Malaysian Journal of Sustainable Environment
Special issue, June (2025), 41-60
doi: 10.24191/myse.v12i2.7056

A SYSTEMATIC REVIEW OF MULTIMODAL ANALYTICAL TECHNIQUES FOR ENHANCING CYBER RESILIENCE IN SMART CITIES

Zul-Azri Ibrahim[1]*, Saiful Adli Ismail[2] & Fiza Abdul Rahim [3]
***Corresponding Author**

[1]College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Putrajaya Campus, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia,
[2,3]Faculty of Artificial Intelligence, Universiti Teknologi Malaysia (UTM), Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia

*zulazri@uniten.edu.my, saifuladli@utm.my
fiza.abdulrahim@utm.my
Received: 14 July 2024
Accepted: 11 October 2024
Published: 30 June 2025

ABSTRACT

In recent years, the advancement of smart city technologies has necessitated a rethinking of urban resilience strategies, especially in the face of escalating cyber threats. Enhancing cyber resilience in smart cities is crucial due to their reliance on interconnected digital infrastructures that manage essential services such as transportation, energy, healthcare, and public safety. Cyber attacks on these systems can disrupt services, compromise data, and undermine public trust. This systematic review examines 27 selected articles on the application of multimodal analytical techniques to enhance the cyber resilience of smart cities. By evaluating these studies based on their analytical frameworks, techniques used, and research contexts, the researcher identifies the benefits and challenges of integrating multimodal data such as text, audio, video, and sensor data into urban cyber resilience strategies. The findings reveal that various multimodal strategies, including machine learning, data fusion, and real-time monitoring, significantly contribute to the robustness of smart urban infrastructures against cyber attacks. This review provides a comprehensive understanding of the application of multimodal analytics in safeguarding smart city infrastructures, highlighting best practices, technological advancements, and future research directions.

Keywords: *Cyber security, Multimodal analysis, Urban resilience, Machine Learning, Systematic Review*

INTRODUCTION

With the growing digitization of the world, smart cities have emerged as a viable solution to tackle the problems caused by urbanization and improve the quality of life for citizens (Chaudhuri & Bozkus Kahyaoglu, 2023). However, the integration of diverse technologies and systems in smart cities also poses significant cybersecurity risks (Kalinin et al., 2021), emphasizing the need for robust and comprehensive approaches to guarantee cyber resilience.

Multimodal analysis involves combining different types of data—such as text, audio, video, and sensor data—to develop a thorough understanding of complex systems (Kohout et al., 2021). This method holds significant promise for enhancing the resilience of smart cities against cyber incidents. Although these strategies have potential benefits, their actual implementation and application still pose a significant challenge (Juliana & Arafah, 2018). A multifaceted approach to resilience is necessary due to the intricate nature of urban systems and the diverse range of cyber threats.

This study systematically reviews the literature on multimodal analytical techniques used in smart city environments. The objective is to analyze selected articles to synthesize how multimodal analytics directly contribute to protecting smart cities by evaluating their data sources and techniques. This paper contributes to the ongoing discussion on improving urban resilience through multimodal analytical techniques by identifying challenges, successful implementations, and future research opportunities. This study is significant, given the growing digitalization of urban areas and the increasing complexity of cyber threats.

LITERATURE REVIEW

Smart cities represent a transformative approach to urban development, leveraging technology and data to enhance the efficiency, sustainability, and quality of life for residents. These cities integrate information and communication technologies (ICT) into their infrastructure and services, creating interconnected networks that optimize everything from traffic management to energy usage. By utilizing advanced sensors, data analytics, and automation, smart cities aim to provide more responsive and effective public services, reduce environmental impact, and foster economic growth (Gracias et al., 2023).

Moreover, the integration of smart technologies in urban planning and development can lead to more inclusive and participatory governance. Citizens can engage with city authorities through digital platforms, providing feedback and suggestions on various issues. This collaborative approach not only improves service delivery but also fosters a sense of community and ownership among residents (Bastos et al., 2022). Overall, the significance of smart cities extends beyond technological advancements; it encompasses a holistic vision of sustainable and resilient urban living.

