Review

Bibliometric Analysis of Missing Value Imputation (MVI) Research Using Python and VOSviewer: Trends and Future Directions

Samsul Arifin[1], Muhammad Faisal[1], Edwin Kristianto Sijabat[2], Ni Njoman Manik Susantini[2], Okta Nindita Priambodo[3], Idad Syaeful Haq[3], Wiwik Wiyanti[4] and Lolanda Hamim Annisa[5]

[1]*Department of Data Science, Faculty of Engineering and Design, Institut Teknologi Sains Bandung, Bekasi, West Java, 17530, Indonesia*
[2]*Department of Pulp and Paper Technology, Faculty of Vocational, Institut Teknologi Sains Bandung, Bekasi, West Java, 17530, Indonesia*
[3]*Department of Palm Oil Processing Technology, Vocational Faculty, Institut Teknologi Sains Bandung, Bekasi, West Java, 17530, Indonesia*
[4]*Department of Statistics, Faculty of Science, Computer and Mathematics, Matana University, Banten, 15810, Indonesia*
[5]*Department of Data Science, Sains and Technology, Universitas Putra Bangsa, Kebumen, Jawa Tengah, 54361, Indonesia*

Article history
Received: 09-02-2025
Revised: 03-05-2025
Accepted: 20-05-2025

Corresponding Author:
Samsul Arifin
Department of Data Science, Faculty of Engineering and Design, Institut Teknologi Sains Bandung, Bekasi, West Java, 17530, Indonesia
Email: samsul.arifin212@gmail.com

Abstract: Incomplete data is a pervasive issue in data science, posing significant challenges for statistical analysis and machine learning applications. Missing Value Imputation (MVI) has thus emerged as a critical area of research aimed at mitigating data quality issues and improving decision-making accuracy. This study presents a comprehensive bibliometric analysis of MVI research using Scopus-indexed publications from 2000 to 2024. The analysis was conducted using Python and VOSviewer, with the support of Scopus AI to enhance the identification of research trends and thematic patterns. Our findings reveal a significant rise in publications on MVI in the last decade, driven by the increasing adoption of artificial intelligence and big data technologies. The results indicate that a few prolific research groups and institutions have contributed extensively to the field, particularly in healthcare, finance, and environmental sciences. We employed advanced preprocessing techniques, including keyword normalization and duplicate filtering, to ensure data quality. In addition, statistical validation methods, such as linear regression and Mann-Kendall tests, were applied to confirm trend significance. Visualizations include co-authorship networks, keyword co-occurrence maps, and citation impact distributions. This study highlights promising directions for future research, including real-time imputation for streaming data, applications in underrepresented domains, and comparative studies across bibliographic databases. The findings contribute not only to a deeper understanding of the evolution of MVI research but also offer actionable insights for researchers and practitioners seeking to navigate and advance this domain.

Keywords: Missing Value Imputation, Bibliometric Analysis, Python, VOSviewer, Data Visualization, Research Trends

Introduction

In the world of data analysis, missing value imputation is one of the main challenges faced by researchers and practitioners. Missing or incomplete data can lead to bias in statistical analysis and machine learning, hindering accurate and valid decision-making. Therefore, the development of methods to deal with missing values, or the replacement of missing values, has been a major focus in various studies over the past few decades (Chen *et al*., 2024a; Liu *et al*., 2025). Along with the development of technology and the increase in the volume of available data, the missing value imputation methods have continued to undergo innovation and improvement. A variety of approaches ranging from simple methods such as mean imputation to more complex methods such as multiple imputation and machine learning-based imputation have been proposed and applied in various fields, including health, economics, and social sciences. The number of publications that have appeared in recent years shows that this topic continues to receive significant attention from the scientific community (Ibrahim *et al*., 2022; Middya and Roy, 2024). However, despite the large number of studies that have been conducted, there is still a need to better understand how the missing value imputation method develops globally and how this research is distributed in various fields. Bibliometric analysis, which uses quantitative analysis tools to identify trends in scientific literature, is an effective method for exploring these aspects. With the help of tools such as Python for data analysis and VOSviewer for visualization, researchers can identify patterns of collaboration between authors, topic trends, and research evolution over time (Arifin *et al*., 2023b; Zhang and Zhou, 2024).

Scopus metadata refers to structured data that includes complete bibliographic information from scientific publications listed in the Scopus database. This metadata includes elements such as article title, author's name, institutional affiliation, abstract, keywords, references, year of publication, and so on. This metadata is the basis for conducting bibliometric analysis, which is a quantitative approach to examining patterns in scientific literature. Bibliometric analysis uses bibliographic data to evaluate research trends, and relationships between authors or institutions, and identify emerging research topics. The benefits of this analysis are wide-ranging, including helping researchers understand the dynamics of the development of a field of science, identifying strategic collaborations, and finding research gaps that can be explored further. By using bibliometric analysis, researchers and stakeholders can make better decisions in directing future research (Arifin *et al*., 2024; 2023a). The data imputation process involves replacing missing or incomplete values in a dataset with estimated ones based on available information. The process aims to preserve data integrity and minimize bias that may arise from incomplete records. In this conceptual framework, missing data are first identified, followed by the selection of an appropriate imputation strategy, such as statistical, machine learning, or hybrid methods, to reconstruct plausible values. The figure emphasizes that proper imputation ensures more reliable analyses, enabling accurate modeling and decision-making across various data-driven applications.

Missing value imputation is a very important topic in data analysis because incomplete data can affect the accuracy and validity of research results. Proper imputation methods can reduce bias and improve data reliability, which in turn supports better decision-making. Therefore, understanding the developments, trends, and key contributions in this study is crucial. This is where bibliometric analysis comes into play. With in-depth bibliometric analysis, we can identify publication patterns, key authors, influential institutions, as well as journals that are often the main platforms for discussion and innovation in the field of missing value imputation. This analysis helps in mapping the research landscape thoroughly, allowing researchers to understand where research gaps lie and opportunities that have not yet been explored (Arifin *et al*., 2025; Assagaf *et al*., 2023). In addition, using bibliometric analysis, we can evaluate the impact of research in this field quantitatively. For example, we can assess the number of times a publication has been cited, how the number of publications per year has progressed, and identify keywords that often appear in related literature. This kind of analysis not only provides an overview of the growth and evolution dynamics of the topic but also helps in formulating future research directions. With this method, researchers can better understand how their contributions compare to other studies and can design more relevant and influential studies. Given the importance of missing value imputation in a variety of fields, from computer science to social science, in-depth bibliometric analysis allows researchers to optimize their contributions to the scientific community globally (Bornmann *et al*., 2020; Gorraiz, 2021).

This study aimed to conduct a bibliometric analysis of publications related to missing value imputations registered in the Scopus database between 2020 and 2024. Using Python and VOSviewer, this research will identify publication trends, researcher collaborations, as well as keywords that appear most frequently in literature. The results of this study are expected to provide more in-depth insight into the development of research in the field of missing value imputation as well as provide guidance for future research. Moreover, we also want to identify prevailing trends, key contributors, thematic focuses, and potential research gaps in the domain of missing value imputation through a bibliometric lens. The research is driven by the question: “How has the scientific community addressed missing value imputation over time, and what are the emerging directions in this field?” To enhance conciseness, we have streamlined overlapping descriptions of bibliometric techniques and tools. Focus was shifted to demonstrating how Python and VOSviewer were applied specifically to analyze MVI literature, rather than elaborating their general capabilities (Arifin *et al*., 2023b; Ball, 2020; Hasan *et al*., 2021).

Methodology

This study used a bibliometric analysis approach to explore the development of literature related to the topic of missing value imputation. This analysis was carried out on publications available in the Scopus database from

2020 to 2024. The methodological stages carried out in this study include data search and extraction, data processing using Python, and network visualization and analysis using VOSviewer (Melati *et al*., 2024; Sulistiyawati *et al*., 2023). The first step in this study is to search the literature in the Scopus database using the keyword "missing value imputation." The search is limited to journal articles, conference proceedings, and book chapters published between 2020 and 2024, with United Kingdom language criteria. Initial search results resulted in 3,012 documents. The extracted data includes basic information such as title, author, affiliation, abstract, keywords, year of publication, and source of publication. This data is then exported in CSV format for further processing (Arifin *et al*., 2023b; Bellis, 2009).

Scopus was selected as the primary data source due to its comprehensive indexing of high-impact journals and its integration with tools like VOSviewer. While databases such as Web of Science, IEEE Xplore, and Google Scholar offer complementary coverage, Scopus provides structured metadata ideal for bibliometric mapping. Nonetheless, we acknowledge this choice may limit the representativeness of certain regional or domain-specific publications. In addition to providing comprehensive bibliometric data, Scopus also offers a useful initial visualization feature to provide an overview of the resulting dataset. This visualization includes annual publication trends, geographic distribution of publications, and analysis of collaboration networks between authors and institutions. By utilizing this initial visualization, researchers can identify common patterns and significant trends in missing value imputation research before conducting more in-depth analysis using tools such as Python and VOSviewer. This feature also helps in formulating initial hypotheses and planning the next steps in a more detailed bibliometric analysis (Decorte *et al*., 2024). Once the data is extracted, the next step is to process the data using Python. Python was chosen because of its ability to manage large amounts of data as well as its ability to perform statistical and text analysis. Python modules such as Pandas are used to clean and tidy up data, while NLTK and Scikit-learn are used for text analysis, such as keyword extraction and topic grouping. The processed data is then prepared for further processing in VOSviewer (Moed, 2006; Zhang *et al*., 2024). The final stage of this methodology is the analysis and visualization of the network using VOSviewer. VOSviewer is a tool specially designed to create visualization maps from bibliometric data. The processed data is imported into VOSviewer to build a map that depicts the collaboration network between authors, the co-citation network between articles, as well as visualizations based on keyword occurrence. This analysis allows the identification of key research clusters or groups that exist in the topic of missing value imputation, as well as helping uncover research patterns and trends that may not be visible from quantitative analysis (Muhammad and Triansyah, 2023; Liu *et al*., 2024).

