IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 7 December 2023, accepted 31 December 2023, date of publication 3 January 2024,
date of current version 10 January 2024.

Digital Object Identifier 10.1109/ACCESS.2023.3349379

SURVEY

A Systematic Literature Review of Digital Twin Research for Healthcare Systems: Research Trends, Gaps, and Realization Challenges

MD. DOULOTUZZAMAN XAMES, (Member, IEEE), AND TAYLAN G. TOPCU
Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

Corresponding author: Taylan G. Topcu (ttopcu@vt.edu)

ABSTRACT Using the PRISMA approach, we present the first systematic literature review of digital twin (DT) research in healthcare systems (HSs). This endeavor stems from the pressing need for a thorough analysis of this emerging yet fragmented research area, with the goal of consolidating knowledge to catalyze its growth. Our findings are structured around three research questions aimed at identifying: (i) current research trends, (ii) gaps, and (iii) realization challenges. Current trends indicate global interest and interdisciplinary collaborations to address complex HS challenges. However, existing research predominantly focuses on conceptualization; research on integration, verification, and implementation is nascent. Additionally, we document that a substantial body of papers mislabel their work, often disregarding modeling and twinning methods that are necessary elements of a DT. Furthermore, we provide a non-exhaustive classification of the literature based on two axes: *the object* (i.e., product or process) and *the context* (i.e., patient's body, medical procedures, healthcare facilities, and public health). While this is a testament to the diversity of the field, it implies a specific pattern that could be reimagined. We also identify two gaps: (i) considering the human-in-the-loop nature of HSs with a focus on provider decision-making and (ii) implementation research. Lastly, we discuss two challenges for broad-scale implementation of DTs in HSs: improving virtual-to-physical connectivity and data-related issues. In conclusion, this study suggests that DT research could potentially help alleviate the acute shortcomings of HSs that are often manifested in the inability to concurrently improve the quality of care, provider wellbeing, and cost efficiency.

INDEX TERMS Digital twin, healthcare systems, sociotechnical systems, health information technology, systematic literature review, systems engineering.

I. INTRODUCTION

Digital twin (DT), a concept first introduced by Grieves in 2002 as a 'conceptual ideal' for product life cycle management, defines the triad of (i) a physical system, (ii) its virtual representation, and (iii) the bilateral information flow that links the physical and the virtual counterparts together [1]. Over the past two decades, DT research has matured with the advances in artificial intelligence (AI), machine learning (ML), and the internet of things (IoT) [2]; and provides a multitude of capabilities through synergistic use of modeling, real-time data collection, and data-analytics [3]. These capabilities are increasingly being implemented in practice, for purposes such as system health prognostics [4], early detection of anomalies [5], and predictive maintenance [6]. Additionally, DTs provide organizations with a high-fidelity digital ecosystem to safely explore "what-if" scenarios regarding operational fluctuations or potential changes, without interfering with the ongoing operations of their physical assets [7]. This ability is particularly valuable, as it allows for policy analysis, identification of bottlenecks [8], and proactive mitigation of potential operational perturbations [9]. Given their promise and abundance of vast streaming data, DTs

The associate editor coordinating the review of this manuscript and approving it for publication was Zhenliang Zhang.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution 4.0 License.
For more information, see https://creativecommons.org/licenses/by/4.0/

are predominantly being designed for, and implemented in, engineering applications such as aerospace, civil, energy, manufacturing, and mechanical systems [10], [11].

In addition to their success in engineering applications, DT research offers a plethora of opportunities for healthcare systems (HSs), particularly regarding the shortcomings in concurrent improvement of patient care, physician wellbeing, and facility operations – sustained global challenges that have led to numerous unaddressed calls from the World Health Organization and the National Academies [12], [13], [14], [15]. However, this is easier said than done. Compared to engineered systems such as aerospace and manufacturing, leveraging DTs for HSs is more challenging because of their inherent sociotechnical complexity [16] that originates from their dependence on (i) biological processes, (ii) human decision-makers in the loop (e.g., physicians and nurses), and (iii) provider-technology collaboration for service delivery. These heterogeneous characteristics render it more challenging to monitor HS processes with low latency, create sufficiently representative virtual replicas, and predict their behavior [17], [18].

Despite these challenges, DT research in HSs is expanding rapidly in scope, depth, and breadth of applications [19], [20]. Nevertheless, so far, only a few rapid review studies have been published [21], [22], [23], and a systematic literature review of DT research in HSs is nascent. Although rapid or mini reviews are valuable in their own right, they differ from systematic reviews as they (i) do not follow a formalized screening methodology [24], [25], (ii) are non-exhaustive in their coverage of articles [26], and (iii) provide only a high-level discussion of existing articles [27]. Consequently, they may overlook or disregard certain bodies of knowledge [28], [29]. To that end, this study differs from previous review papers by presenting the first-ever systematic literature review of DT research in HSs by using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) approach [30], a broadly accepted systematic literature review methodology in healthcare research. Moreover, this study establishes a benchmark of the literature structured around the following research questions that are essential for supporting the growth of this community:

RQ1: What are the current trends of DT research in HSs?

RQ2: What are the gaps and opportunities for DT research in HSs?

RQ3: What are the realization challenges of DT research in HSs?

The contribution of this research is four-fold. First, to the best of our knowledge, this paper presents the first systematic literature review of DT research in HSs. In the absence of a systematic review, it is difficult for the research community to identify relevant research streams and position their work, particularly given that this is an emerging interdisciplinary research area that is fragmented across different communities.

Second, the findings benchmark the state-of-the-art for the research community and identify current trends. We document common misconceptions about the DT terminology and find a strong concentration of research on the conceptual design of DTs with a lack of attention on implementation and validation. We then provide a summary of DT methods and techniques used in other application areas to draw attention to the analytical capabilities that could be leveraged for HS research. We also juxtapose the existing HS literature in terms of its utilization of these methods. Additionally, we provide a taxonomy of the current literature that categorizes DT research in HSs in two dimensions: the *twinning object* (such as products and processes) and the *twinning context* (such as the patient's body, medical procedures, facilities, and public health).

Third, we identify two fruitful research gaps that are currently being overlooked by the existing literature: (i) consideration of the human-in-the-loop nature of HSs and (ii) implementation research. Here, the former suggests that DT research could more explicitly explore, capture, manage, and leverage how provider interactions with the patients, collaborators, technology, and the broader HS translate into the quality of care and overall HS performance. The latter calls for a push towards implementation, test, evaluation, and validation research to help translate the existing bulk of conceptual DT research into the implementation of HS operations.

Finally, based on our review, we provide a rich discussion on realization challenges for DT research in HSs, covering a range of technical and data management-related issues that hinder the translation of theoretical and conceptual research into HS operations. We document ongoing data-related challenges such as collection, synthesis, and privacy; and summarize the limited available research on these critical topics.

The rest of the paper is organized in the following manner. Section II provides an overview of various DT definitions used in the literature, characterizes DT archetypes with an emphasis on their differences, and takes stock of DT modeling and twinning technologies by borrowing from other application areas. We consider this necessary before presenting our systematic review because there is significant ambiguity and misconceptions in the community regarding DT terminology. Section III presents the systematic literature review methodology along with a discussion of the process. We present the search procedure, the selection of the filtering criteria, and their justification. In Section IV, we summarize our findings in terms of current trends, gaps and opportunities, and realization challenges of DTs for HS research. In Section V, we discuss the broader implications of this research based on our overall findings. Finally, Section VI concludes with a summary of the major takeaways.

II. A PRIMER ON DIGITAL TWIN TERMINOLOGY AND ENABLING TECHNOLOGIES

A. DT DEFINITIONS AND ARCHETYPES

This subsection glances over various definitions of DTs, discusses the discrepancies between different perspectives, and highlights some common misconceptions. The term DT

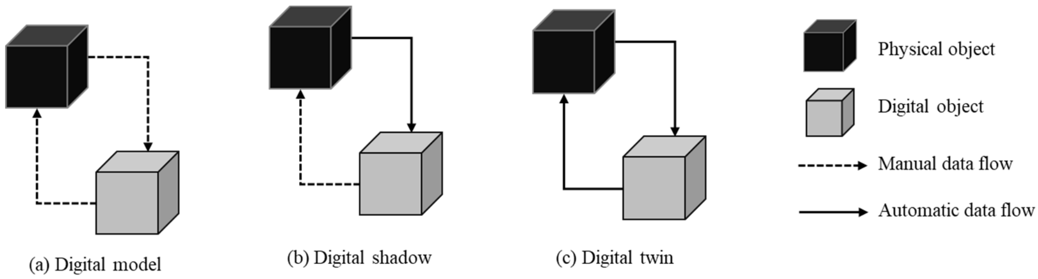


FIGURE 1. A visualization of the terminology adapted from Kritzinger et al. [35]. (a) Digital model, (b) Digital shadow, and (c) Digital twin.

is used quite generously and its definition varies between application areas. For instance, Grieves defined DTs as "a set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level" [9]. On the other hand, according to NASA, "DT is an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin". Whereas the manufacturing community outlines the following definition [31]: "The DT consists of a virtual representation of a production system that is able to run on different simulation disciplines that are characterized by the synchronization between the virtual and real system, thanks to sensed data and connected smart devices, mathematical models and real-time data elaboration."

So why is the discrepancy between these definitions? According to Barricelli et al. [32], the definitions of DTs vary based on the application area. We concur and would like to add that there is also an intended use case effect. For instance, NASA employs DTs for a range of purposes, such as certification and maintenance, monitoring the well-being of spacecraft, managing missions, and conducting in-situ diagnosis and prognosis [33]. Therefore, NASA's definition of DT encapsulates their purposes of using DTs across the lifecycle of a space mission. In manufacturing, the leading use case is real-time monitoring and control of manufacturing processes [34]. To achieve this objective, the DT of a production system requires real-time synchronization with its physical twin through sensors and other connected devices, as reflected in their definition of DT [31]. So, we argue that the differences in DT definitions are somewhat contingent upon the use cases that are of course, often determined by the application area.

On another note, we observe some confusion that manifests itself in inaccurate labeling of DT research in HSs, which could originate from a lack of clarity regarding the necessary characteristics of DTs. In the past, the terms digital model, digital shadow, and digital twin were often used interchangeably; however, these terms represent different things [35]. As portrayed in Fig. 1, the key difference between these terms lies in the richness of data integration between the physical and virtual objects. A digital model, shown in Fig. 1(a), allows for manual data exchange between the physical and virtual realms. Digital models allow designers to perform offline simulations and analyses. They are particularly useful when automatic data integration is not required or not possible. In contrast, a digital shadow, shown in Fig. 1(b), has unidirectional data flow solely from the physical to the digital domain. Finally, a DT portrayed in Fig. 1(c), necessitates bidirectional data flow between the physical and virtual objects. Furthermore, in process control applications of DTs, this bidirectional data exchange involves a significantly rich information transfer and thus must be automated and facilitated by some control software. This ensures that any change in the state of the physical object almost immediately – pertaining to minor transfer delays – translates into a corresponding change in the state of the digital object, and vice versa.

We contend that it is essential to recognize the distinction between these terms, as a lack of understanding can lead researchers to mislabel or misposition their work in the literature, such as denoting digital shadows or digital models as DTs [36]. This mislabeling trend previously emerged in engineering applications [35] and as we discuss further in Section IV–A, we find it to be an ongoing issue for the DT research in HSs.

Before proceeding into our literature review, we contend that it is useful to share a primer on modeling and twinning methods for DTs that are employed in engineering systems. Here, modeling methods refer to how a digital replica is created from a physical artifact. Whereas twinning methods capture how the physical and virtual worlds are linked to one another. These techniques could be creatively applied to HSs problems or could potentially lead researchers to develop new methods for unaddressed HSs needs. We provide a brief overview of these methods in the following two subsections.

B. DT MODELING METHODS

DT modeling methods are used for creating a sufficiently representative virtual replica of a physical entity, process, or object, as shown in Fig. 2. These methods are quite diverse and offer a range of alternatives depending on the nature of the physical artifact and the desired level of detail, fidelity, and functionality. According to Thelen et al. [10], DT modeling methods can be classified into five: geometric modeling, physics-based modeling, data-driven modeling, physics-informed ML modeling, and systems modeling.

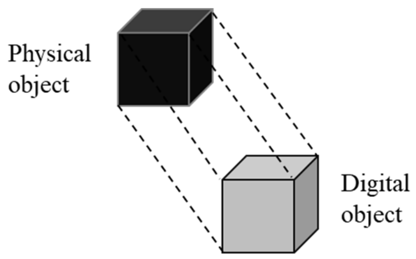


FIGURE 2. DT model creation.

