

Navigating Data Overload: A Bibliometric Analysis of Trends and Challenges in Real-Time Data Processing (2020–2026)

Ruziana Mohamad Rasli^{1,*}, Roznim Mohamad Rasli², Juhaida Abu Bakar³, Sobihatun Nur Abdul Salam¹, Hafizah Abdullah @ Ali⁴, Emy Hazlinda Mohammad Ridzwan⁵, Mime Azrina Jaafar⁶

¹ School of Multimedia Technology and Communication, Universiti Utara Malaysia, Sintok, Malaysia

² Faculty of Computing and Meta-technology, Universiti Pendidikan Sultan Idris, Tanjong Malim, Malaysia

³ School of Computing, Universiti Utara Malaysia, Sintok, Malaysia

⁴ Kolej Komuniti Bandar Darul Aman, Jitra, Malaysia

⁵ Politeknik Ungku Omar, Ipoh, Malaysia

⁶ Politeknik Sultan Abdul Halim Mu'adzam Shah, Jitra, Malaysia

Email: ^{1,*}ruziana@uum.edu.my, ²roznim@meta.upsi.edu.my, ³juhaida.ab@uum.edu.my, ⁴sobihatun@uum.edu.my,

⁵hafizah@kkbda.edu.my, ⁶emyhazlinda@puo.edu.my, ⁷mimeazrina@polimas.edu.my

Correspondence Author Email: ruziana@uum.edu.my

Submitted: 04/01/2026; Accepted: 19/01/2026; Published: 20/01/2026

Abstract—The continuous growth of digital technologies has resulted in the generation of massive volumes of data that must be processed and acted upon in real time. Applications such as smart cities, healthcare monitoring, financial systems, industrial automation, and intelligent transportation increasingly rely on real-time data streams to support timely decision-making. However, this dependence on real-time data has intensified the problem of data overload, where the volume, velocity, and complexity of incoming data exceed the processing capabilities of systems or the cognitive capacity of human users. This paper provides a comprehensive discussion of real-time data processing in the context of data overload, focusing on technical and human-centered challenges and emerging solutions such as edge computing, artificial intelligence, and streaming analytics. A total of 193 documents published between 2020 and 2026 were retrieved from the Scopus database and were identified based on predefined search criteria. The documents are then analyzed using VOSviewer using keyword co-occurrence and the results generated shows that the publication trend analysis reveals a clear upward trajectory in research output from 2020 to 2026, with a particularly sharp increase after 2021. Based on the visualization results, in summary, there are four clusters generated which reveal a multi-layered research structure where Cluster 1 focuses on intelligent algorithms and cognitive capabilities, Cluster 2 provides the distributed infrastructure and system efficiency, Cluster 3 demonstrates domain-specific, real-time experimental applications, and Cluster 4 emphasizes human interaction, data integration, and information use. This growth indicates heightened scholarly interest and reflects broader technological and societal shifts toward intelligent and real-time systems. By providing a structured overview of the field's intellectual landscape, this study supports informed decision-making and lays the foundation for future research in artificial intelligence-driven real-time systems.

Keywords: Real-Time Data; Data Overload; Streaming Analytics; Scopus-Indexed Publications; Bibliometric Analysis

1. INTRODUCTION

The increase of connected devices, sensors, and digital platforms has transformed how data are generated and consumed (Allam et al., 2022). Real-time data processing enables systems to respond to events as they occur, supporting critical applications such as healthcare monitoring, autonomous systems, and financial trading (Kwok et al., 2025; Elhadad et al., 2022). Despite these benefits, real-time environments are increasingly affected by data overload. When excessive data are presented without adequate filtering or prioritization, system performance and human decision-making may reduce significantly (Pandey et al., 2025). Addressing this challenge has become a priority in both academic research and system design.

Recent studies show that incorporating edge computing techniques significantly alleviates real-time data overload by processing data closer to its source, thereby reducing latency and network bandwidth pressure (Qian, 2025). By deploying edge-based data handling frameworks, systems can filter, aggregate, and analyze data locally before forwarding only meaningful information, improving throughput and reducing the volume of raw data transmitted to central servers. This architectural shift is particularly beneficial in high-density Internet of Things (IoT) environments where continuous data generation can overwhelm traditional cloud-centric models.

