Information Fusion 103 (2024) 102134

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Multi-modality approaches for medical support systems: A systematic review of the last decade

Massimo Salvi [a], Hui Wen Loh [b], Silvia Seoni [a], Prabal Datta Barua [c,d], Salvador García [e], Filippo Molinari [a], U. Rajendra Acharya [f,g,*]

[a] *Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy*
[b] *Department of biomedical informatics, National University of Singapore, Singapore*
[c] *School of Business (Information System), University of Southern Queensland, Toowoomba, QLD 4350, Australia*
[d] *Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia*
[e] *Department of Computer Science and Artificial Intelligence, Catedrático de Universidad, Granada, Spain*
[f] *School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia*
[g] *Centre for Health Research, University of Southern Queensland, Australia*

ARTICLE INFO

Keywords:
Data fusion
Deep learning
Multi-modality
Fusion methods
Diagnosis and prognosis
Healthcare

ABSTRACT

Healthcare traditionally relies on single-modality approaches, which limit the information available for medical decisions. However, advancements in technology and the availability of diverse data sources have made it feasible to integrate multiple modalities and gain a more comprehensive understanding of patients' conditions. Multi-modality approaches involve fusing and analyzing various data types, including medical images, biosignals, clinical records, and other relevant sources.

This systematic review provides a comprehensive exploration of the multi-modality approaches in healthcare, with a specific focus on disease diagnosis and prognosis. The adoption of multi-modality approaches in healthcare is crucial for personalized medicine, as it enables a comprehensive profile of each patient, considering their genetic makeup, imaging characteristics, clinical history, and other relevant factors. The review also discusses the technical challenges associated with fusing heterogeneous multimodal data and highlights the emergence of deep learning approaches as a powerful paradigm for multimodal data integration.

1. Introduction

Healthcare has traditionally relied on single-modality approaches, where medical decisions are based on analyzing a single type of data, such as radiology images or clinical data [1]. However, advancements in technology and the availability of diverse data sources have made it increasingly feasible and promising to integrate multiple modalities [2,3]. First, medical scanners are producing higher resolution digital images across modalities like MRI, CT, and PET [4]. Second, electronic health records now compile diverse clinical data in structured formats [5]. Finally, advanced analytics methods like deep learning are capable of modeling complex multi-modal relationships [6]

Multi-modality approaches involve fusing and analyzing various data types, including medical images, biosignals, clinical records, and other relevant sources, to gain a more comprehensive understanding of patients' conditions [7]. Each modality reveals a particular aspect of physiology and pathology, and effectively aggregating and analyzing multimodal data presents both unique opportunities and challenges [8]. For example, in Alzheimer's disease diagnosis, relying solely on structural MRI scans or speech analysis [9] results in approximately 80% detection accuracy. However, by also incorporating complementary modalities like audio features, speech transcript, genomic and clinical assessments, multi-modality models have achieved over 90% diagnosis accuracy [10,11].

In this paper, our objective is to comprehensively review the state-of-the-art multimodal approaches in healthcare, analyze their benefits and limitations, and outline opportunities to address current challenges. While single-modality approaches have their merits, they often provide limited information and may not capture the full complexity of diseases. In contrast, multi-modality approaches leverage the complementary nature of different data sources, enabling a more holistic assessment of patients' health. The adoption of multi-modality approaches in

* Corresponding author.
E-mail address: Rajendra.Acharya@usq.edu.au (U.R. Acharya).

https://doi.org/10.1016/j.inffus.2023.102134
Received 13 September 2023; Received in revised form 5 November 2023; Accepted 6 November 2023
Available online 10 November 2023

healthcare is crucial for the advancement of personalized medicine [7].

Each patient is unique, and their healthcare journey involves dynamic interactions between various biological, environmental, and lifestyle factors. By integrating diverse modalities, healthcare practitioners can obtain a comprehensive profile of each patient, considering their genetic makeup, imaging characteristics, clinical history, and other relevant factors. This comprehensive understanding enables the delivery of personalized healthcare interventions that are tailored to individual needs, ultimately improving patient outcomes and quality of life.

Multi-modality approaches harness the richness of complementary information across modalities. However, developing approaches to fuse heterogeneous, high-dimensional multimodal data poses unique technical challenges due to data incompatibility [12]. Data types can vary extensively in dimensionality, representation, scale, structure, and spatial-temporal characteristics. Genetics data may contain thousands of features while medical images constitute high resolution, multi-channel pixel arrays. Variability in how each data type is acquired and pre-processed further exacerbates these incompatibilities. As a results, directly aggregating raw data becomes intractable due to differences in representation, scale, structure, and dimensionality. Traditional machine learning techniques also struggle with such complexity. As a result, deep learning approaches have emerged as a powerful paradigm for multimodal data integration. Deep neural networks can learn hierarchical, abstract data representations that embed modality-specific topology and statistical dependencies [13]. Fig. 1 illustrates how multimodality is applied in healthcare. Multimodal data, including both images and non-image information from the same patient, are utilized to develop AI systems for diagnostic and prognostic purposes.

This review paper aims to provide a comprehensive overview of the current state of multi-modality approaches in healthcare. The primary objectives are as follows: (1) Provide an overview of the methods for fusing healthcare data; (2) Categorize and analyze the types of data utilized in multimodal approaches, including bioimaging, biosignals, and clinical data; (3) Summarize key applications where multi-modality is applied; and (4) Discuss limitations and outline directions for future work. The review synthesizes insights from 81 relevant primary studies published between 2012–2022 and aims to serve as a useful reference for readers to appreciate progress, select suitable methods, and identify avenues for new research.

2. Methods

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to select relevant articles on multi-modality in healthcare.

2.1. Related reviews

The topic of multi-modality is highly relevant in the field of healthcare, and three recent reviews have been published in this area. However, these reviews have certain limitations in terms of their scope and focus:

- Zhou et al. [14] "A review: Deep learning for medical image segmentation using multi-modality fusion": This review provides an overview of deep learning-based approaches for multi-modal medical image segmentation. However, it solely focuses on deep learning methods and does not cover the broader aspects of multi-modality.
- Huang et al. [15] "A Review of Multimodal Medical Image Fusion Techniques": This review specifically concentrates on multimodal medical image fusion methods, highlighting recent advances in fusion techniques based on mathematics and deep learning. However, it only partially addresses the topic of multi-modality and is limited to the field of medical imaging.
- Cui et al. [16] "Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review": This review describes current multimodal learning workflows in disease diagnosis and prognosis. Unfortunately, it only analyzes 34 studies and overlooks the challenges associated with biosignals.

The objective of our review is to provide a comprehensive overview of all multimodality approaches for disease diagnosis and prognosis in healthcare. This will be accomplished by incorporating both imaging data (2D, and 3D images) and non-imaging data (biosignals, demographics, and clinical data). In addition to presenting the current landscape of multimodality approaches, this work will also address the benefits and challenges associated with multimodality, suggesting potential solutions and avenues for future development. Fig. 2 shows the comparison between our review paper and the previous literature reviews.

2.2. Literature search strategy

This review focuses on articles published between 2012 and 2022, specifically within the last decade. A comprehensive search for relevant journal articles was conducted using scientific repositories such as PubMed, the Institute of Electrical and Electronics Engineers (IEEE), and Scopus. The keywords combined terms related to multimodality and healthcare. The search strategy employed a Boolean approach, combining various keywords such as "Multimodal," "Machine learning," "Deep learning," "Detection," "Classification," "Prediction," "Diagnosis," "Medical," "Healthcare," "Mental," and "Health" in different

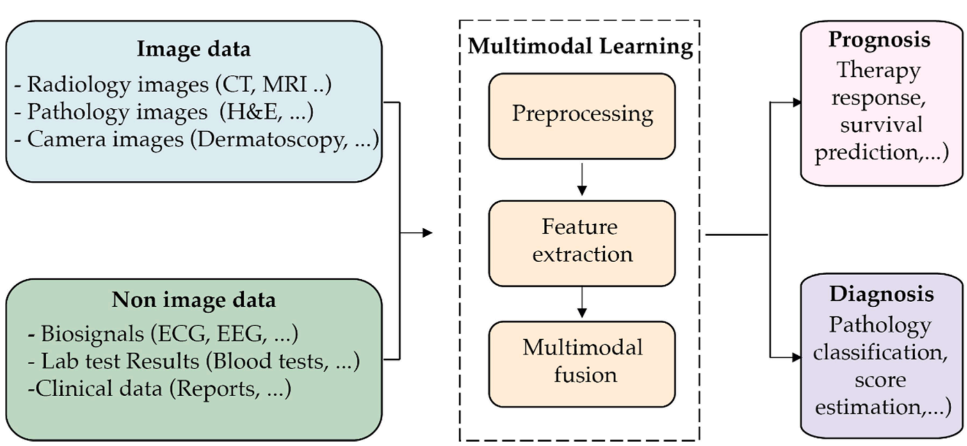

Fig. 1. Overview of multimodality in healthcare: Multiple data types, such as images and non-image data, are gathered from patients. Through multimodal learning, this data is processed via steps like preprocessing, feature extraction, and fusion to enable diagnosis and prognosis.

Fig. 2. Comparison of our review paper with existing literature reviews. DL: deep learning.

