the second category utilizes DT to generate replicas of public systems or facilities. Precision (topic 3) and personalization (topic 7) are the two main requirements for personal DT copies and the urgent need for future healthcare.⁷² In cardiology, Corral-Acero et al.⁵⁰ reviewed the present attempts of DT application in precision cardiology. The combination of mechanical and statistical models in developing future patient copies has the potential to identify valuable data, guide treatment decisions, and evaluate prognoses. The construction of DT for various cancers showed the directions for future cancer treatment.⁷³ Virtual twins that integrate high-performance

computing and modeling capabilities can fully imitate the physical properties of entities. They can repeatedly perform trajectory prediction on physical conditions and treat simulations, with a precise focus on preventing and monitoring cancer.⁷⁴

Specifically, dynamic twins for personalized healthcare make people calmer about disease manifestations. The personal DT presents the physiological state of the individual,⁷⁵ and the concepts of "normal" and "health" will be transformed into specific regular periodic changes in individuals because of DT's meticulous observation and detection of the human body.²⁸ When deviations include increased Body Mass Index, rising blood pressure, and changing biological and clinical parameters, the tailor-made life rhythm and diet plan can quickly help humans correct the parameters.⁷⁶ Athletes and bodybuilders can use DT to accurately judge whether exercise posture is standard and scientifically calculate the amount of training to achieve the best effect.²¹ Another critical issue for personalized DT is its potential privacy and ethical implications. The monopolization of DT by capital can create an unequal situation where not everyone can access or use it, resulting in wider social inequality and discrimination against groups without DT access.²⁸ Additionally, the reliance on big data for DT necessitates privacy considerations. An active, comprehensive DT life cycle privacy protection system is essential to protect interests without compromising privacy. Ensuring transparent data flow is crucial for DT's personalized application.²⁷

The last main idea for DT application in healthcare is creating DT for copies of public systems or facilities. The possible application scenarios include hospitals, elderly communities, rehabilitation centers, medical resource allocation centers, etc. In such DTs, people/medical resources play the random variables, and the environment becomes the object of DT. The ultimate goal of DT for the above facilities is achieving optimized management, including efficient-reasonable resource allocation and rapid processing (topic 5). Karakra et al.³⁶ proposed a hospital DT to simulate discrete events and emergencies for better hospital management and resource distribution. Chen⁷⁷ centered on the needs of older people and built a DT of disease and health management for older people. With more system-based applications continuously developed, the future of healthcare system management is expected to embrace comprehensive informatization, intelligence, and enhanced efficiency.

This section presents the specific content of the eight topics identified through STM analysis. "Conception of DT (topic 1)" and "Framework development (topic 8)" discuss the conceptualization and framework development of DT in healthcare. "Healthcare digitalization (topic 2)" and "AI disease diagnosis (topic 4)" focus on the utilization of IoT technologies and the role of AI in healthcare. The remaining topics turn into various specific applications of DT in healthcare. Figure 15 gives a schematic diagram of the main discussion in this section. DT provides a platform

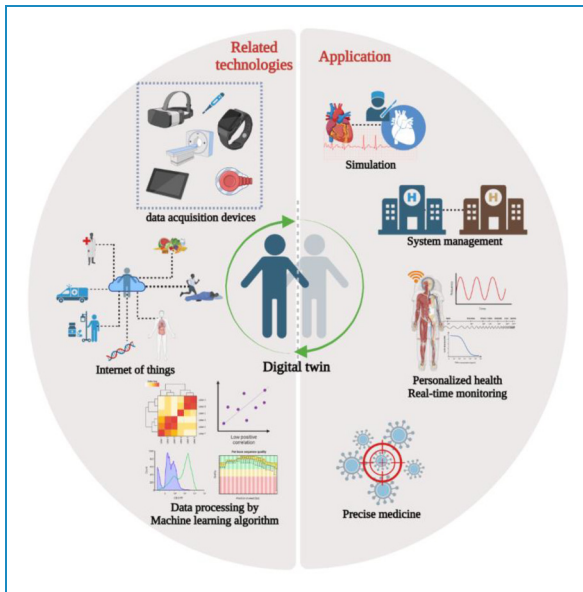


Figure 15. Concerned content for STM analysis.

that integrates modern technologies such as intelligent sensors, IoT, and ML, offering pathways for future healthcare. Real-time monitoring of bodily data will expose potential anomalies, while precisely tailored lifestyle plans and dietary programs will assist individuals in maintaining constant well-being. The DT public facilities enable better resource allocation. The rare disease data generated by DT can provide valuable references for treatment while ensuring patient privacy protection. Simulated surgeries based on DT have significantly reduced training costs and improved the success rate of operations. Additionally, the utilization of DT in medical image and data processing methods has demonstrated superior accuracy and speed compared to traditional approaches.

The interconnected evolution of research hotspots: Understanding topic distribution and topic interaction over time

The diversity of DT applications in healthcare has led to the exploration of various topics, resulting in identifying eight distinct areas of discussion. In the convergence of cross-technologies in digital twin, fully penetrating the healthcare industry, the detailed contents of these topics have been extensively discussed. The following section will delve into the trends and preferences associated with each topic. This analysis will be based on the number of papers related to each topic within the entire dataset and its correlations with other topics.

The world's attention to DT has prompted researchers to explore the diverse possibilities of DT concepts based on the unique characteristics of their subjects. Generally

speaking, the development and application of emerging technologies require feasibility verification and discussions on limitations. This helps determine whether DT can be realized under existing technical conditions or in the foreseeable future. Moreover, careful considerations are essential to avoid causing harm to human development or causing unnecessary resource waste during the implementation of DT.⁷⁸ Currently, verifications and discussions are underway regarding the integration of DT with established technologies and its innovative applications within the healthcare domain.^{79,80} As a result, papers involving the concept of DT (topic 1) presented a developing popular trend. Topic 1 reached a peak in popularity between 2020 and 2021, with the highest distribution reaching 30%, as shown in Figures 13(a) and 14.

