Journal of Medical and Health Studies (JMHS)
ISSN: 2710-1452
DOI: 10.32996/jmhs
Journal Homepage: www.al-kindipublisher.com/index.php/jmhs

JMHS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

RESEARCH ARTICLE

The progress of smart healthcare based on Multimodal Data fusion technology

Yining Zhang*[1] and Yisu Zhang[2]

[1] *School of Heaith Economics and Management, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China*
[2] *Reading Academy, Nanjing University of Information Science and Technology, Nanjing, Jiangsu, China*

ABSTRACT

Smart healthcare is moving towards precision, personalization, and intelligence. Multimodal data fusion technology, as a core means to break medical data silos and improve clinical decision-making efficiency, has received extensive attention. This paper systematically reviews the mainstream methods and their performance in typical medical tasks from four levels: data level, feature level, decision level, and hybrid fusion. Based on this, the current commonly used medical multimodal datasets and evaluation criteria are introduced, and the experimental results of various fusion methods on public datasets are summarized. Further, this paper analyzes the challenges of existing methods in terms of data heterogeneity, modal absence, interpretability, and privacy protection, explores the deficiencies in the dataset construction and evaluation system, and proposes corresponding solutions. Finally, it looks forward to future research directions, emphasizing the synergy of medical-engineering integration, privacy computing and explainable models to provide references for the in-depth research and application of multimodal fusion technologies in smart healthcare.

KEYWORDS

Smart healthcare; Multimodal data fusion; Data-level fusion; Feature-level fusion; Decision-level fusion; Hybrid fusion

ARTICLE INFORMATION

ACCEPTED: 15 June 2026 **PUBLISHED:** 31 July 2026 **DOI:** 10.32996/jmhs.2026.7.9.12

I. Introduction

In the context of the digital transformation of healthcare, although there has been a massive accumulation of multi-source heterogeneous medical data, barriers between systems have formed serious data silos. Single-modal data can no longer meet the demands of precision diagnosis and treatment. Multimodal data fusion technology, by integrating and analyzing data from different dimensions, has become the key to breaking down barriers and promoting the development of smart healthcare.

From a clinical perspective, the integration of imaging, pathology and electronic medical records can create a comprehensive health profile of patients, significantly improving diagnostic accuracy and reducing the risk of misdiagnosis; Combining genomic and phenotypic data can provide precise decision support for targeted therapy and personalized intervention, facilitating the implementation of precision medicine. From the perspective of industry development, the technology can facilitate the downward flow of high-quality medical resources and enhance the diagnosis and treatment capabilities at the grassroots level through intelligent analysis systems. From the perspective of technology research, the smart healthcare scenario provides a complex application field for data fusion, and the challenges it faces, such as data heterogeneity, modal absence and privacy protection, are driving collaborative innovation in technologies such as feature fusion, federated learning and privacy computing.

To sum up, the research on multimodal data fusion technology is not only a practical need to address the pain points of the medical industry, but also an important support for achieving precise and intelligent medical care, with significant clinical, industrial and academic value.

II、Smart healthcare based on data-level fusion

Data-level fusion is the most fundamental level in a data fusion system, referring to the direct processing and integration of raw data collected by sensors, with the aim of improving the accuracy, completeness, and reliability of the data. In the field of smart healthcare, with the prevalence of the Internet of Things (IoT) and wearable devices, various physiological signals (such as heart rate, body temperature, blood pressure, respiratory rate, etc.) are continuously collected, forming massive, multi-source, heterogeneous data streams. Effective fusion processing of these raw data is the basis for achieving health status perception, early disease warning, and personalized medical services.

Dautov et al. proposed a hierarchical data fusion architecture for smart healthcare, emphasizing the initial fusion of raw sensor data directly at the edge layer of IoT systems, such as smart beds and wearable devices. The study points out that the traditional cloud-based centralized processing model has problems such as high latency, high bandwidth pressure, and high data security risks, while delegating the data-level fusion task to edge devices can significantly improve the system's response speed and reliability. For example, by analyzing real-time changes in patients' body temperature, heart rate and blood pressure, edge devices can quickly identify signs of acute illness and trigger warnings, thus buying valuable time for medical intervention. The study also introduced complex event processing (CEP) technology to support real-time detection of complex patterns in data streams, further enhancing the expressive power and practicality of data-level fusion[i] .