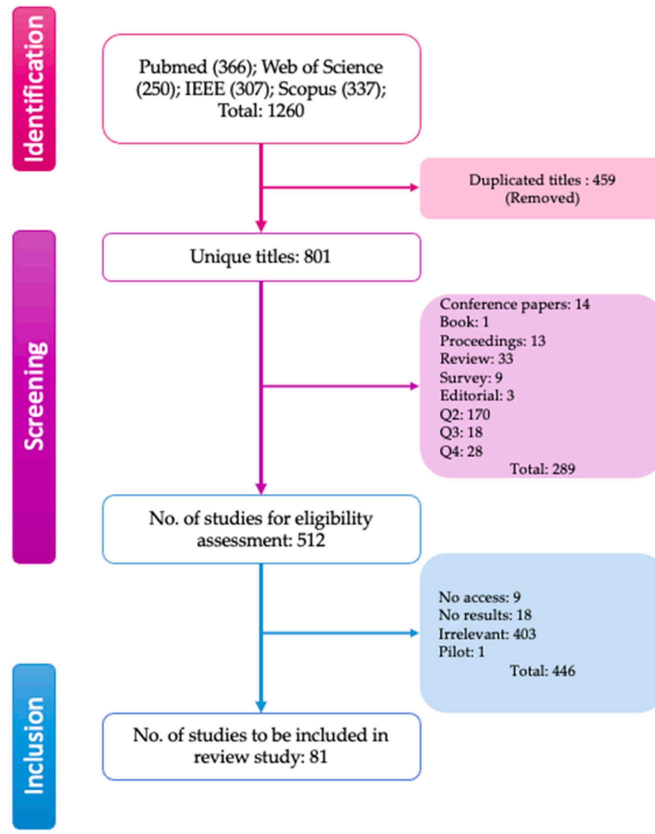

Fig. 3. Selection of relevant articles based on PRISMA guidelines.

combinations. The search process was carried out in June 2023.

Initially, a total of 1260 articles were identified. After removing duplicate articles (n=459) and excluding books, abstracts, and conference proceedings, the remaining studies were further refined by excluding articles not published in journals ranked in the top quartile (Q1) based on their impact factor. The assessment of the remaining studies was then based on the following criteria:

(i) The articles described multi-modal methods for detecting or predicting diseases in healthcare.
(ii) The articles described methods based on machine learning or deep learning models.
(iii) The articles were published in peer-reviewed journals.
(iv) The articles were written in English.

Articles that did not meet these criteria, such as those written in languages other than English, pilot studies, studies published before 2012, or articles not available in full text, were excluded. For each selected paper, we extracted key information including participant demographics, data modalities, fusion techniques, analytical methods, and performance. After a thorough examination, a total of 446 articles were excluded based on these criteria. After screening over 1200 results published over the past decade based on defined inclusion criteria, 81 high impact studies were selected through a rigorous process adhering to PRISMA guidelines. This allowed us to systematically review the latest advancements across healthcare domains, providing unique breadth compared to existing works focused on multimodal medical imaging or specific disease. Fig. 3 showcases the utilization of the PRISMA guideline for article selection.

2.3. Multi-modality approaches

Approaches to multimodal data integration in healthcare can vary depending on the specific clinical application and available modalities [16]. In general, there are two main categories of approaches for fusing information from multiple modalities:

- *Feature-level fusion*: This method involves combining features extracted from different modalities to create a unified representation. Features are extracted from each modality separately, capturing their unique characteristics, and then combined to form a joint feature representation. Alternately, the features are fused at an intermediate stage after initial modality-specific processing [6]. In feature-level fusion, data from modalities like MRI scans and blood tests are processed separately into distinct feature sets capturing their unique characteristics. These feature sets are then aggregated into a combined representation for final classification. This fusion approach aims to capture complementary information from different modalities and enhance the overall representation.
- *Decision-level fusion*: In this method, decisions or predictions made independently by each modality are combined to reach a final decision. Each modality provides its own decision based on its specific analysis, and these decisions are aggregated using various techniques such as voting, averaging, or weighted combinations [17]. For example, each modality like an MRI scanner and genetics lab would produce its own benign/malignant classification. These individual classifications would then be combined via methods like averaging to arrive at an overall diagnosis. Decision-level fusion aims to leverage the diversity of information provided by different modalities and improve the overall decision-making process. This allows flexibility in choosing modalities without needing end-to-end training. However, it does not model interactions between modalities like feature-level fusion.

Feature-level fusion can capture interactions between modalities but requires compatible feature sets. Decision-level fusion is simpler to implement but may miss relationships between modalities. Regardless of the specific application or available modalities, most multimodal fusion methods fall into one of these two categories based on the stage at which fusion takes place. Fig. 4 depicts an overview of the feature-level and decision-level fusion approaches.

These fusion techniques continue to advance through modern innovations. For instance, graph neural networks show promise for learning optimal combinations of multimodal data in a latent space [18]. Dynamic routing methods adaptively aggregate modality

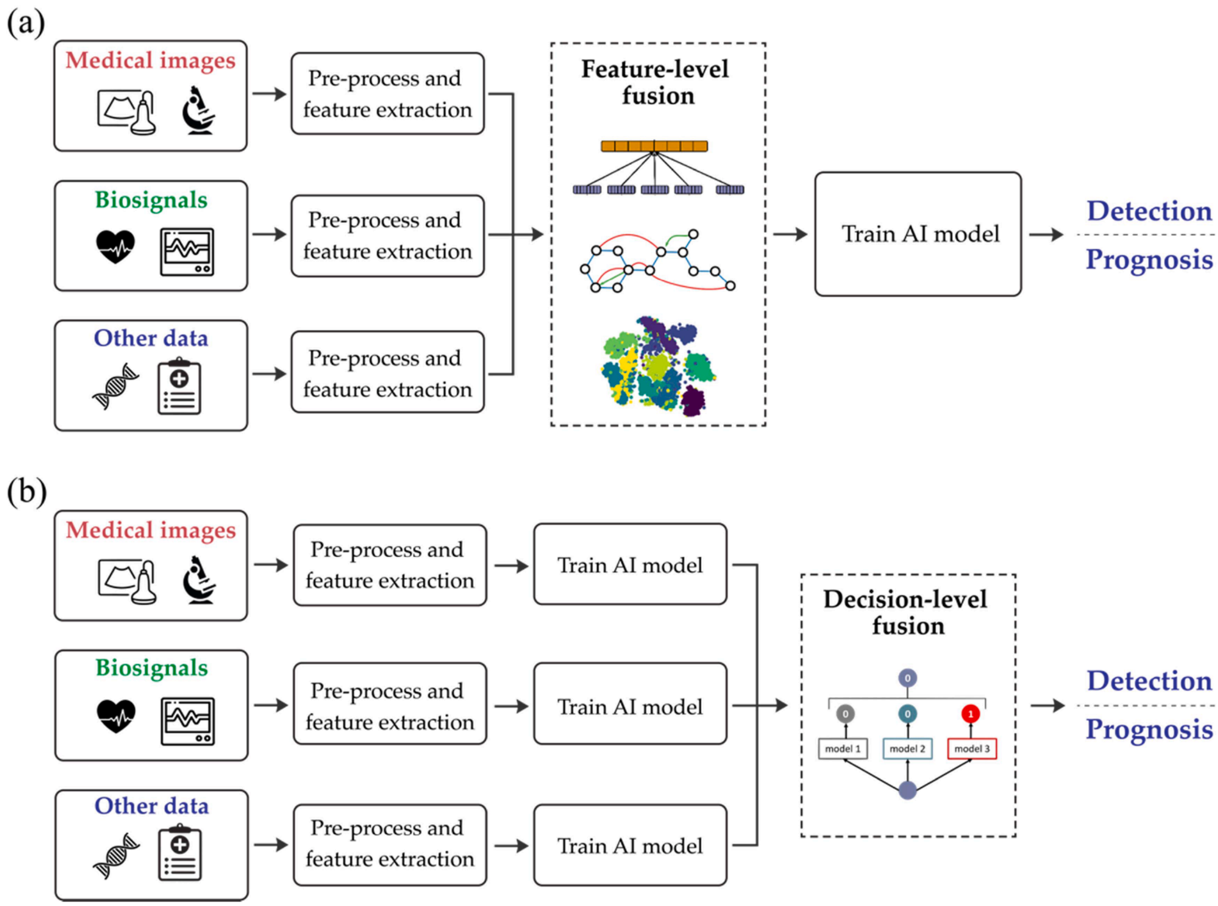

Fig. 4. Data fusion approaches in healthcare. (a) Feature-level fusion: This approach involves extracting features independently from each modality using techniques like deep learning. They are then aggregated into a joint feature representation before being input to a model for prediction. (b) Decision-level fusion: each modality is input to a separate model which makes an independent prediction. These predictions are then fused using methods like averaging or weighted sums to reach the final output.

representations [19]. Deep canonical correlation analysis extracts shared representations [20]. Overall, data fusion is an active research area as investigators explore sophisticated techniques to fully leverage complementary multimodal data.

2.4. Applications

In our study, we have classified the data type of multi-modality approaches into three main categories:

1 Bioimaging: This category involves the fusion of different types of medical images. Examples include combining functional images like PET scans with anatomical images such as CT and MRI scans.
2 Biosignals: This category includes approaches that integrate physiological monitoring signals. For instance, it involves merging measurements of electrophysiological signals like photoplethysmography (PPG), electroencephalography (EEG), and electrocardiography (ECG).
3 Mixed: This category encompasses approaches that integrate a combination of images, signals, and additional clinical data sources. For example, it involves combining medical images with physiological measurements, as well as incorporating electronic health records that contain structured clinical data.

Together, these three categories cover the common data types used in healthcare, including anatomical, functional, and clinical data sources. The selection of specific modalities to incorporate depends on the clinical question at hand and the availability of patient data. Fig. 5 illustrates how healthcare data from multiple sources can be aggregated when developing multimodal AI systems.

In addition to the three categories based on data modalities, we further categorized the multimodal data integration approaches based on the clinical application or disease domain that each study focused on. A total of 9 categories were identified: cognitive impairment, mental disorders, sleep health, cardiac diseases, COVID-19, oncology, ophthalmology, pediatric disorders, and other studies (miscellaneous applications).

This additional layer of categorization by clinical domain provides more context regarding the specific patient populations and healthcare challenges that each approach aims to address. Together with the three categories based on data modalities, this dual-axis categorization framework offers a more comprehensive perspective on trends in this research field.