Cyber Resilience

Cyber resilience refers to the ability of an organization or system to withstand, respond to, and recover from cyber incidents. In the context of smart cities, cyber resilience is critical due to the increasing dependence on digital infrastructure and interconnected systems. A cyber-resilient smart city can maintain its essential functions and services even in the face of cyber threats, ensuring the safety and well-being of its residents.

Urban infrastructures, such as transportation systems, energy grids, and communication networks, are particularly vulnerable to cyber attacks. These infrastructures rely heavily on ICT, making them attractive targets for cybercriminals (Kim et al., 2023). A successful cyber attack on a smart city's energy grid, for example, could disrupt power supply, affecting homes, businesses, and critical services like hospitals. Similarly, attacks on transportation systems could lead to significant disruptions, causing delays and potential safety hazards.

Given these risks, enhancing the cyber resilience of urban infrastructures is paramount. This involves implementing robust security measures, such as encryption, intrusion detection systems, and continuous monitoring, to protect against cyber threats. Additionally, it requires developing comprehensive response and recovery plans to minimize the impact of incidents when they occur. By prioritizing cyber resilience, smart cities can safeguard their infrastructures, ensuring uninterrupted services and maintaining public trust.

Multimodal Analysis

Multimodal analysis in smart cities integrates data from various sources like text, audio, video, and sensors, offering a comprehensive view for improved decision-making in areas like security and transportation. This approach is crucial for enhancing cyber resilience, as it allows for effective detection and response to cyber threats by combining diverse data such as network logs, surveillance footage, and social media. Techniques such as machine learning, data fusion, and real-time monitoring are key, with machine learning identifying patterns and anomalies, data fusion creating unified datasets for richer analysis, and real-time systems ensuring prompt threat responses (Ahmadi-Assalemi et al., 2020).

However, implementing these techniques involves challenges such as managing vast, diverse data streams which require sophisticated algorithms and significant computational resources. Data privacy is another major concern; integrating data raises issues regarding personal information security and compliance with regulations like General Data Protection Regulation (GDPR). Additionally, the dynamic nature of urban environments demands that analytical models be scalable and adaptable to continuously evolving conditions and new data types (Almeida, 2023). Developing solutions that address these complexities is essential for leveraging multimodal analysis to fortify cyber resilience in smart cities.

METHODOLOGY

The systematic literature review (SLR) undertaken in this study follows the procedure by Kitchenham and Charters (Kitchenham & Charters, 2007), to identify, evaluate, and synthesize existing research on multimodal analytical techniques for enhancing cyber resilience in smart cities. The systematic review process involves a structured and comprehensive approach to literature search and analysis, ensuring the inclusion of relevant studies and minimizing bias. Initially, the researcher defined the research questions and established the criteria for article selection to guide the review process.

Search Strategy

To gather a comprehensive dataset of relevant studies, the researcher conducted a thorough search of multiple academic databases, including Web of Science (WoS) and ScienceDirect. The search was performed using a combination of keywords and phrases related to smart cities, cyber resilience, multimodal analysis, machine learning, data fusion, and real-time monitoring. The researcher also employed Boolean operators to refine the search and ensure the retrieval of pertinent articles. The search was limited to peer-reviewed journal articles, conference papers, and reputable industry reports published within the last decade to capture the most recent advancements in the field.

("multimodal analysis" OR "multimodal data" OR "multi-modal" OR "multimodal modeling") AND ("smart city" OR "smart cities" OR "urban" OR "sustainable").

Selection Process

The selection criteria for this systematic review were carefully defined to include studies that use multimodal analytical techniques to enhance cyber resilience in smart cities. Articles selected include if they addressed the use of multimodal analytical techniques for enhancing cyber resilience in smart cities with a focus on integrating multiple data types, like text, audio, video, and sensors, and applying analytical methods for detecting, preventing, or mitigating cyber threats. Exclusions were made for studies which focused on single-modal analysis, unrelated to urban cyber resilience, or lacking empirical data and rigorous methodology. Studies were selected from peer-reviewed journals or reputable conferences to ensure academic rigor. A methodological quality assessment, considering the clarity of objectives and robustness of methods, identified high-quality articles that contribute valuable insights into multimodal analysis for cyber resilience

Quality Assessment

To ensure the validity of the findings, a thorough quality assessment of the selected studies was conducted, based on criteria such as research objective clarity, methodological robustness, and the reliability and validity of data sources and analytical techniques. Studies were evaluated based on research design transparency, data collection appropriateness, and data analysis robustness. Those aligning well between objectives, methods, and findings scored higher, with consistency and replicability being key quality indicators. Studies that failed to meet these standards were excluded to maintain review integrity. This stringent assessment ensured that only high-quality studies were included, providing credible insights into applying multimodal analytical techniques to enhance cyber resilience in smart cities.