In addition to edge strategies, hybrid frameworks that integrate blockchain with edge computing have been proposed to enhance data integrity and real-time performance simultaneously (Dong et al., 2022). Such systems enable reliable decision-making with reduced delays, achieving high classification accuracy and robust throughput even with massive IoT-generated datasets. This combination not only improves real-time responsiveness but also addresses data security and privacy concerns while managing high-velocity data streams.

Beyond technical infrastructure, advancements in real-time analytics have shown that rapid automated decision-making can significantly enhance operational intelligence across industries. Real-time analytics platforms now support immediate insights for strategic decisions, enabling organizations to respond to changes in market conditions, customer behavior, or system performance without time lags (Grandhe, 2025). These innovations help organizations extract actionable insights from continuously streaming data, thereby mitigating overload through prioritized and context-aware information processing.

Human factors also play a crucial role in managing data overload (Arnold et al., 2023). Research on information overload and decision quality indicates that when data exceeds human cognitive capacity, performance deteriorates

sharply, emphasizing the need for effective filtering and information prioritization mechanisms (James, 2025). Adaptive user interfaces and intelligent decision support systems can reduce cognitive load by presenting relevant data contextually, improving decision accuracy and reducing fatigue in high-stakes, real-time environments.

Furthermore, real-time sensor networks, particularly in industrial and IoT-enabled systems, continue to demonstrate substantial gains in efficiency and scalability when combined with advanced communication protocols and analytics (Abdulhussain et al., 2025). These integrated systems enhance process control and decision-making by improving the collection, processing, and interpretation of high-frequency data streams (Pandey et al., 2025). However, challenges such as energy efficiency, security, and scalability remain key areas for future research to fully realize the potential of real-time data systems. Therefore, there is a need to do bibliometric analysis since currently, there is still lack of quantitative mapping of the overall intellectual landscape. The results can help to identify clusters and future directions in Real time data processing.

2. RESEARCH METHODOLOGY

This section outlines the systematic approach used to retrieve, filter, and analyse the bibliometric data related to data overload in real-time data processing, limited to publications from 2020 to 2026. This analysis examined the publications collected from the Elsevier Scopus database based on documents categorized by conference papers, article papers and book chapters and other related documents types as in Figure 1.

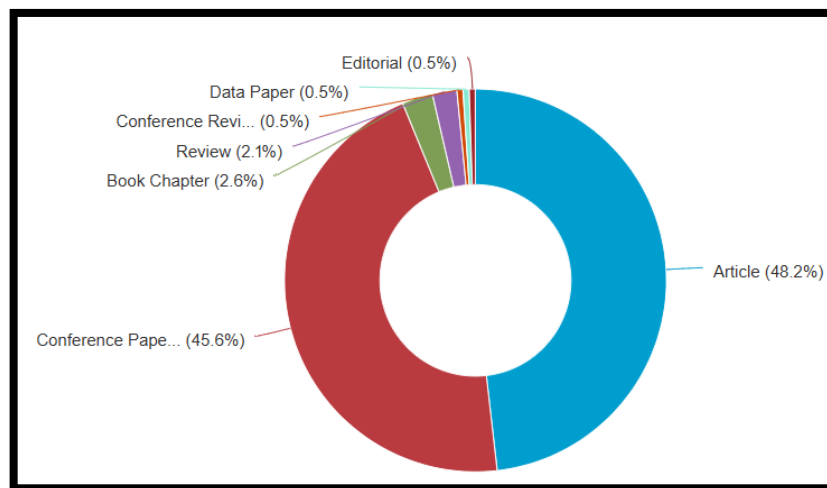


Figure 1. Distribution of documents

Figure 1 shows that journal articles contributed to the largest share (48.2%), reflecting a strong emphasis on mature, peer-reviewed research outputs. This suggests that the topic has attracted sustained academic interest and that many studies have progressed to a level suitable for archival journal publication. Closely following are conference papers (45.6%), which account for nearly half of the total publications. This high proportion highlights the dynamic and evolving nature of the research area, where researchers actively disseminate early findings, prototypes, and emerging ideas through conferences. Such a pattern is common in fast-moving domains related to real-time data, computing systems, and digital technologies. The remaining publication types contribute relatively small proportions. Book chapters (2.6%) indicate limited but notable inclusion of the topic in scholarly books, often providing conceptual or integrative perspectives. Review articles (2.1%) suggest that comprehensive syntheses of the field are still emerging, implying opportunities for further systematic reviews. Minor categories such as conference reviews (0.5%), data papers (0.5%), and editorials (0.5%) collectively represent marginal contributions, mainly offering commentary, curated datasets, or insights rather than primary research.