3. Results

3.1. Cognitive impairment

Cognitive impairment refers to deterioration in cognitive abilities like memory, attention, judgment, and processing speed. Studies in this category aim to predict the diagnosis or progression of Alzheimer's disease, dementia, and mild cognitive impairment using multimodal data. Different data modalities provide insights into the structural, functional, and molecular changes related to cognitive functions and changes in the brain. Combining sparse clinical features describing symptoms with high-dimensional time series data from neuroimaging

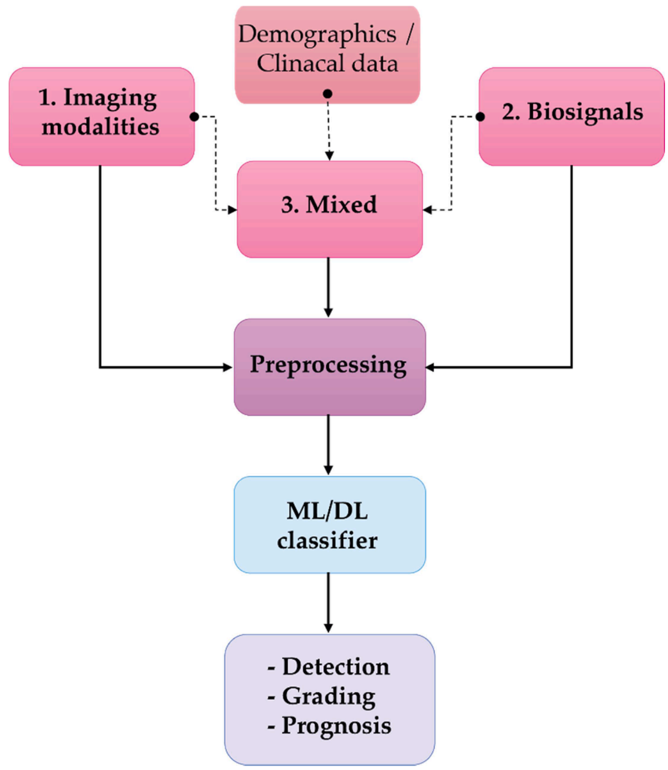

Fig. 5. Different approaches used by researchers to aggregate healthcare data in the development of multimodal AI systems. ML: machine learning; DL: deep learning.

allows capturing pathological changes occurring at the molecular, cellular, systems, and behavioral levels. Table A1 summarizes the studies reported in this section.

One of the most common applications of multimodality is in the study of Alzheimer's disease. Zhang et al. [21] conducted a study where they integrated structural MRI, PET, and cognitive tests to develop a multivariate classification system for predicting the diagnosis and progression of Alzheimer's disease. Liu et al. [22] proposed a multimodal neuroimaging feature learning framework using stacked auto-encoders for diagnosing different stages of Alzheimer's disease. They fused MRI and PET data using a zero-masking strategy and achieved improved classification performance compared to single-modality methods. Shi et al. [23] proposed a multimodal stacked deep polynomial network for diagnosing Alzheimer's disease. Their method extracted high-level features separately from MRI and PET scans before fusing them, resulting in state-of-the-art multiclass classification accuracy. Bhagwat et al. [10] integrated longitudinal MRI, genetic, and clinical data to predict clinical symptom trajectories in Alzheimer's disease. Their model effectively combined multimodal data from two timepoints and provided high prediction accuracy. Wang et al. [24] proposed a deep learning framework that jointly learns Alzheimer's diagnosis and predicts mild cognitive impairment by combining structural MRI, demographic, neuropsychological, and genetic data. Spasov et al. [25] developed a deep learning model that combines structural MRI, demographic, cognitive, and genetic data to predict the conversion from mild cognitive impairment to Alzheimer's disease. Pelka et al. [26] fused sociodemographic data, APOE genotype, and MRI data using an LSTM model to improve the detection of amnestic mild cognitive impairment. Augmenting MRI with clinical data and enhancing color information boosted performance, resulting in 90 % classification accuracy.

Muhammed et al. [27] proposed a multimodal approach using EEG signals, MRI, and CSF biomarkers for early diagnosis of Alzheimer's disease, achieving better performance than unimodal methods. Wang et al. [28] developed a multi-kernel adaptive sparse representation classification model using MRI, PET, and CSF data that outperformed single-kernel methods for diagnosing Alzheimer's disease. Syed et al. presented an ensemble learning approach using audio, text, and neuropsychological tests for diagnosing Alzheimer's disease and predicting MMSE scores. The model integrates modalities through feature-level fusion and classifier-level fusion, demonstrating improved performance over individual modalities. Song et al. [29] proposed a multimodal neural network model using speech, language, and neuropsychological tests to quantify Alzheimer's disease severity, outperforming methods based on individual modalities. The model captures complex interactions between modalities using co-attention modules and provides insight into multimodal integration. El-Sappagh et al. [30] proposed a two-layer random forest model for Alzheimer's diagnosis and progression detection that integrates 11 modalities and provides multiple explanations using SHAP models. Gao et al. [31] proposed a deep learning framework for the imputation of missing PET images, along with a path-wise transfer dense convolution network for multimodal classification of Alzheimer's disease. Using MRI and PET images, the method shows superior performance in PET image imputation and diagnosis compared to other state-of-the-art methods.

Zhang et al. [32] developed a multimodal classification model that combines fMRI, PET, MRI, and neuropsychological assessments to capture intra-modality and inter-modality relationships using cross-modal interactions. The method achieves better prediction of different stages of Alzheimer's disease compared to single-modality methods. Ilias et al. [33] proposed a multimodal deep learning model that combines speech, language transcripts, and neuropsychological tests to detect dementia and predict MMSE scores. Qiu et al. [34] developed a multimodal deep learning method that combines MRI, PET, CSF, genetics, and cognitive assessments to predict Alzheimer's disease progression. The model captures both inter-modality and intra-modality relationships and demonstrates superior prediction compared to single-modality models. Velazquez et al. [35] developed a multimodal deep learning method using speech, language transcripts, and neuropsychological tests to detect Alzheimer's disease. The model integrates modalities using gated multimodal units and attention mechanisms and demonstrates the benefit of multimodal learning. Golovanevsky et al. [36] developed a multimodal deep learning framework that integrates imaging, genetic, and clinical data to detect Alzheimer's disease and mild cognitive impairment. The model uses cross-modal attention to capture interactions between modalities and outperforms previous state-of-the-art models. El-Sappagh et al. [37] proposed an ensemble learning framework for Alzheimer's disease progression detection that integrates heterogeneous base learners using stacking and achieves superior performance compared to state-of-the-art techniques.

Two studies investigating Parkinson's disease using multimodal approaches have been identified. Papadopoulos et al. [38] proposed a deep learning framework that combines accelerometer data capturing hand tremors with typing data capturing fine-motor impairment during natural smartphone use to detect Parkinson's disease and its symptoms, including tremor and fine-motor impairment. The results suggest that passively captured multimodal smartphone data can serve as an enhanced medium for Parkinson's disease screening. Makarious et al. [39] developed a machine learning model that integrates genomics, transcriptomics, and clinical data to make improved predictions of Parkinson's disease risk, which were validated in an external cohort.

In addition, four other studies have utilized a multimodality framework to investigate cognitive impairments such as dementia and brain aging. Feis et al. [40] employed MRI data to train classification models that distinguish presymptomatic dementia carriers from controls. Their model, combining diffusion and anatomical MRI features, successfully detected dementia mutation carriers. Kassani et al. [41] proposed a sparse machine learning method that utilizes multimodal fMRI data, including task-based and resting-state scans, to predict adolescent brain age. The experiments demonstrated that multimodal approaches achieve higher classification accuracy compared to unimodal methods, with task-based and resting-state fMRI providing the most effective combination. Kang et al. [42] developed machine learning models using multimodal data, including neuropsychological tests, MRI, and clinical data, to predict amnestic mild cognitive impairment in patients. The multimodal model achieved higher accuracy than the model without MRI data, highlighting the effectiveness of the multimodality approach. Ko et al. [43] developed a deep learning framework that represents MRI and SNP data and utilizes their learned joint representation for diagnosing Alzheimer's disease and mild cognitive impairment.

The authors of these papers faced several challenges in developing multimodal approaches to study cognitive impairment. A key challenge was integrating diverse data types, including genetics, neuroimaging, sensor data, and clinical assessments, which have very different formats and dimensionality. This required developing sophisticated data integration and feature selection techniques to identify complementary information across modalities. For neurological disorders that are heterogeneous in their presentation like Parkinson's and Alzheimer's, capturing different symptom dimensions added complexity. The black-box nature of deep learning models also posed interpretability challenges that required novel analysis methods to extract biological insights. Despite these hurdles, the authors demonstrated the power of thoughtfully combining modalities to improve disease prediction and subtype differentiation.

3.2. Mental disorders

These studies focus on the diagnosis, stratification, and treatment response prediction of various mental disorders including schizophrenia, bipolar disorder, depression, and anxiety disorders using multimodal data. Integrating neuroanatomical data, functional information, genetic biomarkers, and clinical factors from history and examinations helps obtain a more comprehensive understanding of these complex disorders with heterogeneous etiologies. Table A2 summarizes the studies reported in this section.

We have identified three studies that utilize multimodal approaches to assess depression. In the work by Patel et al. [44], multi-modal MRI measures, along with age and cognitive scores, were used as input features for machine learning methods to develop prediction models for late-life depression diagnosis and treatment response. Ding et al. [45] recorded EEG, eye-tracking, and galvanic skin response data from both depression patients and control subjects during emotional tasks. They extracted features from the multi-modal data and trained machine learning classifiers, showing that combining modalities improved classification accuracy compared to using individual modalities alone. Zhu et al. [46] employed an EEG-eye tracking synchronized acquisition system to simultaneously collect EEG and eye movement signals from subjects with mild depression and control subjects. By utilizing feature fusion strategies, they achieved up to a 7% improvement in classification accuracy compared to using unimodal approaches.

