Vol. 6, No. 3, 2024

Oleh Basystiuk[1], Zoriana Rybchak[2], Iryna Zavushchak[3], Uliana Marikutsa[4]

[1] Department of Artificial Intelligence, Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: oleh.a.basystiuk@lpnu.ua, ORCID 0000-0003-0064-6584
[2] Department of Information Systems and Networks. Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: zoriana.l.rybchak@lpnu.ua, ORCID 0000-0002-5986-4618
[3] Department of Information Systems and Networks. Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: iryna.i.zavushchak@lpnu.ua, ORCID 0000-0002-5371-8775
[4] Department of Virtual Reality Systems. Lviv Polytechnic National University, 12, S. Bandery str., Lviv, Ukraine, E-mail: uliana.b.marikutsa@lpnu.ua, ORCID 0000-0002-9514-7413

EVALUATION OF MULTIMODAL DATA SYNCHRONIZATION TOOLS

Received: November 04, 2024 / Revised: November 20, 2024 / Accepted: November 25, 2024

https://doi.org/10.23939/cds2024.03.104

Abstract. The constant growth of data volumes requires the development of effective methods for managing, processing, and storing information. Additionally, it is advisable to apply multimodal approaches for knowledge aggregation to extract additional knowledge. Usually, the problem of efficient processing of multimodal data is associated with high-quality data preprocessing. One of the most critical preprocessing steps is synchronizing multimodal data streams to analyze complex interactions in different data types. In this article, we evaluate existing approaches to synchronization, focusing on strategies based on real-time classifiers, which are based on comprehensive platforms for data integration and management. After the synchronization of multimodal sets, the key stage is data fusion, data identification in different channels, such as text, video, and audio. The results demonstrate the feasibility of the proposed synchronization approach for revealing subtle relationships between various data sets. An architectural solution was also suggested to integrate the proposed method into existing multimodal data processing pipelines. This work contributes to developing synchronization tools for multimodal data analysis in dynamic real-world scenarios.

Keywords: multimodal data, data analysis, synchronization tools, real-time application, machine learning.

Introduction

Multimodal data are data sets that contain multiple types of information, such as text, images, audio, video, and other formats. In natural language processing (NLP), multimodal data often includes text descriptions, photos, videos, or audio recordings related to specific objects or phenomena being analyzed. The complexity of processing such data is related to their diverse formats and characteristics, which requires specialized methods and techniques for practical analysis and integration [1].

Artificial intelligence (AI) plays a crucial role in solving the problems of analyzing multimodal data. AI technologies, such as speech recognition, are central to this field. Speech recognition systems analyze and integrate voice, audio, and text data, which is essential to multimodal data processing. AI capabilities go beyond recognition, enabling automation and optimization in various industries, from natural language processing to autonomous driving and facial recognition [2, 3]. AI training processes can be compared to traditional modeling methods such as decision trees. However, these processes are designed for autonomous learning. After training, the system's performance is tested on new data sets. This iterative approach ensures continuous improvement in accuracy and reliability.

Machine translation is an example of AI applied to multimodal natural language tasks. These systems mimic the work of a translator, requiring an understanding of grammatical structures and contex-

tual nuances. Their performance mainly depends on domain-specific knowledge, as terminology and contextual constraints vary across domains [4, 5]. Unlike single-word translation, machine translation focuses on phrases or sentence structures to effectively convey complex ideas. A unique feature of machine translation is its ability to process data of varying lengths, achieved using recurrent neural networks (RNNs). ANNs are well suited for sequential processing of data, enabling accurate translation and other NLP tasks.

The field of multimodal data recognition is rapidly evolving due to the development of deep learning and neural network architectures. Using transformational approaches for preprocessing and analyzing multimodal data improves the accuracy and relevance of system results. Integrated multimodal models simultaneously process different data formats – text, audio, images, and video – contributing to a holistic understanding of information. Fusion architectures combine data streams at different processing levels, and attention mechanisms prioritize critical modalities during analysis to obtain more effective results. Special emphasis is placed on developing quality metrics that ensure semantic accuracy, grammatical correctness, and naturalness of system results, further expanding the industry's capabilities [6].