Geometric modeling techniques involve creating a 3D geometric representation of a physical object by capturing its shape, size, and/or spatial relationships of its parts. These techniques include solid modeling, laser scanning, augmented reality (AR)/ virtual reality (VR), and mixed reality (MR) modeling – these terms are sometimes collectively referred to as Extended Reality (XR) [37]. Solid modeling involves the creation of computer-aided design models. These are usually incorporated into simulation models for increased fidelity and leveraged to explore specific system-level objectives such as assembly line layout optimization [38] or training of a robotic arm in a realistic operational environment [39]. Solid modeling is particularly useful in studying human-robot interactions as this allows researchers to explore how humans can be seamlessly integrated into DTs. For example, modeling an offshore oil drilling platform to facilitate training and testing of human-robot collaboration scenarios such as inspection and emergency response [40]. Solid modeling and tolerancing/geometry variation were explored for adaptive optimization of cutting parameters in CNC machining [41]. On the other hand, laser scanning techniques are broadly used for extracting surface-level features of physical assets with the help of laser light, enabling their high-resolution incorporation into a digital environment. These techniques are primarily used in manufacturing DT for geometry and surface quality monitoring and control of additively manufactured parts [34], [42]. Laser scanning technologies can quickly map and model large structures and buildings making them popular for civil engineering applications [5], [43], [44], [45].

Increasingly, XR-based data visualization and interaction technologies are gaining popularity among DT researchers in part due to their ability to synergistically connect real-time users, assets, and data streaming. AR blends digital information and content into the physical environment, while VR creates an immersive digital environment for its users that is isolated from the physical environment. MR blends physical and digital environments, enabling them to interact with each other. Some notable applications of these technologies in DT research for HSs include MR-assisted DT for safety-aware human-robot collaboration [45] and an AR-based robotic arm DT for additive manufacturing [46]. Matulis and Harvey [39] presented an integrated MR system that uses DT and deep learning for safety-aware human-robot collaboration. The system accurately measures the minimum safe distance in real-time by incorporating 3D offset-based safety distance calculation based on the robot's DT and provides task assistance to the human operator through MR glasses. Others utilized AR to enhance communication between a reconfigurable additive manufacturing system consisting of robotic arms and its DT for toolpath planning and simulation [40]. Their proposed methodology allows for efficient retrieval of layout information from the physical system into the DT, enabling optimized layout deployment in the physical system.

Physics-based modeling involves developing mathematical models consisting of partial differential equations to describe the underlying physics and/or initial and boundary conditions for a physical phenomenon. Physics-based techniques have a wide range of applications in DT modeling that include the use of finite element [47], [48], [49], computational fluid dynamics [50], multiphysics simulations [51], etc. On the other hand, data-driven DT modeling techniques are used when the underlying physics is not fully understood or when the required simulations are computationally expensive. These techniques can be classified into two categories – statistical models and ML models. Some popularly used statistical models are Markov process modeling, Poisson process modeling, inverse Gaussian process modeling, auto-regressive integrated moving average models, etc. The conventional ML models include artificial neural networks [52], support vector machines [53], gaussian process regression [54], etc. The relatively popular deep learning models include convolutional neural networks [55], long short-term memory [56], autoencoder [57], etc. When both physics-based simulated data and physical system data are available, hybrid physics-informed ML models are used for the DT modeling of the physical systems [58]. There are several approaches to such hybrid modeling such as physics-informed loss functions [59], transfer learning [60], data augmentation [61], and delta learning [62].

Finally, systems modeling techniques originate from the systems engineering community and prioritize the functional and formal (component) interactions among the elements of a system organized in layers of hierarchy. Some widely used techniques for systems interaction modeling are unified modeling language (UML) [63] and systems modeling language (SysML) [64]. UML and SysML are general-purpose modeling languages of model-based systems engineering [65]. They enable the creation of meta-models of system elements, their functional, formal, and behavioral interdependencies; and interactions in context [66]. These usually rely on some ontology that offers a common language to uniquely refer to each possible object (i.e., component or sensor) in the physical asset. This increases the flexibility in explaining the evolution of DTs when they are subject to postproduction changes or modifications over their life cycle [67]. Next, we present a synopsis of the DT twinning methods.

C. DT TWINNING METHODS

Twinning methods are vital for DT research as they are used to establish physical-to-virtual (P2V) and virtual-to-physical (V2P) connections – the bilateral automated data interconnections as shown in Fig. 3.

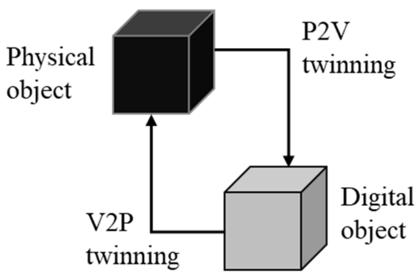


FIGURE 3. Twinning connections in DTs.

We start our discussion with P2V twinning techniques. P2V twinning techniques may vary based on the choice of DT modeling method [10]. These techniques include – using physical measurements as inputs to the virtual space [66], probabilistic model updating [58], ML model updating [68], fault diagnostics and failure prognostics [69], and ontology-based reasoning [70].

The simplest way of connecting a physical and a virtual object is to collect physical measurements and transfer them as inputs to the virtual space. Usually, this is achieved in two ways [10]: (i) using measurements to directly update digital models or (ii) by using streaming data as an input for physics-based analysis models. Probabilistic modeling, such as Bayesian filtering, is preferred when the uncertainty and noise in the physical data are paramount and render it infeasible to use physical updating. The third P2V method is ML model updating, which is usually formulated as a parameter estimation problem, expressed as a discrete-time state-space model. The key difference is that ML models are updated continuously to ensure their performance does not degrade or adapt to the changing physical entity. Fault diagnostics and failure prognostics P2V twinning techniques are particularly popular in manufacturing applications. These techniques involve the use of raw sensor data and ML or deep learning techniques to capture and identify current health states (diagnostics) and explore future possible failures in operation (prognostics). Finally, ontology-based P2V twinning techniques rely on knowledge graphs and ontology maps and are primarily implemented using ontology markup languages such as web ontology language [71] and XML schema [72]. Some ontology techniques are logic rules-based reasoning [73], distributed representation-based reasoning [74], and neural network-based knowledge reasoning [75].

Regarding V2P linkages, there are several mechanisms such as system reconfiguration, structure optimization, model predictive control (MPC), predictive maintenance scheduling, and production planning [10]. Among these, MPC and predictive maintenance scheduling are the most popular techniques. MPC is used to predict the future behavior of processes and to determine the optimal control action within a set of constraints [76]. An MPC framework comprises several key elements: a process model, an objective function, process measurements, constraints, and sampling points [77]. The process model represents the mathematical representation of the controlled system, allowing predictions of its future behavior. The objective function defines the desired performance criteria that the controller aims to optimize. Process measurements refer to the data collected from the system in real-time, providing information about its current state. Constraints are conditions or limits imposed on the system variables to ensure safe and stable operation. Sampling points are specific instances in time at which the controller updates its predictions and makes control decisions. In short, an MPC structure combines these components to achieve effective control of dynamic systems by continuously adjusting inputs based on predictions and optimization.

Predictive maintenance is an increasingly popular V2P method for proactive equipment maintenance that involves three main steps. First, it focuses on recognizing patterns in sensor data that indicate changes in the condition of the equipment. Second, it aims to predict when a machine, part, or component is likely to fail. Finally, it involves scheduling maintenance tasks before the equipment fails, specifically during planned periods of downtime. The conventional ML pipeline for constructing and deploying a predictive maintenance solution comprises four key stages: gathering representative data samples, preprocessing the data to enhance its quality, training ML models using the prepared data, and optimizing decision-making related to maintenance activities. Predictive maintenance is predominantly used in the automotive, aviation, and manufacturing industries. In manufacturing, DTs enable continuous monitoring, predicting machine failures, and scheduling maintenance work during planned downtime [78], [79]. In the case of the automotive and aviation industry, DTs enable to utilization of massive data collected from distributed operational physical artifacts for customized scheduling of maintenance activities or to inform postproduction design changes such as batch software updates. DT's efficacy in leveraging data from distributed operational units to inform customizable decisions is a thread that attracted significant attention in the HSs community, which we discuss later in Section IV.

Other V2P twinning approaches such as system reconfiguration, structure optimization, and production planning are currently utilized mostly in manufacturing applications. To facilitate manufacturing system reconfiguration, DT was combined with a knowledge graph-based approach to propose a novel framework [80]. This approach could enable the system to find flexible yet optimized configurations with different criteria. For system reconfiguration of automated manufacturing systems, a DT-driven open architecture machine tool platform was developed [75]. This state-transfer architecture-based IoT platform enables rapid reconfiguration of control and sensor networks without the need for human intervention. Other notable works on

TABLE 1. Inclusion and exclusion criteria for literature screening.

	Inclusion Criteria	Exclusion Criteria
Article language	English	Non-English
Article type	Journal articles, conference articles, conference proceedings, book chapters, and articles in press	Preprint, review papers, editorial, commentary, letter to the editor, perspective, opinion, viewpoint, and brief communication

DT-based system reconfiguration can be found in [82], [83], and [84].

DT-based structure optimization studies were conducted in several application domains including mechanical systems [85], aerospace systems [86], [87], and power systems [88]. In these cases, P2V connectivity requires optimization algorithms to design and optimize the physical system. Lastly, production planning is a popular V2P twinning method in manufacturing or production systems. While there are several instances, a framework was proposed for achieving such connections which is composed of four key components: the production system, a trigger function, an optimization model, and a simulation model [89]. In this context, the trigger function monitors the production system and initiates production planning based on simulation-based optimization.

D. SYNTHESIS

We began Section II by addressing the discrepancies in DT definitions and elaborated on how the application area and use cases influence these disparities. Next, we clarified the common misconceptions about the terminology of digital model, digital shadow, and digital twin; and highlighted the necessary characteristics of a DT. We also provided a concise overview of the DT modeling methods and DT twinning methods. We contend that this discussion would help the broader audience comprehend the findings of our systematic review with improved clarity. Next, we proceed to our systematic literature review methodology.

III. METHODOLOGY

A. RESEARCH DESIGN

The methodology for this systematic review follows the PRISMA framework that provides guidelines for conducting systematic literature reviews, critical literature analyses, and meta-analyses. PRISMA framework is broadly accepted in the HSs community and its guidelines are commonly used to assess and evaluate the quality and validity of a systematic literature review. The framework emphasizes the formulation of inclusion and exclusion criteria, enabling a systematic assessment of the selected literature to determine its eligibility for inclusion or exclusion.

More specifically, we tailored our review around three research questions posed earlier, to focus on current trends, gaps and opportunities, and realization challenges in DT research in HSs. Our systematic review involves three key steps in accordance with the PRISMA framework: identification, screening, and inclusion. Fig. 4 provides an overview of our procedure along with the selection processes. We discuss each of the steps below.