Data Coding

Finally, the relevant information from the 27 selected articles was systematically coded using a Microsoft Excel spreadsheet, which included: (a) author(s), (b) year of publication, (c) data source used for multimodal analysis, (d) multimodal techniques used, (e) domain, (f) how directly multimodal analytics are used to safeguard smart city infrastructure, (g) challenges, (h) future research, and (i) how multimodal analysis is used to improve cyber resilience of smart cities. To ensure reliability and accuracy, a second coder cross-checked the coding results. Any discrepancies between the two coders were resolved through discussion, resulting in full agreement on the coded information. Figure 1 shows the flowchart of the systematic review process adapted from (Kitchenham & Charters, 2007) that illustrates the steps followed in the systematic review and coding process as described in the methodology section.

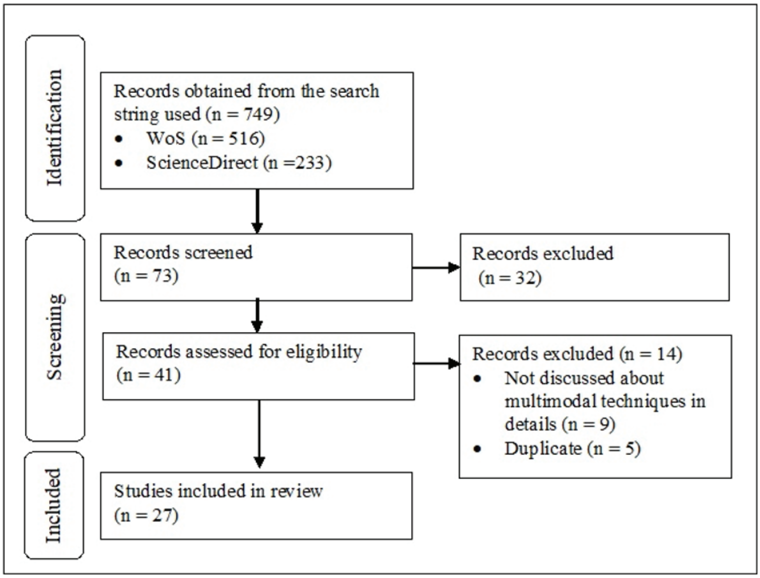

Figure 1. The Systematic Review Process Flowchart

RESULTS AND DISCUSSION

The systematic review included 27 articles on multimodal analytical techniques for enhancing cyber resilience in smart cities. Many studies utilized remote sensing imagery, like satellite and aerial images, combined with ground-level data, such as street view images and sensor data, to improve urban analytics and infrastructure management. For instance, Suel et al. (2021) and Srivastava et al. (2019) integrated satellite and street-level imagery with socioeconomic data to measure urban inequalities and land use. Similarly, Su et al. (2024) and N. Zhang et al. (2022) used high-resolution remote sensing images with Point of Interest (POI) data and building footprints to identify urban functions.

Several studies focused on urban planning and infrastructure management through multimodal data fusion. Yan et al. (2024) and Taubenböck et al. (2022) developed models for urban land use mapping and urbanization measurement, essential for sustainable development. Real-time monitoring and event detection are other key applications. Barthélemy et al. (2019) and Tang et al. (2024) employed video analytics and cross-modal event detection for traffic monitoring and urban safety, while Sharma et al. (2024) used multimodal data and federated learning for fire detection.

High computational complexity and resource demands are frequently cited challenges. Ma et al. (2020) and Qin (2023) highlighted the need for efficient algorithms and models to manage large-scale data. Safety and security applications are also significant, with Kumari et al. (2020) addressing cyberbullying detection using text and images, and Rasouli et al. (2022) developing pedestrian action prediction models for intelligent driving. Future research should focus on improving data integration and model robustness, with Carneiro et al. (2021) and Su et al. (2024) suggesting automation and integration of diverse data types to enhance multimodal fusion techniques.