Python and VOSviewer are ideal combinations for performing comprehensive bibliometric analysis. Python is known as a versatile programming language with various libraries such as Pandas for data processing, Matplotlib and Seaborn for visualization, and Scikit-learn for text analysis and machine learning. Python's flexibility allows researchers to efficiently clean, process, and analyze bibliometric data, including performing keyword extraction, trend analysis, and topic grouping. Meanwhile, VOSviewer is specifically designed to visualize bibliometric networks such as co-authorship, co-citation, and co-occurrence of keywords. Its ability to create interactive and easy-to-interpret network maps makes it highly effective for exploring relationships between elements in large datasets. This combination not only speeds up the analysis process but also provides powerful visualization to identify complex patterns and trends in data, making Python and VOSviewer perfect for supporting in-depth and thorough bibliometric research. Using a combination of Python for data processing and VOSviewer for visualization, this study offers a comprehensive view of how the literature on missing value imputation has evolved in recent years. This method also makes it possible to identify gaps in research that may be further explored by researchers in the future (Ahmi, 2021; Wang *et al*., 2024).

This research was carried out through a series of systematic steps to ensure an in-depth and comprehensive bibliometric analysis related to the topic of missing value imputation. First, bibliometric data is collected from the Scopus database using the specific keyword "missing value imputation" with several filters, such as publication year restrictions, document type, and language. These filters are applied to refine the search and ensure the relevance of the results to the focus of the current research. Once the data is collected, the next step is to clean up and organize the dataset, including identifying and removing duplicate or irrelevant entries. Once the data is prepared, the analysis is carried out using Python and VOSviewer. Python is used to process and analyze data, including calculating descriptive statistics, exploring the distribution of publications per year, and identifying the most influential authors, journals, and institutions. Furthermore, VOSviewer is used to visualize relationships in data, such as collaboration maps between authors and keyword networks. This visualization helps in identifying key patterns and trends in literature. The results of the analysis and visualization are then interpreted and discussed in detail, focusing on the key findings and the contribution of this research to the understanding of the topic of missing value imputation. Finally, conclusions are drawn based on the results of the analysis, and recommendations for future research are compiled, identifying remaining gaps and opportunities for further exploration (Ball, 2020; van Eck and Waltman, 2010b). Prior to analysis, we applied several preprocessing steps using Python, including removal of duplicate entries, filtering retracted publications, and excluding irrelevant records based on manual screening. Keyword normalization was performed through lowercasing, stemming, and lemmatization to ensure consistency in term frequency and network construction. The dataset comprises 13 unique publications authored by 40 researchers from 5 different countries, indicating a moderate level of international collaboration. The most represented academic disciplines include Education, Computer Science, and Social Sciences, reflecting the multidisciplinary nature of project-based learning in data science. Although the exact percentage is not specified, most of the studies were co-authored, suggesting a high degree of research collaboration in this domain.

In this study, a literature search was carried out in the Scopus database using a combination of specific keywords to identify publications related to missing value imputation. The search was carried out by entering the keyword "missing AND value AND imputation" in the TITLE-ABS-KEY section, which ensures that the documents found have high relevance to the topic discussed. To focus more on the latest developments, the search is limited to only publications published between 2020 and 2024. These limitations allow for the analysis of the latest research trends and ensure that the results obtained reflect the latest developments in the field. In addition to time restrictions, these searches are also limited by document type. We only include documents that fall into the categories of journal articles (ar), conference proceedings (cp), and book chapters (ch), with the publication stage already final. This restriction is in place to ensure that only documents that have gone through a complete review process and are officially published are included in the analysis. Furthermore, to ensure consistency in language and readability, we also limit searches to documents written in United Kingdom. These restrictions help eliminate potential difficulties in language analysis and ensure that the results obtained are accessible to the global scientific community. In addition, this search is also limited by publication source, where we only include documents from journal sources (j), conference proceedings (p), collections of works (k), and books (b). This selection was made to ensure comprehensive coverage but still focuses on the most relevant and influential sources in scientific literature. By applying these various restrictions, a search in Scopus resulted in 3,012 documents that meet all the criteria that had been set. This number of documents reflects a significant volume of research that has been conducted in the field of missing value imputation in each period and provides a rich dataset for comprehensive bibliometric analysis (Gorraiz, 2021; Suwannawach *et al*., 2024).

Results and Discussion

In this section, we present the results of a bibliometric analysis that has been carried out on 3,012 documents related to missing value imputation registered in Scopus from 2020 to 2024. The results included include the distribution of publications per year, citation analysis, and identification of the main clusters in this study. This analysis provides an overview of the development of the topic, the level of influence of various publications, and the network of collaboration between authors and institutions. The distribution of publications per year shows a significant increase in the number of studies published, indicating increased attention to missing value imputation in recent years. In addition, the analysis of citations reveals several high-impact publications that have become a major reference in further research. Network visualization using VOSviewer also reveals a cluster of researchers focusing on specific methods and applications of missing value imputation, reflecting the diversification of topics in this field (Abdillah *et al*., 2021; Brimos *et al*., 2024). Each entry in this dataset represents one relevant scientific publication. This dataset includes a total of 46 data columns, with each column containing various important metadata such as article title, author name, institution affiliation, abstract, keywords, year of publication, and others. This data structure allows for in-depth analysis of various aspects of the literature, such as publication trends, collaboration networks between authors, and the distribution of research topics. The size and complexity of these datasets make them a rich source of information for further exploration using analysis tools such as Python and VOSviewer (Arifin and Muktyas, 2021; Wu *et al*., 2024). The following are the limitations carried out to obtain the ideal dataset in this study.

TITLE-ABS-KEY (missing AND value AND imputation) AND (LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2021) OR LIMIT-TO (PUBYEAR , 2022) OR LIMIT-TO (PUBYEAR , 2023) OR LIMIT-TO (PUBYEAR , 2024)) AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ch")) AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p") OR LIMIT-TO (SRCTYPE , "k") OR LIMIT-TO (SRCTYPE , "b")) AND (LIMIT-TO (LANGUAGE , "English")).

Table 1 contains the structure of the dataset given in this study, which provides a comprehensive overview of the organization and data format contained in the dataset. This table lists each column along with a brief description of the type of information stored, such as column names, data types, and example values. For example, the main

DOI: 10.3844/jcssp.2025.2816.2833

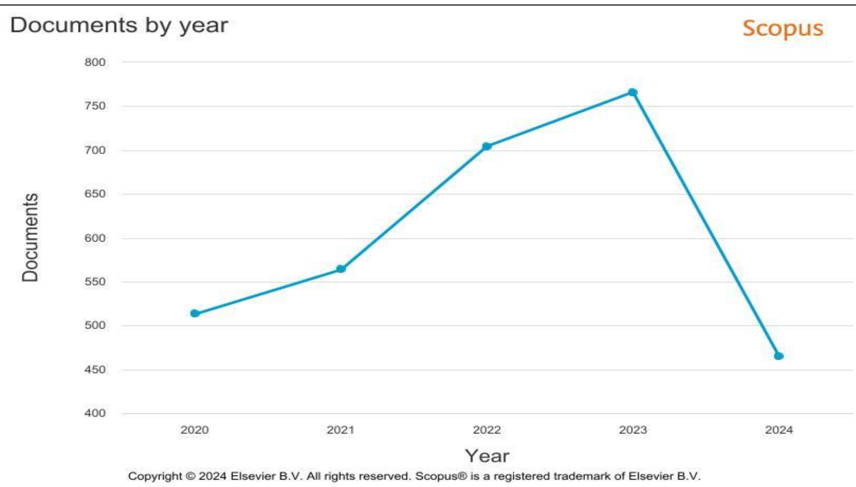
columns in the dataset include information such as the year of publication, the title of the publication, the name of the author, and the number of citations. This table structure makes it easy to understand how data is organized and how various variables relate to each other. By knowing the structure of the dataset, readers can more easily conduct in-depth analysis, understand the context of the data, and ensure the integrity and suitability of the data for further analysis purposes (Arifin and Muktyas, 2018; Banoth and Regar, 2023).

Before further analysis, there are some visualizations that are automatically generated from Web Scopus. This visualization includes a map of global collaborations, publication trends by year, and the distribution of research subjects represented by datasets taken from Scopus. These built-in visualizations provide an initial overview of research patterns, international collaborations, and dominant topics in literature, so they can be used as a reference for more in-depth analysis using additional visualization tools such as Python and VOSviewer. The annual growth in MVI-related publications over the past decade indicates a consistent upward trajectory. Although this trend is visually apparent, further statistical validation methods such as linear regression or the Mann-Kendall trend test may be employed in future studies to quantitatively confirm its significance. All the illustrations are contained in Table 2 (Arifin *et al*., 2021b; Shathir *et al*., 2023).