King et al. systematically reviewed the application status of data fusion technology in wearable health monitoring, with particular attention to the implementation of signal-level fusion. The study noted that data-level fusion typically employs algorithms such as weighted averaging, Kalman filtering (KF), particle filtering (PF), etc., for state estimation and noise suppression of raw signals from multiple sensors. For example, in gait analysis and fall detection, human postures and motion trajectories can be estimated more accurately by fusing accelerometer and gyroscope data. The study also highlights that the output of data-level fusion can be used for higher-level fusion tasks, such as feature extraction and decision support, to build a complete hierarchical fusion system [ii].

In summary, the application of data-level fusion in smart healthcare not only enhances the quality and availability of data but also lays a solid foundation for subsequent intelligent analysis and decision-making. With the development of edge computing and low-power communication technologies, data-level fusion is expected to play a key role in more medical scenarios with high real-time requirements.

III、Smart healthcare based on feature-level fusion

Feature-level fusion is an intermediate level in the data fusion system. It refers to extracting features from multi-source data, then combining, transforming or selecting these feature vectors to form a more discriminative joint feature representation, which is then input into the classifier for decision-making. Compared with data-level fusion, feature-level fusion can reduce the data dimension while retaining the key features of the original information; Compared with decision-level fusion, it can more fully explore the intrinsic correlations and complementarities between different modal data, and thus has been widely applied in the field of smart healthcare.

Zasim Uddin et al. proposed a deep learning-based hybrid fusion framework for dermatological diagnosis, which simultaneously achieves feature-level fusion (FLF) and decision-level fusion (DLF). In the featment-level fusion part, the study uses three models - improved DenseNet201, VGG19, and visual converter (ViT) - to extract deep features, fuses multi-model features by point-by-point addition, and inputs shared classification heads for end-to-end classification. Experimental results showed that feature-level fusion achieved accuracies of 99.3%, 92.7%, 86.7%, and 94.5% on the four benchmark datasets of PH2, HAM10000, ISIC 2018, and ISIC 2019, respectively, significantly outperforming a single model. The study fully validated the effectiveness of multi-model feature fusion in improving the performance of dermatological diagnosis [iii].

Rajalingam et al. proposed a technical framework that fuses traditional and hybrid strategies for the problem of multimodal medical image fusion in clinical disease analysis and conducted systematic research and practice on feature-level fusion. In the feature-level fusion section, the research uses core image features such as edges and textures as processing units, and relies on techniques such as curvilinear transform, ridge transform, and rotational wavelet transform to complete feature extraction of multimodal medical images. It achieves effective fusion of feature layers through subband decomposition and geometric feature analysis, and then completes image reconstruction based on the fused features. At the same time, the neural fuzzy algorithm is used to fuse and optimize features such as wavelet coefficients. Experimental results show that the feature-level fusion strategy effectively retains anatomical and functional feature information on multimodal medical images such as CT, MRI, PET, SPECT, etc., improves local contrast and edge sharpness of the fused images, and provides high-quality image support for feature recognition and analysis in subsequent clinical diagnosis Its feature representation ability is significantly superior to that of a single pixel-level fusion method [iv].

IV、Smart healthcare based on decision-level fusion

Decision-level fusion is the highest level of a data fusion system that combines independent decision results from multiple classifiers, algorithms, or data sources to generate more accurate and reliable final judgments. Unlike data-level and feature-level fusions, decision-level fusions deal with information that has been preliminarily analyzed and abstracted, thus offering greater fault tolerance and flexibility, and are particularly suitable for comprehensive interpretation scenarios of multi-source heterogeneous data in the medical field.

Nazari et al. systematically reviewed the application status of decision fusion technology in healthcare and pointed out that this technology can significantly improve diagnostic accuracy and reduce the risk of misjudgment by integrating the outputs of multiple classifiers. The study categorizes decision fusion methods into three major types: probabilistic methods (such as Bayesian theory), evidence-based reasoning methods (such as Dempster-Shafer theory), and intelligent methods (such as fuzzy logic, neural networks), and elaborates on their applications in medical subfields such as image processing, physiological signal analysis, and clinical decision support systems. The study shows that decision-level fusion has higher robustness and generalization ability than a single classifier in tasks such as early cancer diagnosis, cardiovascular disease prediction, and neurological disease identification, especially when dealing with incomplete or noisy data [v].