Appendix A summarizes the main findings on multimodal techniques from the analyzed articles, illustrating their application in various domains and how directly multimodal analytics are used to safeguard smart city infrastructure.

DISCUSSION ON CHALLENGES AND LIMITATIONS

The similarities and differences in the challenges and limitations of dealing with multimodal data are summarized as found in the reviewed articles:

1.Integration and Fusion of Diverse Data Sources
Integrating and fusing diverse data sources is a significant challenge across many studies. For instance, Cheng et al. (2023) and Xiao et al. (2023) discuss the difficulties in combining electro-optical, infrared, radio frequency, laser sensors, LiDAR, radar, and visual data for accurate object detection and tracking. Similarly, Su et al. (2024) and Zhang et al. (2022) highlight challenges in integrating remote sensing images (RSIs) with POI data due to semantic discrepancies and noise from equipment failures. Sharma et al. (2024) and Raptis et al. (2023) emphasize the complexity of integrating heterogeneous data sources while ensuring data privacy and security. These complexities demand advanced algorithms and robust frameworks to achieve effective

integration.

2.Computational Complexity and Real-Time Processing

High computational complexity and the need for real-time processing are pervasive challenges. Ma et al. (2020) and Qin (2023) discuss the difficulties in processing large-scale multimodal data efficiently, especially for applications requiring real-time performance. Barthélemy et al. (2019) and Das et al. (2024) address the real-time processing demands in smart city applications, where latency and computational resources are critical constraints. Wang et al. (2022) also highlights the resource-intensive nature of processing complex CAD models in additive manufacturing, pointing to the need for optimized algorithms.

3.Variability and Noise in Data

Managing variability and noise in multimodal data is a common challenge. Su et al. (2024) and Zhang et al. (2022) struggle with noise from cloud occlusion and data inconsistency, which impacts the robustness of deep learning models. Yang et al. (2023) and Xia et al. (2023) face difficulties in managing noisy and redundant data in applications such as defect detection and weed resistance assessment. Ensuring the accuracy and reliability of predictions in the presence of such noise is a critical issue that needs to be addressed.

4.Data Resolution and Quality

Data resolution and quality pose significant challenges in multimodal analytics. Srivastava et al. (2019) and Suel et al. (2021) discuss the variability in data resolution and the difficulty in maintaining consistency across different sources. High-resolution, high-quality data is crucial for applications like urban planning and environmental monitoring, where precise measurements are necessary for effective decision-making. The need to manage high-resolution data while ensuring its accuracy and consistency is a recurring theme.

5.Ensuring Accuracy and Robustness

Ensuring the accuracy and robustness of multimodal data fusion techniques is a critical concern. Carneiro et al. (2021) and Zhang et al. (2024) emphasize developing reliable predictive models that handle large-scale data and provide accurate recommendations. Taubenböck et al.

(2022) and Yan et al. (2024) focus on the need for accurate urbanization metrics and land use maps, which are essential for sustainable urban development and planning. The robustness of these models against noise and data inconsistencies is crucial for their successful application.

6.Scalability, System Integration, and Future Research

Scalability and system integration are major concerns for implementing multimodal data fusion frameworks. Raptis et al. (2023) and Barthélemy et al. (2019) emphasize the need for scalable solutions that handle large volumes of data and integrate seamlessly with existing urban infrastructure. Ensuring these systems remain reliable and efficient as they scale is critical. Future research directions often focus on enhancing data integration, improving model robustness, and expanding applications. Carneiro et al. (2021) and Su et al. (2024) suggest further automation and integration of diverse data types, while Yan et al. (2024) and Wang et al. (2022) recommend exploring additional data modalities and refining fusion techniques to enhance accuracy across various contexts.

Potential Utilizations of Multimodal Analysis to Improve Cyber Resilience

Multimodal analysis can significantly enhance the cyber resilience of smart cities by integrating various data sources and leveraging advanced analytical techniques. There are several potential utilizations of multimodal analysis to improve the resilience of smart cities against cyber attacks:

1.Enhanced Situational Awareness

•Real-time Monitoring

Techniques used for real-time monitoring and management of urban infrastructure (Fuentes Reyes et al., 2023), urban mapping (Taubenböck et al., 2022), and land use (Yan et al., 2024) can be adapted to monitor critical cyber-physical systems within a smart city. Real-time monitoring enables the detection of anomalies or irregularities that may indicate a cyber attack, allowing for immediate response. For example, automated building inspections can help in quickly identifying structural vulnerabilities (Yang et al., 2023), which, if left unchecked, could be exploited in cyber-physical attacks.

