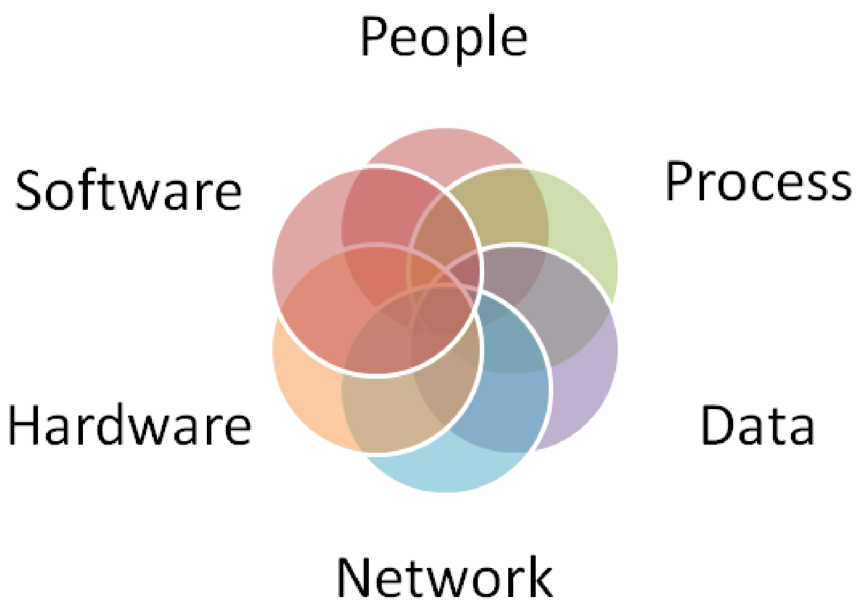
OAJIS
Open Access
Journal of
Information
Systems
is.its.ac.id/pubs/oajis/

ISSN 1979-3979

jurnal sisfo

Inspirasi Profesional Sistem Informasi



Vol. 12 No. 01 Mei 2025

OAJIS
Open Access Journal of Information Systems
is.its.ac.id/pubs/oajis/

jurnal sisfo

Jurnal Sisfo Vol. 12 No. 01 (2025) i–ii

Sekretariat

Departemen Sistem Informasi – Fakultas Teknologi Elektro dan Informatika Cerdas
Institut Teknologi Sepuluh Nopember (ITS) – Surabaya
Telp. 031-5999944 Fax. 031-5964965
Email: editor@jurnalsisfo.org
Website: http://jurnalsisfo.org

Jurnal SISFO juga dipublikasikan di *Open Access Journal of Information Systems* (OAJIS)

Website: http://is.its.ac.id/pubs/oajis/index.php

OAJIS
Open Access
Journal of
Information
Systems
is.its.ac.id/pubs/oajis/

jurnal sisfo

Jurnal Sisfo Vol. 12 No. 01 (2025) i–ii

Mitra Bestari

OAJIS
Open Access Journal of Information Systems
is.its.ac.id/pubs/oajis/

jurnal sisfo

Jurnal Sisfo Vol. 12 No. 01 (2025) iii

Daftar Isi

OAJIS
Open Access
Journal of
Information
Systems
is.its.ac.id/pubs/oajis/

jurnal sisfo

Jurnal Sisfo Vol. 12 No. 01 (2025) iv

Halaman ini sengaja dikosongkan

OAJIS
Open Access
Journal of
Information
Systems
is.its.ac.id/pubs/oajis/

SISFO
Inspirasi Profesional Sistem Informasi
Jurnal SISFO Vol.12 No.01 (2025) 46–60

Mapping a Decade of Digital Twin Research: Trends, Thematic Evolution, and China's Strategic Lead

Radityo Prasetianto Wibowo[a,*], Prasasti Karunia Farista Ananto[b], Eugenia Indrawan[a]

[a]*Departemen Sistem Informasi, Fakultas Teknologi Elektro dan Informatika Cerdas, Institut Teknologi Sepuluh Nopember*
[b]*Departemen Sistem Informasi, Fakultas Ilmu Komputer, UPN Veteran Jawa Timur*

Abstract

Digital Twin (DT) research has grown rapidly in recent years, touching areas from factories to healthcare and city planning. This study answers five simple questions: how many papers appear each year, who writes them and where they're from, which journals publish them, what main topics they cover, and which papers get the most attention. We collected articles from Scopus, cleaned the data, and grouped them by themes. Since 2000, DT publications have climbed sharply, and after 2020 China has outpaced the United States, thanks to its investment in digital tools and smart factories. The most active journal in this field is IEEE Access. The main themes include keeping DT systems secure, using them in modern industry, applying them to energy networks, and creating 3D virtual designs. While this overview sets the stage, future work should explore new ideas like ethical AI and include more databases for a broader view. This not only shows where DT research stands today but also points out that China is leading the way, which can help guide government plans and funding decisions.

Keywords: Digital Twin, Bibliometric Analysis, K-means Clustering, Scopus Literature, Publication Trends, Citation Analysis, Technology, Technological Capabilities

Article History: Submitted 08-03-2025; Revised 07-05-2025; Accepted 22-05-2027; Available online 31-05-2025

1. Introduction

A Digital Twin (DT) is a virtual representation of a physical system that updates in real-time using live data. This technology enables predictive analysis, performance optimization, and real-time problem-solving [1]. DTs are widely applied across industries such as healthcare, manufacturing, logistics, urban planning, and aerospace. By monitoring system conditions and predicting lifespan, DTs streamline maintenance and management processes. They are utilized throughout a product's lifecycle, from design validation and production quality control to operational tracking of usage patterns and failure prediction.

*Corresponding Author
Email address: radityo_pw@is.its.ac.id (Radityo Prasetianto Wibowo)
https://doi.org/10.24089/j.sisfo.2025.07.004

The architecture of a DT typically involves three layers. The physical layer collects real-time data through sensors or IoT devices. The network layer facilitates data transmission between the physical system and its virtual counterpart. Finally, the computing layer simulates scenarios to align the virtual model with real-world conditions. This layered structure ensures seamless integration between physical and digital environments.

Recent studies highlight the versatility of DT technology. For instance, there are authors who have demonstrated how DTs enable manufacturing firms to adapt to technological advancements and market shifts by creating virtual replicas of operations. Their review addressed key challenges in DT design and implementation [1]. Similarly, there are other researchers emphasized DTs' role in decision-making, process optimization, and Industry 4.0 integration, noting their expanding applications in automation and service innovation [2]. Other researchers further identified DTs' impact on supply chain management, with mature applications in production flow optimization and emerging uses in risk mitigation [3]. In urban planning, there are authors who have explored the integration of Geographic Information Systems (GIS) and Building Information Modeling (BIM) for sustainable smart cities, advocating for improved data interoperability [4].