Furthermore, multimodal approaches have shown benefits in the study of other mental disorders. Han et al. [47] developed a stacked autoencoder model to identify autism spectrum disorder in children by combining EEG and eye-tracking data. They used separate auto-encoders to learn features from each modality and then concatenated them to learn joint representations. This multimodal approach outperformed unimodal and feature-concatenation methods, taking advantage of the complementary nature of neurophysiological and behavioral data. Jiang et al. [48] presented an IoT-based hierarchical solution for stress monitoring that merged edge and cloud computing. They employed lightweight machine learning at the edge for real-time inference before selectively offloading data to the cloud for advanced analysis. This system achieved significantly reduced latency and energy costs compared to cloud-only approaches, benefiting from the integration of local and remote resources. Pan et al. [49] proposed a multi-scale adaptive multi-channel fusion graph convolutional network, which demonstrated significantly improved accuracy over baselines in the classification of autism spectrum disorder and major depressive disorder. Rahaman et al. [50] developed a deep multimodal framework that integrated structural and functional MRI and genomic data to predict schizophrenia. Their multimodal approach outperformed unimodal and multimodal baselines in schizophrenia prediction, and the analysis of salient features helped identify critical neural and genetic factors related to schizophrenia. Soundararajan et al. [51] proposed a biosensor network for analyzing Parkinson's symptoms, integrating acoustic sensors, microphones, and multi-sensor units to monitor biosignals. Their system achieved improved detection and prediction of Parkinson's compared to existing systems.

A major challenge in developing multimodal techniques for mental disorders is the heterogeneity and complexity of the data from different modalities like neuroimaging, biosignals, genetics, and behavioral measures. Effectively integrating complementary data from various modalities to improve the diagnosis and monitoring of mental disorders requires overcoming challenges in data collection, feature engineering, modelling correlations, managing heterogeneity, and model optimization. The works attached make good progress on these fronts using autoencoders, attention mechanisms, graph networks and other deep learning innovations.

3.3. Sleep health

This category focuses on integrating multimodal data to investigate sleep disorders and their impact on overall health. The goal is the diagnosis or assessment of sleep disorders through multimodal sleep monitoring signals fusion. Combining physiological biosignals, clinical symptoms reported in questionnaires, and sleep logs provides a more holistic characterization compared to a single modality. Table A3 summarizes the studies reported in this section.

Sano et al. [52] focused on multimodal ambulatory monitoring of physiological signals for assessing sleep patterns. The authors used LSTM RNN on smartphone and wearable data combining multi-modality. This outperformed actigraphy and achieved near-real-time detection with slightly reduced performance. Zhou et al. [53] integrated multi-modality by using the Fourier decomposition method (FDM) on biomedical signals, specifically electroencephalogram (EEG), electromyogram (EMG), and electrooculogram (EOG) data, to classify different stages of sleep. They demonstrated that utilizing multiple modalities improved the classification results compared to using single-channel EEG data alone, suggesting the potential for enhanced sleep disorder assessment and monitoring. Jia et al. [54] developed SleepPrintNet, a method for sleep stage classification that integrated multiple modalities, including electroencephalogram (EEG) data. They demonstrated that their approach achieved better classification results compared to existing algorithms, suggesting the potential for improved sleep stage assessment and monitoring. Fatimah et al. [55] proposed a multi-modal assessment of sleep stages using adaptive Fourier decomposition and machine learning. They utilized electroencephalogram (EEG), electromyogram (EMG), and electrooculogram (EOG) data to improve the classification of sleep stages. Their method showed better classification results compared to other algorithms and highlighted the potential for low-cost sensor-based setups for continuous patient monitoring and feedback.

The authors in the papers faced a few key challenges in developing automated sleep stage classification systems using multi-modality physiological data. One major challenge was how to efficiently extract relevant features that capture the discriminative characteristics of different sleep stages from EEG signals. While time-domain information provided morphological patterns, frequency-domain and spatial patterns also contained useful information. Another challenge was how to integrate other modalities like EOG and EMG that provide supplementary cues while capturing their discriminative signatures. Existing works often overlooked the unique traits of different modalities and shared representations across them. Developing subject-independent models posed additional difficulties due to inter-individual variations in sleep patterns. Real-time operation with portable devices brought further constraints of efficient computation. The authors in the reviewed papers attempted to address these challenges by techniques like adaptive signal decomposition, hierarchical neural networks, modality-specific processing, and ensemble classifiers to leverage multi-modality for improving automated sleep stage classification.

3.4. Cardiac diseases

This category involves integrating multimodal data to improve the diagnosis and management of cardiac diseases, such as heart failure and arrhythmias. The importance of Multimodal integration combines clinical data, electrocardiography (ECG), echocardiograms, cardiac imaging, and genetic information to assess cardiac function, identify risk factors, and guide treatment decisions. Table A4 summarizes the studies reported in this section.

Puyol-Antón et al. [56] presented an automated diagnostic pipeline for classifying dilated cardiomyopathy. The pipeline combined a multimodal cardiac motion atlas created from both 3D MR and US data with multi-view machine learning algorithms. This approach significantly improves accuracy compared to the single modality. Zhang et al. [57] presented a method for detecting coronary artery disease using multi-modal feature fusion and hybrid feature selection. By combining multiple types of features, they aimed to improve the accuracy of disease detection. Kim et al. [58] developed a multimodal and integrative decision support system for predicting postoperative cardiac events in multiple surgical cohorts. They incorporated multimodal features derived from physiological signal processing techniques and electronic health records (EHR) data, achieving promising results with AUROC values. The study demonstrated the potential of combining EHR data and physiological waveform data for early detection of postoperative deterioration events.

There were several challenges faced by the authors of the papers related to cardiac diseases. First, integrating information from multiple data sources, such as physiological signals, and electronic health records, posed technical difficulties. The data modalities varied in dimensionality and representation, requiring advanced processing techniques for fusion. Second, patient populations across various cohorts showed heterogeneous disease manifestations, making generalizable modeling complex. Third, cardiac conditions have multifactorial origins, and determining etiology in individual cases remains challenging. Diverse feature extraction methods captured complementary information. While progress has been made, larger and more diverse databases will likely further enhance the capability to stratify cardiac conditions and personalize management.

3.5. COVID-19

This category focuses on integrating multimodal data to understand COVID-19, its impact on various organs, and the development of effective treatment strategies: Multimodal integration combines clinical data, laboratory results, chest imaging (such as X-rays or CT scans), genomic data, and patient outcomes to study disease progression, identify risk factors, and predict patient outcomes. Table A5 summarizes the studies reported in this section.

Chen et al. [59] presented a multimodality machine learning approach to differentiate between severe and non-severe COVID-19 cases. By combining clinical and laboratory features, the authors develop a random forest model that achieves high predictive accuracy, shedding light on the comprehensive understanding of COVID-19 and providing insights for evaluating disease severity based on common medical features. Sait et al. [124] developed a novel deep fusion strategy called CovScanNet to predict COVID-19 using breathing sounds and CXR images, achieving 98.72% accuracy on the validation set. Their implementation through a smartphone application enabled convenient and easily accessible screening. Zheng et al. [60] focused on the integration of multi-modal knowledge graphs in doctor-patient dialogues. The authors propose a method that combines textual and visual information to enhance the understanding and effectiveness of medical conversations, leading to improved healthcare outcomes. Manocha et al. [61] proposed a preliminary COVID-19 screening method that analyzed cough recordings via FFT and classified them using CNN. They additionally collected breathing sounds to address gaps in their dataset and explored the multi-modal fusion of audio and visual features, obtaining 95.64% accuracy.

The authors faced several challenges when developing deep learning models for COVID-19 diagnosis and prognosis using multi-modal data. One major challenge was how to effectively integrate and fuse information from different data modalities, such as combining text descriptions with visual medical images to enhance clinical understanding. Integrating diverse data types increased computational complexity, posing implementation challenges for real-time applications. Finally, collecting and integrating data from multiple centers introduced further challenges around data collection, standardization, and ensuring privacy and security when sharing information.

3.6. Oncology

The studies reported in this section aim to enhance cancer diagnosis, treatment selection, and monitoring through multimodal data integration. Specifically, multimodal integration combines different data sources such as medical imaging, histopathological analyses of tumor tissues, and clinical patient records. The fusion of these modalities aims to improve the accuracy of cancer detection and characterization of molecular subtypes. It also seeks to support more personalized treatment planning and monitoring of treatment responses. Table A6 summarizes the studies reported in this section.

Karim et al. [62] proposed an adversary-aware multimodal convolutional auto-encoder model to predict cancer susceptibility from multi-omics data. The model demonstrated high confidence in predicting cancer types and robustness against adversarial attacks. Kanwal et al. [63] introduced a multimodal deep learning framework for cancer prognosis prediction using clinical, copy number, and RNA data. The experiments demonstrated high accuracy in prognosis prediction and improvement through multimodal fusion.

Regarding prostate cancer, three authors proposed a multimodal approach based on MRI imaging. Molina et al. [64] developed an incremental learning system using SVM classifiers and multimodal MRI features, including T2-weighted, dynamic contrast-enhanced MRI, and texture analysis, to classify prostate cancer. Le et al. [65] proposed a multimodal CNN model to diagnose prostate cancer in multi-parametric MRI, using apparent diffusion coefficient and T2-weighted images. Rossi et al. [66] proposed a siamese neural network model to retrieve prostate MRI images that are diagnostically similar in terms of lesion PIRADS scores. All authors achieved higher diagnostic metrics compared to current approaches and monomodal methods.

For breast cancer assessment, multimodal approaches were used in the fields of radiology and pathology. Mokni et al. [67] proposed a multimodal CAD system that fused mammography and DCE-MRI for breast cancer diagnosis. They extracted texture features using the GLIP descriptor from each modality and achieved higher AUC compared to mammography and MRI individually. Yang et al. [68] developed a multimodal deep learning model that integrated H&E histological images and clinical data to predict recurrence risk in HER2+ breast cancer. They achieved a high AUC on independent blind test data.

Multimodal approaches have also been used for the study and analysis of brain tumors such as gliomas. Lu et al. [69] developed a machine learning model using MR radiomic features to classify gliomas into five molecular subtypes based on histology, IDH mutation status, and 1p/19q codeletion status. The model improved the classification performance by over 15% compared to MR alone. Li et al. [70] developed an improved 2D U-Net model for brain tumor segmentation using multimodal MRI. The model generates segmentation maps slice-by-slice and achieves higher segmentation performance compared to mono-modality. Wang et al. [71] proposed a 3D multitask CNN model to jointly segment glioma lesions and predict IDH mutation status from multimodal MRI. They achieved high classification performance compared to previous models, and experiments showed that lesion segmentation helps predict gene mutation. Xiao et al. [72] proposed a neural architecture search method called DLS-DARTS to discover optimal CNN architectures for classifying glioma grades using multimodal intraoperative imaging. Their approach outperformed manually designed networks and demonstrated the potential of multimodality in glioma grading.