2.1 Conceptual Background

Real-time data processing is defined by low-latency requirements, continuous data streams, and high system responsiveness (Liu et al., 2025). Unlike batch processing, real-time systems must handle data immediately upon arrival, often under strict timing constraints (Stonebraker et al., 2005). Data overload extends the concept of information overload by emphasizing operational and cognitive impacts. In real-time contexts, even moderate data volumes can cause overload when relevance, timing, and quality are not properly managed (Jauhari & Vobugari, 2025). One major challenge in real-time systems is computational saturation, where processing pipelines are unable to keep pace with incoming data streams (Pulivarthy et al., 2026). This can result in increased latency, dropped data, or incorrect system responses (Banoth et al., 2025). From a human-centered perspective, excessive real-time information can overwhelm operators and reduce situational awareness, particularly in safety-critical domains such as healthcare and transportation (Schmager et al., 2025).

2.2 Search Strategy

In order to retrieve relevant documents, a Boolean search string was developed that captured the intersection of data overload in real time data processing. The query was applied to the title “Data Overload in Real Time Data Processing” with a filter of year from 2020-2026. Figure 2 shows an illustration of a flowchart of the methodology used for conducting this bibliometric analysis of the topic of “Data Overload” and “Real Time Data Processing” using the Scopus database.

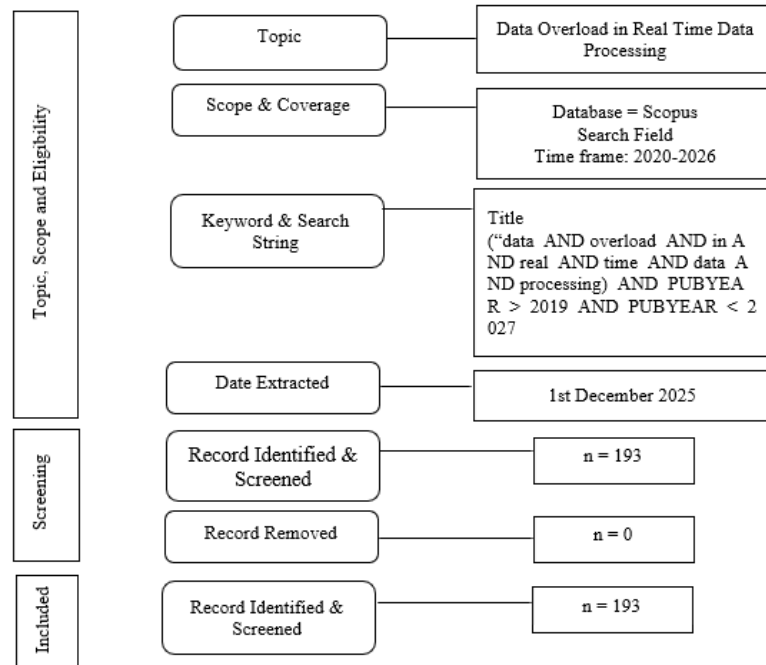


Figure 2. Flow Diagram of the Strategy (Source: Ishak et al. (2023) as cited from Zakaria et al. (2024))

Figure 2. consists of three main sections which are 1: Topic, Scope and Eligibility, 2: Screening and 3: Included. The first section outlines the foundational aspects of the research where the topics used are stated, the scope and coverage of the paper, the title of the search and also when the documents are extracted. It can be seen in the second section, Screening where the total amount of documents was listed out which is 193 documents. In this search, there is no removed documents in this search process because the keywords are specifically filtered in the searching process on Scopus database. The documents that have been searched are in detail already, therefore there is no record removed and it indicates that all the records initially identified were considered relevant to the analysis. This analysis comprises of three components listed below:

- Pattern in citations is examined to discover cited publication, influential authors, and related research groups
- The themes and concepts on real time data processing is analyzed based on co-occurring keywords.
- The emerging trends, areas and limitations of the area are discussed to offer a view of the research landscape and guidance.

This study offers insight into fundamental knowledge frameworks, research patterns, and trends of the technology used in visualizing real time data.