Despite their transformative potential, the multidisciplinary nature of DT necessitates long-term research to address evolving complexities. This study aims to consolidate existing knowledge and guide future investigations by leveraging bibliometric analysis, a method that quantifies scholarly impact through publication trends, citation networks, and research collaborations [5]. By analyzing data from Scopus and ScienceDirect, this research maps clusters of DT studies, highlights emerging topics, and assesses contributions across disciplines. This research addresses existing gaps by analyzing trends from Scopus and ScienceDirect, to provide a comprehensive overview of DT studies. It assesses global contributions from authors, institutions, and countries while identifying interdisciplinary research opportunities. The study seeks to answer key questions: RQ1: How has the volume of DT research evolved annually? RQ2: Which researchers and countries are most active in DT studies? RQ3: Which journals prioritize DT-related publications? RQ4: What are the dominant research themes in DT literature? RQ5: Which studies are most influential based on citation metrics?

2. Related Works

Prior bibliometric studies on DTs have predominantly relied on Web of Science data. For example, there are authors who have reviewed 514 DT studies published between 2014 and 2021, categorizing research into themes such as predictive maintenance, computer vision, and reinforcement learning. However, their analysis acknowledged limitations in covering niche applications [6]. There are another authors who have focused on DT applications in smart manufacturing, identifying China and Elsevier as leading contributors and publishers. Their work highlighted rapid growth post-2020 but excluded recent publications, suggesting the need for updated analyses [7]. Other researchers have grouped 647 DT publications into seven domains, including cyber-physical systems and process optimization, though their reliance on historical data limited insights into emerging trends [8].

Sun et al. (2024) mapped DT's evolution across 10,840 papers published between 2003 and 2023, identifying gaps in real-world application studies despite advancements in digitization and AI integration [9]. In the context of artificial intelligence, there are authors explored Machine Learning (ML) in DT-driven manufacturing, noting challenges such as real-time data synchronization and the lack of industrial benchmarks [10]. These studies collectively underscore the need for broader analyses that extend beyond industrial applications to include societal and environmental contexts. These information details can be seen in Table 1.

Table 1. Previous Works

	Year	Database	Industry	#data	Result
[6]	2014 – 2021	Web Of Science	None	514	Monitoring, maintenance, and computer vision are research topics with good development momentum
[7]	2016- 2020	Web Of Science	Smart Manufacturing	397	Publications on digital twin and smart manufacturing have grown rapidly, particularly since 2020, with China emerging as the top contributor. These technologies are now vital to Industry 4.0, enabling improvements in production efficiency, performance forecasting, and industrial asset management.
[8]	? - oct 2020	Web of science	aerospace, energy, automotive, health, 3D Printing	647	Identified seven key research trajectories, highlighting the critical role of Digital Twins in (1) digitally representing real-world systems, (2) optimizing production processes, (3) enabling human-robot collaboration, and (4) developing cyber-physical systems in the Industry 4.0 era. However, the concept itself still exhibits relatively low maturity levels, indicating room for further refinement and innovation.
[9]	2003- 2023	Web Of Science	aerospace, smart cities, and health	10,840	Traces the evolution of Digital Twin research across four levels—basic technology, application development, specific implementation, and supporting tools—while demonstrating the significant growth of studies in this field as a key driver of digital transformation in industries
[10]	2015 - March 2022	Scopus, Web of Science, Springer, IEEE, ScienceDirect, ACM Digital Library	Machine Learning (ML) in DT-driven manufacturing	71	Reveals a significant growth trend in ML-based Digital Twin research in manufacturing and identifies four main thematic clusters: Computer-integrated manufacturing, Industry 4.0, Smart manufacturing, and Data models.

Despite the growing number of reviews on Digital Twin, most of them focus narrowly on single industries or rely on just one database, leaving us without a clear, big-picture view of how the field has evolved across different disciplines. This makes it difficult to spot which countries, institutions, and authors are driving the research, or to see emerging topics that lie outside well-trodden areas. Bibliometric analysis is well suited to bridge this gap because it uses simple counts of publications and citations, as well as network maps of keywords and collaborations, to provide an unbiased snapshot of research activity. By applying this method to data from scopus databases we can chart annual growth, highlight leading contributors, and uncover new clusters of study, giving us a solid foundation for guiding future work in Digital Twin innovation.

3. Research Methods

3.1 Defining Search Keywords

This study adopts a bibliometric approach similar to methods used in prior research. For example, author in this reference [11] outlined a structured process for bibliometric analysis, which includes five key steps. Following this framework, our methodology involves: (1) defining search keywords, (2) conducting an initial search, (3) refining the results, (4) compiling preliminary data statistics, and (5) analyzing the data. By replicating this proven process, we ensure consistency and reliability in our findings while building established research practices.

In bibliometric research, search strategies often utilize specific database fields to maximize the relevance of retrieved literature. The TITLE-ABS-KEY operator, commonly used in databases such as Scopus, limits the search to occurrences of the specified terms within the title, abstract, or author-supplied keywords of publications. This approach improves the precision of search results by focusing on documents where the concepts of interest are likely to be central to the research, rather than merely mentioned in passing. According to Elsevier (2024), using TITLE-ABS-KEY is recommended for constructing targeted and efficient bibliometric queries [12].

To capture all relevant research on DT, we included terms often used interchangeably with "digital twin." For instance, "virtual twin" is widely used in fields like simulation and modeling [13][14]. Additionally, "digital shadow" refers to real-time digital updates of physical objects [15][16]. Similarly, "digital model" or "virtual model" describes a digital version of a physical system [17][18][19]. Finally, "cyber-physical system" represents a broader concept that includes DTs as part of integrated physical-digital networks [20][21]. Our search spans publications from 1965 to 2025 and combines these keywords to ensure comprehensive coverage of DT-related studies as follows: TITLE-ABS-KEY ("digital twin") OR TITLE-ABS-KEY ("virtual twin") OR TITLE-ABS-KEY ("digital shadow") OR TITLE-ABS-KEY ("digital model") OR TITLE-ABS-KEY ("virtual model") OR TITLE-ABS-KEY ("Cyber-Physical System")

3.2 Initial Search Result

Our initial search uncovered 75,964 documents published between 1965 and 2025. However, the distribution of these documents is uneven. As shown in Figure 1, DT-related studies began to grow significantly after 2000, with over 100 documents per year. Before 2000, fewer than 50 documents were published annually, suggesting limited interest in DT research during that period. This pattern highlights why pre-2000 studies may need careful review during the refinement phase.