B. SYSTEMATIC LITERATURE REVIEW PROCESS

1) IDENTIFICATION

The process of selecting relevant articles involved identifying keywords, followed by conducting a search to find similar and related terms based on the existing literature. On 30 June 2023, a systematic search was performed on all four databases (Scopus, Web of Science, PubMed, and Engineering Village) using the search string mentioned in section III–A. The selected databases were chosen because of their well-established reputation for delivering high-quality and interdisciplinary content, distinguishing them from other databases. The following sample search string was used to search for literature in the selected databases:

({digital twin} OR {digital thread*}) AND ({health} OR {healthcare} OR {medical} OR {medicine} OR {biomedical} OR {disease*} OR {surger*} OR {hospital*} OR {emergency department*} OR {procedure*} OR {patient*} OR {provider*})*

The search string has two components connected with a Boolean AND. The first part captures keywords specific to digital twins and the second part accounts for the healthcare-related keywords. If an article in its title/abstract/keywords has one of the terms of the first part and one of the terms from the second part, our search string would retrieve the article in the search results. Note that, an asterisk symbol followed by a term encompasses all possible variations of that term starting with that word. For example, the term "twin*" will automatically consider terms such as twin, twins, twinning, and twinned. Collectively, the string cast a wide net to cover all publications in our area of interest. As a result, during the identification stage, the number of retrieved articles from Scopus, Web of Science, PubMed, and Engineering Village databases were 268, 263, 212, and 202, respectively, totaling 1051 articles. Next, we eliminated the duplicate papers, resulting in the exclusion of 494 articles. An additional 318 papers were removed after title screening. In the title screening phase, we evaluated the relevance of each article to our research questions based on their titles. It is crucial to emphasize that this was an iterative procedure, wherein only those papers deemed clearly irrelevant to our research questions were excluded at this stage. To mitigate the inadvertent omission of potentially relevant articles, those suspected of even marginal relevance were retained

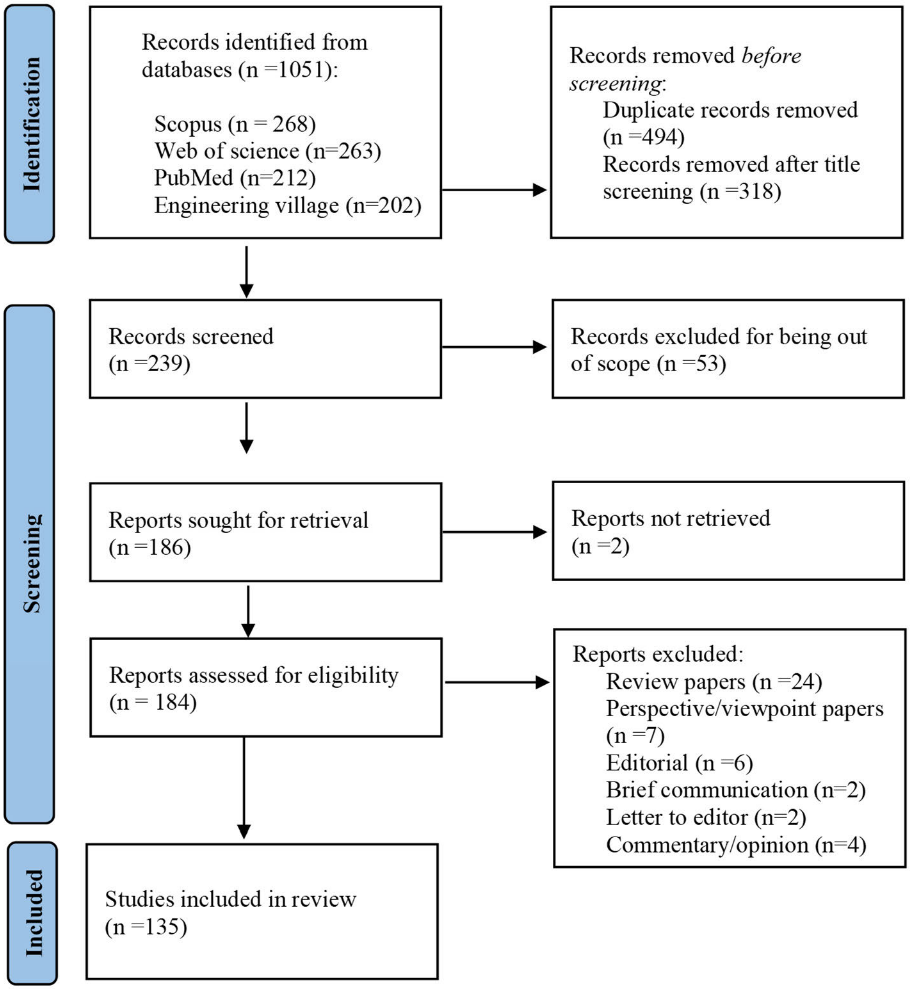


FIGURE 4. Reporting items for the systematic review (adapted from [90]).

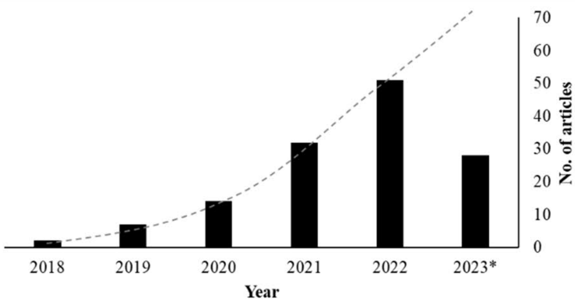
for further evaluation in the subsequent stage of abstract screening. The title screening left us with 239 papers for full-text screening.

2) SCREENING

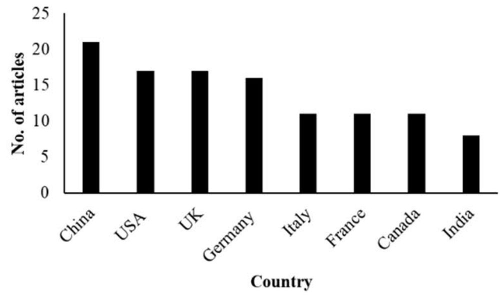
In the second stage, we conducted a procedure composed of manual abstract screening followed by full-text retrieval and categorical evaluation. After the abstract screening, we found 53 papers to be out of the scope of our research focus and therefore eliminated them. These out-of-scope papers include DT applications in ''health'' monitoring of – a machine in a production facility or an infrastructure/building that is not HS. The remaining 186 papers were sought for retrieval of full texts. We were unable to retrieve full texts for 2 papers. Then, we conducted the final round of screening using full-texts of the remaining 184 papers. The remaining 184 articles were filtered based on predefined inclusion and exclusion criteria, which intended to focus on peer-reviewed research instead of technical communications or preprints as summarized in Table 1. The inclusion criteria encompassed English language journal articles, conference articles, conference proceedings, book chapters, and articles in press. Since our review focused on research papers, we excluded preprints, editorials, commentaries, letters to the editor, and brief communications.

3) INCLUSION

As shown in the summary of reporting items in Fig. 4, at the end of the screening stage, we excluded 24 review papers, 7 perspective/viewpoint papers, 6 editorials, 2 brief communications, 2 letters to editor, and 4 commentary/opinion. The justification for excluding review papers is as follows. The objective of this study is to benchmark the state of the art in DT research in HS. To that end, while review papers are useful for establishing the status quo and for discussion purposes; they do not provide any methodological advances or innovation in terms of applications of DTs. Thus, the exclusion of review papers allows us to address our research questions solely based on analysis of advances in research, ensuring objectivity and independence of our findings. Finally, we were able to include 135 papers for final review. This final set of 135 papers was fully reviewed and data extraction was completed in the next step.



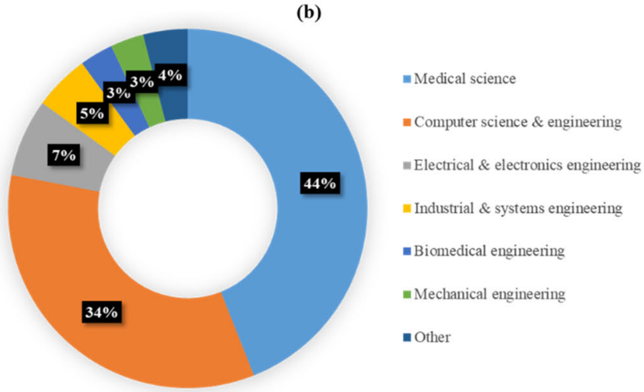
(a)



(b)

Journal	No. of articles
IEEE Access	12
Journal of Medical Internet Research	6
npj Digital Medicine	5
Sensors	4
IEEE Internet of Things Journal, Journal of Personalized Medicine	3 (each)
IEEE Internet Computing, IEEE Journal of Biomedical & Health Informatics, Frontiers in Oncology, JMIR Medical Informatics, Applied Sciences, IFAC-PapersOnLine, Studies in Health Technology & Informatics	2 (each)

(c)



(d)

FIGURE 5. **Lay of the land for DT research for HSs. (a) Number of articles and the growth trend; (b) Countries of origin; (c) Research outlets; (d) Disciplinary background of researchers function of the applied field.**

IV. FINDINGS

In this section, we present our findings from our systematic review organized around the three research questions. Regarding *RQ1*, the current trends, we take stock of the literature and provide a summary of the state of the art. We characterize the research community and outlets, map the focus of current work on DT lifecycles, discuss modeling & twinning techniques used, and finally provide a taxonomy of the literature. With respect to *RQ2*, gaps, and opportunities, we find that there is a prioritization of patient outcomes and process management with little attention to the performance or well-being of healthcare providers. We also note a lack of research on DT verification, validation, and implementation. Finally, regarding *RQ3*, realization challenges, we discuss two issues that hinder the implementation of the DT technology in HS operations: technical challenges that pertain to the extension of modeling and twinning techniques to HSs and the dismissal of data collection and privacy issues.

A. RQ1: CURRENT TRENDS

1) SOURCES, OUTLETS, & COMMUNITIES

We start our discussion with the time series trend. There is an exponential growth trend for DT research for HSs as shown in Fig. 5(a). The number of research articles drastically increased from only 2 in 2018 to 51 in 2022. Around the time we compiled this review, halfway through 2023, there were 29 published research papers, suggesting the trend is poised to continue.

Next, we investigated the countries of origin of publications (based on the first author's affiliation) to establish a status quo of the research community engaged in this field as shown in Fig. 5(b). Fig. 5(b) suggests a trend of global engagement. The leading countries in the research domains include China (21 articles), the USA and the UK (tied with 17 articles), Germany (16 articles), Italy, France, Canada (tied with 11 articles), and India (8 articles). While not shown in the figure, other notable sources of origin countries are the Netherlands, Australia, and South Korea, among others. These findings indicate that the potential of DTs to revolutionize healthcare outcomes has been recognized across the globe, and is in the process of spreading out.

Third, we examined the journals and conferences that are publishing relevant articles to establish a benchmark of the outlets that are disseminating this research. As summarized in Fig. 5(c), the leading journals are IEEE Access (12 articles), Journal of Medical Internet Research (6 articles), npj Digital Medicine (5 articles), Sensors (4 articles), IEEE Internet of Things Journal (3 articles), and Journal of Personalized Medicine (3 articles). Other notable journals are IEEE Internet Computing, IEEE Journal of Biomedical & Health Informatics, and Frontiers in Oncology, among others, with 2 articles each. Overall, IEEE journals are the leading publisher, with IEEE Access being the most popular, and six other IEEE journals have collectively contributed to a total of 10 articles. Moreover, IEEE conferences emerged as the most common avenue with 26 conference proceedings,

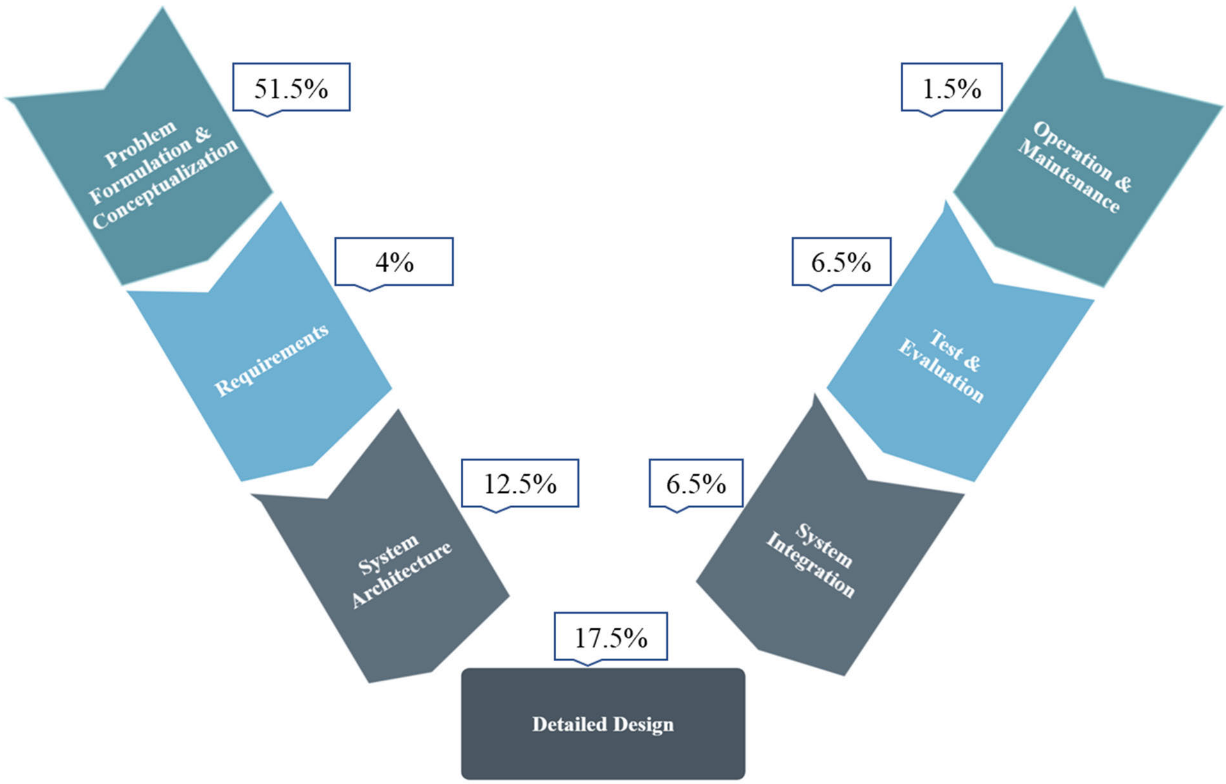


FIGURE 6. Distribution of the focus of the articles along the systems engineering V-model.

further indicating the role of IEEE as the leading community in this research area.