•Enhanced Data Quality

Reliable and high-quality data from IoT sensors (Das et al., 2024) can significantly improve the detection and response to cyber attacks. High-quality data ensures that the information used for threat detection is accurate and timely, reducing the chances of false positives and negatives. Similarly, accurate urban mapping data (Taubenböck et al., 2022) can provide detailed insights into the urban environment, enhancing the ability to detect and respond to threats against urban infrastructure.

•Proactive Decision Making

By integrating multimodal data, city planners and security teams can make informed decisions to enhance the overall security posture of the city. Detailed urbanization patterns (Taubenböck et al., 2022) support better-informed decisions for preventing and mitigating cyber attacks. Furthermore, timely detection and repair of defects prevent further deterioration, maintaining the robustness of urban infrastructure against potential cyber-physical threats (Yang et al., 2023). Up-to-date and accurate geographic data aids in making informed decisions to safeguard urban infrastructure against potential cyber threats (Zhang et al., 2022).

•Optimized Resource Utilization

Efficient data processing and sensor management in IoT systems (Das et al., 2024) can reduce the load on cybersecurity systems, allowing them to function more effectively. By optimizing resource utilization, cities can ensure that their cybersecurity measures are both effective and sustainable, even as the volume of data grows. Findings related to urban structure optimization (Fuentes Reyes et al., 2023) can be leveraged to design resilient urban infrastructures that are less vulnerable to cyber attacks. Ensuring that critical infrastructure components are physically secure and have redundant systems in place can mitigate the impact of potential attacks. Additionally, resilient urban planning can involve the strategic placement of sensors and other monitoring devices to maximize coverage and minimize blind spots, further enhancing the city's ability to detect and respond to cyber threats.

2.Data Fusion and Integration

•Comprehensive Data Analysis

Studies that focus on integrating and aligning multimodal data sources, such as those by Su et al. (2024) and Zhang et al. (2022), can be applied to combine cybersecurity data (e.g., network traffic, system logs) with physical infrastructure data (e.g., sensor readings, surveillance footage). This comprehensive view allows for a more holistic analysis of potential threats, enhancing the ability to detect sophisticated cyber attacks that may span multiple domains. For instance, combining network traffic data with surveillance footage can help identify coordinated attacks that involve both cyber and physical components.

•Improved Detection Algorithms

Advanced data fusion techniques, as discussed by Yu et al. (2023) and Carneiro et al. (2021), can enhance the accuracy and efficiency of cyber threat detection algorithms by incorporating diverse data types and sources. By leveraging the strengths of different data modalities, such as the high resolution of images and the temporal resolution of sensor data, detection algorithms can become more robust against various types of attacks. This can lead to earlier detection and more accurate identification of threats, reducing the potential damage from cyber attacks.

3.Simulation and Training

•Synthetic Data Generation

The ability to generate synthetic data for training models, as demonstrated by Fuentes Reyes et al. (2023), can be extended to create realistic cyber attack scenarios. This can help in training machine learning models to detect and respond to various types of cyber threats, improving their effectiveness in real-world situations. Synthetic data can simulate a wide range of attack vectors and conditions, providing a comprehensive training environment that prepares cybersecurity systems for diverse threats.

•Domain Adaptation

Techniques used for domain adaptation in multimodal analysis, such as those by Carneiro et al. (2021) and Zhang et al. (2024), can be applied to adapt cybersecurity models to different environments or conditions. This improves the robustness and reliability of these models, making them more effective across various scenarios.

4.Infrastructure Optimization

Findings related to urban structure optimization (Fuentes Reyes et al., 2023) can be leveraged to design resilient urban infrastructures that are less vulnerable to cyber attacks. Ensuring that critical infrastructure components are physically secure and have redundant systems in place can mitigate the impact of potential attacks. Additionally, resilient urban planning can involve the strategic placement of sensors and other monitoring devices to maximize coverage and minimize blind spots, further enhancing the city's ability to detect and respond to cyber threats.