2.3 Data Preparation in VOSViewer

The Scopus-exported CSV file was converted to a format compatible with VOSviewer. The files were exported with metadata such as in Figure 3 below:

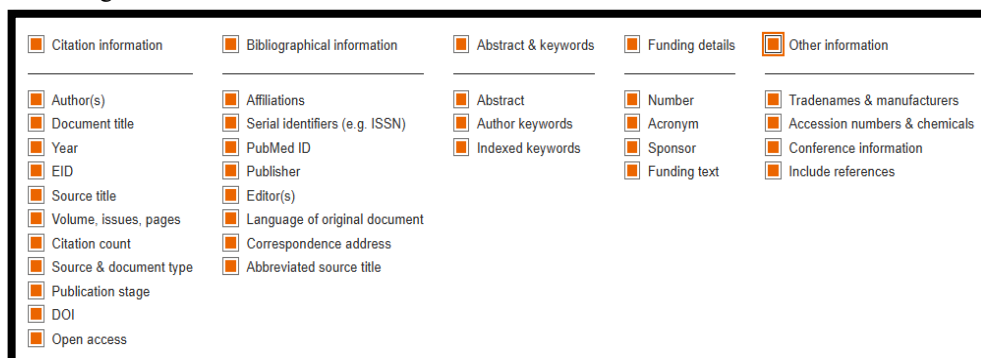


Figure 3. Metadata information exported from Scopus database

Figure 3 shows the various categories of metadata fields commonly used to represent a different type of bibliographic information that helps identify, classify, and analyze research outputs. These categories is important for researchers, especially when conducting literature reviews, citation analyses, or bibliometric studies. The following figures show the basic setting of threshold, number of keywords and the selected keywords.