Problem Statement

The main goal of this project is to develop an intelligent system for recognizing multimodal data in information systems, detecting anomalies, and filtering irrelevant data to improve the overall accuracy of the data. This will be achieved by creating an iterative data processing system consisting of an agent module for monomodal data that will classify the data type and a second model to search for anomalies. Finally, we will use data fusion techniques and evaluate the effectiveness of one of three fusion methods – late fusion, early fusion, and hybrid fusion – for detecting and analyzing multimodal data (including text, metadata, and input images) [7].

Using these fusion techniques, the system seeks to identify critical anomalies in the data, ensuring that only relevant and accurate data is retained [8]. The improved data quality will enhance and optimize the synchronization process in the following data processing stage, supporting more efficient decision-making and system performance. The project aims to improve the efficiency of multimodal data integration and anomaly detection to increase information systems' reliability and accuracy [9].

Unimodal data (single modality) [10] and processing systems that handle text, images, audio, or video, focusing solely on understanding and analyzing one input type. Unimodal systems are more straightforward in architecture and are tailored to specific data types, while multimodal systems are more complex, requiring synchronization and integration of heterogeneous data sources. The initial problem of the multimodal system is data integration and synchronization of data from multiple modalities simultaneously [11].

The scope of multimodal systems is narrower, addressing multi-stream challenges. In contrast, multimodal systems aim to provide a more comprehensive understanding by leveraging the interplay between different and analyze, such as combining text, images, and audio to derive richer insights. A general description of unimodal and multimodal processing systems is presented in Fig. 1.

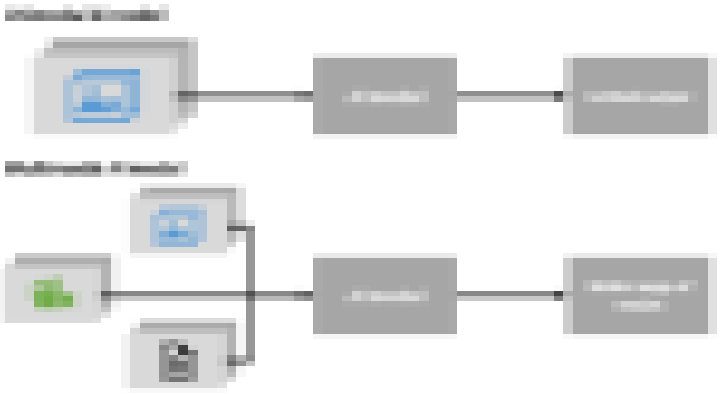

Fig. 1. Schematic representation of multimodal and unimodal handling process

The project aims to evaluate and develop an approach for effective synchronization of multimodal data streams, addressing challenges in processing heterogeneous datasets from diverse sources. The outcomes of this research are expected to contribute to the advancement of synchronization methods by ensuring precise alignment across modalities such as audio, video, and sensor data. This will improve data reliability, coherence, and utility in subsequent analytical processes.

Key outcomes include:

1. With overview synchronization tools, the project seeks to enhance the precision of multimodal data alignment, reduce inconsistencies, and ensure that information from various streams is temporally and contextually coherent.

2. Provide architectural solutions that enable seamless integration of multimodal synchronization methods into existing systems. This includes tools capable of handling the variability of real-world data scenarios while maintaining robust performance.

3. Evaluated existing and proposed tools to process large volumes of heterogeneous data effectively. They will support scalable and adaptable approaches for synchronizing diverse modalities, from low-sampling-rate sensors to high-frequency audio and video streams. The tools leverage machine learning to maximize the information gained from multimodal observations and improve the overall aggregation accuracy.

To sum up, this work highlights the importance of developing advanced multimodal data synchronization systems. The synchronization technique presented in this paper is flexible and applicable to various data streams, offering opportunities to leverage commercial off-the-shelf tools for research purposes. By enabling precise synchronization of multimodal data, the technique expands the usability of systems not initially designed for research into scientific and methodological studies.

This development creates significant potential for applications in psychology, game studies, cognitive sciences, and experiential-based multimodal research. Addressing current limitations and leveraging modern techniques will provide a foundation for creating reliable, efficient, and unified solutions applicable to natural language processing, cognitive sciences, and user behavior analysis.