Researchers are increasingly inclined to integrate traditional model simulation (topic 6) with DT, resulting in a sustained and popular trend, as depicted in Figure 13(f). DT fully expands the traditional idea of “virtual copy,” and its brilliant performance explains the trend of this topic. With the aid of DT, virtual copies can offer real-time feedback to their real-world counterparts, facilitating a two-way interaction of reality-virtuality. This breaks away from the conventional notion of people controlling the virtual objects' natural environment, aiming to ultimately achieve a seamless coexistence of reality and virtuality.⁸¹ Exoskeletons have emerged as valuable tools for aiding patients in their recovery process. When integrated with DT, they can significantly improve human-computer information interaction, enable the generation of patient-specific gaits, and enhance treatment autonomy and safety.³⁷ The distribution of topic 5 (System simulation) in Figure 14 bears a resemblance to the shape of topic 6, albeit with a relatively smaller overall proportion. Both topics involve simulations, but topic 5 focuses on a more specific target, primarily concerning real-time monitoring and management of physical entities and systems. Due to the similarity in the distribution patterns and shared characteristics in topic content, the interpretation of the popularity of topic 5 aligns with that of topic 6.

The emergence of 5G technology pushed the future communication requirement of Massive Machine Type Communications (mMTC),⁸² which provides great support and possibilities for IoT. With more connected devices, sensors, and seamless communication, healthcare digitalization (topic 2) has been greatly accelerated. In 2017, the international organization for standardization International Telecommunication Union and 3rd Generation Partnership Project put forward communication requirements for 5G technology.⁸³ Under the guidance of specific rules, 5G-based healthcare digitalization and IoT technology had more specific goals and standards.⁸⁴ As a result, the 2018 popularity peak of topic 2 in Figure 13(b) may be attributed to the standardization of 5G technology, while the second peak in popularity for this topic in 2020 can

be attributed to the official commercialization of 5G technology in 2019. With the widespread adoption and use of 5G technology, there has been a steady increase in the trends related to “Healthcare digitalization (topic 2).” This indicates the growing importance and impact of digitalization in the healthcare sector, facilitated by advancements in communication technology such as 5G.

“Precision medicine (topic 3, Figure 13(c))” and “Personalized healthcare (topic 7, Figure 13(g)),” as the only pair of topics with a positive correlation, both focus on individual differences and customized treatment/healthcare options. Precision medicine delivers precise and targeted interventions. Personalization may not always be precise, while precision has decided personalization by its implementation process.⁸⁵ However, personalization becomes more significant as healthcare evolves to encompass all aspects of life, such as offering personalized guidance from personal trainers and nutritionists. It works in conjunction with precision medicine to enhance the quality of human life. The interconnectedness is why topics 3 and 7 are regarded as distinct yet interrelated subjects.

For individual trends of the two topics, “Precision medicine (topic 3)” and “Personalized healthcare (topic 7)” show different shapes in Figure 14, resulting in different reasons for the shapes. As a high-tech-oriented concept, policy promotions of DT have undoubtedly given researchers a hotspot target. The prevalence of “Precision medicine” seems related to US Food and Drug Administration’s (FDA) policies. The FDA is recognized globally as a leading regulatory agency for food, drugs, and medical devices, with a reputation for science-based regulation and oversight. The standards for healthcare-related products and a vast amount of statistical data released by the FDA are served as motivation and incentive for researchers. The Digital Health Software Precertification pilot program launched by the FDA in 2017 aimed to optimize the review process of medical software and promote the use of new technologies.⁸⁶ Various countries also actively explored the possible application of DT and have introduced various policies to promote the development of DT.^{87,88} The policies and programs greatly accelerated the application of DT and explained the soaring popularity of the topic of “Precision medicine” in 2019.

Developments in omics have made “Personalized health” popular. Searched in WoS, the growth rate of publications concerning omics in 2021 came to 47.24%, while the growth rate in 2020 was only 21.57%. Omics is a discipline that emerged from the advances in genome sequencing technologies over the past two decades,⁸⁹ focusing on studying the composition and interactions of biological molecules. It encompasses various “omes,” including genome, proteome, transcriptome, immunome, epigenome, metabolome, and microbiome. Each has its specificity and value demonstrated in clinical research.^{90,91} The development and

breakthroughs in omics and personalized health-assisted DT are mutually reinforcing, offering analytical methods at the microscopic level and expanding access to diverse analytical data. Traditional medical data, personal gene-related information, recent lifestyle data, relevant environmental data, and all kinds of data can be included as the analysis object. Thanks to the combination of omics and DT in healthcare, the popularity peak of “Personalized healthcare (topic 7)” in early 2022 could be explained.

Through Figure 13(h), the topic “Framework development (topic 8)” reached two peaks of popularity in late 2019 and late 2021, respectively. The previous peak of popularity could be attributed to the popularity of DT. Papers on “Framework development (topic 8)” are only discussions of the development of the DT framework without actual practice. Although it encompasses DT, healthcare, and many related technologies, the trends of this topic have consistently remained lukewarm. The last peak could be attributed to the COVID-19 pandemic. The pandemic pushed people to explore smart frameworks to avoid infection⁵⁴ and deal with the shortage or disturbing confusion of medical resources caused by the surge of COVID-19 patients.³² The pandemic also promoted the popularity of “AI disease diagnosis (topic 4).” AI disease diagnosis can help mitigate the spread of COVID-19 by reducing human-to-human contact.²³ Furthermore, it eliminates the need for patients to physically visit hospitals, saving them valuable time and ensuring a certain level of privacy protection.

Moreover, the latest studies have raised new possibilities for healthcare DT development. One concept that holds great promise is the Metaverse. Metaverse represents a comprehensive framework that integrates environment, user interface, interaction, and social value.⁹² DT provides basic components for constructing Metaverse, while Metaverse provides DT with a broader platform to promote more diverse and convenient interactions. Moztarzadeh et al.⁹³ developed a DT identification method for cervical vertebral maturation (CVM) using ML algorithms. Compared with traditional identification methods, the DT CVM identification method in Metaverse has the advantages of low cost and high accuracy. Moztarzadeh et al.⁹⁴ and Jamshidi et al.⁹⁵ used a variety of ML algorithms to analyze breast cancer patient data and construct DT, aiming to achieve real-time stable diagnosis in Metaverse. The simulation results showed the applicability and simplicity of DT and are promised to be used in generating credible assisted treatment decisions.