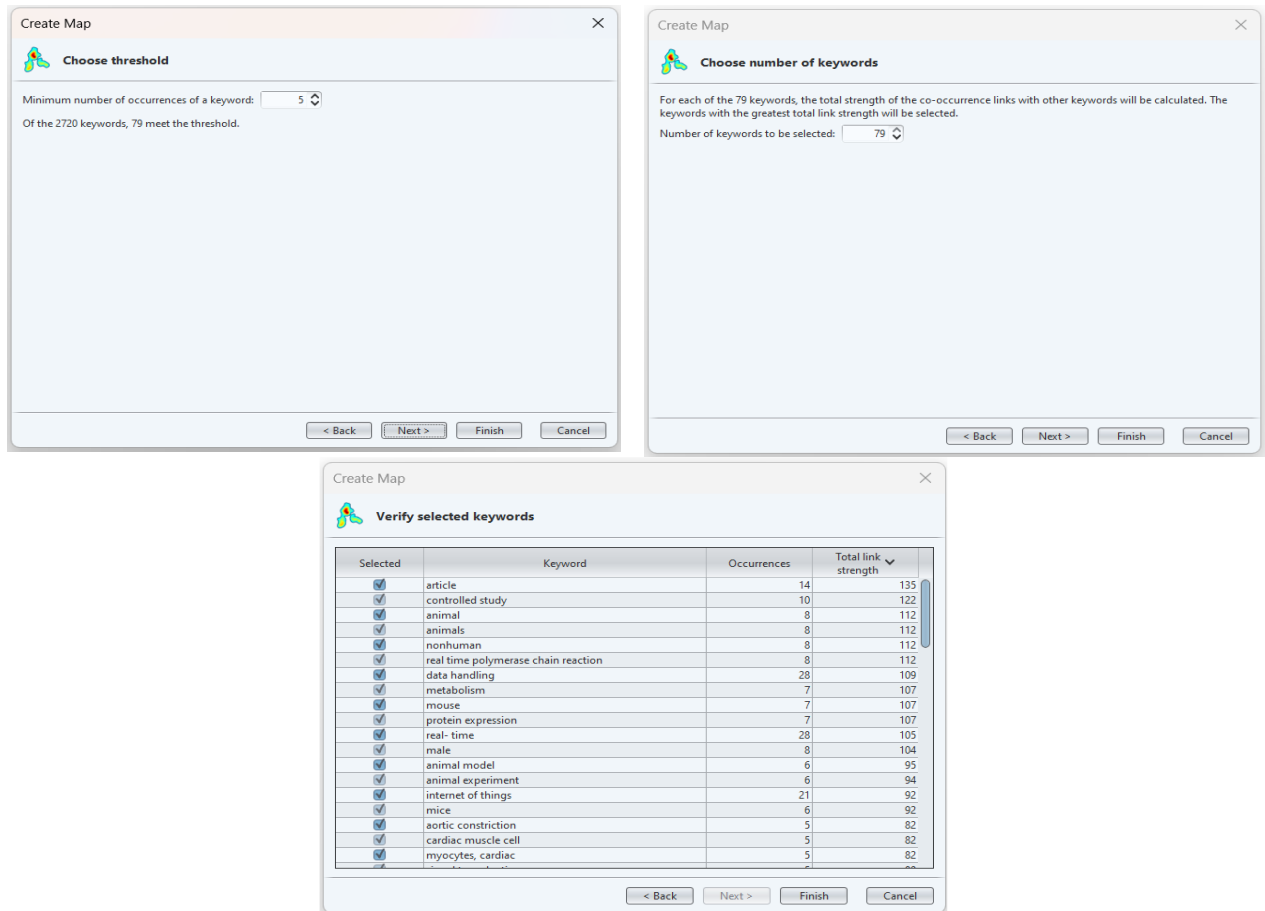


Figure 4. Basic setting of threshold, number of keywords and generated keywords

4. RESULTS AND DISCUSSION

Figure 5 below shows the keyword co-occurrence network visualization generated using VOSviewer, revealing the conceptual structure and thematic relationships within the research domain. In this network, node size represents keyword frequency, link thickness indicates the strength of co-occurrence, and colors denote distinct thematic clusters. The resulting map highlights a highly interconnected and interdisciplinary research landscape centered on real-time data processing and intelligent systems.

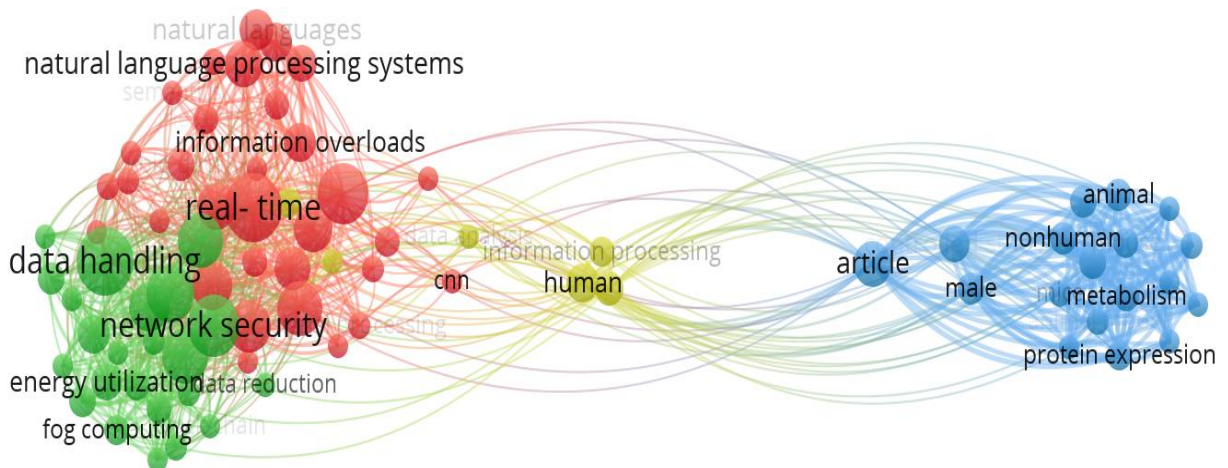


Figure 5. Network visualization generated from VOSviewer

Based on Figure 5, the central region of the network is dominated by keywords such as artificial intelligence, machine learning, deep learning, computer vision, image processing, and real time, indicating that intelligent data analytics constitute the core of the research domain. The dense interconnections among these terms suggest that data-driven perception, learning, and automated decision-making are foundational to addressing challenges associated with high-velocity and high-volume data streams. The presence of terms related to information overloads and decision making further emphasizes the role of intelligent algorithms in filtering, prioritizing, and interpreting real-time information.

A prominent cluster on the left side of the network represents the system and infrastructure layer supporting real-time intelligence. This cluster includes keywords such as cloud computing, edge computing, fog computing, network security, quality of service, resource allocation, and energy efficiency. The strong links between this cluster and the central AI-driven themes highlight the dependency of intelligent applications on scalable, low-latency, and energy-aware computing architectures. These findings indicate that managing data overload is not solely an algorithmic challenge but also a system-level concern requiring efficient data handling and resource management.