Review of Modern Synchronization Approaches

The problem of data synchronization is a critical issue in knowledge processing systems that require accurate integration of data from multiple sources, especially in wireless sensor networks and multimodal data fusion scenarios [12, 13]. Time synchronization is a well-established and extensively studied problem, particularly in wireless sensor networks where subsequent sensor data fusion is required. Traditional solutions often involve aligning the clocks of physically distributed nodes by exchanging specific synchronization messages. This network-wide synchronization method is standard in systems with bidirectional communication channels. Modern data synchronization tools are based on two main approaches:

1. Hardware-based methods typically achieve synchronization by aligning the clocks of physically distributed nodes through dedicated signaling mechanisms [14]. These methods often rely on direct communication between devices, exchanging specific messages or signals to ensure precise alignment. These methods are accurate and reliable, usually achieving sub-millisecond accuracy [15]. However, they are costly, require specialized equipment, and may need help in scalability for large networks or multimodal systems.

2. Software-based methods offer a more flexible alternative, focusing on relative synchronization without modifying source clocks [16]. These methods embed data signatures or events within data streams, leveraging their inherent characteristics for alignment. The main idea of this approach is to search for the correlation of sensor data streams to identify shared events, such as detecting shared movements across devices carried by the same person or using acoustic signals for indoor localization [17]. These methods provide scalable and adaptable solutions but may introduce errors in event detection.

3. Neural network-based methods, based on limitations of existing methods, this research introduces a novel software-based synchronization approach utilizing a neural network-based encoder. This method converts diverse multimodal data into a unified one-dimensional vector representation, enabling precise

stream alignment. Leverages the power of machine learning to align diverse data streams with precision, bridging the gap between traditional hardware and software methods.

Modern synchronization approaches have made significant strides in aligning multimodal data streams. While hardware-based methods remain the gold standard for precision, their high cost and complexity often make them impractical. Although less precise, software-based methods provide flexibility and cost-efficiency [18]. The proposed neural network-based approach combines the strengths of software solutions with enhanced accuracy, offering a promising path forward for multimodal data synchronization.

Main Material Presentation

Synchronization of multimodal recordings is crucial for ensuring accurate analysis and knowledge synthesis from various data streams and monitoring systems. Accuracy issues arise due to variability in sampling rates, clock drift, and transmission delays of data sets from different devices. Both hardware and software solutions address the problem of synchronization of multimodal data, each with advantages and disadvantages. Hardware triggers are ideal for applications requiring the highest accuracy and reliability, while software systems offer flexibility, scalability, and cost-effectiveness for dynamic and distributed systems. However, future developments will focus on creating hybrid approaches that combine the accuracy of hardware solutions with the flexibility of software methods. Typically, these solutions are based on using neural networks to reduce the dimensionality of the data, which accordingly provides feature extraction in each of the unimodal sets and reliable synchronization in various research scenarios during further data processing.

In this study, we propose architectural solutions using an autoencoder-based neural network architecture to solve the synchronization problems of multimodal data streams. The method efficiently and accurately aligns different modalities by encoding multidimensional input data into a unified vector representation. Additionally, for better feature detection and extraction, we propose to use attention mechanisms, the effectiveness of which has been evaluated in previous studies [19]. To ensure the system works, it is essential to go through five key stages designed to process various data sources while maintaining reliable synchronization:

1. Maintaining unimodal data from different sources. Each data set, text, image, video, metadata, or other types, is considered a separate independent input stream. Such a modular design provides flexibility in the input and further processing of data sets.
2. Preprocessing of each data source, a mandatory step for all sets, involves normalization, standardization, noise reduction, and, if necessary, dimensionality reduction for extensive multivariate data.
3. Using an autoencoder to transform data into a feature vector, a separate neural network-based model is trained for each modality, which additionally uses feature extraction methods and attention mechanisms to determine similarities in different data sets. The main task of this stage is to capture the main characteristics of the input data and convert them into unique numerical features, further reducing their dimensionality while preserving the integrity of the available information. This ensures that different modalities are transformed into feature spaces amenable to comparison and further synchronization.
4. Fusion and synchronization: A hybrid fusion approach combines the resulting encoded vectors from all modalities. Building on previous research, this method combines early fusion (when data integration occurs during encoding) and late fusion (when integration occurs after individual processing). This combination optimizes the trade-off between accuracy and computational efficiency, allowing for synchronizing multimodal data streams into a single vector representation. See the authors' previous work for a more detailed comparison of the approaches [20].
5. After synchronization, the resulting vectors are stored in a format ready for further tasks such as anomaly detection, classification, or further multimodal analysis of the datasets. The stored data can be provided in JSON formats as an API interface or stored in relational or non-relational databases; the key element is to preserve the temporal dependence and contextual consistency of the data, ensuring the convenience of collecting and using the obtained processing results.