Finally, we investigated the disciplinary backgrounds of the researchers to benchmark the expertise of the research community based on their affiliations. As shown in Fig. 5(d), the majority of the researchers are from either medical science (44%) or computer science (CS) (34%) backgrounds, with significantly low participation from relevant disciplines such as electrical & electronics engineering (EE), biomedical engineering, and industrial and systems engineering. Although the community exhibits some variety, the distribution is highly skewed towards medical science and CS. We found this trend a bit counterintuitive. While it is expected to have a broad representation from medical science and CS, we contend that increased participation from other relevant disciplines could greatly contribute to the growth of the field. We proceed to a discussion of in which stages of the DT lifecycle the research is being concentrated.

2) FOCUS ON EARLIER STAGES OF SYSTEM LIFECYCLE

Every system has a lifecycle [91], [92]: roughly starting from problem formulation to conceptualization, design & development, integration, operations, sustainment, and retirement. Most commonly represented by the systems engineering V-model [93]. Here, the left side of the "V" depicts the early stages of the lifecycle, involving problem formulation, conceptualization, requirements, architecture, and decomposing these into manageable functions and physical elements, which are then planned, designed, and developed. On the right side of the "V", the focus is on the integration of elements, verification, test and evaluation, validation, and eventually deployment, operations, sustainment, and retirement [94].

Going back to Grieves' initial idea, a DT was conceived as a dynamic conceptual model for managing a product's lifecycle, one that adapts as the system evolves. Indeed, to be able to fully realize the potential benefits of DTs for HSs, research is needed for every stage of the system lifecycle [95]. However, as we were conducting our review, we found that the majority of the articles were gravitating towards the left side of the "V", with a strong emphasis on conceptual design. While this could be attributed to DT research in HSs being a recently emerging research thrust, research on implementation, test and evaluation, and validation were nascent, which we found to be an interesting insight that is worth sharing with the community. In Fig. 6, we are providing a categorization of research around system lifecycles, which highlights the concentration of DT research. We used the following classification criteria:

- *Problem formulation and conceptualization:* frames, develops, or proposes novel ideas or theories on the use case of DTs for a particular HS application.
- *Requirements:* defines, documents, and discusses technical requirements e.g., enabling technologies and algorithms, required to develop DTs for HSs.
- *System architecture:* outlines the function to form mapping of DT, along with the necessary interfaces.

Articulating which elements will perform which roles for the broader DT objectives.

- *Detailed design:* discusses detailed design of DT subsystems or components.
- *System integration:* discusses the integration of subsystems and components to form a functional DT or to implement DTs into the operations of a HS.
- *Test and evaluation:* assessment of DT technology through verification and validation to check whether or not the proposed design meets the intended objectives and design assumptions.
- *Operation and maintenance:* research on implementing DTs in real-world operations of HSs and maintaining their performance over time.

As captured in Fig. 6, the vast majority of the articles (51.5%) are focused on one specific life cycle phase: problem formulation and concept development. This suggests that the community is greatly invested in the early stage of the development of DTs; often without an articulate consideration of the necessary stages to facilitate the successful realization of DTs. There is a minor concentration (12.5%) of research around developing system architectures and evaluation of elements to formulate DTs. For example, a distributed DT architecture was proposed for a hemodialysis unit of a hospital [96]. Others have developed an architecture for a healthcare facility management DT [97]. The proposed architecture consists of a data acquisition layer, transmission layer, integrated data middle office layer, service layer, and target layer. Also, a higher portion of research (17.5%) focuses on detailed design. These studies explicitly discuss development subsystems and components specifically for use in DTs, suggesting that there is some impetus towards the right side of the V. For example, they delve into the intricate aspects of crafting hardware [98], [99] or software, including formulation of algorithms [100], [101] that are essential for DTs. Additionally, certain instances involve the exploration and development of DT prototypes [102], [103], [104].

As illustrated by the right side of the V in Fig. 6, there is a lack of implementation research with only 14.5% of total research publications. For instance, only a handful, 6.5% of papers investigate how DTs could be integrated into HS operations [105], [106], or investigate test and/or evaluation of DTs for HSs [107], [108]. These are phases of the system lifecycle that often illuminate previously overlooked compatibility issues and lead to cost and schedule overruns for development programs thus we find the lack of attention concerning [109]. Operations and maintenance research articles [110], [111] take up only 1.5% of the broader pool; however, document useful evidence. For instance, Peng et al. [110] observed DT operations for a hospital building over a year and reported evidence of significant efficiency gains. Additional research focusing on this stage could play an important role by providing empirical evidence regarding return on investment and we expect there would be numerous opportunities in the near future for focusing on operations and maintenance of DTs.

3) MAPPING OF RESEARCH COMMUNITIES TO SYSTEM LIFECYCLE PHASES

HSs are complex sociotechnical systems with multi-level interactions between their social and technical counterparts [17], [112]. Thus, DT for HS research requires interdisciplinary research perspectives. Given this need, we wanted to understand whether researchers were actually forming interdisciplinary teams; and if there was a tendency among the different research teams to pursue different aspects of the DT lifecycle.

To investigate this, we classified the author groups based on their backgrounds into four: disciplinary engineering teams, multidisciplinary engineering teams, medical science teams, and interdisciplinary teams. Here, a disciplinary engineering team is composed of researchers from the same engineering background, e.g., all authors are from computer science or electrical engineering. Multidisciplinary engineering teams are composed of researchers from two or more engineering disciplines, e.g., a mix of researchers from biomedical engineering and industrial engineering. A medical science team is only composed of researchers from medical science backgrounds, e.g., the school of medicine, the school of surgery, the school of dentistry, etc. Finally, an interdisciplinary team has a mix of researchers, including at least one from engineering and one from a medical science background. In Fig. 7, we present a Sankey diagram that maps the focus of these research teams into system lifecycles.

We notice that most of the research is conducted by disciplinary engineering teams (58 articles), followed by interdisciplinary teams (37 articles), medical science teams (27 articles), and multidisciplinary engineering teams (13 articles). While we find it a bit surprising to observe disciplinary engineering teams dominate an interdisciplinary research area, we attribute this trend to the ''first come first serve'' effect. Recalling Fig. 5(d) most research teams consist of CS or EE researchers, and since DT research originated from engineering applications before they expanded to HSs, these groups of researchers are naturally ahead of other research teams. We observe that the disciplinary engineering teams exhibit the broader lifecycle trends discussed in Fig 6, covering all phases of the DT lifecycle with a concentration on earlier phases. Concept development (28 articles), detailed design of system elements (12 articles), and system architecture (9 articles) are the leading focus areas.

The second leading contributors are interdisciplinary research teams that bring together researchers from medical sciences with one or more engineering disciplines. The interdisciplinary teams have been contributing fairly to almost every life cycle phase but one (operation and maintenance), with a strong representation in system integration and test & evaluation – they conduct more than half of the research in these phases. Since DTs for HSs research require interdisciplinary collaboration, one could expect an increased representation from these research groups in the future.

Medical science teams are third. They contribute to all phases except requirements, with more than half of their work focusing on problem formulation and conceptualization.

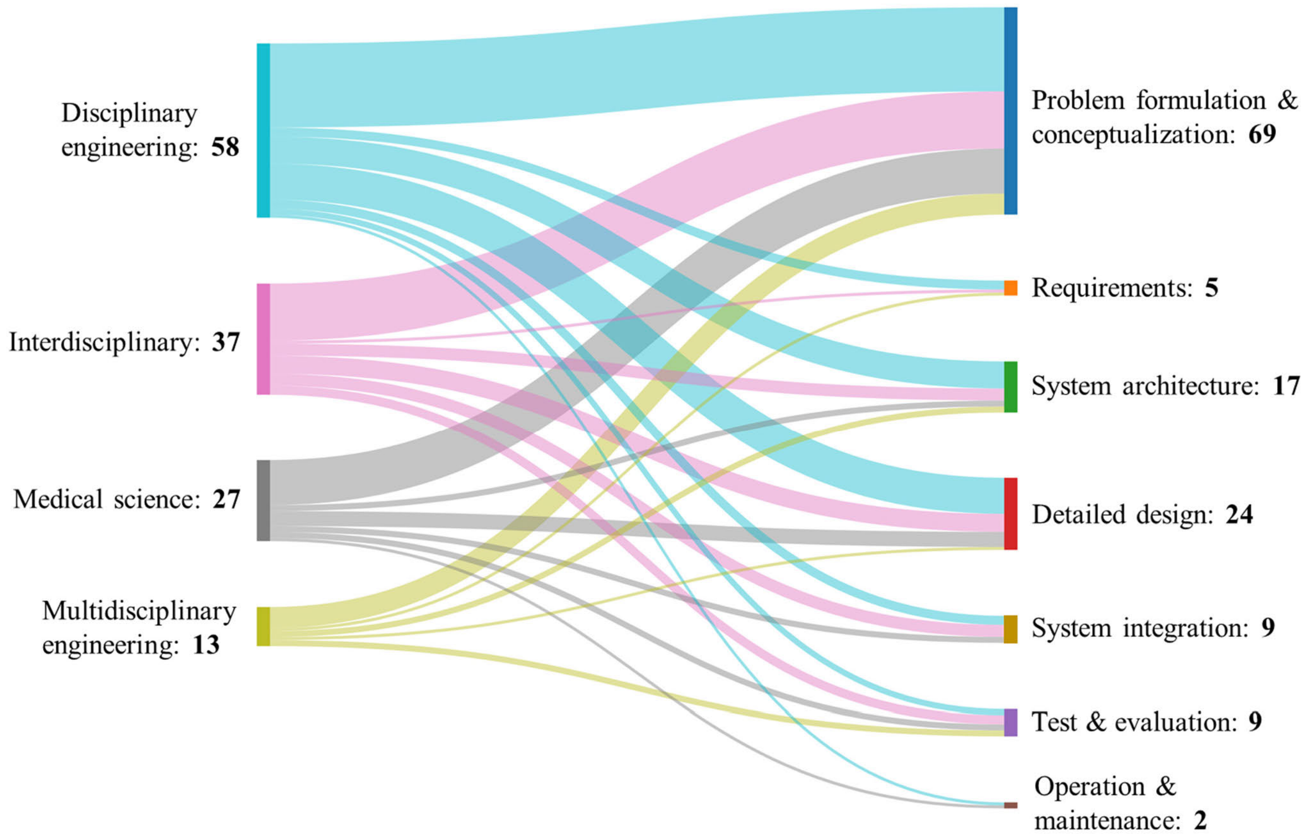


FIGURE 7. Sankey diagram showing the relationship between the researchers' background and the focus of their articles.

Some of their major contribution lies in the concept development of DTs for various diseases' diagnosis and treatment purposes.

Finally, multidisciplinary engineering teams are the least common contributor. Despite constituting a relatively small chunk of the broader pool, we observe that these teams have a tendency to pursue test & evaluation research, proportionally more than others, but interestingly do not pursue system integration. While this might be a limited sample size effect, this could also be originating from their lack of having a medical science contributor (who would have access to an operational HS). We expect multidisciplinary engineering knowledge to contribute more heavily to DT implementation and integration in the future.

4) MODELING & TWINNING TECHNOLOGIES USED

We evaluated the literature in terms of which twinning and modeling methodologies are utilized. As previously discussed in Section II, DT modeling methods are technologies that are used to create the virtual replica and twinning methods are the mechanisms that enable establishing bilateral (P2V and V2P) automated data/information flow. We use the modeling and twinning classification scheme proposed by Thelen et al. [10] to categorize the reviewed articles.