Multimodality has been employed for the study of tumors in other body regions, such as the abdominal area. Yang et al. [73] proposed an integrated deep learning model using ultrasound images and radiologist domain knowledge to classify thyroid nodules. Chen et al. [74] proposed a deep learning scheme using multi-modal MRI and model constraints for the noninvasive prediction of TP53 mutations in pancreatic cancer. Fu et al. [75] introduced a PET-CT tumor segmentation model with a multimodal spatial attention module that enhances tumor regions, improving segmentation performance over state-of-the-art methods. Menegotto et al. [76,77] proposed a multimodal deep learning model combining CT images and structured clinical data to diagnose hepatocellular carcinoma, achieving promising results, including 89.6% precision and 86.9% accuracy, approaching specialist performance. Gao et al. developed a semantic segmentation model using multi-modal ultrasound images to diagnose cervical lymphadenopathy related to COVID-19 vaccines, achieving accurate tissue detection and lymph node diagnosis. Hao et al. [78] proposed a deep learning model called SurvivalCNN, which utilizes CT images and clinical data to predict overall and progression-free survival for gastric cancer patients. Ye et al. [79] developed a machine learning model using multi-modal MRI data to classify germinoma tumors in the pineal region. By evaluating models trained on different combinations of MRI sequences, their model achieved an AUC of 0.88 for germinoma classification.

In oncology, effectively fusing the multi-modal data is difficult - simply concatenating features from different modalities may not fully utilize the complementary information. The authors propose various techniques to integrate the modalities such as constraint terms in the loss function, attention mechanisms, and architectural designs to share features. The multimodal data also increases the model complexity, requiring more computational resources. Additional challenges include utilizing 3D spatial information from volumetric scans efficiently in 2D networks, generating usable augmented data, and integrating radiologist domain knowledge and annotations. Despite these obstacles, the works demonstrate that combining modalities can improve performance over single-modal approaches for tasks like tumor segmentation, diagnosis, and survival prediction.

3.7. Ophthalmology

This category includes studies that focus on utilizing multimodal data integration to examine eye diseases. Diseases such as age-related macular degeneration and diabetic retinopathy are among the key areas of investigation. The goal of multimodal integration is to fuse different ophthalmic data modalities, such as retinal imaging techniques like fundus photography and optical coherence tomography scans. These images are combined with genetic profiles, clinical examination findings, and patient medical records. Table A7 summarizes the studies reported in this section.

Chai et al. [80] developed a Bayesian deep learning model for glaucoma diagnosis. By considering uncertainty and integrating information from multiple modalities, such as medical indicators, images, and texts, their model achieved better performance in glaucoma detection compared to other methods, offering potential benefits for managing the diagnosis of glaucoma. Hervella et al. [81] proposed a novel self-supervised multimodal reconstruction pre-training that explicitly taught networks to recognize common and exclusive characteristics between modalities. Jin et al. [82] presented a multimodal deep learning framework with feature-level fusion for retinal imaging. By integrating multiple modalities, including fundus images, the proposed model achieved improved accuracy in identifying choroidal neovascularization activity, which is crucial for diagnosing and managing diseases such as age-related macular degeneration. Liu et al. [83] proposed a deep learning approach using multimodal fundus images for predicting visual impairment in retinitis pigmentosa. They integrated different modalities of fundus images and used a deep learning model to accurately predict visual impairment, which can aid in early detection and management of the disease. Hervella et al. [84] proposed a novel self-supervised pre-training approach, called multimodal image encoding, which learned representations from unlabeled retinal image pairs. It improved grading on multiple datasets compared to supervised pre-training.

The papers related to ophthalmology highlighted several research challenges faced by their authors. A key challenge was effectively exploiting multi-modality data for self-supervised deep learning approaches. Traditionally in this field, features were extracted separately from each modality using deep networks and then combined downstream into a classifier model. However, this ignored interactions between modalities. Another significant challenge was addressing uncertainty in diagnosis while leveraging the multiple sources of information routinely collected in clinical practice, such as medical indicators, images from different modalities, and patient-reported outcomes/texts. Fully leveraging these diverse yet complementary data sources could help improve diagnostic accuracy and effectiveness.

3.8. Pediatric disorders

This category includes studies that apply multimodal data integration to further the understanding and management of disorders and other health conditions that affect children. The fusion of structural and functional information derived from multimodality aims to improve early detection capabilities for various pediatric disorders. Researchers also seek to support more personalized intervention planning and longitudinal monitoring of treatment responses/developmental progress through a holistic perspective achieved via multimodal data integration [85]. Table A8 summarizes the studies reported in this section.

Petrozziello et al. [86] developed multimodal convolutional neural networks (MCNNs) combining fetal heart rate, uterine contractions, and signal quality data to predict fetal compromise during labor. However, their deep learning models were not suitable for detecting severe fetal injury without acidemia, suggesting hybrid approaches incorporating clinical knowledge are needed. Gao et al. [87] proposed a multimodal AI system combining abdominal radiographs and clinical data to diagnose necrotizing enterocolitis and predict the need for surgery. By integrating radiological and clinical data, they achieved improved diagnostic accuracy and surgical predictive ability compared to either modality alone. Salekin et al. [88] developed a multimodal deep learning approach using video, audio, and physiological signals to assess neonatal postoperative pain. The integration of facial expressions, body movements, and crying sounds enabled more accurate and continuous pain monitoring compared to models relying on a single modality. Guez et al. [89] constructed a multimodal machine learning model combining magnetic resonance enterography and biochemical biomarkers to predict Crohn’s disease endoscopic activity in the ileum. The fusion model demonstrated better predictive performance than either imaging or biomarkers alone and was comparable to experienced clinicians.

Integrating data from multiple sources, including brain imaging, and clinical assessments, requires addressing challenges related to data quality, privacy, and standardized assessment tools for children.

Challenges also include a limited number of samples, high costs of pediatric imaging and lack of standardized collection methods for different age groups.

3.9. Other studies

This section collects multimodal studies that do not fall into one of the applications described above. The authors in each of the following paper developed innovative multimodal approaches and leveraged the strengths of combining different data modalities. Table A9 summarizes the studies reported in this section.

Milosevic et al. [90] used accelerometry and surface EMG to detect tonic-clonic seizures, achieving robust detection of short and non-stereotypical seizures with their multimodal approach. Yao et al. [91] improved infection screening at airports by using multiple vital signs, resulting in higher accuracy than standalone thermography in identifying seasonal influenza patients. Tiulpin et al. [92] proposed a multimodal machine learning approach for knee osteoarthritis progression prediction, demonstrating improved prediction performance by combining radiographs and clinical data. Huang et al. [93] developed a multimodal fusion model for classifying Pulmonary Embolism, showing superior performance compared to single-modality approaches when combining CT scans and clinical patient data. Xue et al. [94] presented a multimodal approach for staging liver fibrosis using ultrasound imaging, demonstrating increased classification performance by including elastography. Subramaniam et al. [95] proposed an automated pain recognition system that emphasizes the importance of multi-modality in pain assessment, utilizing physiological signals and a hybrid deep learning network. Mattia et al. [96] used multimodal brain MRI to classify patients with anoxic brain injury, achieving improved performance with a 3D CNN model.

Tang et al. [97] developed a hybrid deep learning model for lung nodule classification, showing the advantage of fusing structured and unstructured multimodal patient data. Ming et al. [98] proposed a multimodal deep learning framework for cervical cancer classification, integrating PET/CT images and achieving improved recognition accuracy compared to single-modality approaches. Wu et al. [99] introduced a deep multimodal learning network for classifying lymph node metastasis, outperforming single-modality networks in analyzing ultrasound images and elastogram modality.

4. Discussion

4.1. Summary of main findings

Initially, AI tools were limited to working with a single modality. However, healthcare practitioners take a multimodal approach when making a diagnosis, as they evaluate multiple aspects and do not rely solely on a single biosignal or bioimage. In recent years, AI tools have also begun to adopt a multimodal approach, allowing them to emulate the decision-making process of a physician, improve their performance, and provide precise diagnoses.

Fig. 6 shows the number of papers divided by machine learning (ML) and deep learning (DL) relative to the publication year. From the Fig., it can be observed that initially multimodal approaches were exclusively based on machine learning (2012–2017). However, starting in 2018, DL approaches began to emerge and within a year surpassed the number of ML-based papers, reaching nearly five times as many by 2022.

A ML-based multimodal approach is simpler as it only requires aggregating hand-crafted features extracted from images and/or signals with various clinical data. This is supported by the fact that 74% of papers (n = 20) using ML employ feature concatenation as a data fusion technique. With the increasing complexity of problems analyzed and the pursuit of higher performance, there has been a gradual shift towards DL approaches, which are now considered state-of-the-art for multimodal approaches [23]. In recent years, there has been a prevalence of DL use and more sophisticated data fusion techniques like graph neural network [29], cross-modal attention [36], and dynamic fusion strategy [61]. This has also posed challenges, as researchers have had to develop reliable techniques for effectively extracting features computed by deep neural networks [8].

Fig. 7 illustrates the trend over the years for the various data types (biosignal, bioimaging, mixed) used in the multimodal approaches reported in this review. In the last two years, there has been a prevalence of studies utilizing a mixed approach, integrating patient clinical data with diagnostic imaging or electrophysiological tests. In 2022, the number of mixed approaches greatly exceeded the other two modalities combined (24 papers vs. 5 papers). The sharp increase in mixed methods highlights the recognition that combining different sources of patient data, including bioimaging, biosignals, genetics, and clinical data, is key to fully capturing disease heterogeneity [100]. Overall, healthcare is progressing toward sophisticated multimodal techniques that integrate diverse, complementary data types to enable enhanced diagnosis, prognosis, and precision medicine across various disorders. This review reveals the great potential of fused, multiparametric data analysis to unlock new insights and improvements in healthcare.