Nazari et al. further quantified the research distribution of decision fusion in the medical field through a scope review and found that approximately 80% of the applications were concentrated at the disease diagnosis level, with image processing (29%), sensor data analysis (22%), and signal processing (15%) being the main application directions. The study noted that intelligent methods (44%) were the most widely used fusion strategies, while evidence-based reasoning methods (7%) were less widely applied due to their higher computational complexity. The study also highlights that decision-level fusion can effectively manage the uncertainty, heterogeneity and scalability of medical data, providing technical support for personalized treatment and precision medicine [vi].

In a specific application case, Susyanto et al. applied decision-level fusion to inpatient and outpatient patient classification tasks, using four classifiers: logistic regression, support vector machine, random forest, and gradient boosting, and integrating the predictions of each model through majority voting and fractional fusion. Experimental results showed that in scenarios where the false negative rate (FNR) was controlled at 1% and 5%, the majority voting fusion method performed best in reducing high-risk misjudgments (i.e., inpatients being misjudged as outpatients), with accuracies reaching 88.88% and 91.72% respectively, outperforming any single classifier. The study fully demonstrated the practical value of decision-level fusion in clinical risk management [vii].

To sum up, decision-level fusion, by integrating multi-source decision information, effectively enhances the accuracy and reliability of medical diagnosis and shows broad application prospects in areas such as computer-assisted diagnosis, clinical decision support, and patient risk management.

V、Smart healthcare based on hybrid fusion

Hybrid fusion refers to the approach of combining multiple fusion levels (such as data level, feature level, decision level) or multiple algorithms (such as deep learning and traditional feature engineering) in the same system with the aim of leveraging their respective strengths to make up for the shortcomings of a single approach, thereby enhancing the overall performance, robustness, and reliability of the smart healthcare system. With the increasing number of multi-source heterogeneous data generated by Internet of Things (IoT) medical devices, a single fusion strategy often fails to simultaneously meet requirements such as accuracy, real-time performance, and interpretability. Therefore, hybrid fusion strategies have become a research hotspot.

Zhang et al. proposed a hybrid feature fusion method for the classification problem of wearable electrocardiogram signals in the Internet of Things. The method fuses deep features automatically extracted by convolutional neural networks (CNNS) with manual features based on R-wave detection, including temporal features and signal quality indices, and uses support vector machines (SVMS) as classifiers. Experimental results showed that the hybrid fusion strategy improved the average F1 score of four types of signals - normal rhythm, atrial fibrillation, other arrhythmias, and noise - from 80.06% to 84.34% compared with the pure CNN method, especially in a few types such as atrial fibrillation and noise (F1 score for atrial fibrillation increased by 4.46%, noise increased by 8.48%). The study fully demonstrated the complementarity between deep features and manual features, as well as the advantages of hybrid fusion in dealing with imbalanced data and noise interference [viii].

Abdelmoneem et al. reviewed multi-sensor decision-level fusion techniques in Internet of Things healthcare and systematically compared the application of methods such as fuzzy logic, Bayesian inference, Markov processes, and Dempster-Shafer evidence theory in different medical scenarios. Although each of these methods is effective, a single technique often has limitations: fuzzy logic is difficult to handle the dependencies between sensors, Bayesian methods are highly dependent on prior knowledge, and Dempster-Shafer has high computational complexity. Hybrid fusion can further improve the accuracy and robustness of tasks such as anomaly detection and fall monitoring by combining multiple decision-level techniques (such as combining HMM with fuzzy rules) or fusing outputs at different levels (such as feature-level and decision-level fusion). For example, combining everyday behavior modeling based on hidden Markov models with physiological anomaly detection based on fuzzy logic can provide a more comprehensive assessment of a patient's health status [ix].

In summary, hybrid fusion significantly enhances the decision-making capabilities of smart healthcare systems through multi-dimensional and multi-level complementarity and integration, and shows broad application prospects in complex medical scenarios. Future research could further explore adaptive hybrid fusion strategies to balance computational overhead and performance enhancement.

VI、Dataset

Research on multimodal fusion in smart healthcare is constrained by issues such as privacy of medical data and high costs of collection and annotation. The existing core datasets are divided into three categories: clinical medical, health behavior of wearable devices, and medical scene synthesis, which respectively support research in directions such as clinical diagnosis, health monitoring, and scene understanding of medical robots. The following introduces typical achievements in recent years.