Another cluster bridges technical and application-oriented themes through keywords such as human, information processing, and information use. This cluster underscores the importance of human-centered considerations in real-time data environments, where excessive information can lead to cognitive overload and reduced decision quality. Its intermediary position within the network suggests that effective real-time systems must balance automated intelligence with usability and interpretability for human users.

On the right side of the network, a distinct cluster associated with biomedical and experimental research applications is observed. Keywords such as animal model, mice, cardiac muscle cell, protein expression, and real time polymerase chain reaction indicate the application of real-time data processing techniques within life sciences. Although this cluster is more peripheral, its connections to the central real-time and AI-related keywords suggest the transfer of computational methods into biomedical research, where managing high-frequency and sensitive data is critical.

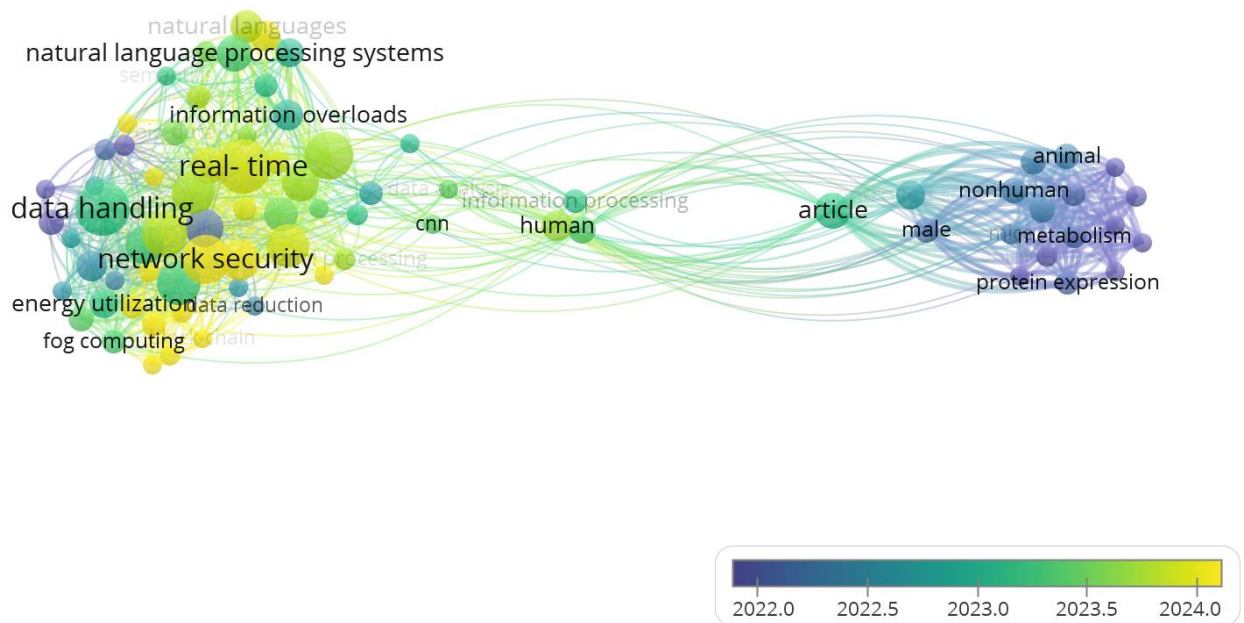


Figure 6. Overlay visualization generated from VOSviewer

The VOSviewer overlay visualization presents a keyword co-occurrence network that highlights both the thematic structure and temporal evolution of the selected Scopus dataset. Two dominant research streams can be observed. The first stream, located on the left side of the map, is centred on computing-related topics such as real-time, data handling, network security, and fog computing, which appear in warmer colours (green to yellow), indicating a more recent concentration of publications around 2023–2024.

This pattern suggests a growing research emphasis on real-time data processing, efficient system performance, and cybersecurity concerns in modern digital environments. In contrast, the second stream, positioned on the right side of the visualization, is composed mainly of biomedical-oriented terms such as animal, nonhuman, metabolism, and protein expression, which are largely represented in cooler colours (blue to purple), reflecting earlier publication activity around 2022. The presence of bridging terms such as human, male, and article indicates overlapping indexing categories that connect these two thematic clusters. Overall, the overlay mapping demonstrates a noticeable shift in research focus from earlier life-science-related topics toward more recent developments in real-time and security-driven computing studies, reflecting changing scholarly priorities over time.