Fig.s 8 and Fig. 9 illustrate the applications and approaches currently employing multimodality. Specifically, Fig. 8 shows the various disease areas utilizing multimodal techniques, while Fig. 9 focuses on the types of data integrated in multimodal studies across applications. Approaches are categorized as bioimaging, biosignal, or mixed based on the modalities combined.

In terms of applications, 31% of the studies reviewed (n = 25) use a multimodal approach for cognitive impairments, as diagnosis involves neurological tests, imaging, and audio recordings. Thus, the authors adopt multimodality to optimize performance. In oncology, there is a tendency to favor bioimaging over mixed approaches (11 vs. 7 papers), even though integrating clinical data could provide a more

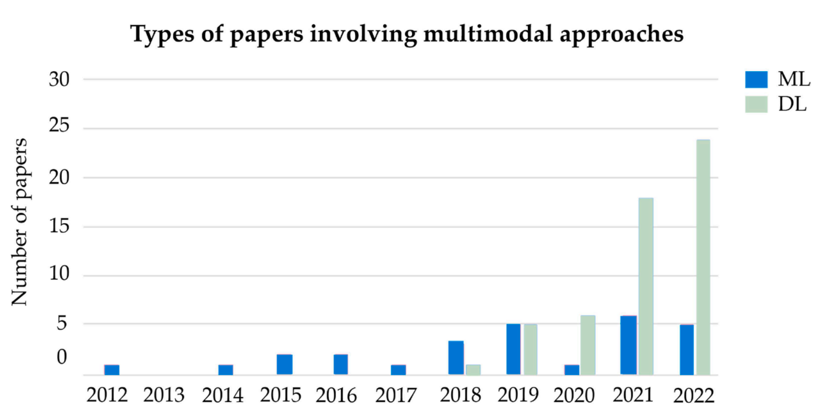

Fig. 6. Reviewed papers categorized by use of machine learning (ML) or deep learning (DL) approaches, divided by publication year.

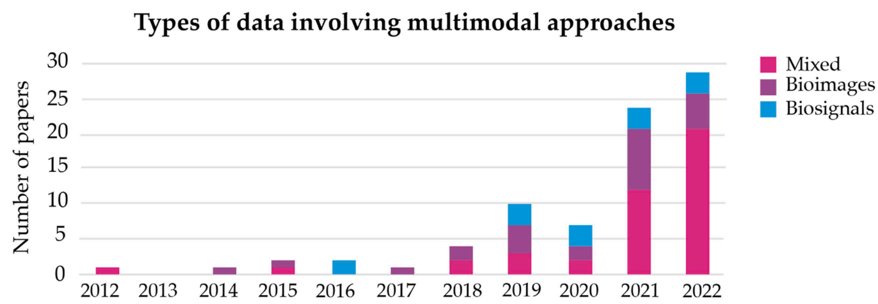

Fig. 7. Trends in data types (biosignals, bioimaging, mixed) used in the multimodal approaches reviewed, shown by year.

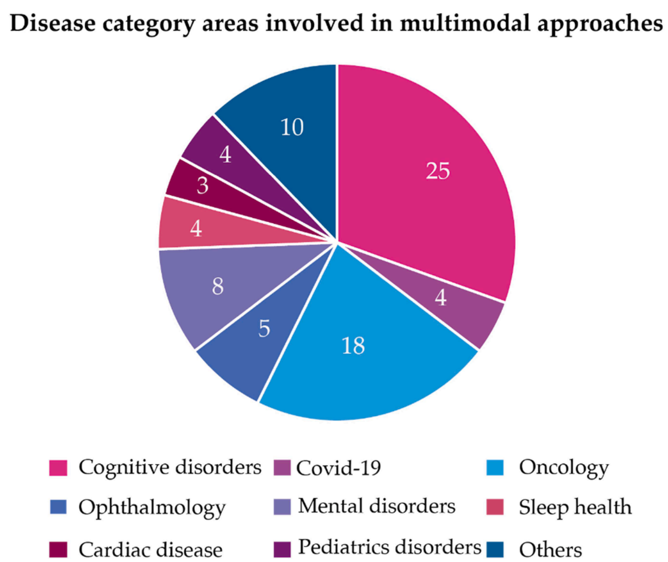

Fig. 8. Disease areas involved in the multimodal studies reviewed.

comprehensive analysis. For COVID-19, mixed methods predominate, as radiology and symptoms are key for diagnosis. In mental health, mixed and biosignal approaches are equally employed (n=4 each), highlighting the utility of physiological and clinical data. Finally, for sleep health, only biosignals-based approaches are utilized, with no consideration for clinical or demographic patient data.

The performance of multimodal techniques varied across clinical applications. For cognitive disorders, most studies showed significant gains, with 5-10% improved classification accuracy over single modalities [23,29]. However, multimodality only provided marginal improvements of 2-3% in cardiac diseases [58]. Factors impacting performance include heterogeneity of modalities, availability of large datasets, and suitability of deep learning techniques for the application.

In Table 1, we have summarized all the open-source datasets used in the multimodal studies in this review paper. Notably, several datasets are available for cognitive impairments, especially for Alzheimer's disease research. The ADNI (Alzheimer's Disease Neuroimaging Initiative) database is used by 14 different studies. For mental disorders, open-source datasets facilitate studying diseases like schizophrenia. In oncology, most studies use the TCGA (The Cancer Genome Atlas Program) dataset, which provides diverse tumor imaging and genetic data. However, many studies (43 of 81) use private datasets, restricting research progress on some diseases. Private datasets also hinder performance comparison across approaches for the same application.

Cognitive impairments and oncology are the two most common healthcare applications employing multimodal approaches. Multimodality is critical for studying cognitive disorders for several reasons. First, cognitive disorders like dementia and Alzheimer's disease are highly heterogeneous in their presentation and progression. Using modalities like genetics, neuroimaging, and cognitive assessments allows capturing different aspects of this heterogeneity. Second, different modalities provide complementary information that on their own may not be sufficient for accurate diagnosis and prognosis. Combining modalities provides a more complete picture. Third, multimodal models can integrate data acquired at different time points, enabling tracking of disease progression by leveraging longitudinal modalities.

Among cognitive diseases, Alzheimer's disease (Fig. 10) is the most extensively studied using multimodal approaches, often requiring the integration of diverse examinations for accurate diagnosis. Authors have applied multimodal techniques to achieve superior performance compared to single modalities, attaining excellent results for both Alzheimer's disease detection [115] and progression estimation [10]. For Alzheimer's diagnosis and monitoring, a mixed approach combining multiple data types has primarily been employed, as reliably detecting, and tracking this condition based solely on bioimaging or biosignals alone [37] has proven insufficient. Genetics, biomarkers, neuroimaging,

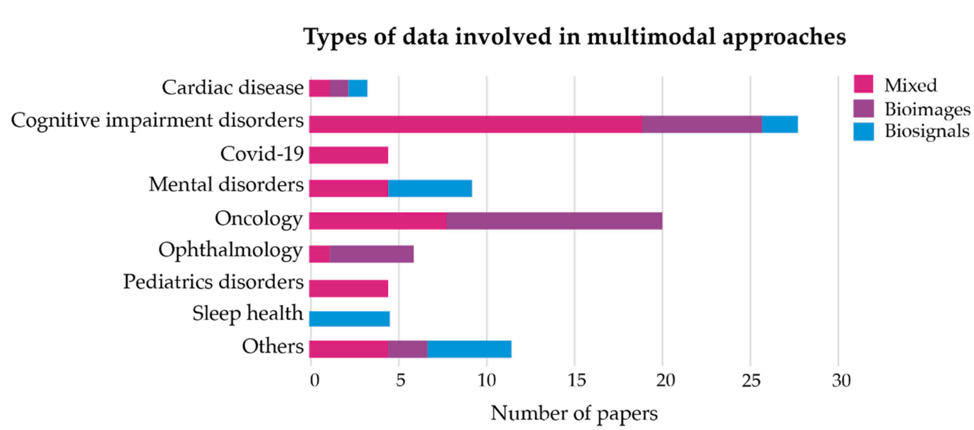

Fig. 9. Data types integrated into the multimodal studies categorized by application area.

Table 1
Open-access dataset used in multimodal studies in healthcare.