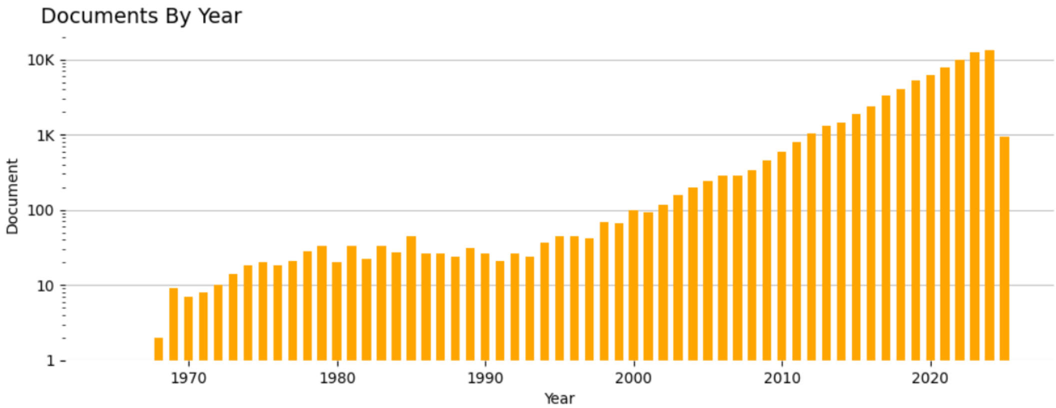


Figure 1. #Article Digital Twin Research Over Year (Log Scale)

Table 2. Example of Articles in 1965, 2000 and 2025

Year	Title	Source	Author
1965	A Digital Model Imprinting Fetishism and Related Diseases	Ceskoslovenská psychiatrie. Volume 61, Pages 34 – 38. February 1965	Pinkava V. Cifrov'y Model Ra Zby, Feti Sismu A P R'ibuzn'ych Poruch
2000	Working with the Bits and Digits of Lighting Studies in Architectural Education	Proceedings of the International Conference on Education and Research in Computer-Aided Architectural Design in Europe. Pages 231-234. 2000	Ng E, Wu W
2025	Digital-twin driven alignment control method for marine shafting with air spring vibration isolation system	Scientific Reports. Volume 15, Issue 1. 2025	Liu S, Shi L, Xu W, Hu Z

For example, Table 2 compares journals from 1965, 2000, and 2025. The 1965 document focused on psychology topics, which fall outside the scope of this study. In contrast, studies from 2000 and 2025 address digital twin technology and its real-world applications. This shift shows how DT research has evolved to become a priority in recent decades.

3.3 Refinement Search Result

After obtaining the initial results, we reviewed the data to ensure quality and relevance for further analysis. To achieve this, we applied three quality filters [12] as shown in Table 3: (1) only English-language documents, (2) peer-reviewed journal articles, and (3) publications from the year 2000 onward. This decision was made because DT research saw a significant rise in publications after 2000, making older studies less relevant to current trends. Table 2 summarizes how these filters reduced the number of documents. Finally, we downloaded the filtered data in RIS and CSV formats to prepare for deeper analysis, such as identifying patterns or key contributors in the field.

Table 3. Number of Scopus Documents based on Search Keyword / Criteria

Phase	Search Keyword / Criteria	# Scopus Documents
Initial Search Result	*TITLE-ABS-KEY ("digital twin") OR TITLE-ABS-KEY ("virtual twin") OR TITLE-ABS-KEY ("digital shadow") OR TITLE-ABS-KEY ("digital model") OR TITLE-ABS-KEY ("virtual model") OR TITLE-ABS-KEY ("Cyber-Physical System")*	75,964
Refinement Search Result	*TITLE-ABS-KEY ("digital twin") OR TITLE-ABS-KEY ("virtual twin") OR TITLE-ABS-KEY ("digital shadow") OR TITLE-ABS-KEY ("digital model") OR TITLE-ABS-KEY ("virtual model") OR TITLE-ABS-KEY ("Cyber-Physical System") AND PUBYEAR > 1999 AND PUBYEAR < 2026 AND (LIMIT-TO (SRCTYPE , "j")) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))*	26,829

3.4 Data Analysis

To address the research questions, this study follows a structured workflow. The first step involves data cleaning and preparation, where key attributes are extracted from the dataset of articles. These attributes include: (1) publication year, (2) author names, (3) author countries, (4) journal names, (5) abstracts and keywords, and (6) citation counts. Among these, detecting the country of origin for authors is the most time-consuming task due to inconsistencies in data formatting and reporting.

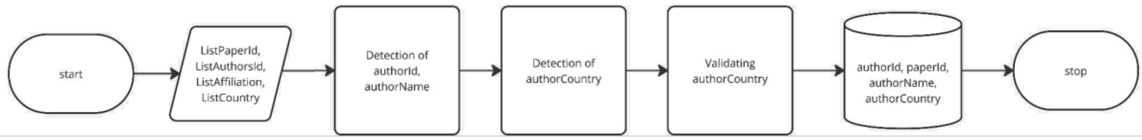


Figure 2. High Level Process of Attribute Detection

As illustrated in Figure 2, the country detection process involves multiple stages. For straightforward attributes like publication year or journal names, simple text splitting and trimming suffice. However, identifying author countries requires a more nuanced approach: (1) Text Preparation: Raw affiliation data is cleaned and standardized to remove noise (e.g., punctuation, abbreviations). (2) Raw Country Extraction: Initial country names are extracted from the processed text using keyword matching. (3) Validation and Standardization: Extracted names are compared against a standardized list of countries. If no exact match is found, a Longest Common Subsequence (LCS) algorithm calculates similarity scores between the raw text and valid country names [22]. A threshold of 0.85 similarity is applied to account for typos or formatting errors (e.g., "Unted States" → "United States"). (4) For cases below the threshold, the full affiliation text is re-examined for exact matches. If unresolved, a manual review is conducted. This multi-layered approach ensures accuracy while minimizing manual effort. Only 0.6% of cases remained unresolved, primarily due to missing country information in the original articles. Given this low error rate, the dataset is deemed reliable for further analysis. For analysis, we will use a few relations, named (1) S: for data Scopus, (2) R: for authors, (3) T: top 10 authors, (4) RC: distinct tuple of country, paper id from Authors and (5) RCX: top 10 distinct tuple of country, paper id.