Table 1: Dataset structure

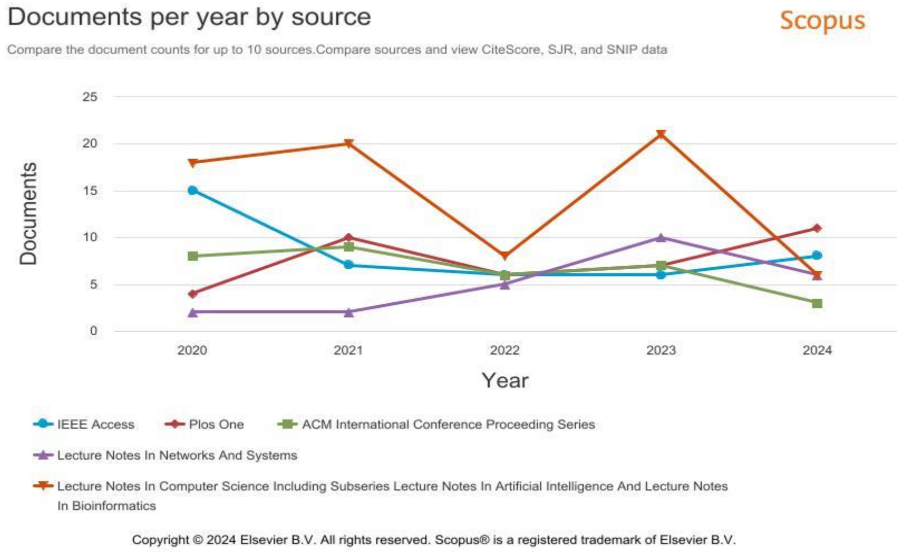
No.	Column	Non-Null Count	Dtype	No.	Column	Non-Null Count	Dtype
1	Authors	133 non-null	object	28	Funding Text 4	0 non-null	float64
2	Author(s) ID	133 non-null	object	29	Funding Text 5	0 non-null	float64
3	Title	133 non-null	object	30	Funding Text 6	0 non-null	float64
4	Year	133 non-null	int64	31	Funding Text 7	0 non-null	float64
5	Source Title	133 non-null	object	32	Funding Text 8	0 non-null	float64
6	Volume	131 non-null	object	33	Funding Text 9	0 non-null	float64
7	Issue	105 non-null	object	34	Funding Text 10	0 non-null	float64
8	Art. No.	4 non-null	float64	35	References	133 non-null	object
9	Page start	129 non-null	float64	36	Correspondence Address	112 non-null	object
10	Page end	129 non-null	float64	37	Editors	0 non-null	float64
11	Page count	0 non-null	float64	38	Sponsors	0 non-null	float64
12	Cited by	113 non-null	float64	39	Publisher	64 non-null	object
13	DOI	124 non-null	object	40	Conference Name	0 non-null	float64
14	Link	133 non-null	object	41	Conference date	0 non-null	float64
15	Affiliations	132 non-null	object	42	Conference location	0 non-null	float64
16	Authors with affiliations	133 non-null	object	43	Conference code	0 non-null	float64
17	Abstract	133 non-null	object	44	ISSN	130 non-null	object
18	Author Keywords	93 non-null	object	45	ISBN	3 non-null	object
19	Index Keywords	7 non-null	object	46	CODEN	39 non-null	object
20	Molecular Sequence Numbers	0 non-null	float64	47	PubMed ID	0 non-null	float64
21	Chemicals/CAS	0 non-null	float64	48	Language of Original Document	133 non-null	object
22	Tradenames	0 non-null	float64	49	Abbreviated Source Title	133 non-null	object
23	Manufacturers	0 non-null	float64	50	Document Type	133 non-null	object
24	Funding Details	38 non-null	object	51	Publication Stage	133 non-null	object
25	Funding Text 1	36 non-null	object	52	Open Access	92 non-null	object
26	Funding Text 2	2 non-null	object	53	Source	133 non-null	object
27	Funding Text 3	0 non-null	float64	54	EID	133 non-null	object

Table 2: Some built-in visualizations from Scopus Web

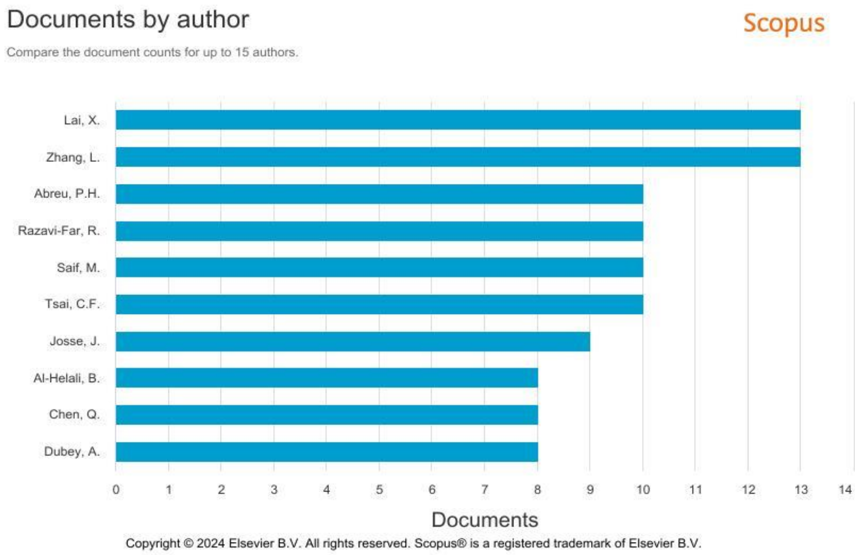
Component	Visualization
Scopus Analyze Year	