Overall, the keyword co-occurrence analysis reveals a multi-layered research structure integrating intelligent algorithms, distributed computing infrastructures, human-centered information processing, and domain-specific applications. The strong inter-cluster connections demonstrate that real-time data and data overload challenges are

addressed through a holistic and interdisciplinary approach, reflecting the growing maturity and convergence of research in this field. Detailed keywords on each clusters are Cluster 1 which is divided into 32 items, cluster 2 which is divided into 23 items, cluster 3 which is divided into 18 items, cluster 4 which is divided into 6 items and cluster 5 is divided into 7 items. Table 1 shows the detailed keywords based on each clusters listed above.

Table 1. Detailed keywords based on clusters

Cluster 1	Cluster 2	Cluster 3	Cluster 4
a. Artificial Intelligence	a. Balancing	a. Animal	a. Data analysis
b. Big data	b. Block-chain	b. Animal experiment	b. Data integration
c. CNN	c. Cloud computing	c. Animal model	c. Human
d. Cognitive systems	d. Cloud-computing	d. Animal tissue	d. Humans
e. Computer vision	e. Computer architecture	e. Animals	e. Information processing
f. Convolutional neural network	f. Data communication system	f. Aortic constriction	f. Iot
g. Data mining	g. Data handling	g. Article	
h. Decision making	h. Data reduction	h. Cardiac muscle cell	
i. Deep learning	i. Data transfer	i. Controlled study	
j. Digital Twin	j. Digital storage	j. Male	
k. Embedded systems	k. Edge computing	k. Metabolism	
l. Fault decision	l. Energy efficiency	l. Mice	
m. Image processing	m. Energy utilization	m. Mouse	
n. Information overloads	n. Fog computing	n. Myocytes, cardiac	
o. Information systems	o. Green computing	o. Nonhuman	
p. Information use	p. Information management	p. Protein Expression	
q. Language processing	q. Internet of things	q. Real time polymerase chain reaction	
r. Learning systems	r. Network architecture	r. Signal transduction	
s. Machine learning	s. Network security		
t. Machine-learning	t. Performance		
u. Metadata	u. Quality of service		
v. Natural language processing	v. Quality-of-service		
w. Natural language processing systems	w. Resource allocation		
x. Natural languages			
y. Power			
z. Program processors			
aa. Real time systems			
bb. Real time			
cc. Recommender systems			
dd. Semantics			
ee. Signal Processing			
ff. Virtual reality			

In summary, these four clusters reveal a multi-layered research structure where Cluster 1 focuses on intelligent algorithms and cognitive capabilities, Cluster 2 provides the distributed infrastructure and system efficiency, Cluster 3 demonstrates domain-specific, real-time experimental applications, and Cluster 4 emphasizes human interaction, data integration, and information use.

Therefore, this structure reflects a holistic research ecosystem addressing real-time data and data overload from algorithmic, infrastructural, application, and human-centered perspectives, highlighting the interdisciplinary nature and growing maturity of the field.

5. CONCLUSION

This paper has examined the challenges associated with real-time data processing in the presence of data overload. By combining technological and human-centered perspectives, the study highlights pathways toward scalable, reliable, and user-centered real-time systems. The VOSviewer overlay visualization presents a keyword co-occurrence network that highlights both the thematic structure and temporal evolution of the selected Scopus dataset. Two dominant research streams can be observed. The first stream, located on the left side of the map, is centred on computing-related topics such as real-time, data handling, network security, and fog computing, which appear in warmer colours (green to yellow), indicating a more recent concentration of publications around 2023–2024. This pattern suggests a growing research emphasis on real-time data processing, efficient system performance, and cybersecurity concerns in modern digital

environments. In contrast, the second stream, positioned on the right side of the visualization, is composed mainly of biomedical-oriented terms such as animal, nonhuman, metabolism, and protein expression, which are largely represented in cooler colours (blue to purple), reflecting earlier publication activity around 2022. The presence of bridging terms such as human, male, and article indicates overlapping indexing categories that connect these two thematic clusters. Overall, the overlay mapping demonstrates a noticeable shift in research focus from earlier life-science-related topics toward more recent developments in real-time and security-driven computing studies, reflecting changing scholarly priorities over time. Effective management of real-time data overload has significant implications across domains. In healthcare, intelligent filtering can reduce alarm fatigue and improve patient safety (Cvach, 2012). In smart cities, real-time analytics support responsive traffic management and public safety systems (Hashem et al., 2016). Future research should focus on adaptive and context-aware data processing frameworks that dynamically adjust filtering and prioritization strategies. The integration of explainable artificial intelligence is also critical to improving transparency and trust in automated real-time systems (Doshi-Velez & Kim, 2017).