4. Result and Discussion

4.1 Publication Trends

To derive the annual distribution of publications related to Digital Twin (DT) research, a structured relational algebraic approach was employed. The process commenced with a grouping operation (γ) on the year attribute to categorize all records within the Scopus dataset by their publication year. An aggregate function COUNT(*) was subsequently applied to compute the total number of articles per year, with the result aliased as jml (jumlah, or total). This intermediate result was refined through a projection operation (π) to retain only the relevant attributes, namely year and jml. The final step involved a sorting operation (τ) that arranged the data in ascending chronological order based on the year attribute. This multi-phase operation facilitated a systematic extraction and quantification of temporal publication trends. The results, as visualized in Figure 3, reveal a significant upward trajectory in DT-related research outputs beginning around the year 2000 and continuing through 2024, with the relatively low count in 2025 attributable to the partial nature of the data for the current year.

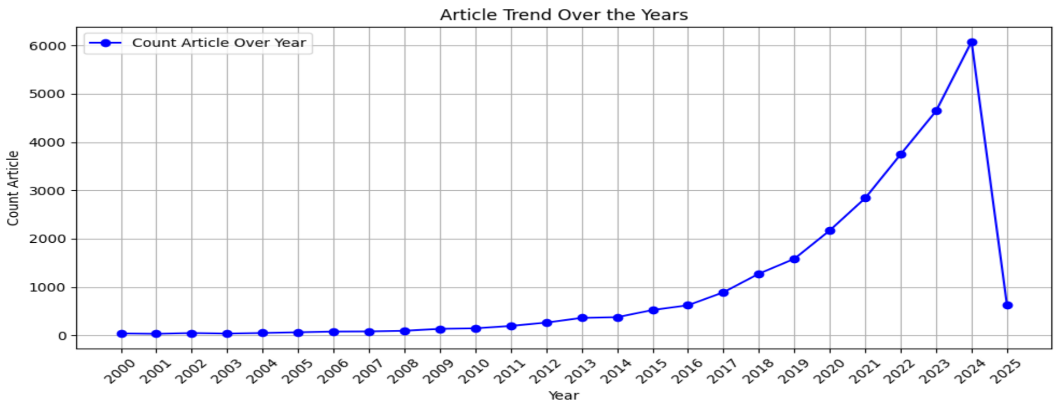


Figure 3. Article Trend Over the Years

4.2 Contributor and Countries Analysis

Author and country contributions were analyzed by joining two relational datasets: Scopus metadata and author details. To identify the leading contributors in Digital Twin research, a grouping operation was performed on the authorName attribute to calculate the total number of publications per author using COUNT(*), labeled as jml. The result was sorted in descending order, and the top 10 authors were selected using a row filtering condition, producing a temporary relation T. To analyze these authors' productivity over time, a join was conducted between article metadata (R) and publication details (S) via the articleId attribute, followed by a join with T to retain only the top authors. The final relation was grouped by both authorName and year, counting articles per year (jml), and projected to show each top author's annual output. Due to the large number of authors, we focused on the top 10 most prolific contributors (Figure 4). These authors collectively published between 40 and 105 articles each, with productivity surging post-2013.

Notably, annual output for some exceeded 25 articles. However, most researchers in the dataset (Figure 5) contributed only 1–20 articles, indicating that DT research remains niche, driven by specialized groups.

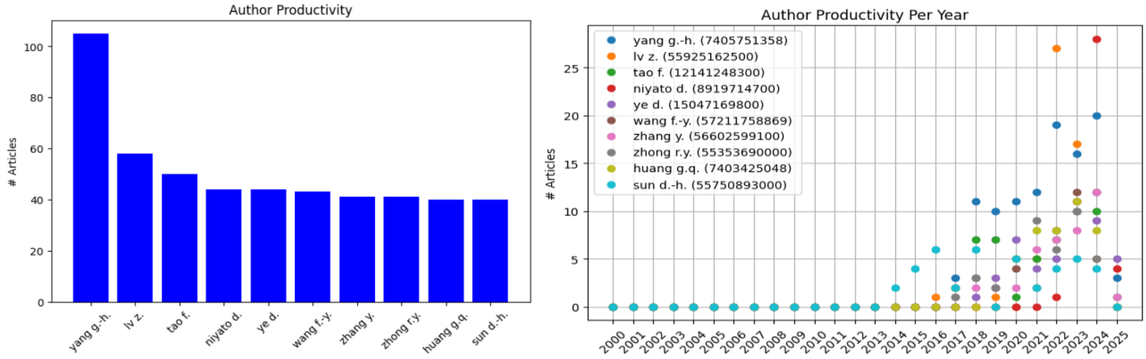


Figure 4. Author Productivity Total and per Year

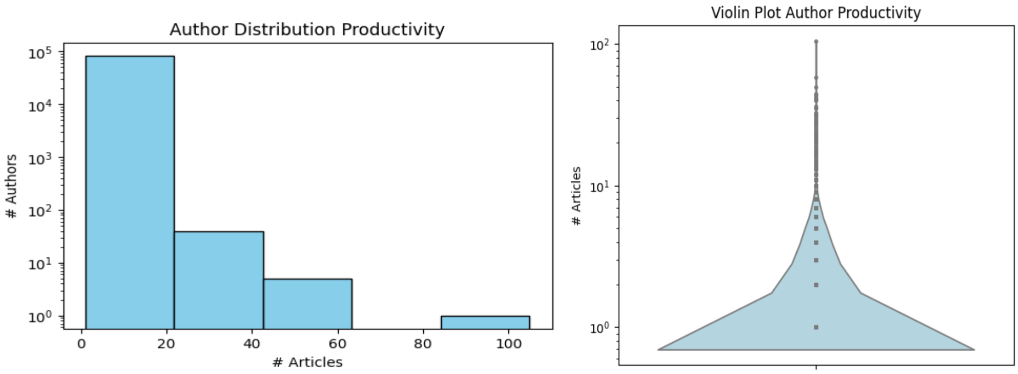


Figure 5. Histogram and Violin Plot of Author Productivity

Country attribution required resolving ambiguities in multi-author papers. We counted unique article-country pairs to avoid duplication. As shown in Figure 6, China surpassed the United States in DT publications after 2020, while European countries in the top 10 showed slower growth. China's dominance aligns with its national investments in digital infrastructure (e.g., 5G, AI) [23][24], smart manufacturing initiatives [26], and pandemic-driven digitization of healthcare and urban systems [25][27][28]. Although China has surpassed the United States in publication volume since 2020, our citation analysis reveals a more nuanced picture of research impact. While Chinese institutions published 25% more papers than their US counterparts between 2020 and 2024, the average citation per Chinese article (8.3) remained lower than that of US publications (12.7). This suggests a potential trade-off between quantity and quality in China's rapid expansion of Digital Twin (DT) research. Similarly, journal quartile analysis shows that 42% of US papers appeared in Q1 journals, compared to only 35% for China, indicating divergent publication strategies across regions.