In Table 2, we present the summary of our findings regarding DT modeling methods. The most popular DT modeling technique is data-driven modeling (n=58), followed by geometric modeling (n=32), and physics-based modeling (n=8). There are a handful of studies that use system modeling approaches (n=4), suggesting there might be some potential for systems engineering approaches in DT modeling. We note zero papers that utilized physics-informed ML modeling techniques. Finally, it is important to note that a considerable amount of the articles (n=33) either do not explicitly mention their DT modeling methods or do not use them at all. We find this concerning because without a modeling approach, by definition, the proposed study cannot be considered as a DT (yet were still labeled as such).

TABLE 2. DT modeling methods.

DT modeling methods	No. of articles	References
Data-driven modeling	58	[99], [101], [103], [105], [113]–[124], [125]–[144], [100], [107], [129], [145]–[161]
Geometric modeling	32	[45], [97], [98], [102], [104], [106], [108], [110], [134], [149], [154], [158], [162]–[178], [111]
Physics-based modeling	8	[179]–[186]
System modeling	4	[187]–[190]
Physics-informed ML modeling	0	-

TABLE 3. DT P2V twinning methods.

P2V twinning methods	No. of articles	References
ML model updating	35	[186], [191], [178], [192], [179], [120], [115], [125], [164], [175], [177] [126], [127], [130], [132], [140]–[142], [193], [144], [113], [121], [122], [128], [129], [131], [147], [151], [154], [155], [160], [176], [101], [119], [124]
Measurements as inputs	28	[97], [110], [171], [121], [133], [134], [156], [157], [162], [194], [195], [99], [106]–[108], [136], [153], [163], [167], [169], [174], [184], [185], [190], [196] [102], [116], [168], [169], [180]
Probabilistic model updating	7	[98], [105], [111], [118], [135], [165], [98], [103], [105], [111], [118], [135], [165]
Fault diagnostics & failure prognostics	4	[110], [132], [134], [171]
Ontology-based reasoning	1	[188]

In Table 3, we summarize our findings regarding DT twinning methods. ML model updating is the most popular twinning method (n=35), this could be due to the fact that ML model updating is usually adopted in conjunction with data-driven modeling, the most popular DT modeling method. Second, measurements of physical systems are directly used as inputs to the virtual model to establish P2V connection in 28 papers. This P2V technique is arguably the most straightforward way of establishing P2V twinning and perhaps that contributes to its popularity. Probabilistic model updating was used in 7 articles. This number is surprisingly low given that medical data, especially, physiological data have noise and measurement uncertainties, and thus could benefit from using probabilistic model updating. There have been a few instances of other methods also. For instance, fault diagnosis DTs were implemented to monitor air handling units of a Chinese hospital [110]. Also, an ontology-based reasoning approach was used in developing DTs for security devices within a hospital environment [188]. However, similar to our previous observation for DT modeling techniques, we observe that only 75 papers out of 135 papers disclose P2V twinning methods. This indicates that researchers either overlook, fail to articulate, or discuss this crucial P2V twinning component. We find this to be interesting because without a P2V linkage, by definition, one cannot formulate a DT.

Finally, we summarize the V2P twinning methods used in Table 4, raising some serious concerns. Out of 135 articles we reviewed, only 21 of them either directly or indirectly discuss V2P connections. MPC was found to be the most popular V2P method (n=16). This dynamic control strategy was used for predictive control of inputs over a definite time horizon with applications such as control of medical microrobots [118] or surgical robots [102] and individual's health condition prediction [99], [116].

TABLE 4. DT V2P twinning methods.

V2P twinning methods	No. of articles	References
Model predictive control	16	[97], [99], [105], [110], [111], [116], [118], [153], [155]–[157], [171], [180], [188], [195], [102]
Structural optimization	3	[98], [169], [186]
Predictive maintenance	2	[110], [171]
System reconfiguration	0	-
Production planning	0	-

Another V2P technique, structural optimization, was implemented particularly for the design optimization of robotic arms [169] or medical devices such as birdcage coils of magnetic resonance imaging (MRI) scanners [98]. Additionally, a predictive maintenance approach was adopted in managing healthcare facilities, primarily hospital buildings [110], [171]. On the other hand, applications of other V2P methods such as system reconfiguration and production planning were absent in the existing literature.

The rest of the articles completely disregard V2P connectivity methods. Given that an entity has to include both P2V and V2P connections to be considered as a DT, we contend that this is indicative of a strong mislabeling trend in the community. From this perspective, one could argue that only ∼15% of the articles that are labeled as DT research could actually be considered as such, and the vast majority of them should be classified as either a digital model or a digital shadow. There are two potential explanations for this observation. First, the gap in acknowledging V2P connections could indicate that there is a lack of understanding in the research community regarding the appropriate definitions of DTs. For a complete representation of a DT, researchers need to acknowledge the *automated* bilateral data flow in DT and discuss in detail how this could be achieved i.e., what technologies/methods could potentially be used to enable the mechanism. Alternatively, these studies could have benefited from more parsimonious framing, as contributing to aspects of a DT is still beneficial for the community and could help the field to advance. Second, this gap could be attributed to a lack of implementation research. Given that V2P connections are used to implement operational decisions in physical systems, one could argue that studies overlook a discussion on V2P connections because they were focusing on the partial conceptualization of DTs. While this is understandable, we contend researchers should at least include some discussion regarding V2P connections or acknowledge the limitations of their work accordingly.

5) CLASSIFICATION OF LITERATURE BASED ON RESEARCH DIMENSIONS

We find that DT research in HSs could be classified based on two dimensions: the *objects* being twinned and the *context* in which the twinning is conducted. In terms of the objects, the literature develops DTs for either a product or a process,

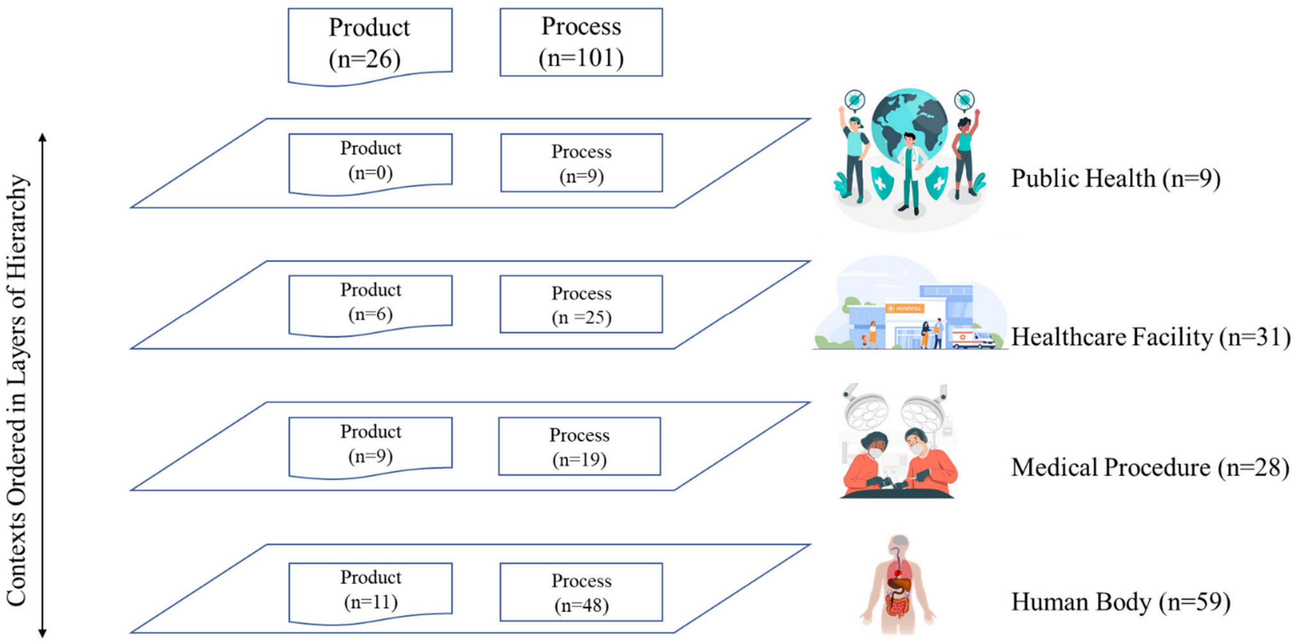


FIGURE 8. Different levels of abstraction in DT for HS modeling, where n denotes the number of articles in that class.

usually with distinct goals, to be operated in some context. Here, the term *product* refers to an engineered system, tool, or artifact, that has been developed for a specific objective to be achieved in a healthcare context. The term *process*, on the other hand, refers to a systematic collection of a series of activities or tasks. For example, disease diagnosis, patient treatment, or other biological or cellular processes.

In terms of the second dimension, the context, there are four main levels: the patient's body, a medical procedure, the healthcare facility, and public health. To elaborate, the patient's body context broadly refers to the biological processes that occur within the boundary of a patient's body. The medical procedure refers to any surgical and non-surgical procedure administered by a healthcare provider to a patient in a healthcare facility or at their homes. The third layer, healthcare facility, refers to places where healthcare is provided and could include hospitals, operating rooms, emergency rooms, inpatient, outpatient, or in-home settings, etc. Finally, public health refers to the health and disease management of communities. Fig. 8 provides a visualization of our categories along with a breakdown of the concentration of the articles. Below we discuss the articles in detail around the object-context categorization. Note that we were unable to fit 8 articles into our classification scheme since those are focused solely on data security, privacy, or ethical issues without designing or developing DTs. We discuss those articles later in Section IV-C.

a: PRODUCT TWINNING

We find that DTs are developed for various products to be operated in three contexts: patient body, medical procedure, and healthcare facility (and currently not in the context of public health). These products exhibit great diversity. They include medical robots for surgery [102], [107], microrobots for drug delivery [118], birdcage radio frequency coils of MRI scanners [98], organ transplantation devices [180], unobtrusive microwave sensors for data collection applications [168], wearable exoskeletons for medical monitoring [196], to name a few.

In the context of a patient's body, DTs have been developed primarily for patient health monitoring [99], [119], disease diagnosis [103], and rehabilitation [190], [197]. For example, a DT for implantable cardioverter defibrillators (ICD) was developed to enable the personalization of device parameters based on patient conditions [119]. This study performed experiments on three virtual patients with evolving heart conditions and showed that the proposed approach could identify ICD parameter settings that can achieve better performance compared to default parameter settings. A DT of a Lymphometer device for early detection and monitoring Lymphedema was explored [103], with the objective of breast cancer prevention.

In the medical procedure context, product DTs were developed to assist providers with surgical procedures [102] and surgical decision-making [184], improve the outcomes of a medical procedure [180], and increase disease diagnosis accuracy [169]. Some examples include internal surgery [118], orthopedic surgery [102], medical imaging [98], and organ transplantation [180]. A magnetic medical microrobot DT was developed to assist with sensitive internal surgeries that could also be used for patient monitoring and precision drug delivery [118]. Similarly, a surgical robot DT was proposed for orthopedic surgeries [102]. A DT of a birdcage RF coil of an MRI scanner for building RF test environments was built that was successfully constructed and verified as a sufficiently accurate description of the physical field [98]. Here, DT provides an alternative for representing the RF test environment. In another research, a DT prototype to predict and improve transplantation outcomes was developed [180]. Some key challenges in designing the DT models such as uncertain process conditions and external forces were also discussed in that work.

Product DTs to be utilized in the context of a healthcare facility are usually developed to help with monitoring facilities and patients within those facilities. For example, Khan et al. [168] developed an unobtrusive microwave sensor and analyzed its performance in data collection in a care home environment. The analysis showed that the position of the sensors is important for sensing the presence of and collecting vital data from patients. Additionally, the use of DTs was illustrated for both access control devices and fire sensors for simulation within the framework [188]. Those researchers have shown further that an architecture based on modern horizontally scalable cloud technologies could be used to realize a useful implementation.

An innovative DT-based real-time solution was introduced that utilizes a robotic device for the remote monitoring of isolated COVID-19 patients' health status [167]. The system allows for efficient navigation and monitoring of patients' conditions from a distance. We limit our discussion to this selected subset of papers in this sub-section and move on to our findings on process twinning DTs.

b: PROCESS TWINNING

The vast majority of DT research in HSs is focused on healthcare processes in four contexts: patient body, medical procedure, healthcare facility, and public health. In these contexts, twinning objectives range from enabling personalized treatment, prognosis, and digital coaching, to precision medicine. Below we elaborate on each.