By utilizing these multimodal analysis techniques, smart cities can improve their resilience against cyber attacks, ensuring the safety and security of urban infrastructure and residents. These approaches not only enhance detection and response capabilities but also support proactive measures and strategic planning to mitigate risks and vulnerabilities.

CONCLUSION

This systematic review highlights the crucial role of multimodal analysis in enhancing the cyber resilience of smart cities. Smart cities managed to achieve increased situational awareness, proactive decision-making, and efficient resource use through the integration of a wide variety of data sources and the application of modern analytical methodologies. The studies reviewed, emphasize the successful implementations of multimodal data fusion, real-time monitoring, and machine learning techniques, which have the potential to detect and mitigate cyber threats. This study identifies the key benefits and challenges of multimodal techniques and outlining best practices and technological advancements in the field. Then critical challenges of data integration complexity, high computational demands, and ensuring data quality have been recognizes, providing a clear agenda for future research by recommending the development of scalable and robust multimodal frameworks capable of adapting to the dynamic nature of urban environments and evolving cyber threats. By overcoming these difficulties, multimodal analysis has the potential to make a significant influence on the safety and capacity to recover of smart cities, thereby ensuring the smart cities' continued viability and protection in the era of digital technology.

FUNDING

The research presented in this paper is a component of the broader project on the Formulation of a Multi-Modal Machine Learning Model for Botnet Detection in Wireless Sensor Networks. This project has been partially funded by the Universiti Tenaga Nasional BOLD Grant for the year 2023.

AUTHOR CONTRIBUTION

Zul-Azri Ibrahim: This author is in charge of gathering relevant data and performing a detailed analysis. He conducted the systematic literature review, collected information on multimodal analytical techniques, and evaluated their application in enhancing cyber resilience in smart cities.
Saiful Adli Ismail: This author is responsible for the primary conceptualization of the research framework and the overall direction of the paper. He identified the key objectives and scope of the study, focusing on the integration of multimodal strategies and their effectiveness in improving urban resilience against cyber threats. Fiza Abdul Rahim: This author focused on writing and refining the paper's content. She ensured clarity, coherence, and academic rigor in presenting the research findings, discussing the challenges and limitations, and proposing future research directions.

REFERENCES

Ahmadi-Assalemi, G., Al-Khateeb, H., Epiphaniou, G., & Maple, C. (2020). Cyber resilience and incident response in smart cities: A systematic literature review. *Smart Cities, 3*(3), 894–927. https://doi.org/10.3390/smartcities3030046

Almeida, F. (2023). Prospects of cybersecurity in smart cities. *Future*

Internet, 15(9), 285. https://doi.org/10.3390/fi15090285.

Barthélemy, J., Verstaevel, N., & Forehead, H. (2019). Edge-computing video analytics for real-time traffic monitoring in a smart city. *Sensors, 19*(9), 2048. https://doi.org/10.3390/s19092048

Bastos, D., Fernández-Caballero, A., Pereira, A., & Rocha, N. P. (2022). Smart city applications to promote citizen participation in city management and governance: A systematic review. *Informatics, 9*(4), 89. https://doi.org/10.3390/informatics9040089.

Carneiro, D., Amaral, A., Carvalho, M., & Barreto, L. (2021). An anthropocentric and enhanced predictive approach to smart city management. *Smart Cities, 4*(4), 1366–1390. https://doi.org/10.3390/smartcities4040072.

Chaudhuri, A., & Bozkus Kahyaoglu, S. (2023). Cybersecurity assurance in smart cities: A risk management perspective. *EDPACS, 67*(4), 1–22. https://doi.org/10.1080/07366981.2023.2165293

Cheng, P., Xiong, Z., Bao, Y., Zhuang, P., Zhang, Y., Blasch, E., & Chen, G. (2023). A deep learning-enhanced multi-modal sensing platform for robust human object detection and tracking in challenging environments. *Electronics (Switzerland), 12*(16), 3423. https://doi.org/10.3390/electronics12163423

Das, A., Singh, N., & Chakraborty, S. (2024). UniPreCIS: A data preprocessing solution for collocated services on shared IoT. *Future Generation Computer Systems,* 153, 543–557. https://doi.org/10.1016/j.future.2023.11.029