Dataset	Data description	Application	Used by
ADNI database (www.adni.loni.usc.edu)	data modalities included cognitive scores, MRI scans, PET images, genetics, medical history, neurological exams, symptoms, lab tests, and physical exams related to Alzheimer's disease	cognitive impairment	[10,21,23–26,28–30,34,35,37,43]
ADNI-TADPOLE [101]	contains multimodal data such as MRI, PET, CSF, genetics, cognitive tests, and demographics for Alzheimer's disease research	cognitive impairment	[27]
EMIF-AD [102]	cognitive test results and plasma, DNA, MRI, or cerebrospinal fluid (CSF) related to Alzheimer's disease	cognitive impairment	[32]
PDBP (www.pdbp.ninds.nih.gov)	contains clinical, genetic, imaging and biomarker data associated with Parkinson's disease	cognitive impairment	[39]
PPMI (www.ppmi-info.org)	includes clinical, imaging, omics, genetic, sensor, and biomarker data related to Parkinson	cognitive impairment	[34,39]
COBRE dataset [103]	contains resting state fMRI, structural MRI, and diffusion MRI data on healthy controls and individuals with schizophrenia	mental disorders	[50]
MRPC [104]	contains structural MRI, resting state fMRI, and DTI data from healthy controls and patients with schizophrenia	mental disorders	[50]
ABIDE [105]	neuroimaging (functional magnetic resonance imaging - fMRI) and corresponding phenotypic data (age, gender, and acquisition site) of 1112 subjects	mental disorders	[49]
sleep-EDF database [106]	197 whole night polysomnographic sleep pattern recordings. The database contains EOG, EEG, chin EMG, and event markers	sleep health	[55]
CinC Challenge [53]	994 polysomnography (PSG) recordings following AASM standards, with EEG, EOG, EMG, and other physiological signals	sleep health	[55]
MASS-SS3 sleep dataset [107]	PSG recordings from 62 healthy subjects. The dataset includes EEG, EOG, and EMG signals	sleep health	[54]
STACOM 2011 [108]	MR and 3D ultrasound (3DUS) data from a dynamic phantom and 15 datasets from healthy volunteers	cardiac disease	[56]
Coswara dataset [109]	dataset containing a diverse set of respiratory sounds (breathing, cough, and speech) and metadata about the patient (demographic information, health information)	COVID-19	[61]
TCGA (www.cancer.gov/ccg/research/genome-sequencing/tcga)	cancer genomics program, molecularly characterized over 20,000 primary cancers and matched normal samples spanning 33 cancer types	oncology	[63,68,69,71,76]
METABRIC dataset [110]	multi-modal data of 1046 breast cancer patients with gene expression profiles, CNA profiles, along with the clinical data or information.	oncology	[63]
STS dataset [111]	FDGPET-CT and MR scans from 51 patients with lung cancer	oncology	[75]
BRATS challenge [112]	multimodal MRI scans for brain tumor segmentation, collected from multiple institutions. It includes T1, T1-contrast enhanced, T2, and FLAIR MRI modalities.	oncology	[70,71]
Isfahan MISP [113]	59 multimodal image pairs consisting of a color retinography and a fluorescein angiography image of the same eye	ophthalmology	[81,84]
BioVid Heat Pain database [114]	dataset with bio-potential signals (ECG, GSR, and EMG at Trapezius muscle) and facial action video signals	other studies	[95]

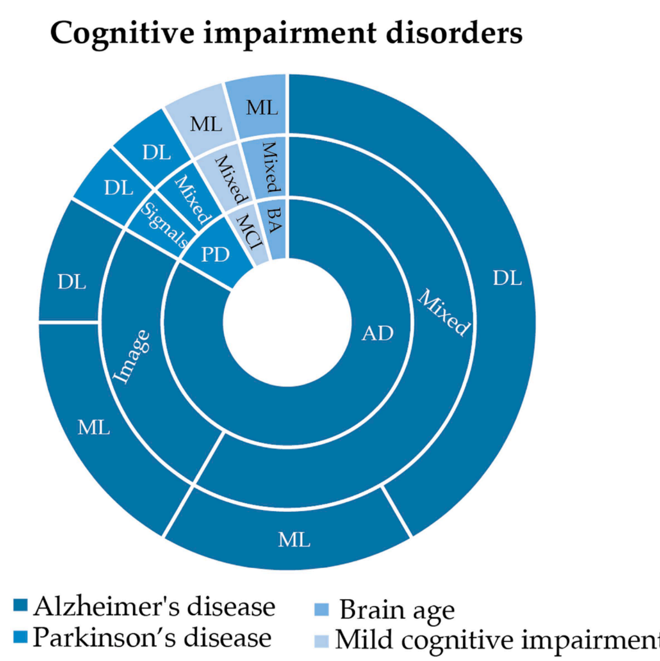

Fig. 10. Reviewed studies on cognitive impairment disorders divided by technique and data type. ML: machine learning; DL deep learning.

and cognitive data each provide an incomplete picture in isolation. Only a multimodal approach combining these sources allows constructing the complete timeline of Alzheimer's onset and progression.

Oncology accounts for 22% (n = 18) of the reviewed research, with highly diverse cancer applications. The most studied are prostate, breast, and brain tumors (Fig. 11). Typically, bioimaging approaches are favored over mixed methods. Many studies combine radiological and histological images to analyze tumor heterogeneity [69]. This multi-modal, multiscale approach is crucial in oncology for several reasons. First, tumors exhibit intricate characteristics across scales, which cannot be adequately assessed through a single imaging method or scale alone. Integrating data from diverse sources like radiology, histology, and

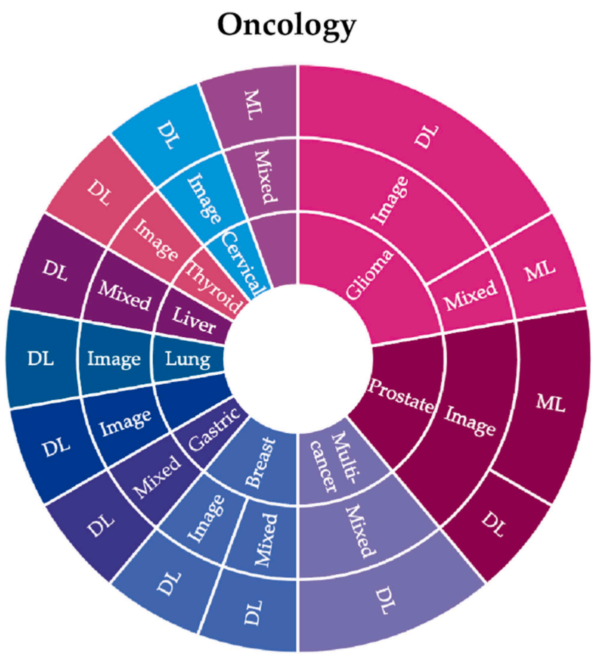

Fig. 11. Distribution of multimodal papers on oncology categorized by technique, data type and tissue under investigation. ML: machine learning; DL deep learning.

genetics provides a more comprehensive tumor perspective. Second, cancer treatment decisions depend on both tumor properties and patient factors like genetics and medical history. Capturing both perspectives allows holistic understanding to guide personalized care. Finally, imaging at multiple scales reveals connections between macroscopic anatomical patterns and microscopic molecular drivers of malignancy. Using fusion techniques to combine multimodal and multiscale data is therefore essential for uncovering cancer biology and mechanisms.

4.2. Benefits and challenges of multi-modality

Several works have highlighted the benefits of adopting a multimodal framework in healthcare. A clinical study on breast cancer diagnosis highlighted the real-world potential of multi-modality integration [67]. By combining mammogram images and MRI scans, the accuracy of distinguishing malignant from benign tumors improved from 93% with MRI alone to 99% with the multi-modality approach. This enabled earlier and more reliable breast cancer detection, allowing physicians to start necessary treatment sooner and potentially improving patient outcomes. The integration of multimodal data has enormous potential to transform many areas of healthcare by providing a more comprehensive view of each patient's unique health profile. Adopting a multimodal approach offers several key benefits:

- *Improved diagnostic accuracy*: Combining data from different modalities allows for more accurate diagnosis compared to single data types alone. Multimodal analysis leverages complementary information, capturing subtleties that may be missed by individual tests or scans. This enables earlier and more reliable detection of disease. Fused multimodal data also provides a clearer picture of the underlying mechanisms, stage, and heterogeneity of the disease, allowing physicians to better select targeted therapies tailored to an individual.
- *Precision diagnosis:* Multimodal techniques facilitate precision diagnosis where patients can be stratified into disease subtypes, severity grades, and prognostic categories based on their multiparametric profile. This is crucial for precision medicine.
- *Patient-centric care models*: By integrating a patient's genetics, lab tests, clinical history, and imaging data, multimodal models can predict optimal courses of treatment and response to therapies. This enables data-driven, personalized care. Multimodal analysis provides a more holistic understanding of the patient as a whole person, not just a disease. This empowers more patient-focused healthcare with improved quality of life.

Overall, thoughtfully combining complementary modalities ushers in a new era of medicine where diagnostics, treatments, and care delivery can be tailored to the individual. Multimodal integration is key to fully realizing the promises of precision, personalized, and patient-centric healthcare.

While integrating multimodal data has immense potential, there remain several key challenges that must be addressed:

- *Reliable and time-efficient data fusion*: Effective data fusion techniques are needed that *aggregate* multimodal features without loss of critical information. In some studies, single modalities outperform multimodal approaches [83], or there is only [88] a small margin of improvement in the multimodal approach [28]. This indicates room for improvement in fusion techniques to fully leverage the complementary strengths of different modalities. These techniques must also be time-efficient, as multimodal pipelines have high computational costs compared to single modalities, which can hinder real-time applications.
- *Optimizing model complexity, interpretability, and transparency*: Combining multiple data sources leads to inherently more complex multimodal models, sometimes at the expense of model interpretability and transparency. Multimodal model complexity reduces interpretability, which is crucial in healthcare [30]. New methods are required to maintain model interpretability in the context of ever more sophisticated multimodal models by imposing appropriate constraints or developing techniques to approximate model mechanics. For critical applications, regulating model complexity or designing inherently interpretable models are important to provide clinical rationale and gain practitioner trust in automated decisions.
- *Multiscale analysis*: combined modalities capture information at very different scales, from molecular and cellular processes to whole organ and organism-level phenomena. For instance, genetics and proteomics provide molecular insight, while medical imaging visualizes macroscopic anatomy and function. Managing the integration of data across such a range of scales with very different semantics is critical for an effective multimodal approach but remains difficult [69].
- *Data standardization*: Normalization and harmonization of data across different centers, scanners, and acquisition devices is essential prior to fusing data. Variability in how data is collected can introduce systematic biases that compromise the multimodal system if not properly addressed [37]. Careful development of data standardization pipelines that account for acquisition differences while preserving biological variance is crucial for multimodal analysis.

In the end, thoughtfully addressing these challenges will be key to unlocking the full potential of multimodal techniques to provide deeper insights into human health. Overcoming the hurdles of multimodality will pave the way for a new era of data-driven healthcare. Various approaches have been proposed to address these challenges, including multi-view learning [116], adversarial training [117], and attention mechanisms [118]. However, substantial work is still required to fully overcome the difficulties of heterogeneous multimodal data integration and realize its possibilities. Fig. 12 outlines the key challenges and potential benefits of multimodal research in healthcare.