6.1 Clinical medical multimodal dataset HAIM-MIMIC-MM

The HAIM-MIMIC-MM dataset constructed by Soenksen et al[x]. which is a fusion of the MIMIC-IV intensive care database and the MIMIC-CXR-JPG chest X-ray database, contains 34,537 samples, 7,279 hospitalization records, and 6,485 patients. It is an important benchmark for clinical multimodal fusion studies. The dataset integrates four core modalities: table, time series, text, and medical image, covering 11 independent data sources, and supports 12 types of clinical tasks such as chest pathological diagnosis and 48-hour mortality prediction; Patient-centered multimodal data matching and integration addressed the problem of heterogeneous clinical data formats and scattered sources, and its public preprocessing and feature extraction methods provided a reference for the construction of clinical multimodal datasets.

6.2 Wearable Device Health Behavior Dataset MMC-PCL-Activity

Hu et al. constructed the MMC-PCL-Activity dataset for the demand of health behavior recognition [xi]. They collected multi-source data from 14 subjects through smartphones and Mi bands, with a cumulative collection of over 251 hours, generating 6,944 valid samples and 6,510 perspective images. Covering 16 types of daily activities. The dataset includes four modalities: sensors (acceleration, gyroscope, heart rate), visual images, environmental information (GPS, weather), mobile app usage, and, for the first time, adds dual labeling of physical (energy/fatigue, etc.) and mental (excellent/average, etc.) states. It is the only health behavior dataset that integrates behavior, physiology, psychology, and environment. The data was collected in an uncontrolled everyday scenario, which is closer to real-world applications. Its publicly available sensor filtering preprocessing method and fusion benchmark experiments provide a basis for multimodal fusion research in wearable devices.

6.3 Synthetic multimodal dataset Syn-Mediverse for medical scenarios

Mohan et al. built the Syn-Mediverse synthetic dataset based on the NVIDIA Isaac Sim simulation platform, generating 48,000 multi-view images and providing over 1.5 million annotations, in response to the limitations of real medical scene data collection [xii]. It is the first multi-task and multi-modal synthetic dataset in the field of medical scene perception. The dataset covers 13 types of medical scenarios including operating rooms and consultation rooms, contains 31 types of semantic objects (27 types of completed instance annotations), and provides high-precision annotations for five types of tasks such as object detection, semantic segmentation, and depth estimation; An industry-standard multi-camera acquisition scheme is adopted to add real medical scene noise to the synthetic images, enhancing the authenticity of the data. The dataset was divided into training/validation/test sets in 9:2:5, and an online evaluation benchmark was established. The transfer learning of the synthetic data improved the mIoU of the medical scene semantic segmentation model by 4.7% to 8.1%, effectively addressing the scarcity of real medical data.

6.4 Characteristics and Existing Problems of the dataset development

The existing multimodal data sets for smart healthcare show a trend of modal fusion, scenario segmentation, and data synthesis, all achieving the fusion of structured and unstructured data, and deeply binding with specific tasks such as clinical diagnosis and health monitoring, with synthetic data becoming an important supplement to real data. At the same time, there are obvious problems: there is a scarcity of specialized disease-specific datasets, most of which focus on general clinical tasks; There is an imbalance in sample distribution in real-world scenario datasets, which affects the generalization ability of the model; In some datasets, only temporal coarse alignment is achieved for multi-source data, and fine-grained cross-modal feature association annotations are missing, increasing the difficulty of fusion.

VII、Conclusion

This paper focuses on the application of multimodal data fusion technology in smart healthcare, systematically reviews the principles, representative literature, and experimental performance of four mainstream methods: data-level, feature-level, decision-level, and hybrid fusion, and introduces common datasets and evaluation criteria. Studies show that current multimodal fusion methods have demonstrated superior performance to single-modal models in tasks such as skin disease diagnosis, depression identification, and electrocardiogram classification, and some methods have made significant progress in accuracy, robustness, and generalization ability.

However, the existing research still faces many challenges: First, at the data level, it is difficult to align multi-source heterogeneous data, modal absence is widespread, and the dataset size is small with poor representativeness; Second, at the methodological level, existing models are not robust enough to modal heterogeneity and deletion, and lack interpretability and clinical acceptability; Third, at the evaluation level, the evaluation metrics are mostly focused on traditional metrics such as accuracy, which are disconnected from the actual clinical needs (such as misdiagnosis risk, uncertainty quantification).

To address these issues, this paper proposes the following: Build an adaptive cross-modal fusion framework to enhance robustness against heterogeneous data and modal missing; Introduce generative models to complete missing modalities and enhance data integrity; Promote the construction of multi-center, all-modal, standardized medical datasets; Develop a clinical-oriented evaluation system that integrates interpretability and performance metrics; Combining federated learning with privacy computing to ensure data security and privacy.