Previous research has elevated DT to a broader platform. Most of the papers reviewed in this review emphasize the importance of DT’s interaction in both the physical and virtual realms. At the same time, within the Metaverse concept, the DT in the research mentioned above highlighted immersive interactions in the virtual world, where individuals engage with their DT avatars. The immersive experience

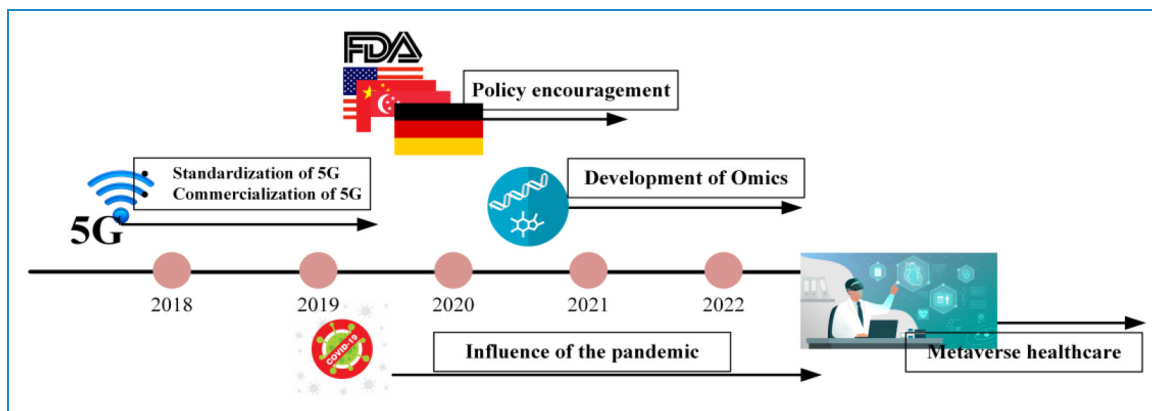


Figure 16. Possible reasons for the popularity of topics.

offered by the Metaverse expedites the sharing of medical resources and facilitates the advancement of remote healthcare. Undoubtedly, the Metaverse is poised to play a significant role in driving the future development of DT.

Regarding the relationship between topics, except for the positive correlation between topic 3 and topic 7, the remaining topics show clear competition, meaning that when a topic appears in one paper, other topics are unlikely to be present. The intense competition among various research topics underscores the distinct delineations in current studies. Each facet addressed in this review offers a comprehensive discourse, providing an articulation of a particular subset pertaining to DT within the healthcare domain. The multiplicity and rivalry of these topics vividly exemplify the potential trajectories for advancing DT in healthcare. Each of these trajectories boasts its own distinctive content, poised to yield further unparalleled insights in the times ahead.

In summary, this section examines the distribution of popularity proportions and the changing trends of each topic over time, as depicted in Figures 9 and 13. It also analyzes the potential reasons behind the observed distribution and changes. Figure 16 demonstrates the main discussion content of this section: 5G communication development, positive policy encouragement of various countries for healthcare DT, advancements in Omics research, and the impact of unexpected epidemics. These four factors contribute to the fluctuations in the popularity of different topics and the distribution of topic popularity within the “DT + healthcare” field. In the future, the development of DT will be diverse. The emergence of the Metaverse and the fast growth of AI will allow DT to progress faster and generate more variants.

Exploring topic preferences and topic words biases

As the two most important research directions in healthcare, the attributes “Health enhancement” and “Disease treatment” covered all the dataset’s papers. According to the relative preference shown in Figure 11, all eight

topics are more inclined toward the direction of “Health enhancement,” where five of eight topics prefer the covariate “Health enhancement” and three of eight topics prefer the covariate “Disease treatment.” In discussing absolute preferences, as shown in Figure 12, the preference of topics concerning covariates is consistent with those presented in Figure 11. It is worth noting that topic 5 (System Management) exhibits a significantly pronounced preference for “Health enhancement.” Mainly built upon Figure 11 and helped by Figure 12, this section will explore the preferences of topics to the covariates.

Topics with the most distinct preferences are “Precision medicine (topic 3, Estimate = -0.07245)” (The introduction to “Estimate” were shown in Topic preference comparison), “Model simulation (topic 6, Estimate = -0.09243),” and “Framework development (topic 8, Estimate = 0.07374).” The preference for “Precision medicine (topic 3)” can be captured from the topic words. The topic words “patient,” “medicine,” “clinic,” and “treatment” can be assigned to the covariate “Disease treatment.” Meanwhile, the word in topic 3 “computational” shows that this topic focuses on using advanced computing methods to promote medical development, using DT to provide patients with excellent treatment. Advanced computational methods will fully use various physiological data collected to generate as accurate a DT as possible to accurately reflect the actual situation of the physical individual.

“Model simulation (topic 6)” prefers the covariate “Disease treatment,” which illustrates that papers on the topic are mainly focused on the medical field. “System” and “patient” are two core words for the “Disease treatment” side, which meant that the simulated drug effects,^{42,67} surgical and rehabilitation procedures,^{37,40} and system deployment^{34,70,96} frequently appear in papers corresponding to this topic. In contrast, papers about health enhancement can seldom appear on this topic. There is significant untapped potential in modeling and simulation for the “Health enhancement” side, including various aspects such as

simulating personal food intake modeling, daily exercise levels modeling, and nutritional planning simulation. The above areas offer ample opportunities for exploration and development.

“Framework development (topic 8)” is preferred to the covariate “Health enhancement” distinctly. Combining concepts such as Healthcare 4.0 and Smart City,^{97,98} papers in this field focus on utilizing DT for information management and supporting decision-making processes. Laamarti et al.⁹⁷ proposed a DT framework based on the ISO/IEEE 11073 standard. This framework aimed to standardize the collection of a large volume of health data from personal devices to enable information fusion and create individual DTs. The DT can improve the quality of personal life and enhance individual well-being, driving the development of smart healthcare in smart cities.

However, some papers investigate frameworks for the “Disease treatment” side. Aloqaily et al.⁹⁹ focused on the development of DT in the context of extended reality (XR) and constructed an XR-DT system framework for the healthcare domain. The framework explored immersive healthcare services, where individual physiological data is presented in holographic 3D models. Such an approach can revolutionize the way medical professionals and patients interact, promote the development of remote healthcare, and enable personalized medical treatments.

The above two frameworks addressed the enhancement of health and treatment aspects, respectively. The DT framework by Laamarti primarily interacts with the physical individual itself, while the Aloqaily proposed framework emphasizes healthcare professionals’ feedback role at the DT interface. Based on the appearance and data presented by DT, suitable treatment methods could be provided as feedback to individuals by medical professionals. Meanwhile, the treatment information could also be transmitted back to the information management layer of the DT to facilitate the evolution of the DT.