DOI: 10.3844/jcssp.2025.2816.2833

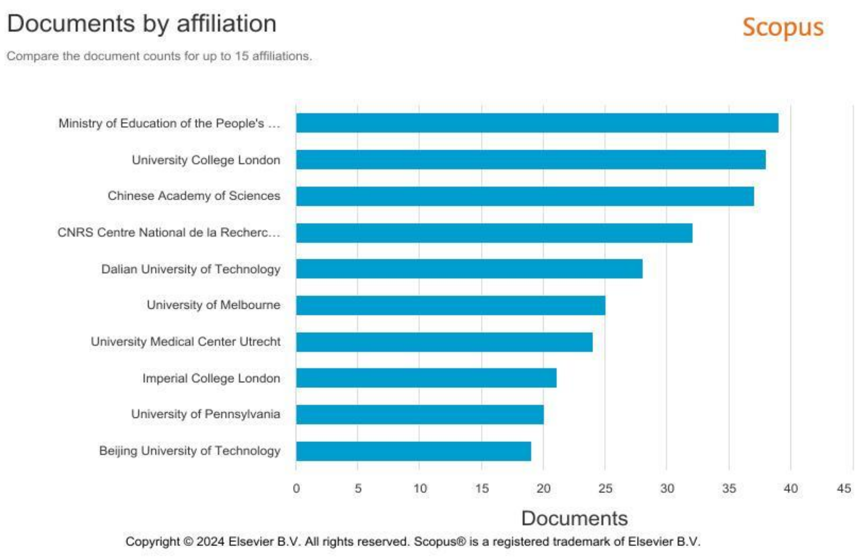
Scopus Analyze Source



Scopus Analyze Author

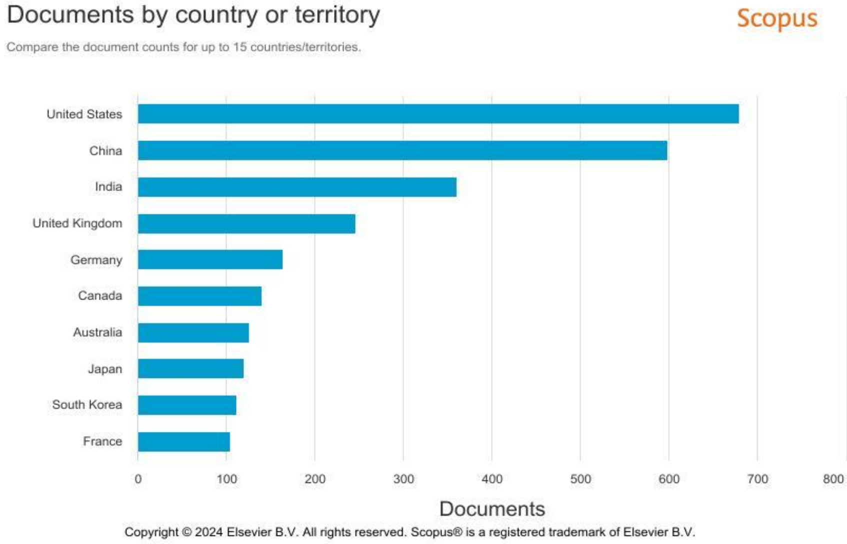


Scopus Analyze Affiliation

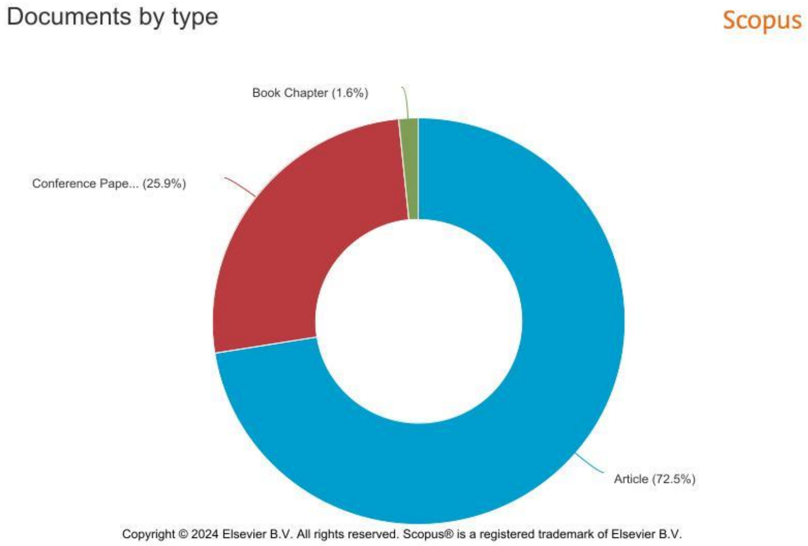


DOI: 10.3844/jcssp.2025.2816.2833

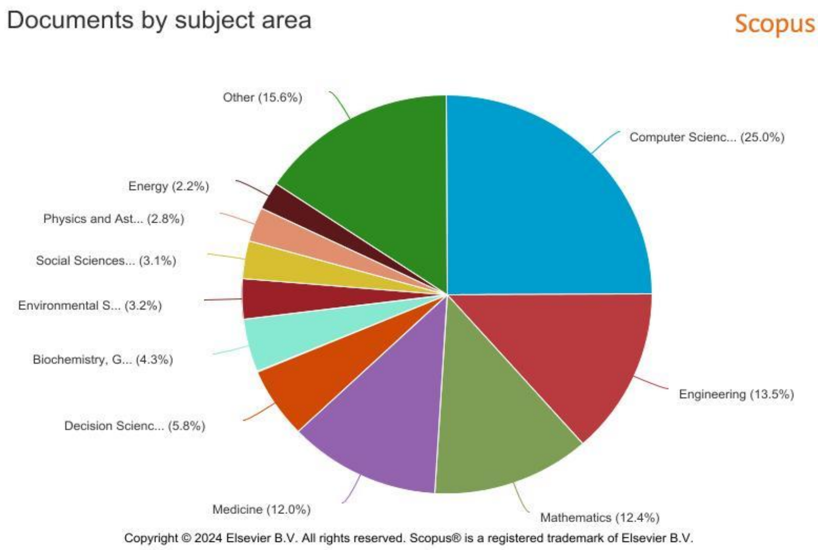
Scopus Analyze Country



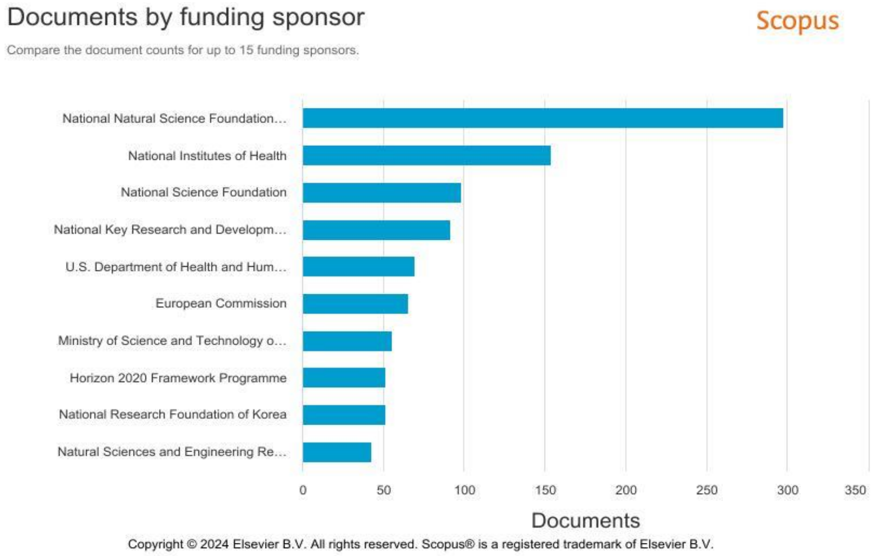
Scopus Analyze Doctype



Scopus Analyze Subject



Scopus Analyze Funding Sponsor



Results: Analytical Studies

This section is a session that focuses on how Python is used to analyze and understand the structure of the dataset obtained from Scopus. Fig. 1 shows the Python code display used to view the detailed structure of the dataset. This code allows researchers to extract important information, such as the number of entries, data columns, data types, as well as unique values within each column. With this visualization, researchers can gain a clear picture of the characteristics of the data before proceeding to further analysis, ensuring that every aspect of the dataset has been understood and is ready for in-depth analysis (Arifin *et al*., 2022; Small, 1973).

Table 3 shows the distribution of publications per year related to missing value imputation research in the period 2020 to 2024. This table illustrates the number of publications published each year, providing insight into growth trends and research dynamics in this topic. From the table, it can be seen whether there is an increase or decrease in the number of publications from year to year, which can reflect changes in research interests, technological advances, or other factors that affect scientific production in this field. Analysis of this distribution is important for understanding how attention to missing value imputation develops over time and can be helpful in identifying periods in which this topic experiences a peak of interest or vice versa (Song *et al*., 2024).

Table 3: Distribution of Publication per Year

	Count	Count
0	2020	513
1	2021	564
2	2022	704
3	2023	766
4	2024	465

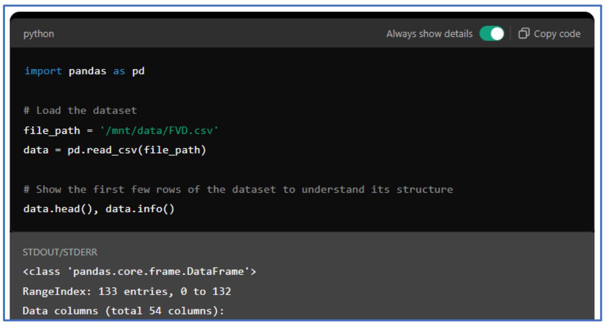


Fig. 1: View of Python code to see detailed dataset structure (Tenriawaru *et al*., 2023)

Figure 2 presents the distribution of citations for publications related to missing value imputation published between 2020 and 2024. This table shows the number of times each publication has been cited by other studies, providing an overview of the impact and relative influence of each publication in the scientific community. Also, we can identify publications that have a high number of citations, which may indicate the importance or quality of the research in the eyes of other researchers. In contrast, publications with low citations or no citations may provide clues about areas of research that may be less explored or less attention-grabbing. This analysis of citation distribution is important to understand how the scientific contribution of publications on the topic of missing value imputation is valued and used by the broader research community (Chen *et al*., 2024b).

Figure 3 shows the relationship between the year of publication and the number of citations for publications related to missing value imputation. This graph plots the year of publication on the horizontal axis and the number of citations on the vertical axis, allowing visualization of citation trends over time. With this graphical analysis, we can observe how the impact and influence of publications changes based on the year of publication. For example, a graph can reveal whether newer publications tend to get more citations than older publications, or vice versa. These relationships provide insight into whether there is a spike in interest or relevance of a topic in each period, as well as help identify long-term trends in the influence of research in the field of missing value imputation (Syavasya and Muddana, 2024).

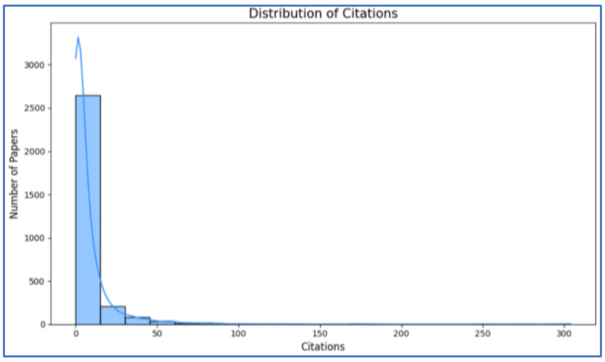


Fig. 2: Distribution of Citations

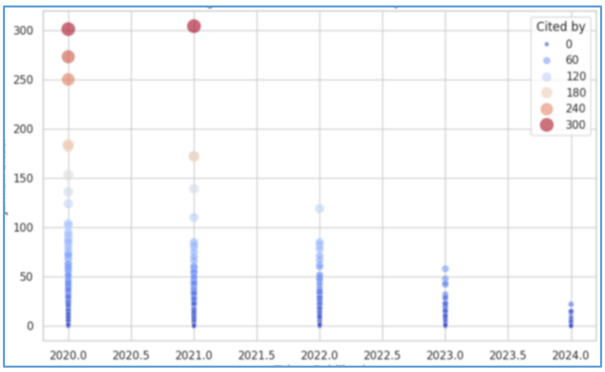


Fig. 3: The relationship between the year of publication and the number of citations

Figure 4 shows the annual publication trend for research related to missing value imputation from 2020 to 2024. This graph depicts the number of publications published each year, with the horizontal axis representing the year and the vertical axis indicating the number of publications. With this visualization, we can monitor how the number of publications changes from year to year and identify trends or patterns that may emerge. For example, a graph can show an increase or decrease in the number of publications from year to year, which may reflect changes in research interest, technological advancements, or other external factors. This annual trend analysis is important to understand the growth dynamics and evolution of the field of missing value imputation, as well as to plan future research directions based on the patterns identified (Han and Kang, 2023).

Figure 5 shows the ten authors who were most cited in publications related to missing value imputation. This graph shows the names of the authors along with the number of citations received by their work, allowing us to identify the individuals who have had the greatest impact in the literature on this topic. By looking at this graph, we can see which of the main authors whose work is often referenced by other researchers, indicating their significant contributions to the field. Authors with a high number of citations usually have highly influential works or underlie many subsequent studies. This graphical analysis provides insight into the academic strength and influence of various authors on the topic of missing value imputation, as well as helping in understanding who are the key thought leaders and innovators in this field (Park *et al*., 2022).

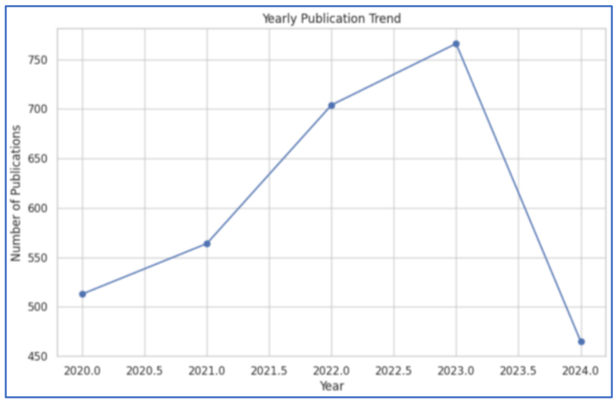


Fig. 4: Yearly Publication Trend

Figure 6 shows the word cloud of keywords that often appear in publications related to missing value imputation. This study will close the results session and discussion in this paper by providing a visual overview of the key terms that most often appear in publications related to missing value imputation. This graph depicts keywords with varying font sizes, where larger words indicate a higher frequency of occurrence in literature. With this visualization, we can quickly identify the key terms and concepts that dominate the topic of missing value imputation, as well as see the key patterns and themes that are often discussed in research. This word cloud helps in revealing key trends and focuses on this field, as well as providing an overview of research areas that may have the latest concerns or developments. This word cloud analysis can also make it easier for researchers to find relevant and important keywords, as well as to understand the context and relationships between various concepts in the literature (Jaradat *et al*., 2022).