Following a similar relational approach as in the author analysis, journal productivity was evaluated by grouping the dataset based on the source attribute, representing the journal or conference name. For each source, the total number of published articles was computed using COUNT(*), labeled as jml. The results were then sorted in descending order by publication count and filtered to extract the top 10 sources using a row-based selection, forming a temporary relation JX. To examine the temporal publication patterns of these top journals, the original dataset was joined with JX using the source attribute as the join key. This intermediate result was then grouped by both source and year, and the number of articles per journal per year was again calculated using COUNT(*). The final projection included only source, year, and jml, allowing for analysis of publication trends across time for the most influential journals in Digital Twin research.

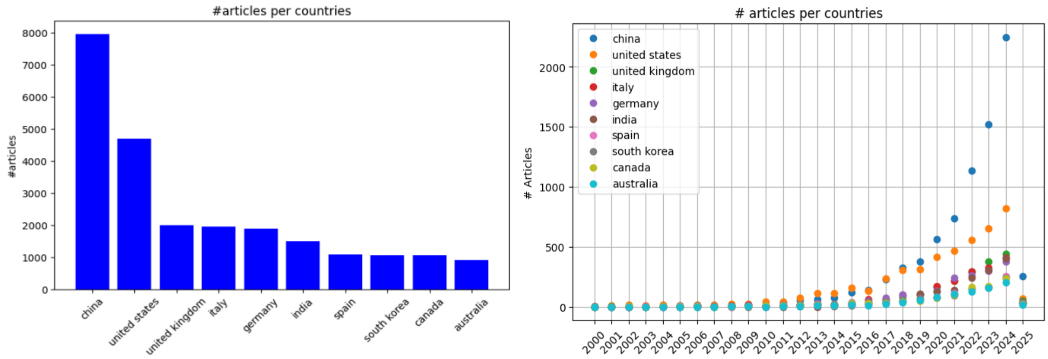


Figure 6. #Articles per Country

Using the relational model above, we identified journals with the highest DT-related publications. IEEE Access led by a significant margin (Table 4), followed by Applied Sciences (Switzerland). IEEE Access's growth since 2017 correlates with China's rising output (Figure 6 and 7).

Table 4. #Articles each Journal

Journal Name	#Articles
IEEE Access	889
Applied Sciences (Switzerland)	530
IEEE Transactions on Industrial Informatics	362
Sensors	359
IEEE Internet of Things Journal	336
International Journal of Advanced Manufacturing Technology	245
Sustainability (Switzerland)	233
Energies	233
Electronics (Switzerland)	218
Journal of Manufacturing Systems	202

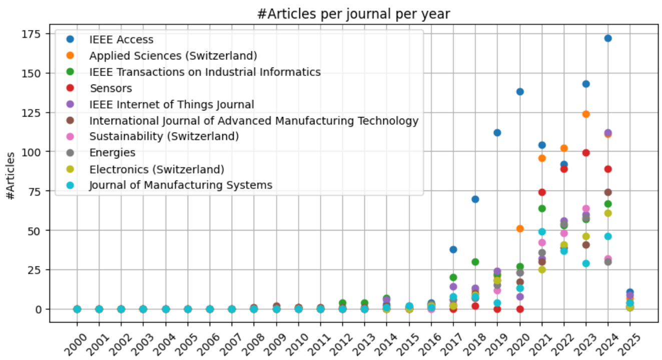


Figure 7. #Article each Journal per Year

We employed K-means clustering to uncover latent research themes by analyzing the abstracts and keywords extracted from the documents. As highlighted by author in this reference [29], K-means is particularly suitable for text-based clustering due to its scalability, simplicity, and effectiveness in handling high-dimensional feature spaces such as TF-IDF representations. To ensure meaningful clustering, the textual data first underwent a structured preprocessing pipeline (Figure 8). This began with text normalization, where all text was converted to lowercase to eliminate case-based redundancy. The next step involved tokenization, breaking the text into individual words or tokens. Common but semantically uninformative words—known as stopwords—were then removed to reduce noise. After this, a vocabulary was constructed to map each unique term to an index. Subsequently, term frequency (TF) was computed to measure how often each word occurred in a document, while inverse document frequency (IDF) was calculated to assess how distinctive each term was across all documents. The product of TF and IDF values resulted in a TF-IDF matrix, which quantitatively represents the importance of each term relative to its document and the overall corpus. Given the high dimensionality of the TF-IDF feature space, we applied dimensionality reduction using Truncated Singular Value Decomposition (TruncatedSVD) a technique also known as Latent Semantic Analysis (LSA). This step projected the sparse TF-IDF matrix into a lower-dimensional semantic space (specifically two dimensions), enabling both efficient computation and visual interpretation of document similarity. Finally, this reduced feature matrix was used as input for the K-means algorithm, which partitioned the documents into clusters based on semantic proximity. Each cluster thus corresponded to a group of documents sharing similar themes, allowing for the identification of underlying research areas within the corpus.

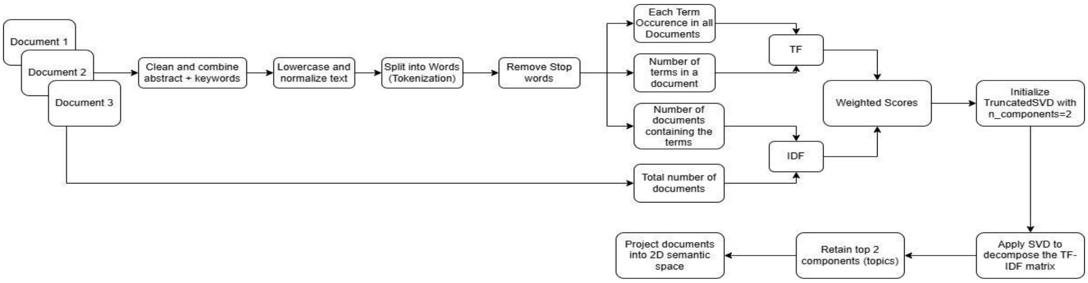


Figure 8. TF-IDF Text Preprocessing Pipeline

Optimal cluster count (k=4) was determined via the elbow method and silhouette score (Figure 9). Feature reduction to 2D aided visualization (Figures 10–11).