“Conception of DT (topic 1, estimate .std = 0.02929)” and “Healthcare digitalization (topic 2, estimate .std = 0.02656)” show relatively weak preferences. Observing the preferred topic words of the two covariates in “Conception of DT (topic 1),” it can be found that papers on topic 1 that prefer “Health enhancement” concerned the human DT. Human DTs often function as personal private doctors to make real-time monitoring and prediction. People can instantly adjust the rhythm of life to prevent problems.^{27,79,100,101} Comprehensive personal data enables better control of common chronic diseases such as obesity and high blood pressure. For the “Disease treatment” side, the papers on topic 1 like to talk about DT applications in medicine and explain the connotation of DT to accelerate digital treatment. The specific content involves trauma management,⁸⁰ remote treatment,¹⁰² and real-time acquisition of patient status.¹⁰³

For the topic “Healthcare digitalization (topic 2),” papers that favor the covariate of “Health enhancement” focus on

how systems are architected and the challenges faced with this emerging technology. While papers on topic 2 that prefer the covariate “Disease treatment” focus on the role of IoT technology, discussed in Digital twins in healthcare: technical components “System management (topic 5, estimate .std = 0.04168)” applies DT to manage various systems or monitor the conditions of medical equipment. In Figure 11, the topic expresses a weak preference, while in Figure 12, the topic expresses a strong preference. DTs in topic 5 care about the operation of the system and resource allocation. From Table 3, papers corresponding to the topic words which prefer the “Health enhancement” side prefer to discuss the management system’s design. In contrast, papers corresponding to the “Disease treatment” side prefer the effect of the DT system.

The remaining two topics are “AI disease diagnosis (topic 4, estimate .std = -0.00773)” and “Personalized healthcare (topic 7, estimate .std = 0.00124).” These topics have the least covariate effects, which makes the two topics’ preferences for the covariates unclear. But in terms of absolute preference, topic 4 showed a preference for the “Disease treatment” side. The topic words of “AI disease diagnosis (topic 4)” who biased toward the covariate “Health enhancement” in Table 3, focusing on building the framework for disease prediction with AI,^{22,73,104} while the topic words biased heading to “Disease treatment” focus on possible risk analysis of the model to determine the safety and reliability of DT.^{105,106} The topic “Personalized healthcare (topic 7),” when biased to the covariate “Health enhancement,” cares about the prevention of diseases through DT.²⁸ When biases to the covariate “Disease treatment,” papers care about the emerging models and model construction for personalized medicine.¹⁰⁷

This section presented preferred research topics for two broad research attributes in healthcare, “Health enhancement” and “Disease treatment.” Figure 17 is given to show the detail discussed in this section. Through the figure, topic 3 and topic 6 only provide the content distribution on the “Disease treatment” side. In contrast, topic 8 only provides the content distribution on the “Health enhancement” side for the three topics that obviously prefer covariates. It can be found that the content of the covariate “Health enhancement” focuses on the concept of DT, implementation methods, and the framework, while the content of the “Disease treatment” was more directional, tightly around the word “healthcare,” or focusing on specific diseases, such as cardiovascular disease, multiple sclerosis, and stroke. The Specific links in the medical system, such as the allocation of resources between operating rooms or wards, are also included in the “Disease treatment” covariate.

Limitations and future work

This study included papers on the subject “DT and healthcare” from 2018 to 2022 to construct a dataset for STM

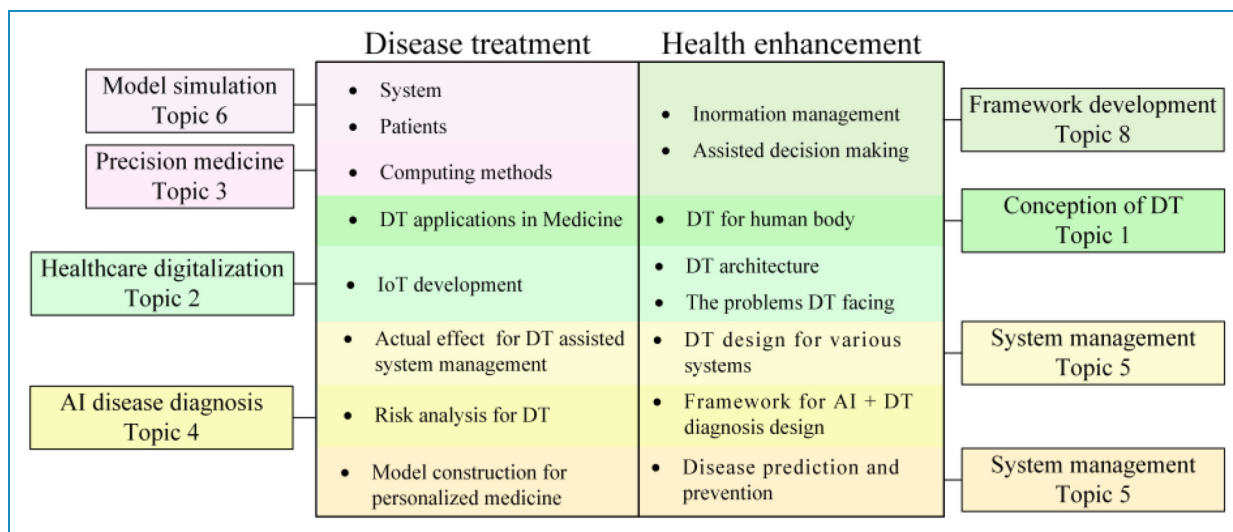


Figure 17. Content distribution for covariates “disease treatment” and “health enhancement.”

analysis. However, it has to be admitted that the sample size of the dataset for the study is not large, and the number of papers is concentrated in 2021–2022. As a concept that is popular in the industrial manufacturing field since 2018, the development of DT in healthcare is still lagging. In this review, three databases and references of included papers were used as data sources to increase the amount of data as much as possible. This collection method made the results presented by STM more diverse and accurate. With the rapid development of DT in healthcare, more publications in this field are promised to appear, and the branches for DT will be more abundant. Diverse results can be presented, which is expected to solve the limited sample size.

In addition, the review progress when using STM still included subjective considerations, which makes the review include human influence. STM has only been used as a method of systematic review in recent years, and there are no clear guidelines and regulations to evaluate the results of this method. Therefore, in addition to using software data as the main outcome indicator, two experts in related fields also formulated the results, reflected in selecting several topics and calibrating covariates. The software can only give a rough range of the number of topics based on principles. As a result, the attributes of papers can only be ensured correctly through manual classification. The above operations can also be supported by reviews that also use STM.^{18,108}

Considering the aforementioned limitations, future STM research can be directed toward enhancing the overall quality. The field of “DT + healthcare” is poised for rapid development, and the increasing number of publications will provide more extensive materials for future quantitative reviews, thus addressing the issue of insufficient sample size. To refine the selection of topics, future research can explore cross-checking or multi-text analysis methods in

conjunction with analytical approaches. The combined approach will help determine the number of topics using objective data-driven techniques and reduce subjectivity in the topic selection process.