In the context of the patient's body, DTs are developed for multiple biological levels of hierarchy ranging from a cell level [136], [198], tissue level [147], organ level (e.g., liver, heart, lung) [144], [170]. Nevertheless, DTs are predominantly developed to understand certain disease trajectories to be able to facilitate early diagnosis [132], prevention [136], and patient-tailored treatment for a disease [154]. Digital twinning in the patient body context has the most impact on modeling cardiovascular diseases [113], [126], [129], [131], [132], [138], [143], [149], [152], [160], [161], [170] and cancer care, especially lung cancer [127], [144], breast cancer [121], [153], liver cancer [106], [128], brain cancer [159], head and neck cancer [117], prostate cancer [193], and uterine cancer [151].

Here, we highlight some studies that use existing medical imaging and diagnostics tools to develop DTs in biological processes related to heart and cancer cells. A proof-of-concept DT was proposed for ischemic stroke biomarker identification using a wearable electroencephalography (EEG) headset [132]. A cardio twin architecture for detecting and preventing ischemic heart disease and stroke was proposed using electrocardiography (ECG) data [138]. Others have also developed DTs that use ECG data to detect heart diseases [131]. Other examples of heart disease modeling include monitoring and treatment of acute coronary syndrome [152] and predicting ventricular tachycardia [149]. A DT framework was proposed for personalized and individualized cancer care planning combining clinical, physiological, and demographic data [199]. The use of the cancer DT was demonstrated in detecting cancer using structured radiology reports in three different organ contexts: lungs, liver, and adrenal gland [128].

There is a body of research that looks into building cell-level, tissue-level, organ-level, and body-level DTs for other clinical purposes including diagnosis of liver infection [140], carotid stenosis severity detection [129], diagnosis of respiratory fungal infection [183], prediction of response to sepsis treatment [135], precision treatment of type 2 diabetes [156], [157], etc. Researchers demonstrated that integration of DT and ML algorithms could improve identification, classification, and tailored treatment in case of liver infections such as hepatitis, cirrhosis, and fibrosis [140]. A human DT model was proposed for detecting the severity of carotid stenosis using a combination of computational models (blood flow & head vibration model) and computer vision applications [129]. A clinical proof-of-concept DT was demonstrated to predict the risk of postoperative portal hypertension [150]. The researchers illustrated the potential of a mathematical model of the entire blood circulation system as a numerically assisted clinical tool to transition medical practices from evidence-based medicine to a revolutionary digital era of advanced surgical techniques. Additionally, it was demonstrated that DT-enabled precision nutrition could effectively reduce hemoglobin A1c in patients with type 2 diabetes [156].

Process twinning in the context of medical procedures includes robotic surgery [104], [182], [185], trauma management [162], [200], vertebroplasty [179], arthroplasty [164], laparoscopic surgery [175], [184], dental surgery [122], transcatheter aortic valve replacement [133], throat cancer treatment [142], among others. With the aim of reducing the surgeon's required skill level and cognitive workload during surgery, the use of DT-based haptic assistance during surgical training was demonstrated [182]. Additionally, a novel DT prototype was developed to analyze the requirements of communication in performing remote surgery [104]. A proof in-concept DT was demonstrated using augmented reality and machine learning with an application in laparoscopic surgery [175]. In another work, a DT was developed for trauma management and simulated various bone healing scenarios using 3D X-ray images of patients [162]. Similarly, a trauma DT was developed to digitalize and support the process of severe trauma management [200]. A medical twin virtual environment based on real-world data obtained from clinical patients was designed to simulate transcatheter aortic valve replacement (TAVR) [133]. TAVR is a minimally invasive procedure that involves the insertion of an aortic valve, similar to stent implantation, through a femoral artery without the need for open-heart surgery. Similar attempts at improving surgical outcomes using DTs have also been explored for other procedures including vertebroplasty [179] and arthroplasty [164].

Twinning in the context of a healthcare facility usually aims to process improvements through staff schedule optimization, capacity planning, and enhancing workflows; in units such as emergency departments [187], [201], [202], intensive care units (ICUs) [134], [173], [203], operating room [163], [171], hemodialysis unit [96], ventilation unit [186], hospitals [97], [110], [166], [192], [194], [195], including services provided within these facilities. An ICU DT was developed for the investigation of the real-time allocation of ICU resources and was validated through real medical ICU data [203]. Similarly, DT was developed for processes in the ICU to facilitate remote monitoring, detect faults and anomalies, and enable interventions at an early stage [173]. In a tele-ICU context, a novel extensive simulation framework for human-robot interaction was proposed [134]. Some researchers discussed the requirements of building DT architecture for an emergency department (ED) of a hospital capable of visualizing the service behavior in quasi-real-time and forecasting ED throughput time [201]. DT applications in the operating rooms (ORs) context were also explored [171], particularly in monitoring the air quality, the performance of the HVAC system, and how they affect the MRI machine performance within the OR. Furthermore, to optimize ventilation systems, an integrated framework was devised combining DT and machine learning techniques [186]. The model was capable of capturing infectious disease-related respiratory emissions such as from COVID-19. In a similar work, a DT that uses distributed interactive simulations of a hemodialysis unit was developed to monitor and assess the spread of COVID-19 [96].

The DT of a hospital was studied and empirical evidence of performance improvement was reported in multiple facets of operations such as a 10% increase in management satisfaction, a 1% decrease in annual energy consumption, and an over 10% decrease in facility faults and repairs [110]. Similarly, improved decision-making with better healthcare service was also reported through a case study of a DT of a Chinese hospital [97]. In addition, how the scheduling and allocation of emergency resources in a hospital could be improved using its DT was also studied [192]. In one of their series of studies on investigating DT applications in healthcare facilities management, Karakra and colleagues [195] investigated the usefulness of DT for hospital management, real-time monitoring of patients' pathways, and predicting future outcomes. In their previous work [194], they discussed the main components, the structure, and the way to synchronize the state and the behavior of the DT with the patients' pathways in the real hospital. Earlier, they demonstrated proof-of-concept of a hospital DT for better planning and improvement of usage of hospital resources [166].

In the context of public health processes, we found that the COVID-19 pandemic stimulated research in DTs for public health systems, generally with the objective of controlling the spread of infectious diseases. We found nine articles that discuss building DTs for public health with a specific emphasis on COVID-19. The applications range from pandemic alerting [204] and predicting the virus spread [145], [155], [205] to monitoring social distancing [148] and studying the effectiveness of intervention measures [146]. A blockchain-based DT was developed for pandemic decisions e.g., alerts, quarantine, lockdown, social distancing, etc. [204]. A DT was used to simulate the different possible strategies and scenarios to predict the spread of the COVID-19 virus and minimize the impacts while ensuring continuity in providing services to citizens [145]. Additionally, a DT was developed for a vaccination process and implemented in a clinic, enabling real-time simulation of patient flow and the vaccination center [206]. They suggested that by analyzing the virtual model, issues can be identified and addressed in the actual vaccination center, thereby enhancing vaccination efficiency.

In another study, a DT was utilized to simulate different hypothetical situations to forecast the transmission of the virus [146]. The aim was to assess the efficacy of potential policy interventions along with a prediction of their possible outcomes, balancing public health, citizen well-being, and economic considerations. A conceptual framework was proposed that integrates a digital system for public health emergencies [207]. This approach involves incorporating multi-paradigm simulation (i.e., integration of agent-based, dynamic, and discrete simulation approaches), to construct a DT. A framework was presented that utilizes blockchain technology and dynamic DT for establishing a resilient network between hospitals to effectively respond to pandemics [208]. Two DTs based on a SEIRS (Susceptible – Exposed – Infectious – Recovered – Susceptible) model were introduced and applied to a hypothetical city [155]. The SEIRS model was adapted to account for spatial variation, and whenever possible, the model parameters were derived from official COVID-19 transmission data from the UK. Very recently, a prototype DT was developed as part of an innovative smart healthcare ecosystem based on IoT to prevent the spread of the virus in a real nursing home [205]. A summary of our classification of the literature based on the object and context taxonomy with a full breakdown of articles is provided in Table 5.

B. RQ2: GAPS & OPPORTUNITIES IN DT FOR HS RESEARCH

Recent research trends indicate a growing interest in the adoption of DTs for HS research, which has the potential to revolutionize HSs in many aspects. Nevertheless, our systematic review has unveiled two major research gaps that are currently being overlooked by the current literature. These gaps are considering the human-in-the-loop nature of HSs and the lack of implementation research.

1) CONSIDERING THE HUMAN-IN-THE-LOOP NATURE OF HEALTHCARE SYSTEMS

Current literature mostly focuses on either the patients or their interests when designing and developing DTs. Indeed, patients are the primary beneficiaries of healthcare

TABLE 5. Classification of the articles based on two research dimensions – twinning object and twinning context.

Twinning object	Twinning context			
	Patient body	Medical procedure	Healthcare facility	Public health
Product	[99], [101], [103], [107], [111], [119], [124], [190], [197], [209], [210]	[98], [102], [118], [169], [174], [180], [184], [196], [211]	[116], [167], [168], [181], [188], [212]	-
Process	[100], [106], [108], [113], [117], [121]–[123], [126]–[130], [130]–[132], [135]–[138], [140], [143], [144], [147], [149]–[154], [156]–[161], [165], [170], [176], [183], [198], [199], [213]–[216] [217]	[104], [114], [115], [120], [125], [133], [139], [141], [142], [162], [164], [172], [175], [177], [179], [182], [185], [189], [200]	[96], [97], [105], [110], [124], [134], [163], [166], [171], [173], [178], [186], [187], [191], [192], [194], [195], [201]–[203], [218]–[222]	[145], [146], [148], [155], [204]–[208]

interventions and they require the highest attention. However, DT research for HSs could be extended in two directions complementary to this goal: focusing on the providers and investigating HS resilience.

First, patient safety and quality of care in HSs [12], [223], [224], are significantly linked to human-in-the-loop decision-makers (physicians and nurses) who actively interact with the patients and make the necessary safety-critical decisions for service delivery. Without providers, HSs cannot sustain their operations. Given the criticality of their role, it is crucial to understand how demands of the broader HS, such as workload and stress, influence their decision-making processes, situational awareness, and consequently, the care they deliver. Furthermore, modern healthcare delivery has evolved into a complex dance between the providers and the technological tools that they use, such as Electronic Health Records (EHR) [225]. In short, the interactions between providers, patients, their technological counterparts, and the rest of the HS are being overlooked in the existing literature. We contend that this point is a critical research gap.

Related to the human-technology interaction aspect of the aforementioned point, over the last decade, the impact of the rapid deployment of technology has brought both tremendous advantages and new challenges for HSs [112], [226]. On one hand, clinical decision support can help assist complex diagnostic decision-making, EHR allows patients better awareness of their care processes, and remote monitoring and telehealth are now routine. On the other, physicians and nurses in modern HSs interact far more with technologies such as EHR than with patients during a typical outpatient visit [227]. What adds to the complexity of this puzzle is that healthcare organizations, at least in the US, purchase thousands of individual technologies, each with its own training, user interfaces, underlying security, and ability to integrate with existing workflows [13]. In short, HSs have to be evaluated with these sociotechnical dependencies in mind, and DT research could help address this problem by looking into how systemic changes influence the provider's daily work and care for patients [13], [112], [226], [228].

Additionally, there is a social, provider-team-organization facet of the puzzle. Providers routinely share information with their peers and interact with the patients, while being influenced by organizational factors, such as hierarchical structures, safety culture, and managerial support. These interactions and interdependencies are very complex and dynamic in nature; often creating difficult-to-predict emergent properties that affect HS performance in terms of quality of care, healthcare costs, and human errors.

These gaps present a wealth of opportunities for DT research, particularly in terms of capturing and leveraging these interactions to impact HS performance. For example, in the provider context, process DTs could investigate the quantification and management of a safely attainable workload that is instrumental in provider well-being as well as patient care [229]. Process DTs could also incorporate macroergonomic factors such as task design, provider-technology interface, and organizational factors for improved workspace design for the providers [14], [112]. Additionally, researchers could explore the development of product DTs tailored for providers that could explicitly capture factors such as anthropometry (i.e., physical measurements of the human body), work physiology, and work biomechanics and incorporate them in DT design to reduce provider safety incidents such as work-related back injuries. Provider-focused DT design could explore the development of assistance tools such as surgical training platforms used to improve surgery outcomes.