Fig. 2 shows the general concept of the multimodal data processing pipeline, which ensures the operation of the proposed synchronization algorithm based on the autoencoder approach. The alignment process provides accurate time consistency between different data sources, demonstrating the ability of the algorithm to handle complex scenarios accurately. It is also well suited to align multimodal data streams by calculating offsets and synchronized outputs; a more detailed comparison of the accuracy of the different systems will be presented in the results section.

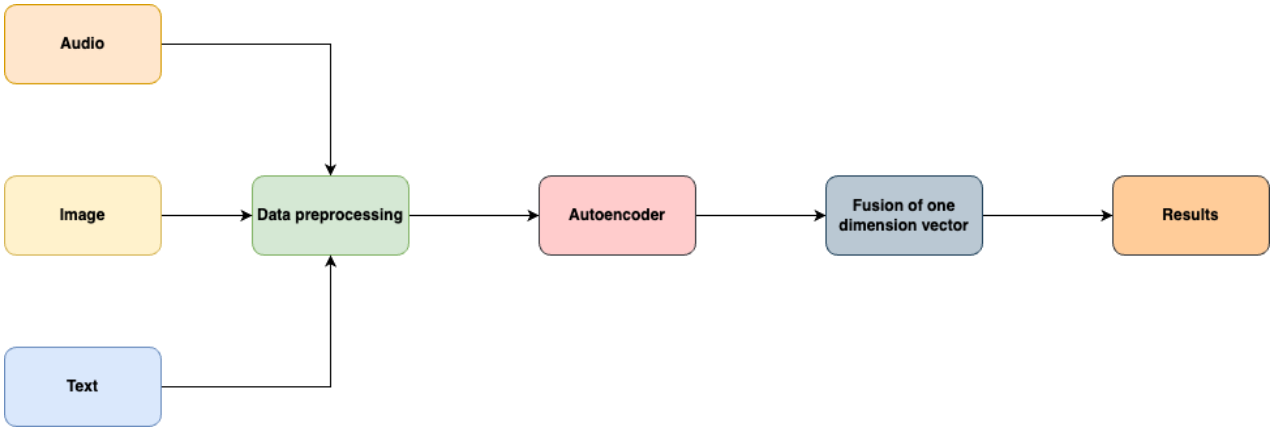

Fig. 2. Multimodal data synchronization approach pipeline

Results and Discussion

This research compares the two most popular synchronization approaches for multimodal data: hardware- and software-based. We propose a novel enhancement to the software-based approach by incorporating a neural network-based encoder. This encoder converts various types of multimodal data into a unified one-dimensional vector, which is then used for synchronization. We aim to improve the performance and accuracy of software-based synchronization by leveraging advanced machine-learning techniques. The evaluation of these synchronization methods is presented in detail in the results section, where we compare the effectiveness of each approach across different modalities.

Table 1

Comparison results of three synchronization method approaches

Modalities	Hardware synchronization accuracy, %	Software synchronization accuracy, %	Proposed synchronization method accuracy, %
Text + Image	98.1	96.2	97.1
Text + Video	98.3	95.5	97.4
Text + Metadata	99.2	94.3	96.7
Metadata + Image	96.6	95.1	97.2
Metadata + Video	98.9	95.7	98.8
Text + Image + Metadata	96.7	93.6	96.3

The evolution of three synchronization approaches (hardware-based, software-based, and the proposed synchronization method) multimodal data across various combinations of modalities, like text, image, video, and metadata, was established and presented in Table 1 and Fig. 3. The results of the evaluation could be described as following points:

1. The hardware synchronization approach utilizes external triggers to synchronize the data streams from different devices. It typically achieves high synchronization accuracy, ranging from 96.6 % to 99.2 %. For example, in the combination of Text + Metadata, hardware synchronization achieved 99.2 %, which was the highest accuracy among the three methods evaluated. However, while it provides accurate synchronization, hardware synchronization requires specific equipment and setups that are often expensive and difficult to implement.