Conclusion

This review employed STM to analyze the changes and focus for the “DT + healthcare” field. The results of STM were visually presented in the form of figures or tables and discussed. The main findings of this review can be concluded in three aspects: (1) Papers in the “DT + healthcare” field mainly cares about technologies integration and practical applications. IoT and AI are technologies that are of great concern. One helps establish a connection between the physical and virtual worlds, and the other makes the DT more intelligent and close to physical entities. The three main types of DT applications in healthcare are diverse: precision medicine, personalized healthcare, and the management of various public health systems. (2) The reasons behind the changes in the popularity of different research topics are different. In addition to the common cause of the widespread popularization of the concept of DT, the promotion of various policies, the COVID-19 pandemic and the development of technologies such as omics, AI, and 5G also impact the popularity of different topics. They are becoming important reasons to promote the accelerated integration of DT into healthcare. (3) The current topics are generally biased toward the attribute of “Health enhancement.” But many topics follow different contents although having clear preferences to an attribute, which indicates that the application of DT in healthcare is comprehensive, with a wide variety of content. Several studies highlighted the challenges and problems like data security

and privacy protection, which were the obstacles that should be overcome.

The development of DT in healthcare found in this review is inspiring. The concept of DT is continuously evolving, along with emerging frameworks and standards. Various technologies can already be integrated into DT, and the applications of DT are also emerging, which greatly help personal health, disease treatment, and resource allocation. Currently, DT faces challenges in data exchange and processing. Acquiring human body data requires smaller sensors, while real-time data transmission and processing necessitate robust communication methods and powerful processing methods. The increasing prevalence of 5G communication, IoT technology, and AI will be crucial in future healthcare DT research. The COVID-19 pandemic has demonstrated the need for greater emphasis on intelligent healthcare to address potential future epidemics effectively. The governments of all countries should actively implement policies to promote the development of DT in healthcare and establish regulations to define data collection boundaries, ensure data security, and protect public privacy.

Acknowledgments: BS is supported by a Shanghai Pujiang Fellowship. The authors would like to thank Jennifer Qiao for her English improvement.

Contributorship: This protocol was written by ZYW with supervision from all five co-authors. All authors contributed to the review process and the manuscript and approved the final version.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical approval: This review is not subject to ethical approval since the review did not have participants (humans or animals).

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article was supported by the Shanghai Pujiang Program, under grant number 21PJ1404000, the National Natural Science Foundation of China, under grant number 62103252, and Shanghai Sports science and technology project, under grant number 22Q003.

Guarantor: JT, CQD.

Data availability: The dataset used and analyzed during the current manuscript is available from the corresponding author upon reasonable request.

ORCID iD: Jing Tao  <https://orcid.org/0000-0003-4106-4492>

Registration and protocol: The review was not registered.

References

1. Ahsan MM and Siddique Z. Industry 4.0 in healthcare: A systematic review. *Int J Inf Manag Data Insights* 2022; 2: 100079.
2. Tortorella GL, Fogliatto FS, Vergara AMC, et al. Healthcare 4.0: trends, challenges and research directions. *Prod Plan Control* 2020; 31: 1245–1260.
3. Grieves M. Origins of the Digital Twin Concept. Published online 2016. DOI: 10.13140/RG.2.2.26367.61609
4. Gartner's Top 10 Technology Trends 2017, <https://www.gartner.com/smarterwithgartner/gartner-top-10-technology-trends-2017>
5. Gartner Top 10 Strategic Technology Trends for 2018. Gartner. Published 2017, <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018>
6. Emerging Technologies: Revenue Opportunity Projection of Digital Twins, <https://www.gartner.com/en/documents/4011590>
7. Fei T, Weiran L, Jianhua L, et al. Digital twin and its potential application exploration. 2018; 24: 1–18. DOI: 10.13196/j.cims.2018.01.001.
8. Manocha A, Afaq Y and Bhatia M. Knowledge-based systems digital twin-assisted Blockchain-inspired irregular event analysis for eldercare. *Knowledge-Based Syst* 2023; 260: 110138.
9. De Benedictis A, Mazocco N, Somma A, et al. Digital twins in healthcare: an architectural proposal and its application in a social distancing case study. *IEEE J Biomed Heal Informatics* 2022: 1–12. DOI: 10.1109/JBHI.2022.3205506.
10. Sun T, He X and Li Z. Digital twin in healthcare: Recent updates and challenges. Published online 2023. DOI: 10.1177/20552076221149651
11. Haleem A, Javaid M, Pratap R, et al. Biomedical technology exploring the revolution in healthcare systems through the applications of digital twin technology. *Biomed Technol* 2023; 4: 28–38.
12. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J* n71. Published online March 29, 2021. DOI: 10.1136/bmj.n71.
13. Halevi G, Moed H and Bar-Ilan J. Suitability of google scholar as a source of scientific information and as a source of data for scientific evaluation—review of the literature. *J Informetr* 2017; 11: 823–834.
14. Haddaway NR, Collins AM, Coughlin D, et al. The role of google scholar in evidence reviews and its applicability to grey literature searching. Wray KB, ed. *PLoS One* 2015; 10: e0138237.
15. Sheng B, Zhang Y, Meng W, et al. Bilateral robots for upper-limb stroke rehabilitation: state of the art and future prospects. *Med Eng Phys* 2016; 38: 587–606.
16. Cretchley J, Rooney D and Gallois C. Mapping a 40-year history with leximancer: themes and concepts in the journal of cross-cultural psychology. *J Cross Cult Psychol* 2010; 41: 318–328.
17. Zhong S, Geng Y, Liu W, et al. A bibliometric review on natural resource accounting during 1995–2014. *J Clean Prod* 2016; 139: 122–132.
18. Chen X, Zou D, Cheng G, et al. Detecting latent topics and trends in educational technologies over four decades using