Second, regarding resilience, HSs are critical infrastructures and at its core, the fundamental HS challenge arises from the tension between the fluctuating demands of the public and the need for continuous operations that require adaptive management of limited resources (e.g., staff schedule, equipment, professional attention, and cognitive load) [230]. Recently, the COVID-19 pandemic has exposed such vulnerabilities worldwide [231], [232]. It has been shown that for a robust and resilient HS, the focus should be on long-term sustainability and building capacity to withstand these external shocks [233], [234]. Regarding this point, DTs offer the ability to manage systems throughout their lifecycle, and smart leveraging of these capabilities can lead to improvements in both patient-level and system-level outcomes. Currently, suboptimal

HS performance, including inefficiencies, safety concerns, physician turnover, and systematic failures such as cost overruns and delays, hamper the overall effectiveness of healthcare delivery. Moreover, providers routinely experience high-intensity interactions with the patients [235], [236] that may asymmetrically influence their stress levels, and when combined with other work-life balance concerns, their overall well-being, leading to systematic issues such as burnout. These pressing concerns have been repeatedly highlighted by National Academies [12], [13], [112], [223], [224], [225], [226], [237], [238] and are an ongoing threat to the robustness and long-term sustainability of HSs.

To address these challenges, the modeling, development, and management of DTs can be extended through the HS lifecycle, enabling the prediction and proactive prevention of undesirable outcomes, system inefficiencies, or failures. Additionally, public health DT applications could be further expanded for community-wide health monitoring. Insights from such prognostics could be utilized to predict and prevent the outbreak of a pandemic. Moreover, DT-based improved user interface design could help reduce diagnosis-, prescription-, and surgical errors, and consequently improve organizational efficiencies. We contend that achieving these goals necessitates comprehensive attention to passive HS stakeholders, including business managers, insurance companies, government bodies, and policymakers, as their decisions significantly shape the long-term functioning of HSs.

2) IMPLEMENTATION RESEARCH

To harness the full potential of DTs and provide the purported value proposition to society, it is imperative that they are translated into real-world HS operations. However, integrating DTs into existing HS processes and workflows is tricky, because of technical, social, and regulatory challenges. These issues may range from standardization concerns to simple user pushback that is commonly observed in HS practice. This trend manifests itself in the concentration of ongoing research with a lack of attention on V2P mapping, as we discussed in detail in Section IV-A. We believe this is a promising avenue for DT research; delving into how DTs can be implemented in practice, and how DT-driven decisions can be effectively translated to change organizations or control operational decisions could significantly transform HS practice.

Related to this point, the adoption and implementation of cutting-edge technologies in healthcare is documented to be challenging [239] and is often accompanied by negative perceptions from its users, e.g., the clinicians [240], [241]. This is partly due to the contextual differences in healthcare facilities, providers, and policymakers. For example, successful healthcare technology adoption is positively correlated with healthcare leadership [242]. DT research can benefit from understanding the nuanced differences among the stakeholders in DT implementation and utilizing that knowledge to maximize overall system performance. Therefore, another significant area of opportunity for DT research is centered around implementation strategies, particularly focusing on obtaining stakeholder buy-in during the design phase to facilitate future adoption and utilization of DTs in HS operations. To achieve this, researchers can employ user-centric design approaches that involve key healthcare stakeholders such as providers from the early stages of DT development, actively seeking their input and feedback over time, incorporating their tacit knowledge and preferences into the design process. Obtaining buy-in from all relevant stakeholders is essential and in this multi-stakeholder collaborative environment, it is imperative to identify and resolve conflicting objectives ahead of time. Balancing the needs, priorities, and interests of diverse stakeholders is crucial to fostering collaboration, maximizing the effectiveness of DTs, and eventually realizing tangible system-level improvements.

In addition, there is a technical aspect of this gap. To pursue implementation-level research, researchers could investigate how commercially available software platforms could be utilized to establish digital connectivity. This could serve as a "low-hanging fruit" by identifying and utilizing existing technologies with high technical maturity to realize some immediate gains. In a notable example, a Microsoft Azure Digital Twin services was used to develop a proof-of-concept DT for an operating room within a healthcare facility [163]. Additionally, a few other studies have delved into the application of Microsoft Azure in constructing DTs for various contexts. These include the creation of DTs for elderly care homes [114], [205] and the development of patient-specific DTs [191]. Other similar platforms such as Predix by General Electric, AWS IoT TwinMaker by Amazon, ThingWorx by PTC, Eclipse Ditto™ by Eclipse Foundation, and Digital Enterprise Suite by Siemens which have already found successful applications in the mechanical and manufacturing sectors [243] could be explored in the healthcare DT context.

C. RQ3: REALIZATION CHALLENGES

The findings of our systematic review indicate that there is a significant lack of research on the realization of DTs and their translation into operations. While various DT applications have been successfully implemented in engineered systems, their widespread adoption in HSs is still nascent. Here, we have identified two challenges that impede the translation of DTs to HS operations. First are the technical challenges i.e., shortcomings associated with defining, modifying, and using DT modeling and twinning techniques for HSs. We contend that one of the factors that create this obstacle is that the vast majority of DT modeling and twinning techniques originate from engineered systems and are not directly transferable to HSs. There is room for growth in the development of these analytical models specifically for HSs. The second set of challenges originates from the collection and use of data, considering privacy, security, and ownership issues. Although these factors necessitate explicit consideration given the sensitive nature of HSs, we find that they are mostly overlooked in the existing literature. We elaborate on both challenges below.

1) TECHNICAL CHALLENGES: MODELING & TWINNING

DT modeling techniques facilitate the creation of virtual representations of physical entities while twinning techniques establish bidirectional data and information connectivity between the physical and virtual twins. Both are integral elements of a DT and are vital for the successful implementation of DTs into HS operations. However, as extensively elaborated in our findings section, roughly 44% of articles included in our systematic review overlook their P2V methodology whereas an astonishing ∼85% ignore V2P connectivity. While this observation could be attributed to the emerging nature of this research domain, we consider this a critical obstacle for the translation of DT technology into public service. So why are we observing this trend and what could be done to bridge this gap?

We start with V2P twinning. In engineered systems, such as manufacturing, establishing V2P links and implementing operational decisions are relatively straightforward as these are often highly automated and electronically controlled processes. However, decision-making processes in HSs differ significantly in part due to the presence of human decision-makers in the loop and in part due to their decentralized organizational architectures. In HSs, interrelated processes (or segments of a given process) could be jointly overseen by different organizational roles governed by different decision authorities. Thus, neither managerial nor physician decisions can be automatically translated into operations through DTs (at least currently). Two approaches could prove useful for this pursuit.

First, recognizing that most V2P methods that originated from engineered systems do not translate well to HSs (we noted only the common use of MPC), innovative V2P methods that are specifically tailored for HSs have to be developed. We contend that hybrid techniques that bring together production control and predictive control could prove useful for this goal; however, we expect their direct infusion into HS products and processes will continue to remain a research challenge. Second, an increased number of interdisciplinary collaborations could be a powerful approach for alleviating some aspects of this problem. While interdisciplinary collaborations already exist in the current literature, these often include a medical expert and an engineer. Bringing in more distant yet relevant experts into these interdisciplinary teams, such as organizational scientists and management scholars, could help identify innovative mechanisms for establishing effective V2P connections. Collectively, these experts could jointly consider organizational barriers with scheduling and process control decisions, and could potentially help circumvent this challenge through a sociotechnical approach.

2) DATA CHALLENGES: COLLECTION, SECURITY & PRIVACY, AND OWNERSHIP

Data-related challenges, including collection, privacy, security, ownership, and ethical issues such as biases, inequalities, and informed consent, are critical for the successful and equitable use of DTs. Unfortunately, we find that the literature generally overlooks these issues. We discuss some of these data-related concerns in the following sub-sections.

a: DATA COLLECTION AND SYNCHRONIZATION

A prevalent issue we observed in conceptual DT studies is the lack of explicit consideration and documentation of the data proposed to be used. A clear description of variables is often absent, leaving gaps in our understanding of the specifics. For instance, information about whether primary, secondary, or a combination of data is proposed to be used is frequently disregarded [244]. Additionally, aspects vital for data fusion and synchronization such as the timing, frequency, and units of data are frequently omitted.

We consider that these require increased attention given that HSs rely on distributed sensors. In HSs, data from various heterogeneous sources, such as medical devices, EHR, wearables, and other sensors, need to be collected and used in conjunction with each other. Here, time is a critical factor in comprehending trends, patterns, and shifts in a patient's health status. Thus, precise timestamps for each datum are essential for establishing a clear temporal relationship between various variables. Furthermore, the timing and frequency of data collection play a pivotal role in harmonizing different data streams. In addition to timing, data can be gathered in various units and formats, depending on the source and the specific medical device used. It is crucial to accurately identify and standardize units of measurement for the sake of precise data integration and interpretation. Inconsistencies in units can introduce errors into the analysis and lead to incorrect conclusions or flawed model predictions. Thus, standardization of measurement units is crucial for ensuring the reliability and accuracy of DTs. In our review of the literature, we found very little attention to these characteristics and we contend that DT research in HSs needs to be more proactive and explicit about these concerns.

b: DATA SECURITY AND PRIVACY

Data security and privacy are key issues to address for any cyber-physical system [245]. However, in the case of DTs for HSs, they are of even greater significance given the confidential and sensitive nature of patient data [154], [246]. This data includes medical records, physiological parameters, diagnostic information, and treatment plans, which are highly valuable and sensitive. As DT technology advances and new capabilities emerge, the privacy, integrity, and confidentiality of sensitive patient information have to be continuously preserved, bringing forth new research needs. These needs are well-recognized by the industry with regulations and standards in place concerning data privacy and data security, such as the General Data Protection Regulation and ISO/IEC 27001 [121]. In Table 6, we provide an overview of the articles that explicitly highlight issues associated with data security and privacy in DTs for HSs and discuss them below.

TABLE 6. List of articles addressing data security and data privacy issues.

Data issues		
Data security	Data privacy	Data security & data privacy
[104], [105], [124], [137], [178], [204]	[143], [151], [166]	[100], [113], [121], [131], [148], [154], [175], [191], [218]

There are plenty of opportunities for research in this area. Research on the design of robust security measures and their implementation into the DT framework could protect sensitive data from unauthorized access, breaches, or misuse. Similarly, research looking into encryption techniques, access controls, and secure communication protocols; and their implementation in a DT framework could safeguard data transmission and storage. Establishing a comprehensive data security framework and regularly auditing and updating security protocols are vital to instilling trust among patients, healthcare providers, and other stakeholders, fostering the successful implementation and adoption of DT technology in healthcare. Nevertheless, we find that only a few articles (n=18) acknowledge and/or address either one of these two issues or both.

Healthcare data security was acknowledged as an important concern in constructing a cloud-based healthcare DT platform [105]. Data security was identified as one of the perceived benefits of combining DTs, blockchain, and data analytics technologies [204]. Data security features of blockchain were utilized in combination with DT's data augmentation features to develop a physical activity monitoring DT framework [137]. A deep learning and XR technology-based medical DT was developed that had neural network-based risky code identification to ensure cybersecurity [178]. Additionally, privacy-preserving similarity query-based healthcare monitoring over cloud computing and DT techniques was designed [100]. As a solution to the data security and privacy concerns, researchers proposed combining cloud computing, edge computing, and federated learning (FL) with DTs in their future research [113]. Other researchers also pointed to addressing data security and privacy issues in their future work [148].

Moreover, FL was combined with DT to ensure the security, trustworthiness, and traceability of data in a medical DT [124]. FL is an ML technique that allows organizations to train AI models using distributed data, eliminating the necessity to centralize or disclose that data [247]. Others integrated blockchain and FL to guarantee privacy and security [154]. Keeping data privacy in mind, a cardiovascular disease prediction algorithm that combines FL and DTs was proposed [143]. In a similar work, FL-based DTs were proposed to protect patient's data privacy [191]. Kaul et al. [151] acknowledged data privacy issues as a key concern in healthcare DT and called for the need to protect DTs from cyberattacks. Similar calls were also made in some other works [131], [175].