2. The software synchronization approach relies on a computer's internal clock or software-based timestamps to align the data streams. Although it is more flexible and cost-effective than hardware, software synchronization accuracy tends to be slightly lower, with accuracy ranging from 93.6 % to 96.6 %. For instance, Text + Image achieved an accuracy of 96.2 % with software synchronization. While it is a viable solution, it may sometimes offer a different level of precision than hardware-based systems.

3. The proposed synchronization method offers a compromise between hardware and software solutions. It utilizes a software-based approach with a neural network layer application to encode multimodal data into a vector representation that is specifically optimized to enhance synchronization accuracy. With accuracies ranging from 96.3 % to 98.8 %, the proposed method provides synchronization results close to or exceeding those achieved with hardware synchronization. For example, in the case of Text + Video, the proposed method achieved 97.4 % accuracy, only slightly lower than hardware synchronization at 98.3 %.

The proposed software synchronization method delivers the best balance of cost and performance. While hardware-based synchronization can achieve higher accuracy, it is costly and sometimes feasible to implement in most setups. Software-based synchronization, although more affordable, can be less accurate. The proposed method, however, offers a cost-effective alternative that achieves high synchronization accuracy without the need for specialized hardware, making it a practical choice for many research and application scenarios.

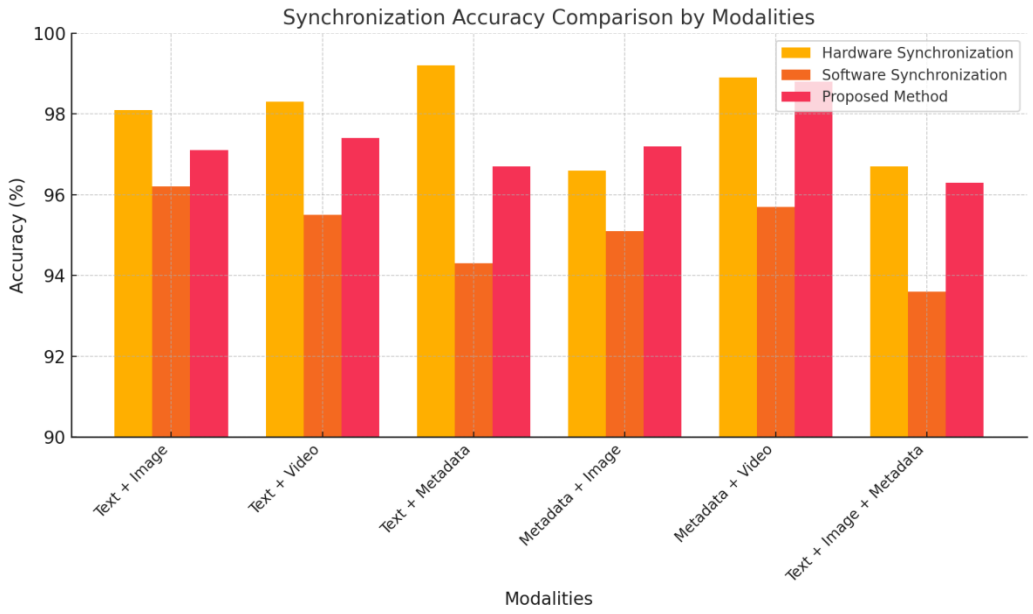

Fig. 3. Comparison of synchronization method approaches

Conclusions

In summary, the research compares multimodal data synchronization tools based on two established approaches – hardware and software synchronization – and proposes a new improvement of the software solution using an autoencoder based on neural networks. The proposed method transforms a wide range of multimodal data into a unified one-dimensional vector, improving synchronization accuracy.

Key project results include a detailed overview of synchronization tools that demonstrate the potential to improve the accuracy of multimodal data reconciliation. Reducing inconsistencies of different streams in time and context is crucial for improving the accuracy of subsequent data analysis.

In addition, the research proposes architectural solutions that facilitate the seamless integration of multimodal synchronization methods into existing systems. These solutions are designed to cope with the variability inherent in real-world data transfer scenarios while ensuring reliable operation with different data types and devices.