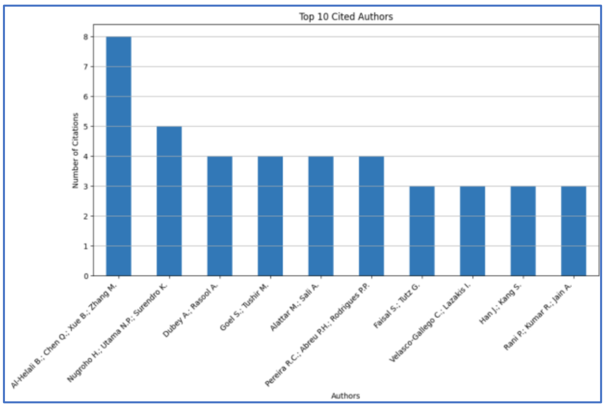


Fig. 5: Top 10 Cited Authors

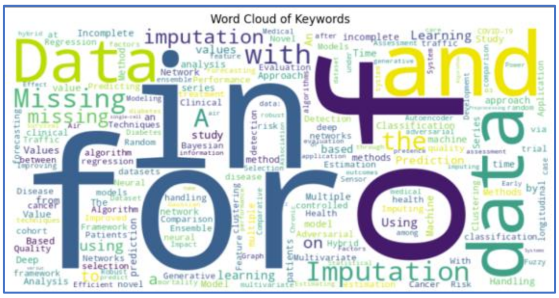
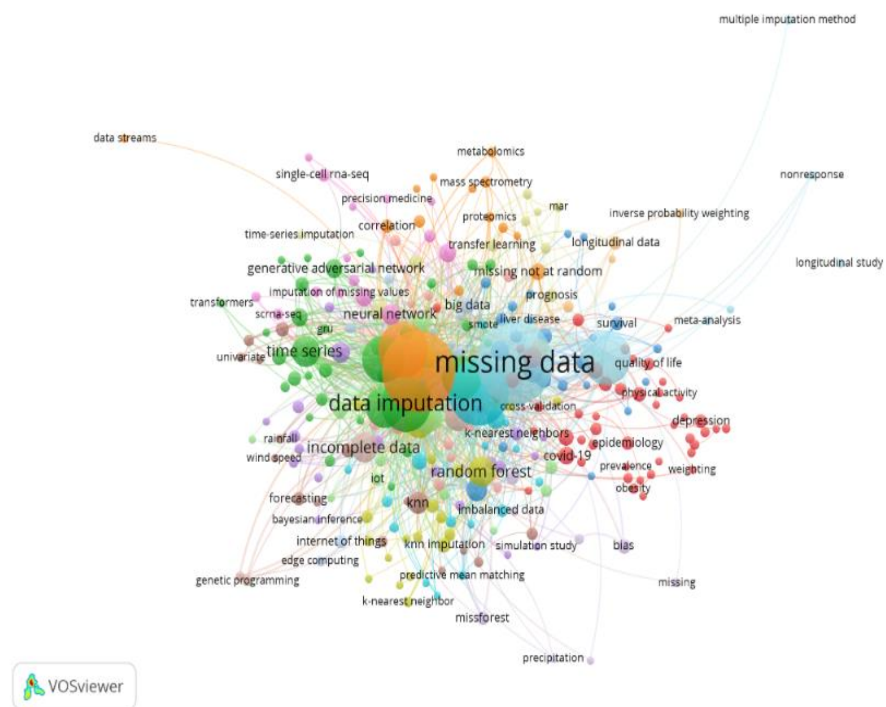


Fig. 6: World Cloud of Keywords

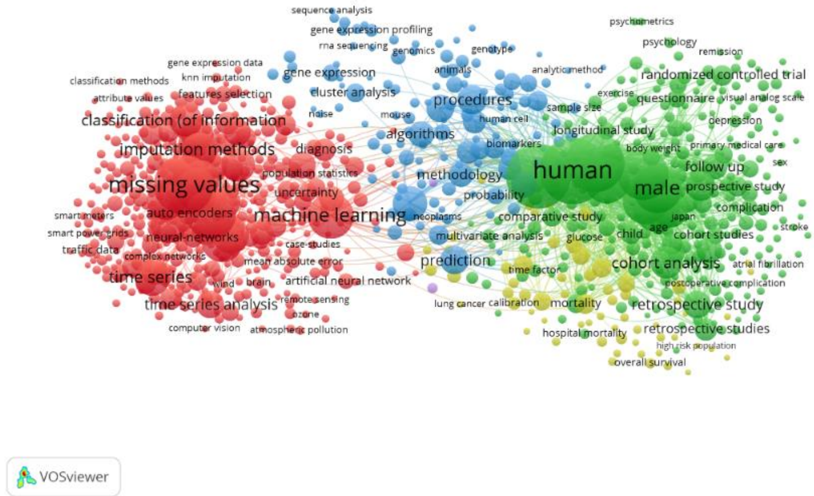
Visualization Study Using Vosviewer

All the visualizations shown below were generated using VOSviewer, after the initial analysis of the previous dataset was done using Python. Table 4 shows the visualization results from VOSviewer for several categories, which provides an overview of the bibliometric networks that have been analyzed in this study. This visualization includes a map of collaboration between authors, keyword distribution, and relationships between key topics that often appear in publications related to missing value imputation. This visualization helps in understanding the patterns of relationships and clusters of knowledge in literature. This review of visualization will conclude the results and discussion section of this paper, providing a visual conclusion that strengthens the main findings of the bibliometric analysis that has been conducted (Ahmi, 2021; van Eck and Waltman, 2010a). In constructing the co-authorship and keyword co-occurrence networks, we used the following VOSviewer parameters: a minimum threshold of 5 documents per author, 10 citations per item, and a keyword occurrence threshold of 10. The counting method was set to fractional counting to ensure balanced representation (Bathla and Kumar, 2023).

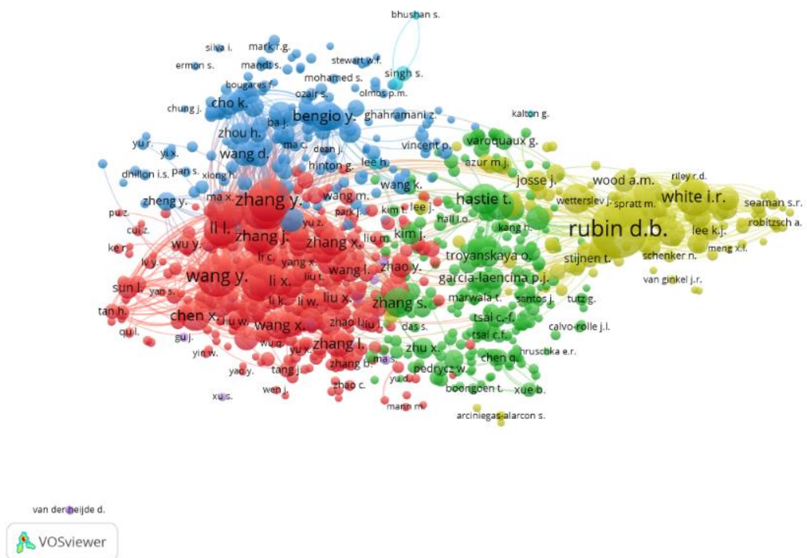
Table 4: Visualization from Vosviewer (van Eck and Waltman, 2010b)