- structural topic modeling: A retrospective of all volumes of Computers & Education. *Comput Educ* 2020; 151. DOI: 10.1016/j.compedu.2020.103855.
19. Newrzella SR, Franklin DW and Haider S. Three-dimension digital twin reference architecture model for functionality, dependability, and life cycle development across industries. *IEEE Access* 2022; 10: 95390–95410.
 20. Gyulai S. Integration of Digital Twin Technology and Health Dialogues. Published online 2022.
 21. Barricelli BR, Casiraghi E, Gliozzo J, et al. Human digital twin for fitness management. *IEEE Access* 2020; 8: 26637–26664.
 22. Mulder ST, Omidvari AH, Rueten-Budde AJ, et al. Dynamic digital twin: Diagnosis, treatment, prediction, and prevention of disease during the life course. *J Med Internet Res* 2022; 24: 1–11.
 23. Ferdousi R. Digital Twin Disease Diagnosis Using Machine Learning. Published online 2021, <https://ruor.uottawa.ca/handle/10393/42773>
 24. El Saddik A, Badawi H, Martinez Velazquez RA, et al. Dtwins : A digital twins ecosystem for health and well-being. *IEEE COMSOC MMTC Commun – Front* 2019; 14: 39–43.
 25. Huang PH, Kim KH and Schermer M. Ethical issues of digital twins for personalized health care service: preliminary mapping study. *J Med Internet Res* 2022; 24: 1–12.
 26. Braun M. Represent me: Please! towards an ethics of digital twins in medicine. *J Med Ethics* 2021; 47: 394–400.
 27. Jeske SJ. Digital Twins in Healthcare Conceptualisation and Privacy Aspects, Aalborg University, 2020.
 28. Bruynseels K, de Sio FS and van den Hoven J. Digital twins in health care: Ethical implications of an emerging engineering paradigm. *Front Genet* 2018; 9: 1–11.
 29. Coorey G, Figtree GA, Fletcher DF, et al. The health digital twin to tackle cardiovascular disease—a review of an emerging interdisciplinary field. *npj Digit Med* 2022; 5. DOI: 10.1038/s41746-022-00640-7.
 30. Corral-Acero J, Margara F, Marciniak M, et al. The “digital twin” to enable the vision of precision cardiology. *Eur Heart J* 2020; 41: 4556–4564B.
 31. Laubenbacher R, Niarakis A, Helikar T, et al. Building digital twins of the human immune system: Toward a roadmap. *npj Digit Med* 2022; 5. DOI: 10.1038/s41746-022-00610-z.
 32. Khan A, Milne-Ives M, Meinert E, et al. A scoping review of digital twins in the context of the COVID-19 pandemic. *Biomed Eng Comput Biol* 2022; 13: 117959722211021.
 33. Patrone C, Lattuada M, Galli G, et al. The role of internet of things and digital twin in healthcare digitalization process. In: Ao SI, Kim HK and Amouzegar MA (eds) *Transactions on engineering technologies. WCECS 2018*. Singapore: Springer, 2020, pp.30–37.
 34. Karakra A, Lamine E, Fontanili F, et al. Hospit’win: A digital twin framework for patients’ pathways real-time monitoring and hospital organizational resilience capacity enhancement. In: *9th Int work innov simul heal care, IWISH 2020*, pp.62–71. Published online 2020. DOI: 10.46354/i3 m.2020.iwish.012
 35. Karakra A, Fontanili F, Lamine E, et al. HospiT ‘ Win : A Predictive Simulation-Based Digital Twin for Patients Pathways in Hospital To cite this version : HAL Id: hal-02292836 HospiT ‘ Win : A Predictive Simulation-Based Digital Twin for Patients Pathways in Hospital. Published online 2019.
 36. Karakra A, Fontanili F, Lamine E, et al. Pervasive Computing Integrated Discrete Event Simulation for a Hospital Digital Twin. In: *Proc IEEE/ACS int conf comput syst appl AICCSA, 2019*, 2018-Novem(October). DOI: 10.1109/AICCSA.2018.8612796
 37. Wang W, He Y, Li F, et al. Digital twin rehabilitation system based on self-balancing lower limb exoskeleton. *Technol Heal Care* 2023; 31: 103–115.
 38. Trobinger M, Costinescu A, Xing H, et al. A dual doctor-patient twin paradigm for transparent remote examination, diagnosis, and rehabilitation. *IEEE Int Conf Intell Robot Syst* 2933–2940. Published online 2021. DOI: 10.1109/IROS51168.2021.9636626
 39. Martinez-Velazquez R, Gamez R and El Saddik A. Cardio twin: a digital twin of the human heart running on the edge. In: *Med meas appl MeMeA 2019 – symp proc*, pp.1–6. Published online 2019. DOI: 10.1109/MeMeA.2019.8802162
 40. Lombardo A and Ricci C. Digital twins federation for remote medical care of De-hospitalized patients. In: *2022 IEEE int conf pervasive comput commun work other affil events, PerCom work 2022*, pp.718–723. Published online 2022. DOI: 10.1109/PerComWorkshops53856.2022.9767417
 41. Bezborodova OE, Bodin ON, Gerasimov AI, et al. Digital Twin technology in medical information systems. *J Phys Conf Ser*. Published online 2020. doi:10.1088/1742-6596/1515/5/052022
 42. Lal A, Li G, Cubro E, et al. Development and verification of a digital twin patient model to predict specific treatment response during the first 24 hours of sepsis. *Crit Care Explor* 2020; 2: e0249.
 43. Roberts ME, Stewart BM and Tingley D. STM: an R package for structural topic models. *J Stat Softw* 2019; 91: 1–40.
 44. Roberts ME, Stewart BM and Airoidi EM. A model of text for experimentation in the social sciences. *J Am Stat Assoc* 2016; 111: 988–1003.
 45. Roberts ME, Stewart BM, Tingley D, et al.. The structural topic model and applied social science, In *Proceedings of the NIPS 2013 Workshop on Topic Models: Lake Tahoe, Nevada, 2013*, 2–5.
 46. Lebyrk T. Introduction to The Structural Topic Model (STM). Mediumm. Published 2021. <https://towardsdatascience.com/introduction-to-the-structural-topic-model-stm-34ec4bd5383>
 47. Hofstra B, Kulkarni VV, Galvez SMN, et al. The diversity–innovation paradox in science. *Proc Natl Acad Sci U S A* 2020; 117: 9284–9291.
 48. Giorgi FM, Ceraolo C and Mercatelli D. The R language: An engine for bioinformatics and data science. *Life* 2022; 12: 48.
 49. Bischof JM and Airoidi EM. Summarizing topical content with word frequency and exclusivity. *Proc 29th Int Conf Mach Learn ICML 2012* 2012; 1: 201–208.
 50. Choi H and Woo JR. Investigating emerging hydrogen technology topics and comparing national level technological focus: Patent analysis using a structural topic model. *Appl Energy* 2022; 313: 118898.