The proposed synchronization method uses machine learning techniques to maximize the information obtained from multimodal observations, improving overall aggregation accuracy and allowing for more accurate multimodal analysis. Overall, this research advances the field of multimodal data synchronization, offering valuable insights and practical solutions for improving the reconciliation and integration of diverse data sources in complex real-world applications.

Acknowledgments

The authors would like to thank the Armed Forces of Ukraine for providing security to perform this work. This work was possible only because of the resilience and courage of the Ukrainian Army.

References

[1] Jun, S. Technology Integration and Analysis Using Boosting and Ensemble. J. Open Innov. Technol. Mark. Complex. 2021, 7, 27. https://doi.org/10.3390/joitmc7010027

[2] Chen, Z., Feng X., Zhang S. Emotion detection and face recognition of drivers in autonomous vehicles in IoT platform, Image and Vision Computing, Vol. 128, 2022. https://doi.org/10.1016/j.imavis.2022.104569.

[3] Yih-Shiuan L., Wang C. (2024). "A Cyber-Physical Testbed for IoT Microgrid Design and Validation", *Electronics*, 13, No. 7: 1181. https://doi.org/10.3390/electronics13071181

[4] Havryliuk, M., Kaminskyy, R., Yemets, K., Lisovych, T. (2023). Interactive Information System for Automated Identification of Operator Personnel by Schulte Tables Based on Individual Time Series. In: Hu, Z., Zhang, Q., He, M. (eds) *Advances in Artificial Systems for Logistics Engineering*, Vol. 180. Springer, Cham. DOI: 10.1007/978-3-031-36115-9_34

[5] Basystiuk, O., Melnykova, N. and Rybchak, Z., 2023, June. Machine Learning Methods and Tools for Facial Recognition Based on Multimodal Approach. In *MoMLeT+ DS* (pp. 161–170).

[6] Strubytskyi R., Shakhovska N., Method and models for sentiment analysis and hidden propaganda finding, *Computers in Human Behavior Reports*, Vol. 12. https://doi.org/10.1016/j.chbr.2023.100328.

[7] Dai, Z., Zakka, V. G., Manso, L. J.; Rudorfer, M.; Bernardet, U.; Zumer, J.; Kavakli-Thorne, M. Sensors, Techniques, and Future Trends of Human-Engagement-Enabled Applications: A Review. *Algorithms* 2024, *17*, 560. https://doi.org/10.3390/a17120560

[8] Chen H., Ma H., Chu X., Xue D., Anomaly detection and critical attributes identification for products with multiple operating conditions based on isolation forest, *Advanced Engineering Informatics*, Vol. 46. https://doi.org/10.1016/j.aei.2020.101139.

[9] Havryliuk, M., Hovdysh, N., Tolstyak, Y., Chopyak, V., & Kustra, N. (2023, November). Investigation of PNN Optimization Methods to Improve Classification Performance in Transplantation Medicine. In *IDDM* (pp. 338–345).

[10] Basystiuk O., Melnykova N., Rybchak Z. "Multimodal Learning Analytics: An Overview of the Data Collection Methodology", IEEE 18th International Conference on Computer Science and Information Technologies, Lviv, Ukraine, 2023, pp. 1–4. DOI: 10.1109/CSIT61576.2023.10324177.

[11] Loaiza-Arias, M.; Álvarez-Meza, A.M.; Cárdenas-Peña, D.; Orozco-Gutierrez, Á.Á.; Castellanos-Dominguez, G. Multimodal Explainability Using Class Activation Maps and Canonical Correlation for MI-EEG Deep Learning Classification. *Appl. Sci.*, 2024, *14*, 11208. https://doi.org/10.3390/app142311208

[12] Su, Q.; Yao, Y.; Chen, C.; Chen, B. Generating a 30 m Hourly Land Surface Temperatures Based on Spatial Fusion Model and Machine Learning Algorithm. *Sensors*, 2024, *24*, 7424. https://doi.org/10.3390/s2423742

[13] Yakovyna V., Shakhovska N. "Software failure time series prediction with RBF, GRNN, and LSTM neural networks", Procedia Computer Science 207(4): 837–847. DOI:10.1016/j.procs.2022.09.139.