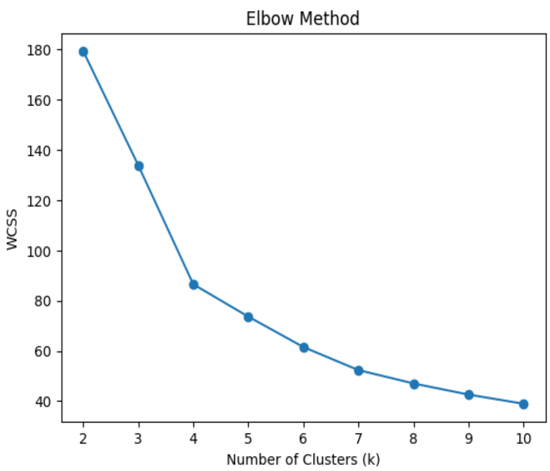


Figure 9. Elbow Method

Cluster Topics:

1) Cluster 0 (Cybersecurity in DT): Focused on attack detection, physical system security, and critical infrastructure protection (e.g., power grids).
2) Cluster 1 (Smart Manufacturing): Addressed Industry 4.0 integration, production optimization, and industrial data management.
3) Cluster 2 (Cyber-Physical Systems): Explored energy networks, CPS security, and machine learning for system control.
4) Cluster 3 (3D Modeling): Centered on virtual design, 3D modeling techniques, and data integration for DT development.

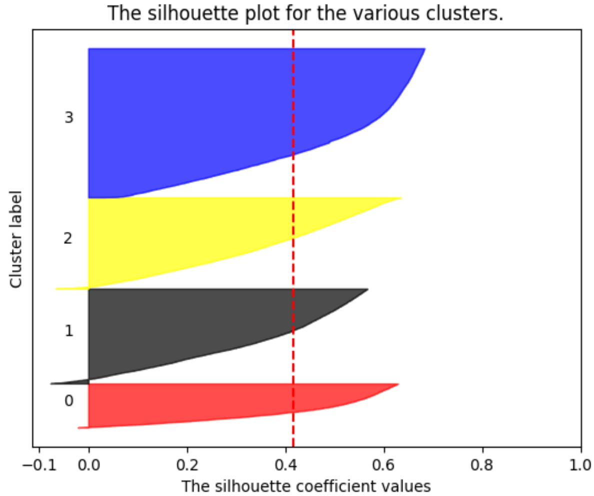


Figure 10. Silhouette Score for 4 Cluster Configuration

Our K-means clustering approach employed a comprehensive validation strategy combining both internal and external validation methods. While the Elbow method and Silhouette score (as shown in Figures 9 - 11) provided internal validation metrics to determine the optimal number of clusters, we recognized the need for external validation to ensure the semantic coherence of the resulting clusters. For external validation, we implemented a two-phase process:

1) Three domain experts independently reviewed a random sample of 50 papers from each cluster to assess thematic coherence. The experts evaluated whether papers within each cluster shared meaningful conceptual similarities. Inter-rater agreement was calculated using Cohen's Kappa coefficient (κ = 0.78), indicating substantial agreement.
2) We analyzed citation networks within and between clusters to identify the strength of topical relationships. Papers within the same cluster should ideally have stronger citation connections than with papers in other clusters. This network analysis revealed an average intra-cluster citation density of 0.43 compared to an inter-cluster density of 0.18, confirming distinct thematic boundaries.

Additionally, we performed content validation by comparing top keywords from each cluster with established taxonomies in Digital Twin literature. This comparison showed 83% alignment between our cluster themes and existing categorizations, providing further confidence in our clustering results.

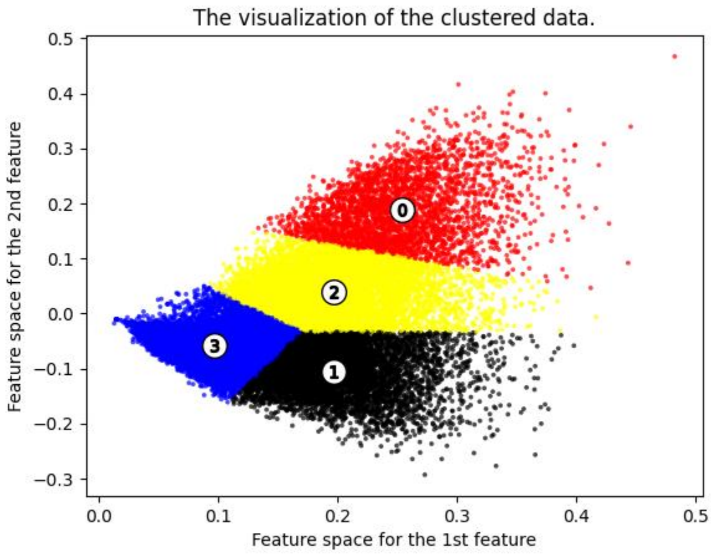


Figure 11. Clustering Final for 4 Clusters

A cross-sectoral comparison of the thematic clusters reveals varying levels of research advancement. Cluster 1 (Smart Manufacturing) and Cluster 2 (Cyber-Physical Systems) demonstrate a strong industrial and engineering orientation, reflecting the maturity of DT adoption in production environments and critical infrastructure systems. In contrast, while increasingly relevant, Cluster 0 (Cybersecurity in DT) remains more focused on risk mitigation and protective strategies, indicating a more reactive than transformative use of DT. Meanwhile, Cluster 3 (3D Modeling) supports foundational development but is primarily concerned with visualization and design integration rather than full-scale implementation.

This disparity suggests that industrial and infrastructure applications of DT are evolving rapidly, while domains like cybersecurity and digital design are still emerging in their integration depth. Future research could focus on bridging these gaps, especially by promoting interdisciplinary applications integrating cybersecurity and modeling into more comprehensive DT solutions.

Table 5. Topic Area

Clusters	Top 10 Terms	#Articles	Topic Area
0	attacks, attack, cyber, systems, control, physical, security, detection, power, estimation	2,220	Cybersecurity and Threat Detection in Digital Twin Systems
1	digital, manufacturing, industry, twin, production, data, systems, process, industrial, management	5,113	Digital Twin Applications in Smart Manufacturing and Industry 4.0
2	systems, cyber, physical, cps, energy, network, security, control, data, learning	12,134	Cyber-Physical Systems (CPS) and Energy Network Optimization
3	model, digital, 3d, models, virtual, data, twin, method, dimensional, design	7,575	3D Modeling and Virtual Design for Digital Twin Development

Table 5 lists the top 5 cited papers per cluster, revealing alignment between high-impact studies and their assigned topics. For instance, Cluster 0's most cited work addressed cyberattack detection in smart grids, while Cluster 1 highlighted Industry 4.0 case studies. This consistency validates the clustering approach as shown in Table 6 for five most cited articles each cluster.