Data privacy is another key factor to address in the context of DT technology for HSs [143], [151], [166]. Data privacy entails an individual's capacity to autonomously decide when, in what manner, and to what extent their personal data is disclosed or made available to others. Given the sensitive nature of patient information involved in DTs, preserving the privacy of individuals becomes paramount. Healthcare organizations must prioritize anonymization and de-identification of patient data to prevent unauthorized disclosure of sensitive personally identifiable information. Implementing privacy-enhancing techniques, such as data masking, aggregation, or differential privacy, can help mitigate privacy risks and ensure that patient identities remain protected. Transparent data governance practices, including obtaining informed consent from patients and clearly communicating the purpose and scope of data collection, are essential for maintaining trust and respecting patient privacy rights. Furthermore, strict access controls and role-based permissions should be in place to limit data access to authorized personnel only.

c: DATA OWNERSHIP

Data ownership in the context of DT technology for HSs is critical to establish who has control and rights over the data. Ownership is protected by layers of legal and regulatory standards. Data ownership in DTs typically involves a complex interplay between patients, healthcare providers, and technology developers. Unfortunately, we find no DT articles that explicitly consider or address data ownership issues. While patients are the primary source of their own health data, healthcare providers are instrumental in capturing, managing, and utilizing this data. At the same time, technology developers contribute to the design and implementation of DT systems that enable data collection and analysis. Thus, to address data ownership concerns across these stakeholders, it is necessary to establish clear guidelines and agreements. This includes defining the purpose and scope of data collection, ensuring patient consent and involvement in decision-making, and establishing frameworks for data sharing and access. Overall, striking a balance between patient empowerment, healthcare provider responsibilities, and technological advancements is crucial for equitable and responsible data ownership in DT technology for healthcare.

d: ETHICAL ISSUES

Ethics are a system of moral principles that should be taken into account in designing DTs for HSs and are vital to preserve patient rights, data, and well-being; as well as to ensure responsible and equitable use of DTs. Ultimately, patient and provider trust in DT technologies depends on how well ethical concerns have been addressed in their design. Unfortunately, we found very little attention on ethical issues in developing healthcare DTs. A few research papers [248], [249], [250], [251], [252], a perspective paper by [253], a viewpoint paper by [254], and an opinion paper by [255] discussed ethical issues of personalized healthcare DTs. Surprisingly, articles listed earlier in Table 5 did not address how ethical considerations should be incorporated in DT design, development, and use. We discussed the handful of articles we were able to identify below.

A process-oriented ethical map for DTs was developed that addresses ethical issues in four processes: data collection, data management, data analysis, and information use [250]. This map could be used by DT developers to identify potential ethical risks in the process of transforming raw data into meaningful information. Four major ethical challenges were identified for HS DTs: fairness (in terms of access to technology), responsibility (in terms of liability and ownership), autonomy (in terms of a false sense of control and manipulation risk), and privacy [251]. Others focused on the major ethical challenge of how a person is represented in their virtual twin, arguing that there should be a provision for ''dynamic consent'' so that the person being represented has control over their digital representation [248].

Socio-ethical risks in healthcare DT were discussed in three broader aspects by Popa et al. [252]. First, they put forward that the quality and ownership of data must be ensured to attain socio-ethical benefits from using DTs. Second, they contended that disruptions in management structures and roles could raise the question of responsibility and accountability. Finally, they argued that DTs might not be accessible to everyone e.g., health insurance might not cover it, and that inequality could widen the already existing socio-economical gap. Other researchers also warned that the potential inequality in access to DT-based healthcare could increase segmentation and discrimination going forward [249].

The potential benefits and risks of DTs of children were explored in a research [253]. While the authors believe that DTs could empower children by providing more precise information on their health, they raise concerns about vulnerability, recognition, and participation issues. A few other researchers also explored the ethical challenges in pediatric DTs [254]. They discussed some ethical challenges that include threats from the level of autonomy such as increased mistrust in the pediatricians, loss of human contact with the pediatricians, and worsening the existing inequalities in care.

V. DISCUSSION

This systematic review indicates that DT research for HSs is accelerating across the globe, with a rich exploration of potential applications in various healthcare objects and contexts. Albeit a diversity of perspectives, current DT research in HSs pursues three main objectives: disease modeling & management, personalized treatment & precision medicine, and process optimization. Thus, we conclude that DTs for HS research are still in their infancy, with a subtle confusion regarding the necessary characteristics of a DT, and a constrained exploration of applicable modeling and twinning techniques.

We documented that the majority of DT studies are conducted on a problem formulation/conceptualization level and a significant chunk of papers aim to improve the quality of care by solely focusing on the patient. While this is a logical and necessary first step, we contend that extending these perspectives into the core value-generating function of HSs and capturing how providers are making their decisions could greatly help with improving the quality of care while fortifying the resilience of HSs. Thus, the key conclusion of this study is that there are plenty of opportunities for DT research to contribute to HSs. We believe that DT research could potentially help resolve some of the sustained problems of the healthcare community, that arise from the tension between balancing quality of care, provider wellbeing, process costs, and system resilience [12], [13], [14], [15].

To that end, there is a need for both theoretical and applied research. There is an absence of theoretical frameworks for developing, testing, and implementing DTs for HSs. There is a pressing need for novel DT modeling and twinning methods, that are specifically tailored for the unique characteristics of HSs. New perspectives on V2P twinning techniques would be particularly welcome. There is also significant room for growth in developing methods for smart data fusion, considering the heterogeneity of sensors and the complexity of regulatory red tape. In this regard, the design and development of novel data collection tools such as sensors and wearable devices could also be investigated. Furthermore, there are plenty of opportunities to conduct applied research on DT integration with existing HS. Development of test and validation frameworks, and extending the scope of applications remains a high-value research target for realizing the full potential of DTs for HS.

Moreover, research on DTs for HSs should also address relevant ethical issues such as biases, inequalities, informed consent, trust, and transparency. Biases in healthcare data and algorithms could disproportionately impact certain patient groups. Similarly, if DTs cannot be made accessible to the entire population regardless of socioeconomic status, race, or other factors, it could potentially exacerbate already existing healthcare disparities. Informed consent is crucial to ensure that the patients/persons have clear and comprehensive information about how their DTs will be used and also to guide them in making informed decisions. Trust and transparency are correlated as the level of transparency, clarity, and openness about the DTs put in use will impact how the patients find the technology to be trustworthy and reliable. Addressing these ethical issues is critical for providing equitable, trustworthy, and effective healthcare.

VI. CONCLUSION

In this paper, we present the first systematic review of DTs for HSs research; characterizing the state-of-the-art in terms of current trends, research gaps, and realization challenges. The current trends suggest that this emerging research area could potentially revolutionize healthcare by identifying novel mechanisms and providing new capabilities through a plethora of applications. Although interdisciplinary research is prevalent in this area, existing work is concentrated around certain siloes that offer only a constrained view of the diverse nature of HSs. We contend that participation is required from other more distant disciplines, for example, organizational scientists and systems engineers, to foster the growth of this community and alleviate some of the existing barriers towards realization. In summary, this paper has four main conclusions.

First, there is a need to move forward from the conceptual design phase and shift focus to integration, testing, evaluation, and validation research. Technical challenges such as modeling and twinning technologies, data-related challenges such as privacy, security, and ownership, and ethical issues such as biases and inequalities are areas that require immediate attention. The development of twinning technologies that are tailored towards HSs, especially V2P twinning, is a key research challenge that hinders the broader adoption of DTs for HSs. We contend that overcoming these challenges may necessitate the synthesis of interdisciplinary knowledge that is siloed in distinct disciplines.

Second, we discover that a substantial portion of existing literature fails to clearly articulate the specific techniques proposed to be used or employed in their DT modeling, which is a necessary element of establishing DTs. This oversight suggests that researchers may be inaccurately labeling their models as DTs or overclaiming the intellectual merit of the proposed work. This observation, supported by incomplete descriptions of the fundamental components of proposed DTs, also raises serious concerns about the reproducibility and usability of existing research. Future work could benefit from an increased awareness of these issues.

Third, we document that the researchers are primarily designing DTs with only the patient's interests in mind, for a *product* or a *process* in four different HS *contexts* – the patient's body, a medical procedure, a healthcare facility, and public health. While this taxonomy does not aim to be exhaustive, it effectively captures the concentration of research and highlights research gaps between and across these layers. Clearly, HSs offer numerous rich socio-technical research questions to investigate and we expect future work to more intricately explore this diverse landscape.

Fourth, we argue that the aforementioned focus on the patient's perspective implicitly creates a research gap by overlooking the critical role of healthcare providers. HSs are composed of an interwoven set of products and processes that are primarily operated by providers whose performance is actively being influenced by the rest of the HS. These issues tend to compound and lead to national problems such as provider burnout or physicians quitting the workforce [12], [13], [14], [15]. Thus, neglecting how provider decision-making and well-being are affected by these interactions leaves a ton of untapped potential for DT research. Future DT modeling efforts could more explicitly consider the human-in-the-loop nature of HSs and explore the integration of provider macro-ergonomics into DT design. This may require the development of novel artifacts, such as wearable sensors, for capturing human factors and communication among decision-makers. Overall, there is a pressing need for a sociotechnical perspective when designing DTs for HSs [17], [112], [226]. The successful translation of DTs in everyday HS management will depend on navigating these challenges.

Nevertheless, these barriers bring forth new opportunities. Our hope is that DTs could facilitate data-driven equitable management of HSs if developers and researchers can effectively capture and incorporate the interests of *all* associated stakeholders into account. For instance, catering to providers' needs, preferences, and constraints would enable them to manage their workload effectively, which may ultimately enhance job satisfaction, overall performance, and the quality of care. Likewise, DTs could provide both local and national organizations a pathway for an efficient workforce and resource management; improving organizational performance by creating win-win situations. This could potentially provide patients and communities with personalized healthcare services for a fraction of the cost [256].

ACKNOWLEDGMENT
The authors would like to thank the support of Virginia Tech's Institute for Society, Culture, and Environment (ISCE).

REFERENCES
[1] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," *White Paper*, vol. 1, pp. 1–7, Mar. 2014.
[2] S. Khan, T. Arslan, and T. Ratnarajah, "Digital twin perspective of fourth industrial and healthcare revolution," *IEEE Access*, vol. 10, pp. 25732–25754, 2022, doi: 10.1109/ACCESS.2022.3156062.
[3] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the digital twin: A systematic literature review," *CIRP J. Manuf. Sci. Technol.*, vol. 29, pp. 36–52, May 2020, doi: 10.1016/j.cirpj.2020.02.002.
[4] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, "Digital twin driven prognostics and health management for complex equipment," *CIRP Ann.*, vol. 67, no. 1, pp. 169–172, Jan. 2018, doi: 10.1016/j.cirp.2018.04.055.
[5] Q. Lu, X. Xie, A. K. Parlikad, and J. M. Schooling, "Digital twin-enabled anomaly detection for built asset monitoring in operation and maintenance," *Autom. Construct.*, vol. 118, Oct. 2020, Art. no. 103277, doi: 10.1016/j.autcon.2020.103277.
[6] R. van Dinter, B. Tekinerdogan, and C. Catal, "Predictive maintenance using digital twins: A systematic literature review," *Inf. Softw. Technol.*, vol. 151, Nov. 2022, Art. no. 107008, doi: 10.1016/j.infsof.2022.107008.
[7] D. M. Botín-Sanabria, A.-S. Mihaita, R. E. Peimbert-García, M. A. Ramírez-Moreno, R. A. Ramírez-Mendoza, and J. D. J. Lozoya-Santos, "Digital twin technology challenges and applications: A comprehensive review," *Remote Sens.*, vol. 14, no. 6, p. 1335, Mar. 2022, doi: 10.3390/rs14061335.
[8] M. Kumbhar, A. H. C. Ng, and S. Bandaru, "A digital twin based framework for detection, diagnosis, and improvement of throughput bottlenecks," *J. Manuf. Syst.*, vol. 66, pp. 92–106, Feb. 2023, doi: 10.1016/j.jmsy.2022.11.016.
[9] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*, F.-J. Kahlen, S. Flumerfelt, and A. Alves, Eds. Cham, Switzerland: Springer, 2017, pp. 85–113, doi: 10.1007/978-3-319-38756-7_4.
[10] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B. D. Youn, M. D. Todd, S. Mahadevan, C. Hu, and Z. Hu, "A comprehensive review of digital twin—Part 1: Modeling and twinning enabling technologies," *Struct. Multidisciplinary Optim.*, vol. 65, no. 12, p. 354, Nov. 2022, doi: 10.1007/s00158-022-03425-4.
[11] P. Armeni, I. Polat, L. M. De Rossi, L. Diaferia, S. Meregalli, and A. Gatti, "Digital twins in healthcare: Is it the beginning of a new era of evidence-based medicine? A critical review," *J. Personalized Med.*, vol. 12, no. 8, p. 1255, Jul. 2022, doi: 10.3390/jpm12081255.
[12] *The World Health Report 2000: Health Systems: Improving Performance*, World Health Organization, Geneva, Switzerland, 2000.