[14] Paterega, I., Melnykova, N. (2024). Imbalanced data: a comparative analysis of classification enhancements using augmented data. *European Science*, *3*(sge28-03), 54–72. https://doi.org/10.30890/2709-2313.2024-28-00-017.

[15] Basystiuk O., Melnykova N., Rybchak Z. "Multimodal Learning Analytics: An Overview of the Data Collection Methodology", *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)*, Lviv, Ukraine, 2023, pp. 1–4. DOI: 10.1109/CSIT61576.2023.10324177.

[16] Merino-Monge, M., Molina-Cantero, A. J., et al. (2020). An easy-to-use multi-source recording and synchronization software for experimental trials. *IEEE Access*, 8, 200618-200634.

[17] Govindarajan, Y., Ganesan, V. P. A., & Ramesh, D. (2024). Multi-modal biometric authentication: Leveraging shared layer architectures for enhanced security. *arXiv preprint arXiv:2411.02112*..

[18] Muhammad, T. (2022). A Comprehensive Study on Software-Defined Load Balancers: Architectural Flexibility & Application Service Delivery in On-Premises Ecosystems. *International Journal of Computer Science and Technology*, 6(1), 1–24.

[19] Zhaoyang N., Zhong G., Yu H. "A review on the attention mechanism of deep learning", *Neurocomputing*, 452 (2021): 48–62.

[20] Basystiuk O., Melnykova N., Rybchak Z. "Detecting Multimodal Data in Information System", CSIT-2024: Computer Science and Information Technologies, 16–19 October 2024, Lviv, Ukraine.

Олег Басистюк[1], Зоряна Рибчак[2], Ірина Завущак[3], Уляна Марікуца[4]

[1] Кафедра систем штучного інтелекту, Національний університет "Львівська політехніка", вул. С. Бандери, 12, Львів, Україна, E-mail: oleh.a.basystiuk@lpnu.ua, ORCID 0000-0003-0064-6584

[2] Кафедра інформаційних систем та мереж, Національний університет "Львівська політехніка", вул. С. Бандери, 12, Львів, Україна, E-mail: zoriana.l.rybchak@lpnu.ua, ORCID 0000-0002-5986-4618

[3] Кафедра інформаційних систем та мереж, Національний університет "Львівська політехніка", вул. С. Бандери, 12, Львів, Україна, E-mail: iryna.i.zavushchak@lpnu.ua, ORCID 0000-0002-5371-8775

[4] Кафедра систем віртуальної реальності, Національний університет "Львівська політехніка", вул. С. Бандери, 12, Львів, Україна, E-mail: uliana.b.marikutsa@lpnu.ua, ORCID 0000-0002-9514-7413

ОЦІНКА ІНСТРУМЕНТІВ МУЛЬТИМОДАЛЬНОЇ СИНХРОНІЗАЦІЇ ДАНИХ

Отримано: Листопад 04, 2024 / Переглянуто: Листопад 20, 2024 / Прийнято: Листопад 25, 2024

Анотація. Постійне зростання обсягів даних вимагає розроблення ефективних методів управління, опрацювання та зберігання інформації. Крім того, доцільно застосовувати мультимодальні підходи агрегації знань для отримання додаткових знань. Зазвичай проблема ефективного оброблення мультимодальних даних пов'язана з високоякісним попереднім обробленням даних. Одним із найважливіших етапів попереднього оброблення є синхронізація мультимодальних потоків даних для аналізу складних взаємодій у різних типах даних. У статті оцінено відомі підходи до синхронізації. Увагу зосереджено на стратегіях, основаних на класифікаторах реального часу, побудованих на комплексних платформах для інтеграції та управління даними. Після синхронізації мультимодальних наборів ключовими етапами є злиття даних, ідентифікація даних у різних каналах, таких як текст, відео та аудіо. Результати демонструють здійсненність запропонованого підходу синхронізації для виявлення тонких зв'язків між різними наборами даних. Також запропоновано архітектурне рішення для інтеграції запропонованого методу в наявні мультимодальні конвеєри опрацювання даних. Дослідження сприяє розробленню інструментів синхронізації для мультимодального аналізу даних у динамічних сценаріях реального світу.

Ключові слова: мультимодальні дані, аналіз даних, інструменти синхронізації, програма реального часу, машинне навчання.