Fuentes Reyes, M., Xie, Y., Yuan, X., d'Angelo, P., Kurz, F., Cerra, D., & Tian, J. (2023). A 2D/3D multimodal data simulation approach with applications on urban semantic segmentation, building extraction and change detection. *ISPRS Journal of Photogrammetry and Remote Sensing,* 205, 74–97. https://doi.org/10.1016/j.isprsjprs.2023.09.013

Gracias, J. S., Parnell, G. S., Specking, E., Pohl, E. A., & Buchanan, R. (2023). Smart cities—A structured literature review. *Smart Cities, 6*(4), 1719–1743. https://doi.org/10.3390/smartcities6040080.

Juliana, J., & Arafah, S. (2018). The multimodal analysis of advertising tagline "Tolak Angin Sidomuncul" through systemic functional linguistics approach. *Journal MELT (Medium for English Language Teaching), 3*(2), 127–137.

Kalinin, M., Krundyshev, V., & Zegzhda, P. (2021). Cybersecurity risk assessment in smart city infrastructures. *Machines, 9*(4). https://doi.org/10.3390/machines9040078.

Kim, K., Alshenaifi, I. M., Ramachandran, S., Kim, J., Zia, T., & Almorjan, A. (2023). Cybersecurity and cyber forensics for smart cities: A comprehensive literature review and survey. *Sensors, 23*(7), 3681. https://doi.org/10.3390/s23073681.

Kitchenham, B., & Charters, S. (2007). *Guidelines for performing Systematic Literature Reviews in Software Engineering.*

Kohout, J., Skarda, C., Shcherbin, K., Kopp, M., & Brabec, J. (2021). A framework for comprehensible multi-modal detection of cyber threats. *CoRR,* abs/2111.05764. https://arxiv.org/abs/2111.05764

Kumari, K., Singh, J. P., Dwivedi, Y. K., & Rana, N. P. (2020). Towards Cyberbullying-free social media in smart cities: a unified multi-modal approach. *Soft Computing, 24*(15), 11059–11070. https://doi.org/10.1007/s00500-019-04550-x.

Ma, Q., Nie, Y., Song, J., & Zhang, T. (2020). Multimodal data processing framework for smart city: A positional-attention based deep learning approach. *IEEE Access, 8*, 215505–215515. https://doi.org/10.1109/ACCESS.2020.3041447

Qin, X. (2023). Traffic flow prediction based on two-channel multi-modal fusion of MCB and attention. *IEEE Access, 11*, 58745–58753. https://doi.org/10.1109/ACCESS.2023.3280068

Raptis, T. P., Cicconetti, C., Falelakis, M., Kalogiannis, G., Kanellos, T., & Lobo, T. P. (2023). Engineering resource-efficient data management for smart cities with Apache Kafka. *Future Internet, 15*(2), 43. https://doi.org/10.3390/fi15020043

Rasouli, A., Yau, T., Rohani, M., & Luo, J. (2022). Multi-modal hybrid architecture for pedestrian action prediction. *IEEE Intelligent Vehicles Symposium, Proceedings*, 2022-June, 91–97. https://doi.org/10.1109/IV51971.2022.9827055.

Sharma, A., Kumar, R., Kansal, I., Popli, R., Khullar, V., Verma, J., & Kumar, S. (2024). Fire detection in urban areas using multimodal data and federated learning. *Fire, 7*(4), 104. https://doi.org/10.3390/fire7040104

Srivastava, S., Vargas-Muñoz, J. E., & Tuia, D. (2019). Understanding urban landuse from the above and ground perspectives: A deep learning, multimodal solution. *Remote Sensing of Environment, 228*, 129–143. https://doi.org/10.1016/j.rse.2019.04.014

Su, C., Hu, X., Meng, Q., Zhang, L., Shi, W., & Zhao, M. (2024). A multimodal fusion framework for urban scene understanding and functional identification using geospatial data. *International Journal of Applied Earth Observation and Geoinformation, 127*, 103696. https://doi.org/10.1016/j.jag.2024.103696