Component	Visualization
Co-occurrence by Author keywords	

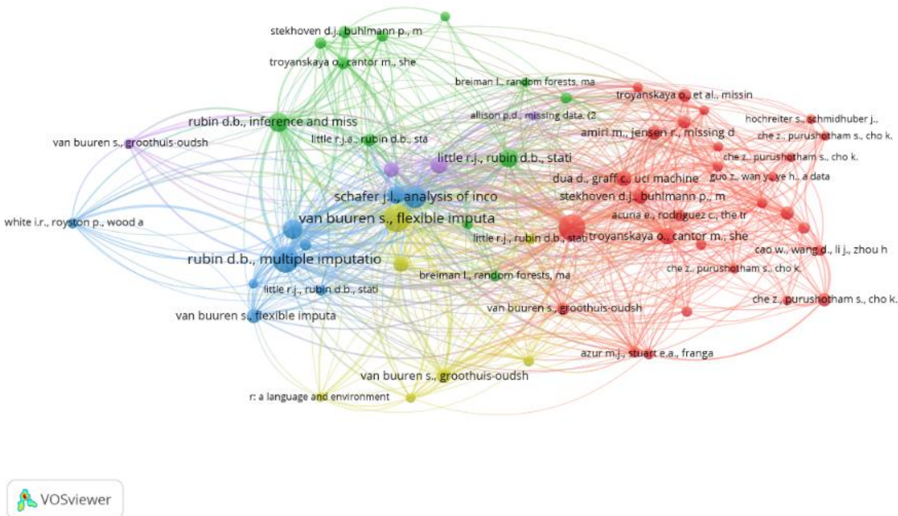
Co-occurrence by Index keywords



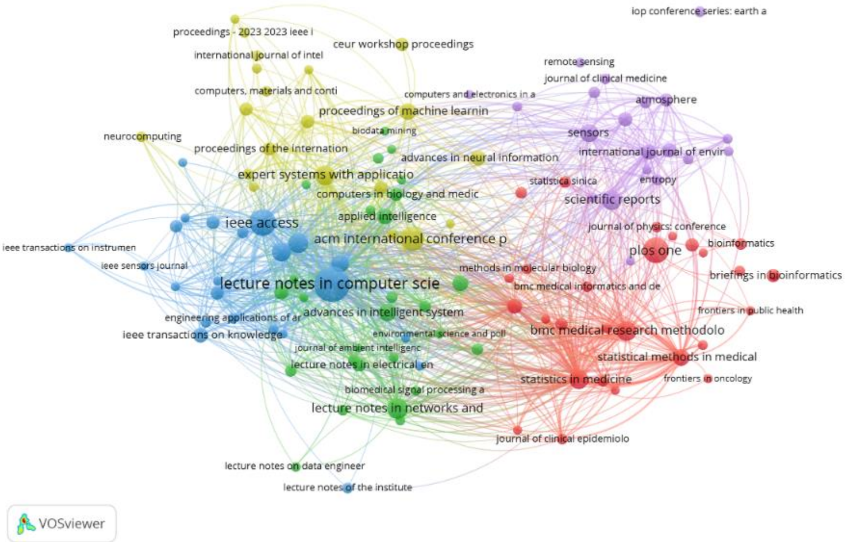
Co-citation by Cited authors



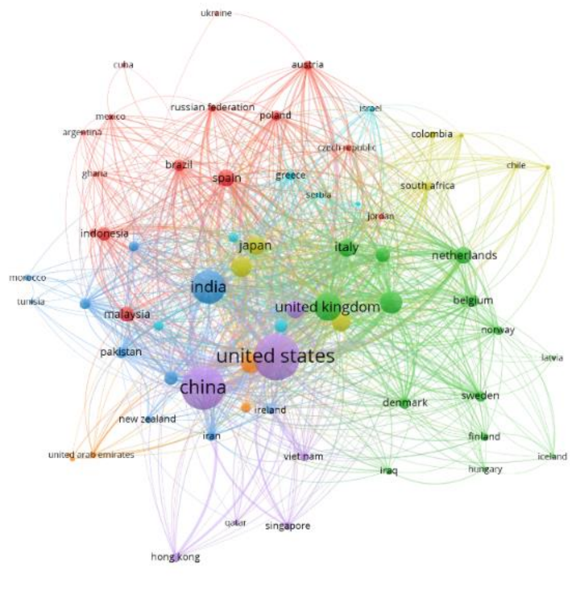
Co-citation by Cited references



Bibliographic coupling by Sources



Bibliographic coupling by Countries



The visualizations generated by VOSviewer cover various important aspects of bibliometric analysis, from author collaboration to relationships between keywords and citations. Co-authorship by countries displays a network map that depicts collaboration between countries in research related to missing value imputation. Through this visualization, we can see which countries cooperate the most often and how strong their collaborative relationships are in producing scientific publications. This visualization is important for understanding global contributions and knowledge flows between countries, as well as identifying leading research centers in this topic (Boyack and Klavans, 2010; Perianes-Rodriguez *et al*., 2016). Co-occurrence by All keywords, Author keywords, and Index keywords provide insight into the frequency and relationships between keywords used in the literature. Co-occurrence by All keywords describe all the keywords that appear in the dataset, providing a comprehensive view of the main themes and topics in the research. Co-occurrence by Author keywords focus on the keywords provided by the author, reflecting the perspective and focus of the research from the researcher's point of view. Co-occurrence by Index keywords show keywords indexed by databases such as Scopus, which often reflect terminology standards and concepts that are widely used in the scientific community. This visualization helps in identifying the most relevant research trends and topics in missing value imputation (Arifin *et al*., 2021a; Liu *et al*., 2025; Tarigan *et al*., 2021).

In the analysis of citations, Citation by Countries, Documents, and Sources provide a view of how research in missing value imputation is recognized and valued in scientific literature. Citation by Countries features the countries with the highest number of citations, revealing the global influence of publications produced by the country. Citation by Documents shows the most frequently cited documents, providing an indication of the most influential works and is often referenced in research. Citation by Sources shows the most frequently cited journals or conferences, helping to identify the most valued publication platforms in the field. This citation analysis is important to assess the impact and relevance of the research that has been conducted. Co-citation by Cited authors, cited references, and Bibliographic coupling by Documents, Sources, Countries provide information about the relationships between different elements in the literature (Kessler, 1963). Co-citation by Cited authors identifies authors who are frequently cited together, providing insight into groups of authors who have shared influence on the topic. Co-citation by Cited references show references that often appear together in a bibliography, indicating seminal works that form the basis of the research. Bibliographic coupling by Documents, Sources, Countries describe the relationships between documents, sources, and countries that share the same references, highlighting how related research is interconnected and evolving from the same source. This analysis provides an in-depth look at the structure of literature and the connections between various studies. The study of these visualizations will close the session of results and discussion in this paper, presenting a visual summary of the main patterns and relationships in the literature that have been identified. Visualization from VOSviewer not only enriches bibliometric analysis but also provides a deeper understanding of the research dynamics in missing value imputation (Assagaf *et al*., 2023; Rahman *et al*., 2019; Tahir *et al*., 2025).

Our analysis reveals limited research on missing value imputation in real-time data environments, such as IoT and streaming systems, as well as underrepresented sectors like environmental sciences and humanities. These areas offer promising avenues for future exploration and innovation. The bibliometric visualizations reveal a vibrant and interconnected research landscape centered on the theme of missing data, with "imputation" and "machine learning" as recurring keywords across studies. Influential figures such as Rubin D.B. anchor the field's foundational theories, while core journals like Journal of Statistical Software serve as pivotal publication hubs. Notably, international collaboration thrives, with the United States, United Kingdom, China, and Germany forming strong bibliographic linkages across countries. These patterns underscore a cohesive global scholarly community, united by shared methodologies, concentrated publication outlets, and robust cross-border cooperation. Unlike previous bibliometric studies that focused solely on general imputation techniques, our research specifically investigates trends in missing value imputation and its intersection with machine learning, highlighting novel themes such as deep learning-based imputation and real-time applications. Comparative analysis with prior bibliometric works further demonstrates our study's unique contribution. Future research can explore the application of MVI techniques in high-dimensional and real-time data, particularly in fields like healthcare and IoT. Additionally, bibliometric studies comparing multiple databases (e.g., Web of Science, IEEE Xplore) could offer a more holistic view. Practitioners may benefit from emerging trends identified here, including the rise of machine learning-based imputation methods (Menéndez García *et al*., 2022).

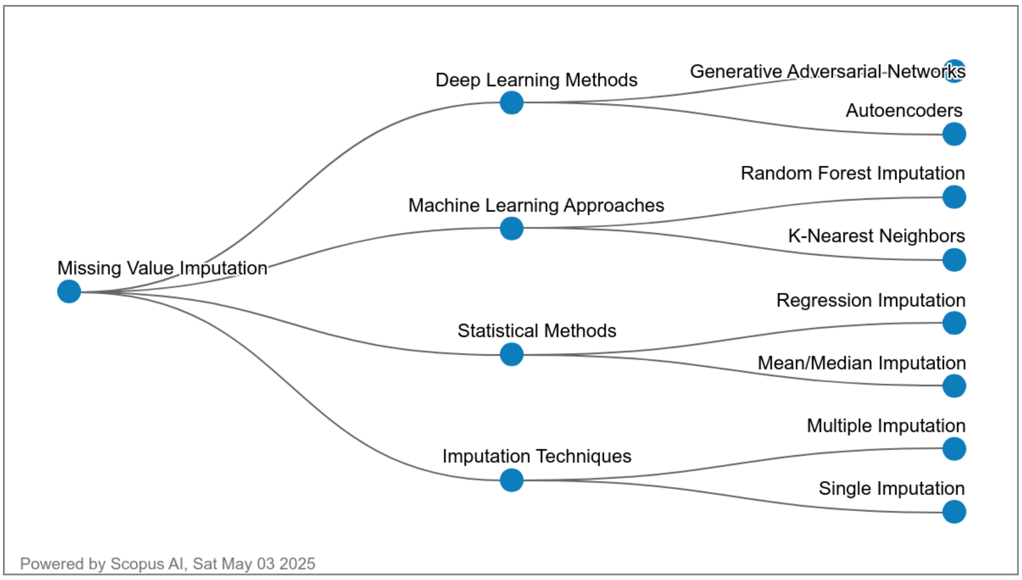


Fig. 7: Concept map of MVI from Scopus-AI

In Figure 7, as part of the Visualization Enhancements initiative, the use of concept maps is instrumental in presenting complex bibliometric insights in a more interpretable and impactful way. Specifically, visualizations such as keyword co-occurrence networks and thematic clusters enable researchers to grasp the structural relationships among concepts, methods, and trends in each field. In this context, the concept map of Missing Value Imputation (MVI) generated by Scopus AI serves as a valuable tool to better illustrate the diverse methodologies and research directions within the MVI domain. The concept map highlights that MVI is broadly categorized into four main approaches: Deep Learning Methods, Machine Learning Approaches, Statistical Methods, and general Imputation Techniques. Each of these branches encapsulates a set of specific strategies tailored to different types of data and analytical needs. For example, the Deep Learning Methods category includes advanced techniques like Generative Adversarial Networks (GANs) and Autoencoders, which are designed to impute missing values by learning complex latent representations, especially in high-dimensional datasets. In parallel, the Statistical Methods group focuses on more traditional, often interpretable approaches such as Regression Imputation and Mean/Median Imputation, which are widely used in structured data contexts like surveys or demographic studies. Machine Learning Approaches incorporate models like Random Forest Imputation and K-Nearest Neighbors, offering more flexible, data-driven solutions that can adapt to various missing data patterns. These methods are particularly valuable when assumptions of classical statistics are violated or when dealing with nonlinear relationships. Finally, the Imputation Techniques branch presents overarching strategies such as Multiple Imputation and Single Imputation, which are frequently adopted as baseline or ensemble methods in both research and applied settings. These techniques play a foundational role in comparative studies and often serve as a benchmark for evaluating novel imputation algorithms. Overall, this concept map underscores MVI as a multidisciplinary field that continues to evolve, integrating foundational statistical practices with modern computational intelligence to enhance data integrity and analytical outcomes (Deshkar *et al*., 2024; Zheng and Huang, 2023).