51. Bajwa J, Munir U, Nori A, et al. Artificial intelligence in healthcare: transforming the practice of medicine. *Futur Healthc J* 2021; 8: e188–e194.
52. Baker SB, Xiang W and Atkinson I. Internet of things for smart healthcare: technologies, challenges, and opportunities. *IEEE Access* 2017; 5: 26521–26544.
53. Ahmed I, Ahmad M and Jeon G. Integrating digital twins and deep learning for medical image analysis in the era of COVID-19. *Virtual Real Intell Hardw* 2022; 4: 292–305.
54. Sahal R, Alsamhi SH, Brown KN, et al. Blockchain-Based Digital Twins Collaboration for Smart Pandemic Alerting: Decentralized COVID-19 Pandemic Alerting Use Case. *Comput Intell Neurosci.* 2022;2022. DOI: 10.1155/2022/7786441
55. Song Y and Li Y. Digital twin aided healthcare facility management: a case study of Shanghai tongji hospital. *Construction Research Congress* 2022; 2022: 1145–1155.
56. Pirbhulal S, Abie H and Shukla A. Towards a Novel Framework for Reinforcing Cybersecurity using Digital Twins in IoT-based Healthcare Applications. *IEEE Veh Technol Conf* 2022; 2022: 1–5.
57. Dash S, Shakyawar SK, Sharma M, et al. Big data in healthcare: Management, analysis and future prospects. *J Big Data* 2019; 6. DOI: 10.1186/s40537-019-0217-0.
58. Khan S, Saied IM, Ratnarajah T, et al. Evaluation of unobtrusive microwave sensors in healthcare 4.0—toward the creation of digital-twin model. *Sensors* 2022; 22. DOI: 10.3390/s22218519.
59. Neghab HK, Jamshidi MB and Neghab HK. Digital twin of a magnetic medical microrobot with stochastic model predictive controller boosted by machine learning in cyber-physical healthcare systems. *Inf* 2022; 13. DOI: 10.3390/info13070321.
60. Gámez Díaz R, Yu Q, Ding Y, et al. Digital twin coaching for physical activities: A survey. *Sensors (Switzerland)* 2020; 20: 1–21.
61. Tai Y, Zhang L, Li Q, et al. Digital-Twin-Enabled IoMT system for surgical simulation using rAC-GAN. *IEEE Internet Things J* 2022; 9: 20918–20931.
62. Elayan H, Aloqaily M and Guizani M. Digital twin for intelligent context-aware IoT healthcare systems. *IEEE Internet Things J* 2021; 8: 16749–16757.
63. Liu Y, Zhang L, Yang Y, et al. A novel cloud-based framework for the elderly healthcare services using digital twin. *IEEE Access* 2019; 7: 49088–49101.
64. Wan HC and Chin KS. Exploring internet of healthcare things for establishing an integrated care link system in the healthcare industry. *Int J Eng Bus Manag* 2021; 13: 1–14.
65. Palozzi G, Binci D and Schettini I. Digital transformation in the healthcare sector: empirical evidences of IoHT benefits and limits on chronic disease management. 2020: 433–449. DOI: 10.1007/978-3-030-30911-4_31.
66. Kaul R, Ossai C, Forkan ARM, et al. The role of AI for developing digital twins in healthcare: The case of cancer care. *WIREs Data Min Knowl Discov* 2023; 13. DOI: 10.1002/widm.1480.
67. Allen A, Siefkas A, Pellegrini E, et al. A digital twins machine learning model for forecasting disease progression in stroke patients. *Appl Sci* 2021; 11. DOI: 10.3390/app11125576.
68. Jamshidi M, Moztarzadeh O, Jamshidi A, et al. Future of drug discovery: The synergy of edge computing, internet of medical things, and deep learning. *Futur Internet* 2023; 15: 1–15.
69. Armeni P, Polat I, De Rossi LM, et al. Digital twins in healthcare: is it the beginning of a new era of evidence-based medicine? A critical review. *J Pers Med* 2022; 12. DOI: 10.3390/jpm12081255.
70. Zhong X, Babaie Sarijaloo F, Prakash A, et al. A multidisciplinary approach to the development of digital twin models of critical care delivery in intensive care units. *Int J Prod Res* 2022; 60: 4197–4213.
71. Possik J, Azar D, Solis AO, et al. A distributed digital twin implementation of a hemodialysis unit aimed at helping prevent the spread of the omicron COVID-19 variant. In: [2022 IEEE/ACM 26th international symposium on distributed simulation and real time applications \(DS-RT\)](#). IEEE; 2022, pp.168–174. DOI: 10.1109/DS-RT55542.2022.9932047
72. Sun T, He X, Song X, et al. The digital twin in medicine: a key to the future of healthcare? *Front Med* 2022; 9: 1–8.
73. Stahlberg EA, Abdel-Rahman M, Aguilar B, et al. Exploring approaches for predictive cancer patient digital twins: Opportunities for collaboration and innovation. *Front Digit Heal* 2022; 4: 1–10.
74. Venkatesh KP, Raza MM and Kvedar JC. Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. *npj Digit Med* 2022; 5. DOI: 10.1038/s41746-022-00694-7.
75. de Boer B, Strasser C and Mulder S. Imagining digital twins in healthcare: Designing for values as designing for technical milieus. *Prometh (United Kingdom)* 2022; 38: 67–81.
76. Gkouskou K, Vlastos I, Karkalousos P, et al. The “virtual digital twins” concept in precision nutrition. *Adv Nutr* 2020; 11: 1405–1413.
77. Hao Lun C. Health management service design based on digital twin—Taking community geriatric health management service as an example, Beijing Institute of Graphic Communication, 2022
78. Kline KJ and Knight AP. Innovation implementation. *Curr Dir Psychol Sci.* 2005;14:243–246.
79. Croatti A, Gabellini M, Montagna S, et al. On the integration of agents and digital twins in healthcare. *J Med Syst* 2020; 44: 1–8.
80. Ricci A, Croatti A and Montagna S. Pervasive and connected digital twins – A vision for digital health. *IEEE Internet Comput* 2022; 26: 26–32.
81. Haoran CLCWTYZSXR& X. Virtual and reality symbiosis: Digital twin (DT) technology and its prospective application in education: also discuss the reconstruction of the ubiquitous intelligence space. *J DISTANCE Educ* 2019; 05: 3–12.
82. ITU-R. M2410 – Minimum requirements related to technical performance for IMT-2020 radio interface(s). *Work Party 5D.* 2017;November(Report ITU-R M.2410-0):1–11, https://www.itu.int/dms_pub/itu-r/opb/rep/R-REPM.%0A2410-2017-PDF-E.pdf.
83. Sun SH, Gao QB, Du Y, et al. Overview on the progress of design and standardization of the 5G Mobile communications