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education (MoHE) of Malaysia through the Fundamental Research Grant Scheme – Early Career (FRGS-EC/1/2024/ICT09/UUM/03/1). The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of MoHE, Malaysia.

REFERENCES

- Abdullahsain, S. H., Mahmmod, B. M., Alwhelat, A., Shehada, D., Shihab, Z. I., Mohammed, H. J., & Hussain, A. (2025). A comprehensive review of sensor technologies in IoT: Technical aspects, challenges, and future directions. *Computers*, *14*(8), 342.
- Allam, Z., Bibri, S. E., Jones, D. S., Chabaud, D., & Moreno, C. (2022). Unpacking the “15-minute city” via 6G, IoT, and digital twins: Towards a new narrative for increasing urban efficiency, resilience, and sustainability. *Sensors*, *22*(4), 1369.
- Arnold, M., Goldschmitt, M., & Rigotti, T. (2023). Dealing with information overload: A comprehensive review. *Frontiers in Psychology*, *14*, 1122200.
- Banoth, S., Vineesha, M., Punna, H. S., Mathiyalagan, P., Prakash, V., & Jasmin, M. (2025). Edge computing architectures for low-latency data processing in Internet of Things applications. In *ITM Web of Conferences* (Vol. 76, p. 03003). EDP Sciences.
- Dong, J., Song, C., Zhang, T., Li, Y., & Zheng, H. (2022). Integration of edge computing and blockchain for provision of data fusion and secure big data analysis for Internet of Things. *Wireless Communications and Mobile Computing*, *2022*(1), 9233267.
- Elhadad, A., Alanazi, F., Taloba, A. I., & Abozeid, A. (2022). Fog computing service in the healthcare monitoring system for managing real-time notification. *Journal of Healthcare Engineering*, *2022*(1), 5337733.
- Grandhe, K. (2025). Impact of real-time analytics on strategic decision-making in large organizations. *International Journal of Scientific and Academic Research*, *16*(4), 1–12. <https://doi.org/10.71097/IJSAT.v16.i4.9032>
- Ishak, N. S., Roslan, M. F., & Abdullah, K. H. (2023). Bibliometric analysis of Malaysian authorship: Trends, patterns, and prospects. *Asian Journal of Research in Education and Social Sciences*, *5*(2), 33–44.
- James, H. (2025). *Information overload and cognitive offloading in the digital age*.
- Jauhari, S., & Vobugari, S. (2025). Delineating the big data era and the information overload problem. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5209835>
- Kwok, J., Li, S., Lohstroh, M., & Lee, E. A. (2025, May). HPRM: High-performance robotic middleware for intelligent autonomous systems. In *Proceedings of the 2025 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 8093–8099). IEEE.
- Liu, J., Liu, X., Qu, M., & Lyu, T. (2025). EITNet: An IoT-enhanced framework for real-time basketball action recognition. *Alexandria Engineering Journal*, *110*, 567–578.
- Pandey, S., Chaudhary, M., & Tóth, Z. (2025). An investigation on real-time insights: Enhancing process control with IoT-enabled sensor networks. *Discover Internet of Things*, *5*(1), 29.
- Pulivarthy, P., Kommineni, M., Aragani, V. M., & Rajassekaran, G. (2026). Real-time data pipeline engineering for scalable insights. In *Machine learning, predictive analytics, and optimization in complex systems* (pp. 83–102). IGI Global Scientific Publishing.
- Qian, F. (2025). Real-time data processing method of IoT based on edge computing. *Journal of Computing and Electronic Information Management*, *17*(1), 6–10. <https://doi.org/10.54097/k2406883>
- Schmager, S., Pappas, I. O., & Vassilakopoulou, P. (2025). Understanding human-centred AI: A review of its defining elements and a research agenda. *Behaviour & Information Technology*, 1–40.
- Zakaria, N., Abdullah, N. A. C., Zahoor, N., Azizan, M., & Syed, Z. A. A. (2024). A bibliometric analysis of quality of work-life: Current status, development, and future research directions. *Pakistan Journal of Life and Social Sciences*, *22*(2), 326–347.