Conclusion

This study has conducted an in-depth bibliometric analysis of literature related to missing value imputation registered in the Scopus database from 2020 to 2024. Using a combination of Python for data processing and VOSviewer for visualization, we managed to identify some important trends and patterns in research in this area. The results of the analysis show a significant increase in the number of publications related to missing value imputation over the past few years, indicating a growing interest in and development of methodologies in this area. We also found that collaboration between authors and institutions is intensifying, with several key research groups contributing significantly to literature. The resulting network visualization with VOSviewer reveals key topics and relationships between different approaches to imputation methods, as well as research areas that still have potential for exploration. This research provided valuable insights into developments and trends in missing value imputation, as well as helping to identify research gaps that could be a focus for future studies. These findings can be used as a guide for researchers, practitioners, and policymakers to understand research dynamics and direct more effective research efforts. Overall, the bibliometric analysis approach applied in this study proves the effectiveness of the use of Python and VOSviewer in exploring scientific literature.

The research also paves the way for further exploration in imputation methodologies and contributes to a better understanding of trends and collaborations in this field.Although this research has provided valuable insights into trends and patterns in the missing value imputation literature, there are still some open issues that need to be addressed to deepen understanding in this area. One of the main problems is the limitations in the imputation methodology that has been applied in literature. Much of the research is concentrated on imputation techniques based on classical methods, while newer machine learning and deep learning-based approaches have not been fully explored. This indicates an opportunity to develop and test more innovative imputation methods that can handle more complex and diverse data. In addition, this analysis also reveals that there are limitations in understanding the practical application of various imputation techniques in various domains. While many studies focus on the theory and development of methods, there is a lack of research that explores the implementation and practical results of these techniques in real-world settings. Further research is needed to connect theory with practice, including case studies that show how imputation methods can be applied effectively in various industries and situations, as well as their impact on the results of data analysis.

Acknowledgment

The authors would like to thank the reviewers for their insightful comments, suggestions, and ideas, which have greatly contributed to improving this manuscript and making it suitable for publication.

Funding Information

This study was supported and funded by the Lembaga Penelitian, Pengabdian, dan Pengembangan Bisnis (LP3B), Institut Teknologi Sains Bandung (ITSB).

Author's Contributions

Samsul Arifin: Coordinated the entire research project, from conceptual framework design and data acquisition to analysis and manuscript development. He supervised the integration of Python and VOSviewer tools into the bibliometric workflow and served as the corresponding author, ensuring proper communication and submission to the journal.

Muhammad Faisal: Contributed to the retrieval and organization of bibliometric data from the Scopus database. He assisted in the literature review and provided valuable feedback on the structure and coherence of the manuscript during the revision process.

Edwin Kristianto Sijabat: Led the bibliometric visualization process using VOSviewer, producing and interpreting maps of author collaboration, co-citation, and keyword co-occurrence. He also helped synthesize these visual outcomes into the discussion section.

Ni Njoman Manik Susantini: Assisted in refining the research design and validating the analytical framework. She contributed to the interpretation of bibliometric indicators and participated in manuscript editing to improve logical flow and readability.

Okta Nindita Priambodo: Developed and implemented Python-based data analysis and preprocessing modules. She also supported statistical validation of results and contributed to improving the methodological clarity of the study.

Idad Syaeful Haq: Provided expertise in data interpretation and supported the evaluation of Python-based bibliometric workflows. He also participated in reviewing and proofreading the manuscript to ensure consistency and academic accuracy.

Wiwik Wiyanti: Offered methodological insights, particularly in statistical reasoning and bibliometric interpretation. She contributed to editing, ensuring accuracy, precision, and adherence to scholarly writing standards throughout the manuscript.

Lolanda Hamim Annisa: Assisted in validating analytical findings and preparing the final version of the manuscript for submission. Her responsibilities included formatting, proofreading, and ensuring compliance with the journal's technical and stylistic requirements.

Ethics

This article presents original research and contains unpublished material. The corresponding author certifies that there are no conflicts of interest associated with this study and that it does not involve any ethical concerns.

References

Arifin, S., & Muktyas, I. B. (2021). Generate a system of linear equation through. *AIP Conference Proceedings*, 020005. https://doi.org/10.1063/5.0041651

Abdillah, A. A., Azwardi, A., Permana, S., Susanto, I., Zainuri, F., & Arifin, S. (2021). Performance evaluation of linear discriminant analysis and support vector machines to classify cesarean section. *Eastern-European Journal of Enterprise Technologies*, *5*(2 (113)), 37–43. https://doi.org/10.15587/1729-4061.2021.242798

Ahmi, A. (2021). *Bibliometric Analysis for Beginners: A starter guide to begin with a bibliometric study using Scopus dataset and tools such as Microsoft Excel, Harzing's Publish or Perish and VOSviewer software*.

Airbyte. (2024). *What Is Data Imputation: Purpose, Techniques, & Methods.* https://airbyte.com/data-engineering-resources/data-imputation

Arifin, S., & Muktyas, I. B. (2018). Membangkitkan Suatu Matriks Unimodular Dengan Python. *Jurnal Derivat: Jurnal Matematika Dan Pendidikan Matematika*, *5*(2), 1–9. https://doi.org/10.31316/j.derivat.v5i2.361

Arifin, S., Anandha, Y. V. A., Setiawan, E., Phoebus, T. D., Jonathan, S., Effendi, M., Setyawan, M. E., & Putra, K. L. O. (2023a). Prospects and Possibilities for Future Research of Fuzzy C-Means (FCM. *International Journal of Intelligent Systems and Applications in Engineering (IJISAE*, *11*(2), 741–751.

Arifin, S., Bayu Muktyas, I., & Iswara Sukmawati, K. (2021a). Product of two groups integers modulo m,n and their factor groups using python. *Journal of Physics: Conference Series*, *1778*(1), 012026. https://doi.org/10.1088/1742-6596/1778/1/012026

Arifin, S., Faisal, M., Manurung, M. M., Siregar, B., Rahadi, A. P., Eli, A., Ramadhan, G., & Fikriansyah, I. (2025). Big Data Analytics (BDA) in the Research Landscape: Using Python and VOSviewer for Advanced Bibliometric Analysis. *Journal of Computer Science*, *21*(2), 347–362. https://doi.org/10.3844/jcssp.2025.347.362

Arifin, S., Manurung, M. M., Jonathan, S., Effendi, M., & Prasetyo, P. W. (2024). Trend Analysis of the ARIMA Method: A Survey of Scholarly Works. *Recent in Engineering Science and Technology*, *2*(03), 1–14.

Arifin, S., Muktyas, I. B., Al Maki, W. F., & Mohd Aziz, M. K. B. (2022). Graph Coloring Program of Exam Scheduling Modeling Based on Bitwise Coloring Algorithm Using Python. *Journal of Computer Science*, *18*(1), 26–32. https://doi.org/10.3844/jcssp.2022.26.32

Arifin, S., Muktyas, I. B., Prasetyo, P. W., & Abdillah, A. A. (2021b). Unimodular matrix and bernoulli map on text encryption algorithm using python. *Al-Jabar: Jurnal Pendidikan Matematika*, *12*(2), 447–455. https://doi.org/10.24042/ajpm.v12i2.10469

Arifin, S., Wijaya, A. K., Nariswari, R., Yudistira, I. G. A. A., Suwarno, & Faisal. (2023b). Long Short-Term Memory (LSTM): Trends and Future Research Potential. *International Journal of Emerging Technology and Advanced Engineering*, *13*(5), 24–35. https://doi.org/10.46338/ijetae0523_04

Assagaf, I., Sukandi, A., Abdillah, A. A., Arifin, S., & Ga, J. L. (2023). Machine Predictive Maintenance by Using Support Vector Machines. *Recent in Engineering Science and Technology*, *1*(01), 31–35. https://doi.org/10.59511/riestech.v1i01.6

Ball, R. (2020). *Handbook Bibliometrics*.