Looking ahead, the multimodal fusion technology of smart healthcare will develop in the direction of "deep integration of medicine and engineering, balancing privacy and performance, unifying interpretability and lightweighting, and wide clinical application", urgently requiring interdisciplinary collaboration and policy support to drive the technology from the laboratory to real clinical scenarios and facilitate the realization of precision medicine and intelligent health services.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

[i] Dautov, Rustem, Salvatore Distefano和Rajkumaar Buyya. 2019. 《Hierarchical data fusion for Smart Healthcare》. *Journal of Big Data* 6(1):19. doi:10.1186/s40537-019-0183-6.

[ii] King, Rachel C., Emma Villeneuve, Ruth J. White, R. Simon Sherratt, William Holderbaum和William S. Harwin. 2017. 《Application of Data Fusion Techniques and Technologies for Wearable Health Monitoring》. *Medical Engineering & Physics* 42(1):1～12. doi:10.1016/j.medengphy.2016.12.011.

[iii] Zasim Uddin, Md., Md. Arif Shahriar, Björn W. Schuller, Md. Nadim Mahamood和Md. Atiqur Rahman Ahad. 2025. 《Skin Disease Diagnosis Using Decision and Feature Level Fusion of Deep Features》. *Frontiers in Digital Health* 7:1478688. doi:10.3389/fdgth.2025.1478688.

[iv] Rajalingam, B., & Priya, R. (2018). Multimodal medical image fusion using various hybrid fusion techniques for clinical treatment analysis. Smart Construction Research, 2(2), 1-20.

[v] Nazari, Elham, Rizwana Biviji, Danial Roshandel, Reza Pour, Mohammad Hasan Shahriari, Amin Mehrabian和Hamed Tabesh. 2022. 《Decision Fusion in Healthcare and Medicine: A Narrative Review》. *mHealth* 8:8～8. doi:10.21037/mhealth-21-15.

[vi] A comprehensive Overview of Decision Fusion Technique in Healthcare: A Systematic Scoping Review. 2020. *Iranian Red Crescent Medical Journal*. doi:10.32592/ircmj.2020.22.10.30.

[vii] Nastiti, Astri Ayu, Laurentinus Anindito Wisnu Susanto, Desi Natalia Muskananfola和Gusti Ayu Dwi Yanti. 2025. 《Decision-Level Fusion on Healthcare》 编辑 Hartono, E. Budi Santoso和H. Pribawanto. *ITM Web of Conferences* 71:01013. doi:10.1051/itmconf/20257101013.

[viii] Zhang, Xianbin, Mingzhe Jiang, Wanqing Wu和Victor Hugo C. De Albuquerque. 2023. 《Hybrid Feature Fusion for Classification Optimization of Short ECG Segment in IoT Based Intelligent Healthcare System》. *Neural Computing and Applications* 35(31):22823～37. doi:10.1007/s00521-021-06693-1.

[ix] Abdelmoneem, Randa M., Eman Shaaban和Abderrahim Benslimane. 2018. 《A Survey on Multi-Sensor Fusion Techniques in IoT for Healthcare》. 页 157～62 收入 *2018 13th International Conference on Computer Engineering and Systems (ICCES)*. Cairo, Egypt: IEEE.

[x] Soenksen, Luis R., Yu Ma, Cynthia Zeng, Leonard Boussioux, Kimberly Villalobos Carballo, Liangyuan Na, Holly M. Wiberg, Michael L. Li, Ignacio Fuentes和Dimitris Bertsimas. 2022. 《Integrated Multimodal Artificial Intelligence Framework for Healthcare Applications》. *Npj Digital Medicine* 5(1):149. doi:10.1038/s41746-022-00689-4.

[xi] Hu, Menghao, Mingxuan Luo, Menghua Huang, Wenhua Meng, Baochen Xiong, Xiaoshan Yang和Jitao Sang. 2023. 《Towards a Multimodal Human Activity Dataset for Healthcare》. *Multimedia Systems* 29(1):1～13. doi:10.1007/s00530-021-00875-6.

[xii] Mohan, Rohit, José Arce, Sassan Mokhtar, Daniele Cattaneo和Abhinav Valada. 2024. 《Syn-Mediverse: A Multimodal Synthetic Dataset for Intelligent Scene Understanding of Healthcare Facilities》. *IEEE Robotics and Automation Letters* 9(8):7094～7101. doi:10.1109/LRA.2024.3418272.