Suel, E., Bhatt, S., Brauer, M., Flaxman, S., & Ezzati, M. (2021). Multimodal deep learning from satellite and street-level imagery for measuring income, overcrowding, and environmental deprivation in urban areas. *Remote Sensing of Environment, 257,* 112339. https://doi.org/10.1016/j.rse.2021.112339

Tang, H., Hu, Y., Wang, Y., Zhang, S., Xu, M., Zhu, J., & Zheng, Q. (2024). Listen as you wish: Fusion of audio and text for cross-modal event detection in smart cities. *Information Fusion, 110*, 102460. https://doi.org/10.1016/j.inffus.2024.102460

Taubenböck, H., Droin, A., Standfuß, I., Dosch, F., Sander, N., Milbert, A., Eichfuss, S., & Wurm, M. (2022). To be, or not to be 'urban'? A multi-modal method for the differentiated measurement of the degree of urbanization. *Computers, Environment and Urban Systems, 95*, 101830. https://doi.org/10.1016/j.compenvurbsys.2022.101830

Wang, K., Song, Y., Huang, Z., Sun, Y., Xu, J., & Zhang, S. (2022). Additive manufacturing energy consumption measurement and prediction in fabricating lattice structure based on recallable multimodal fusion network. *Measurement: Journal of the International Measurement Confederation, 196*, 111215. https://doi.org/10.1016/j.measurement.2022.111215.

Xia, F., Lou, Z., Sun, D., Li, H., & Quan, L. (2023). Weed resistance assessment through airborne multimodal data fusion and deep learning: A novel approach towards sustainable agriculture. *International Journal of Applied Earth Observation and Geoinformation, 120,* 103352. https://doi.org/10.1016/j.jag.2023.103352

Xiao, Y., Liu, Y., Luan, K., Cheng, Y., Chen, X., & Lu, H. (2023). Deep LiDAR-radar-visual fusion for object detection in urban environments. *Remote Sensing, 15*(18), 4433. https://doi.org/10.3390/rs15184433

Xue, F., Lu, W., Chen, K., & Webster, C. J. (2019). BIM reconstruction from 3D point clouds: A semantic registration approach based on multimodal optimization and architectural design knowledge. *Advanced Engineering Informatics, 42*, 100965. https://doi.org/10.1016/j.aei.2019.100965

Yan, X., Jiang, Z., Luo, P., Wu, H., Dong, A., Mao, F., Wang, Z., Liu, H., & Yao, Y. (2024). A multimodal data fusion model for accurate and interpretable urban land use mapping with uncertainty analysis. *International Journal of Applied Earth Observation and Geoinformation, 129,* 103805. https://doi.org/10.1016/j.jag.2024.103805

Yang, X., Guo, R., & Li, H. (2023). Comparison of multimodal RGB-thermal fusion techniques for exterior wall multi-defect detection. *Journal of Infrastructure Intelligence and Resilience, 2*(2), 100029. https://doi.org/10.1016/j.iintel.2023.100029.

Yu, M., Xu, H., Zhou, F., Xu, S., & Yin, H. (2023). A Deep-Learning-Based Multimodal Data Fusion Framework for Urban Region Function Recognition. *ISPRS International Journal of Geo-Information, 12*(12). https://doi.org/10.3390/ijgi12120468.

Yuan, Q., & Mohd Shafri, H. Z. (2022). Multi-Modal Feature Fusion Network with Adaptive Center Point Detector for Building Instance Extraction. *Remote Sensing, 14*(19). https://doi.org/10.3390/rs14194920.

Zhang, G., Li, H., Li, S., Wang, B., & Ding, Z. (2024). MMKG-PAR:

Multi-Modal Knowledge Graphs-Based Personalized Attraction Recommendation. *Sustainability (Switzerland), 16*(5), 1–22. https://doi.org/10.3390/su16052211.

Zhang, J., Liu, X., Liao, W., & Li, X. (2022). Deep-learning generation of POI data with scene images. *ISPRS Journal of Photogrammetry and Remote Sensing, 188,* 201–219. https://doi.org/10.1016/j.isprsjprs.2022.04.004

Zhang, N., Wang, Y., Wang, X., & Yu, P. (2022). A multi-modal fusion network guided by feature co-occurrence for urban region function recognition. *IEICE Transactions on Information and Systems, E105D*(10), 1769–1779. https://doi.org/10.1587/transinf.2021EDP7230.