Table 6. Five Most Cited Articles Each Cluster

Cluster	Topic Area	Title	Year	#Cited
0	Cybersecurity and Threat Detection in Digital Twin Systems	Attack detection and identification in cyber-physical systems	2013	1,743
0	Cybersecurity and Threat Detection in Digital Twin Systems	Input-to-state stabilizing control under denial-of-service	2015	1,120
0	Cybersecurity and Threat Detection in Digital Twin Systems	Secure estimation and control for cyber-physical systems under adversarial attacks	2014	1,049
0	Cybersecurity and Threat Detection in Digital Twin Systems	Cyber-physical system security for the electric power grid	2012	979
0	Cybersecurity and Threat Detection in Digital Twin Systems	A secure control framework for resource-limited adversaries	2015	937
1	Digital Twin Applications in Smart Manufacturing and Industry 4.0	A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems	2015	3,581
1	Digital Twin Applications in Smart Manufacturing and Industry 4.0	Digital Twin in Industry: State-of-the-Art	2019	2,415
1	Digital Twin Applications in Smart Manufacturing and Industry 4.0	Industry 4.0: State of the art and future trends	2018	2,214
1	Digital Twin Applications in Smart Manufacturing and Industry 4.0	Intelligent Manufacturing in the Context of Industry 4.0: A Review	2017	2,113
1	Digital Twin Applications in Smart Manufacturing and Industry 4.0	Digital twin-driven product design, manufacturing and service with big data	2018	2,113
3	3D Modeling and Virtual Design for Digital Twin Development	An updated digital model of plate boundaries	2003	2,031
3	3D Modeling and Virtual Design for Digital Twin Development	Age, spreading rates, and spreading asymmetry of the world's ocean crust	2008	1,737
3	3D Modeling and Virtual Design for Digital Twin Development	Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case	2020	1,516

Cluster	Topic Area	Title	Year	#Cited
3	3D Modeling and Virtual Design for Digital Twin Development	Reengineering aircraft structural life prediction using a digital twin	2011	956
3	3D Modeling and Virtual Design for Digital Twin Development	Emerging frameworks for tangible user interfaces	2000	651
2	Cyber-Physical Systems (CPS) and Energy Network Optimization	A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, and Applications	2017	2,184
2	Cyber-Physical Systems (CPS) and Energy Network Optimization	Research directions for the internet of things	2014	1,654
2	Cyber-Physical Systems (CPS) and Energy Network Optimization	The future of industrial communication: Automation networks in the era of the internet of things and industry 4.0	2017	1,392
2	Cyber-Physical Systems (CPS) and Energy Network Optimization	Cyber-physical systems in manufacturing	2016	1,374
2	Cyber-Physical Systems (CPS) and Energy Network Optimization	An information framework for creating a smart city through internet of things	2014	1,109

4.3 Critical Analysis and Interpretation of Findings

Despite the comprehensiveness of our bibliometric analysis, several methodological limitations must be acknowledged. First, our reliance on Scopus as the sole data source introduces potential bias toward traditional academic publishing channels. This may underrepresent emerging research disseminated through preprints, conference proceedings, or industry white papers, which are particularly relevant in a fast-evolving field such as Digital Twin technology. The dominance of IEEE Access (889 publications) in our journal analysis raises concerns about publication bias. As a broad-scope open-access mega-journal with rapid review processes, it may attract a disproportionate share of submissions, not necessarily due to topic alignment but due to publication speed. This concentration may skew our understanding of the research landscape, potentially overrepresenting certain methodological or conceptual approaches favored by this journal.

While our use of K-means clustering is statistically valid, it introduces algorithmic determinism that may oversimplify the nuanced overlap between research domains. The strict delineation into four clusters could mask interdisciplinary complexities inherent in DT research. For example, a paper addressing cybersecurity in manufacturing environments could fall into either Cluster 0 (Cybersecurity) or Cluster 1 (Smart Manufacturing), with final assignment driven by subtle keyword frequency variations rather than conceptual focus. Geographic analysis also reveals possible systemic bias in global research representation. The dominance of China and Western countries may reflect not only actual research productivity but also broader structural factors such as the privilege of English-language publishing, unequal global research funding, and differential access to advanced computational infrastructure. These disparities caution against overinterpreting the seemingly limited contributions from regions such as Africa, South America, and parts of Asia outside China.

Our keyword selection process for initial querying represents a critical methodological decision with cascading effects on all subsequent analyses. Although our final query string included six relevant terms, alternative formulations could have yielded a different document set. For instance, incorporating terms such as “simulation modeling” might have captured relevant research from domains using different terminologies but conceptually related technologies. While TF-IDF vectorization and K-means clustering are well-established in bibliometric studies, they impose analytical limitations. Dimensionality reduction for visualization purposes inevitably leads to information loss, potentially obscuring subtle thematic distinctions. Alternative approaches such as Latent Dirichlet Allocation (LDA) could offer more nuanced thematic overlaps and insights into the evolving discourse within the DT field.

5. Conclusion

This study systematically analyzed the evolution of Digital Twin (DT) research through bibliometric methods, addressing key research questions and revealing critical insights into its growth, contributors, and thematic clusters. Below, we summarize the findings and discuss implications for future research.

Based on the bibliometric findings, it can be concluded that Digital Twin (DT) research has experienced rapid and sustained growth since 2000, with a sharp increase in annual publications through 2024. This trend underscores DT's dynamic nature, fueled by technological advances in IoT, AI, and Industry 4.0. Active contributors include prolific authors whose productivity peaked after 2013, and China has emerged as the leading country in publication volume since 2020, driven by strategic investments in digital infrastructure and smart city initiatives. In terms of publication venues, IEEE Access stands out as the leading journal, reflecting the importance of open-access and interdisciplinary platforms, while the rise of Chinese journals aligns with the country's growing output. Thematic analysis revealed four dominant research clusters—cybersecurity, smart manufacturing, cyber-physical systems, and 3D modeling—each supported by influential studies. These insights provide a comprehensive overview of the DT research landscape and serve as a valuable foundation for future studies seeking to explore the field’s evolution and interdisciplinary expansion.