- system. *Beijing Youdian Daxue Xuebao/Journal Beijing Univ Posts Telecommun.* 2018; 41: 26–43.
84. Ahad A, Tahir M and Yau KLA. 5G-based Smart healthcare network: Architecture, taxonomy, challenges and future research directions. *IEEE Access* 2019; 7: 100747–100762.
 85. Zhang XD. Precision Medicine, Personalized Medicine, Omics and Big Data: Concepts and Relationships. *J Pharmacogenomics Pharmacoproteomics* 2015; 6: 1–2.
 86. Digital Health Software Precertification (Pre-Cert) Pilot Program. U.S., Food & Drug Administration. Published 2022, <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-pilot-program>
 87. German Federal Ministry of Housing Urban Development and Construction. Urban Digital Twins for urban development of the future. Published 2021, <https://www.connect.edurbantwins.de/en/>
 88. CDBB. The Gemini Principles. Published online 2018. <https://doi.org/10.17863/CAM.32260>
 89. Acosta JN, Falcone GJ, Rajpurkar P, et al. Multimodal biomedical AI. *Nat Med* 2022; 28: 1773–1784.
 90. Karczewski KJ and Snyder MP. Integrative omics for health and disease. *Nat Rev Genet.* 2018; 19: 299–310.
 91. Voigt I, Inojosa H, Dillenseger A, et al. Digital twins for multiple sclerosis. *Front Immunol* 2021; 12: 1–17.
 92. Dwivedi YK, Hughes L, Baabdullah AM, et al. Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *Int J Inf Manage* 2022; 66: 102542.
 93. Moztaarzadeh O, Jamshidi M, Sargolzaei S, et al. Metaverse and medical diagnosis: A Blockchain-based digital twinning approach based on MobileNetV2 Algorithm for cervical vertebral maturation. *Diagnostics* 2023; 13. DOI: 10.3390/diagnostics13081485.
 94. Moztaarzadeh O, Jamshidi MB, Sargolzaei S, et al. Metaverse and healthcare: Machine learning-enabled digital twins of cancer. *Bioeng (Basel, Switzerland)* 2023; 10. DOI: 10.3390/bioengineering10040455.
 95. Jamshidi M, Ebadpour M and Moghani MM. Cancer digital twins in metaverse. *Proc 2022 20th Int Conf Mechatronics – Mechatronika, ME* 2022: 1–6. Published online 2022. DOI: 10.1109/ME54704.2022.9983328.
 96. Lu Q, Xie X, Chen L, Zhen, et al. Creating an inter-hospital resilient network for pandemic response based on Blockchain and dynamic digital twins. In: Winter simulation conference, 2016, pp.1–23.
 97. Laamarti F, Badawi HF, Ding Y, et al. An ISO/IEEE 11073 standardized digital twin framework for health and well-being in smart cities. *IEEE Access.* 2020; 8: 105950–105961.
 98. Janya C, Albert T, Fanwen M, et al. *Healthcare 4.0: Next generation processes with the latest technologies.* 1st ed. Singapore: Palgrave Pivot Singapore, 2019.
 99. Aloqaily M, Bouachir O and Karray F. Chapter 3 - Digital twin for healthcare immersive services: fundamentals, architectures, and open issues. In: Abdulmotaieb El Saddik (ed.) *Digital twin for healthcare design, challenges, and solutions; El Saddik, A.B.T.-D.T. for H., Ed.;* Academic Press, 2023, pp.39–71.
 100. Milne-Ives M, Fraser LK, Khan A, et al. Life course digital twins–intelligent monitoring for early and continuous intervention and prevention (LifeTIME): Proposal for a retrospective cohort study. *JMIR Res Protoc* 2022; 11: 1–8.
 101. Erol T, Mendi AF and Doğan D. The digital twin revolution in healthcare. *Nat Comput Sci* 2021; 1: 307–308.
 102. Doctor E, Keweloh C, Buck C, et al. Towards a Taxonomy of Digital Twin Applications for Towards a Taxonomy of Digital Twin Applications for Telemedical Healthcare. *ECIS 2021 Res Pap.* Published online 2021:6–14, https://aisel.aisnet.org/ecis2021_rip/12
 103. Yuefei C, Sisi W, Mingqi T, et al. Application of digital twins in medical and health fields and related research progress. *Metrol Sci Technol* 2021; 65: 6–9.
 104. Thiong'o GM and Rutka JT. Digital twin technology: The future of predicting neurological complications of pediatric cancers and their treatment. *Front Oncol* 2022; 11: 1–5.
 105. Zhang J, Li L, Lin G, et al. Cyber resilience in healthcare digital twin on lung cancer. *IEEE Access* 2020; 8: 201900–201913.
 106. Sirigu G, Carminati B and Ferrari E. Privacy and security issues for human digital twins. In: **2022 IEEE 4th international conference on trust, privacy and security in intelligent systems, and applications (TPS-ISA).** IEEE, 2022, pp.1–9. DOI: 10.1109/TPS-ISA56441.2022.00011
 107. Sahal R, Alsamhi SH and Brown KN. Personal digital twin: a close look into the present and a step towards the future of personalised healthcare industry. *Sensors* 2022; 22: 1–35.
 108. Bai X, Zhang X, Li KX, et al. Research topics and trends in the maritime transport: A structural topic model. *Transp Policy* 2021; 102: 11–24.

Acronyms and notations

Acronyms	Full form
DT	Digital twin
STM	Structure Topic Modeling
AI	Artificial intelligence
WoS	Web of Science
CNKI	China National Knowledge Infrastructure
FREX	Frequency and Exclusivity
IoT	Internet of things
XR	Extended reality
IoMT	Internet of Medical Things
IoHT	Internet of Healthcare Things
mMTC	Massive Machine Type Communications
FDA	US Food and Drug Administration
CVM	Cervical vertebral maturation
Notations	Meaning
d	Number of documents
k	Number of topics
n	Words within the documents
v	Index of all unique observation words
X	The matrix of topic prevalence
Y	Matrix of topical content covariates
θ	Probability matrix of all documents for defined topics
γ	A normal distribution coefficient matrix

β	Multinomial logistic topic word matrix with covariates	Z	Per topic word assignment
m_ν	The marginal log-transformed rate of term ν	ω	Observed words
κ	The specific deviation of covariates	Estimate	Topic preference parameter for covariates