Banoth, R., & Regar, R. (2023). *Classical and Modern Cryptography for Beginners*. https://doi.org/10.1007/978-3-031-32959-3

Bathla, P., & Kumar, R. (2023). An Efficient Methodology for Brain Stroke Prediction Using Missing Value Imputation Methods. *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS).* https://doi.org/10.1109/icccis60361.2023.10425097

Bellis, N. D. (2009). *Bibliometrics and Citation Analysis*. https://doi.org/10.5040/9798216415817

Bornmann, Lutz, Atkinson, P., Delamont, S., Cernat, A., Sakshaug, J. W., & Williams, R. A. (2020). *Bibliometric Indicators.*

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, *61*(12), 2389–2404. https://doi.org/10.1002/asi.21419

Brimos, P., Seregkos, P., Karamanou, A., Kalampokis, E., & Tarabanis, K. (2024). Deep Learning Missing Value Imputation on Traffic Data Using Self-Attention and GAN-based Methods. *2024 Panhellenic Conference on Electronics & Telecommunications (PACET).* https://doi.org/10.1109/pacet60398.2024.10497055

Chen, X., Sun, B., Bi, S., Yang, J., & Wang, Y. (2024b). IoT Time-Series Missing Value Imputation - Comparison of Machine Learning Methods. *In International Conference on Multimedia Technology and Enhanced Learning* (pp. 406-411). Cham: Springer Nature Switzerland. 406–411. https://doi.org/10.1007/978-3-031-50580-5_37

Chen, F., Yu, L., Mao, J., Yang, Q., Wang, D., & Yu, C. (2024a). A novel data-characteristic-driven modeling approach for imputing missing value in industrial statistics: A case study of China electricity statistics. *Applied Energy*, *373*, 123854. https://doi.org/10.1016/j.apenergy.2024.123854

Decorte, T., Mortier, S., Lembrechts, J. J., Meysman, F. J. R., Latré, S., Mannens, E., & Verdonck, T. (2024). Missing Value Imputation of Wireless Sensor Data for Environmental Monitoring. *Sensors*, *24*(8), 2416. https://doi.org/10.3390/s24082416

Deshkar, P. A., Laghate, K., Ghorpade, A., Padole, D., Shende, H., Kawale, P., & Sakhare, P. (2024). *Data Pre-processing Solution Using Statistical and Data Mining Techniques*. 84–112. https://doi.org/10.1007/978-3-031-70789-6_8

Gorraiz, J. I. (2021). Best practices in bibliometrics & bibliometric services. *Frontiers in Research Metrics and Analytics*, *6*, 771999. https://doi.org/10.3389/978-2-88971-969-3

Han, J., & Kang, S. (2023). Optimization of missing value imputation for neural networks. *Information Sciences*, *649*, 119668. https://doi.org/10.1016/j.ins.2023.119668

Hasan, Md. K., Alam, Md. A., Roy, S., Dutta, A., Jawad, Md. T., & Das, S. (2021). Missing value imputation affects the performance of machine learning: A review and analysis of the literature (2010–2021). *Informatics in Medicine Unlocked*, *27*, 100799. https://doi.org/10.1016/j.imu.2021.100799

Ibrahim, M. A., Arifin, S., Yudistira, I. G. A. A., Nariswari, R., Abdillah, A. A., Murnaka, N. P., & Prasetyo, P. W. (2022). An Explainable AI Model for Hate Speech Detection on Indonesian Twitter. *CommIT (Communication and Information Technology) Journal*, *16*(2), 175–182. https://doi.org/10.21512/commit.v16i2.8343

Jaradat, Y., Alia, M., Masoud, M., Manasrah, A., Jebreil, I., Garaibeh, A., & Al-Arasi, S. (2022). A bibliometric analysis of the International Journal of Advances in Soft Computing and its Applications: Research influence and Contributions. *International Journal of Advances in Soft Computing and Its Applications*, *14*(2), 186–184.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25. https://doi.org/10.1002/asi.5090140103

Liu, J., Qi, X., Wu, Z., Jiang, W., Chen, Z., Jian, Y., & Wen, J. (2024). Research on Imputation Method for Missing Values of Nuclear Power Plants. *Operation Data*, 15–22. https://doi.org/10.3233/FAIA231171

Liu, W.-S., Si, T., Kriauciunas, A., Snell, M., & Gong, H. (2025). Bidirectional f-Divergence-Based Deep Generative Method for Imputing Missing Values in Time-Series Data. *Stats*, *8*(1), 7. https://doi.org/10.3390/stats8010007

Melati, D. N., Umbara, R. P., Astisiasari, A., Wisyanto, W., Trisnafiah, S., Trinugroho, T., Prawiradisastra, F., Arifianti, Y., Ramdhani, T. I., Arifin, S., & Anggreainy, M. S. (2024). A comparative evaluation of landslide susceptibility mapping using machine learning-based methods in Bogor area of Indonesia. *Environmental Earth Sciences*, *83*(3). https://doi.org/10.1007/s12665-023-11402-3

Menéndez García, L. A., Menéndez Fernández, M., Sokoła-Szewioła, V., Álvarez de Prado, L., Ortiz Marqués, A., Fernández López, D., & Bernardo Sánchez, A. (2022). A Method of Pruning and Random Replacing of Known Values for Comparing Missing Data Imputation Models for Incomplete Air Quality Time Series. *Applied Sciences*, *12*(13), 6465. https://doi.org/10.3390/app12136465

Muhammad, I., & Triansyah, F. A. (2023). *Panduan lengkap analisis bibliometrik dengan VOSviewer: Memahami perkembangan dan tren penelitian di era digital*. Penerbit Adab.

Middya, A. I., & Roy, S. (2024). Multiview data fusion technique for missing value imputation in multisensory air pollution dataset. *Journal of Ambient Intelligence and Humanized Computing*, *15*(8), 3173–3191. https://doi.org/10.1007/s12652-024-04816-9

Moed, H. F. (2006). *Citation Analysis in Research Evaluation*.

Park, J., Müller, J., Arora, B., Faybishenko, B., Pastorello, G., Varadharajan, C., Sahu, R., & Agarwal, D. (2022). Long-term missing value imputation for time series data using deep neural networks. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-022-08165-6

Perianes-Rodriguez, A., Waltman, L., & van Eck, N. J. (2016). Constructing bibliometric networks: A comparison between full and fractional counting. *Journal of Informetrics*, *10*(4), 1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Rahman, B., Arifin, S., & Muktyas, I. (2019). All cyclic subgroups in group (Zm × zn, +) using python. *International Journal of Scientific and Technology Research*, *8*(9), 2282–2285.

Shathir, M. K., Ajeena, R. K., & Arif, G. E. (2023). The triple vertex path graph for hill encryption schemes. *AIP Conference Proceedings*. PHYSICAL MESOMECHANICS OF CONDENSED MATTER: Physical Principles of Multiscale Structure Formation and the Mechanisms of Nonlinear Behavior: MESO2022, Tomsk, Russia. https://doi.org/10.1063/5.0170404

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, *24*(4), 265–269. https://doi.org/10.1002/asi.4630240406

Song, Z., Zhou, Y., Cheng, W., Liang, F., & Zhang, C. (2024). Missing Value Imputation for Radar-Derived Time-Series Tracks of Aerial Targets Based on Improved Self-Attention-Based Network. *Computers, Materials & Continua*, *78*(3), 3349–3376. https://doi.org/10.32604/cmc.2024.047034

Suwannawach, P., Jakor, K., Teeravech, N., & Sitikornchayarpong, C. (2024). The Comparison of Missing Value Imputation for Price Index Forecasting Based on ARIMA Model. *2024 21st International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. Khon Kaen, Thailand. https://doi.org/10.1109/ecti-con60892.2024.10595040

Sulistiyawati, S., Kusumah, Y. S., Dahlan, J. A., Juandi, D., & Vos, H. (2023). A Bibliometric Analysis: Trend of Studies in Self-Regulated Learning Over the Past Three Decades. *Indonesian Journal on Learning and Advanced Education (IJOLAE)*, 178–197. https://doi.org/10.23917/ijolae.v5i2.21381

Syavasya, C. V. S. R., & Muddana, A. L. (2024). An efficient missing value imputation and evaluation using GK-KH means and HTR-RNN. *International Journal of Business Intelligence and Data Mining*, *24*(1), 25–46. https://doi.org/10.1504/ijbidm.2024.135130

Tahir, M., Abdullah, A., Udzir, N. I., & Kasmiran, K. A. (2025). A novel approach for handling missing data to enhance network intrusion detection system. *Cyber Security and Applications*, *3*, 100063. https://doi.org/10.1016/j.csa.2024.100063

Tarigan, S., Murnaka, N. P., & Arifin, S. (2021). Development of teaching material in mathematics "Sapta Maino Education" on topics of plane geometry. *AIP Conference Proceedings*, 2331. https://doi.org/10.1063/5.0041650

Tenriawaru, A., Ransi, N., Surimi, L., Saidi, L. O., Mangidi, A. P., & Budiman, H. (2023). Implementation of Haar Cascade classifier method and local binary pattern histogram on face identification. *AIP Conference Proceedings*. 2nd International Conference on Sciences and Mathematics (2nd ICSM), Kendari, Indonesia. https://doi.org/10.1063/5.0138537

van Eck, N. J., & Waltman, L. (2010a). *Vosviewer: A Computer Program for Bibliometric Mapping*.

van Eck, N. J., & Waltman, L. (2010b). Software Survey: VOS Viewer, a Computer Program for Bibliometric Mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

Wang, Y., Xu, X., Hu, L., Fan, J., & Han, M. (2024). A time series continuous missing values imputation method based on generative adversarial networks. *Knowledge-Based Systems*, *283*, 111215. https://doi.org/10.1016/j.knosys.2023.111215

Wu, J., Tang, X., Liu, G., & Wu, B. (2024). An Overview of Graph Data Missing Value Imputation. *Communications in Computer and Information Science: Vol. 2059 CCIS*. Springer, Singapore.

Zhang, X., & Zhou, P. (2024). A transferred spatio-temporal deep model based on multi-LSTM auto-encoder for air pollution time series missing value imputation. *Future Generation Computer Systems*, *156*, 325–338. https://doi.org/10.1016/j.future.2024.03.015

Zheng, H., & Huang, T. (2023). New incomplete data imputation based on k-nearest neighbor type framework. *2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA)*. Shenyang, China. https://doi.org/10.1109/icpeca56706.2023.10075766

Zhang, Z. Yan, X., Zhang, L., Lai, X., & Lu, W. (2024). Fuzzyneuron modeling of incomplete data for missing value imputation. *Information Sciences*, *659*, 120065. https://doi.org/10.1016/j.ins.2023.120065