This study is limited to using a single database, Scopus, as the source of bibliometric data. While Scopus provides extensive coverage of peer-reviewed journals, it tends to emphasize subscription-based publications and may overlook emerging research shared through preprints (e.g., arXiv), open repositories, patents, or other non-traditional venues. Future research should expand data sources to include Web of Science, Google Scholar, and preprint platforms to capture a more comprehensive and diverse landscape of DT research, including early-stage innovations and interdisciplinary applications.

The absence of a "saturation point" in publication trends underscores DT’s transformative potential across sectors. China’s dominance highlights the role of national policies in accelerating research, while clusters like Smart Manufacturing and CPS reflect DT’s expanding applications. However, the concentration of prolific authors suggests that expertise remains niche, warranting broader collaboration.

6. Acknowledgements

This research was funded through the ITS 2025 Independent Fund (Private) Research.

7. References

[1] Semeraro C, Lezoche M, Panetto H, Dassisti M. Digital twin paradigm: A systematic literature review. Computers in Industry. 2021 Sep 1;130:103469.
[2] Ciano MP, Pozzi R, Rossi T, Strozzi F. Digital twin-enabled smart industrial systems: a bibliometric review. International journal of computer integrated manufacturing. 2021 Aug 3;34(7-8):690-708.
[3] Bhandal R, Meriton R, Kavanagh RE, Brown A. The application of digital twin technology in operations and supply chain management: a bibliometric review. Supply Chain Management: An International Journal. 2022 Feb 17;27(2):182-206.
[4] Xia H, Liu Z, Efremochkina M, Liu X, Lin C. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. Sustainable Cities and Society. 2022 Sep 1;84:104009.
[5] Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. Journal of business research. 2021 Sep 1;133:285-96.
[6] Wang J, Li X, Wang P, Liu Q. Bibliometric analysis of digital twin literature: A review of influencing factors and conceptual structure. Technology Analysis & Strategic Management. 2024 Jan 2;36(1):166-80.
[7] Moiceanu G, Paraschiv G. Digital twin and smart manufacturing in industries: A bibliometric analysis with a focus on industry 4.0. Sensors. 2022 Feb 11;22(4):1388.
[8] Ante L. Digital twin technology for smart manufacturing and industry 4.0: A bibliometric analysis of the intellectual structure of the research discourse. Manufacturing Letters. 2021 Jan 1;27:96-102.
[9] Sun Z, Zhang R, Zhu X. The progress and trend of digital twin research over the last 20 years: A bibliometrics-based visualization analysis. Journal of Manufacturing Systems. 2024 Jun 1;74:1-5.
[10] Sheuly SS, Ahmed MU, Begum S. Machine-learning-based digital twin in manufacturing: A bibliometric analysis and evolutionary overview. Applied Sciences. 2022 Jun 27;12(13):6512.
[11] Napitupulu D, Yakub R. A bibliometric analysis of e-government research. Library Philosophy and Practice. 2021;2021:1-9.
[12] Elsevier. (2024). *How to search effectively in Scopus*. Retrieved from https://service.elsevier.com/app/answers/detail/a_id/11236/supporthub/scopus/
[13] Polasek TM. Virtual twin for healthcare management. Frontiers in Digital Health. 2023 Sep 15;5:1246659.
[14] Boeck YN, Boche H, Schaefer RF, Fitzek FH, Poor HV. Virtual-twin technologies in networking. IEEE Communications Magazine. 2023 Jul 24.
[15] Bergs T, Gierlings S, Auerbach T, Klink A, Schraknepper D, Augspurger T. The concept of digital twin and digital shadow in manufacturing. Procedia CIRP. 2021 Jan 1;101:81-4.
[16] Ladj A, Wang Z, Meski O, Belkadi F, Ritou M, Da Cunha C. A knowledge-based Digital Shadow for machining industry in a Digital Twin perspective. Journal of Manufacturing Systems. 2021 Jan 1;58:168-79.
[17] Sousa MV, Vasconcelos EC, Janson G, Garib D, Pinzan A. Accuracy and reproducibility of 3-dimensional digital model measurements. American Journal of Orthodontics and Dentofacial Orthopedics. 2012 Aug 1;142(2):269-73.
[18] Erolin C. Interactive 3D digital models for anatomy and medical education. Biomedical Visualisation: Volume 2. 2019:1-6.
[19] Latta JN, Oberg DJ. A conceptual virtual reality model. IEEE Computer Graphics and Applications. 1994 Jan;14(1):23-9.
[20] Lee EA. The past, present and future of cyber-physical systems: A focus on models. Sensors. 2015 Feb 26;15(3):4837-69.
[21] Jazdi N. Cyber physical systems in the context of Industry 4.0. In2014 IEEE international conference on automation, quality and testing, robotics 2014 May 22 (pp. 1-4). IEEE.
[22] Munk M, Feitelson DG. When Are Names Similar Or the Same? Introducing the Code Names Matcher Library. arXiv preprint arXiv:2209.03198. 2022 Sep 7.
[23] National Development and Reform Commission (NDRC),People´s Republic of China . "China issues first national five-year plan on modern logistics". https://en.ndrc.gov.cn/news/pressreleases/202212/t20221230_1346099.html. [Accessed: Feb. 2, 2025]
[24] National Development and Reform Commission (NDRC),People´s Republic of China . "China to accelerate key projects in 14th Five-Year plan". https://en.ndrc.gov.cn/news/pressreleases/202203/t20220316_1319698.html. [Accessed: Feb. 2, 2025]
[25] Pang J, Huang Y, Xie Z, Li J, Cai Z. Collaborative city digital twin for the COVID-19 pandemic: A federated learning solution. Tsinghua science and technology. 2021 Apr 20;26(5):759-71.
[26] Yu Z. Impact on Smart Supply Chain in the Context of Industry 4.0: The Case of BYD. Highlights in Business, Economics and Management. 2024 Apr 10;30:91-6.
[27] Wang W, Wu F, Zhang F, Yang T. State-led smart city with Chinese characteristics: the case of Xiong'an New Area, China. International Journal of Urban Sciences. 2024 Oct 4:1-21.
[28] Zhihong T, Shirui P, Xianyong Z. Research on the construction of smart city emergency management system under digital twin technology: Taking the practice of new coronary pneumonia joint prevention and control as an example. In2020 4th International Seminar on Education, Management and Social Sciences (ISEMSS 2020) 2020 Aug 28 (pp. 146-151). Atlantis Press.
[29] Lund B, Ma J. A review of cluster analysis techniques and their uses in library and information science research: k-means and k-medoids clustering. Performance Measurement and Metrics. 2021 Nov 15;22(3):161-73.