In addition to the technical aspects, it is crucial to consider the ethical implications of implementing multimodality approaches in healthcare. While integrating multimodal data has immense potential, it

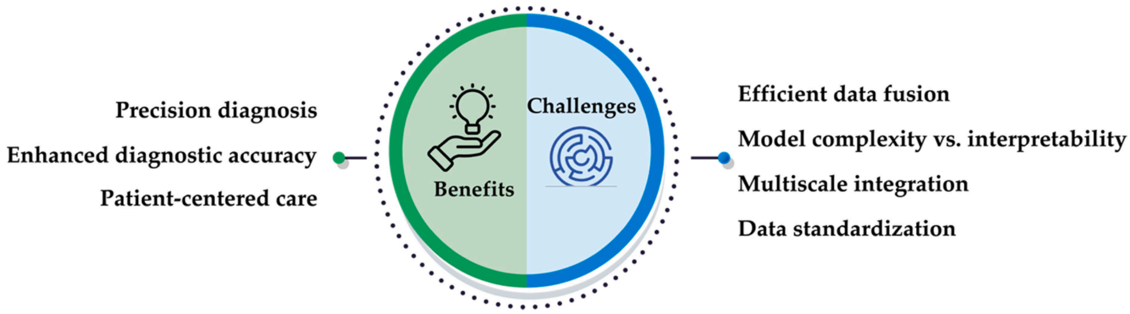

Fig. 12. Challenges and benefits of multimodal healthcare research discussed in this review.

also raises ethical concerns regarding patient privacy, data security, and informed consent that must be carefully addressed. Studies should obtain necessary ethical approvals and patient data must undergo anonymization. Robust data governance frameworks and consent protocols need to be instituted to ensure responsible data handling, protection of patient rights, and proper communication of risks and benefits to patients and healthcare providers.

4.3. Future research directions

As the field of multimodality research continues to evolve, it is important to highlight emerging trends and potential future directions. These may include the integration of wearable devices and Internet of Things (IoT) technologies, the exploration of novel imaging modalities, the application of artificial intelligence in real-time multimodal data analysis, and the development of decision support systems. Several potential improvements and future research opportunities exist to advance multimodality approaches in healthcare:

1 *Enhanced data fusion techniques*: The development of aggregation methods that balance model complexity and information retention is needed. Some studies show minimal performance gains over single modalities [28,88], indicating room for improvement in fusion techniques for certain applications. More advanced integration is required to fully leverage complementary multimodal data for certain applications without unnecessary model complexity. This could be achieved through advanced information fusion techniques that intelligently combine data in latent space through learned transformations rather than simple feature concatenation. Deep learning methods like variational auto-encoders and graph neural networks show promise for learning optimal combinations of multimodal data in embedded space.
2 *Adopt more mixed approaches*: current works often use either bioimaging or biosignals alone. Integrating additional clinical data modalities like lab tests, and clinical data could provide more comprehensive analysis, especially for diseases with complex interactions like cancer. This multimodal integration can better emulate real-world physician evaluation processes.
3 *Improved data standardization*: Robust normalization and harmonization methods are essential for aggregating multi-center, multi-scanner, and multi-dimensional data. Heterogeneous data dynamics pose integration challenges and standardization can lose information. Advances in standardization pipelines tailored to multimodal healthcare data could enhance analysis.
4 *Explainable AI and uncertainty quantification*: While multimodal models achieve strong predictive performance, their complexity reduces interpretability compared to single-modal approaches. Developing explainable AI techniques and quantifying uncertainty estimates for multimodal models could increase clinical adoption where transparency is crucial [124,125]. Probabilistic multimodal frameworks could provide confidence intervals or reveal cases where a model is uncertain or unreliable [119]. Visualization approaches to illustrate model attention mechanisms, feature relevance, and internal representations may also prove useful for practitioners to distill multimodal model behaviors [30].
5 *Expanded application to segmentation tasks*: Most current works apply multimodal fusion to classification problems, with few leveraging it for segmentation tasks [70,68]. Future studies should explore multimodal segmentation approaches, which could benefit applications like surgical planning [72], radiotherapy targeting, and computational pathology [120].
6 *Emerging multimodal opportunities*: Some emerging areas like computational histology [121] and teledermatology [122] could benefit greatly from the strategic fusion of multiple modalities. Integrating dermatological images, patient history, and clinical data could significantly improve remote skin diagnosis to emulate in-person clinician analysis.
7 *Open-access mindset*: Despite the significant benefits offered by multimodal approaches, there is currently a scarcity of open-source multimodal datasets. Research must focus on creating open-source datasets to propel technological advancements and facilitate comparisons among studies working on the same application.

Overall, numerous promising opportunities exist to advance multimodal techniques and expand their capabilities to new applications in healthcare. Thoughtful innovation could enable more integrated, informative, and transparent healthcare analysis.

This review has some limitations to acknowledge. The literature search was restricted to English articles from 2012–2022 and did not include a quantitative meta-analysis. Expanding the search criteria and performing statistical comparisons between findings could provide greater insights into relative multimodal performance. However, this review aimed to provide a comprehensive overview of current approaches and key challenges among different healthcare applications.

5. Conclusion

This systematic review demonstrates the vast potential of multimodal techniques to transform many facets of healthcare by enabling more accurate diagnosis, enhanced treatment planning, and tailored interventions for patients. The integration of data from multiple complementary modalities is transforming medicine by enabling more holistic analysis. Adopting multimodal systems can aid practitioners by providing a more comprehensive perspective of each patient for data-driven, personalized care.

However, realizing this potential in clinical settings will require addressing key challenges around ethical data handling, model interpretability, and seamless integration into practitioner workflows. Each modality provides a different perspective but has its limitations like noise, artifacts, and dimensionality issues. Modalities also possess very different statistical properties and spatial-temporal characteristics. This requires developing sophisticated AI models that can learn joint representations while handling multimodal complexity.

Author Statement

U Rajendra Acharya, MTech, Ph.D., DEng, DSc
Professor, School of Mathematics, Physics and Computing
University of Southern Queensland, Australia
Adjunct Professor, Asia University, Taiwan
Distinguished Professor of Kumamoto University, Japan
Adjunct Professor, University of Malaya, Malaysia
Email: rajendra.acharya@usq.edu.au
Home: https://scholar.google.com.sg/citations?user=8FjY99sAAAAJ&hl=en
ResearchGate: https://www.researchgate.net/profile/U_Rajendra_Acharya
Publons: https://publons.com/researcher/2836800/rajendra-u-acharya/
Scopus: Author ID: 7004510847
Highly Cited Researcher for *seven* consecutive years (2016 to 2022) in Computer
Science http://highlycited.com/
Editor in Chief, Information Fusion

Sub: Submission of manuscript (INFFUS 102134)**to Information Fusion**

Dear EIC,

We are submitting a manuscript (INFFUS 102134) entitled *"Multi-Modality Approaches for Medical Support Systems: A Systematic Review of the Last Decade"* to Information Fusion.

We declare that, there is no conflict of interest in this work.

Thank you very much.

Best Regards, Raj

Declaration

All authors have contributed to an acceptable and satisfactory level.

Declaration of Competing Interest

The authors have declared that no competing interests exist.

Data availability

No data was used for the research described in the article.

Appendix

Table A1, Table A2, Table A3, Table A4, Table A5, Table A6, Table A7, Table A8, Table A9

Table A1
Summary of studies that apply multi-modality approaches for cognitive impairment.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Zhang et al. [21], 2012	186 patients	Mixed: MRI, PET, CSF, Clinical data	kernel combination	ML method: SVM	73.9 % accuracy in mild cognitive impairment classification
Liu et al. [22], 2015	200 patients	Bioimaging: MR, PET	concatenation	ML method: SAE-ZEROMASK	93.32 % accuracy in Alzheimer's detection
Shi et al. [23], 2018	103 patients	Bioimaging: MRI, PET	deep polynomial network	ML method: SVM	97.13 % accuracy in Alzheimer's detection
Bhagwat et al. [10], 2018	1302 patients	Mixed: MRI, genetic, clinical data	concatenation	DL method: LSN	90 % accuracy in Alzheimer's detection
Wang et al. [24], 2019	119 patients	Mixed: sMRI, DWI	concatenation	ML method: random forest	96 % accuracy in Alzheimer's detection
Spasov et al. [25], 2019	785 subjects	Mixed: sMRI, demographic, neurophysical, genetic data	concatenation	DL method: CNN	87.5 % sensitivity in Alzheimer's detection
Pelka et al. [26], 2020	120 patients	Mixed: age, marital status, education and gender, genetic data, MRI	branded image	DL method: LSTM	90 % accuracy in Alzheimer's detection
Muhammed et al. [27], 2021	1737 patients	Mixed: MRI, PET, CSF, age, gender, education	concatenation	ML method: random forest	80 % sensitivity in Alzheimer's classification
Wang et al. [28], 2021	84 patients	Bioimaging: sMRI, fMRI	kernel canonical correlation analysis	DL method: 3DShuffleNet	96 % accuracy in Alzheimer's detection
Syed et. Al [28], 2021	108 subjects	Mixed: audio, textual	concatenation	ML method: ensemble model	89.81 % accuracy in Alzheimer's detection
Song et al. [29], 2021	511 patients	Mixed: MRI + cognitive measures + risk factors	graph neural network	DL method: AMGNN	94.44 % accuracy in Alzheimer's detection
El-Sappagh et al. [30], 2021	1048 patients	Mixed: CS + NB + Genetics	concatenation	ML method: random forest	93.95 % accuracy in Alzheimer's detection
Gao et al. [31], 2021	1308 patients	Bioimaging: MRI and PET	pathwise DCN	DL method: augmentation + CNN	92 % accuracy in Alzheimer's detection
Zhang et al. [32], 2022	881 patients	Mixed: protein level, age, APOE status	concatenation	DL method: DNN	0.831 AUC in Alzheimer's detection
Ilias et al. [33], 2022	156 patients	Mixed: speech, transcript	concatenation	DL method: BERT + ViT + Co-Attention	90 % accuracy in Alzheimer's detection
Qiu et al. [34], 2022	8916 patients	Mixed: demographic, functional assessment, MRI	concatenation	DL method: CNN-CatBoost	77.7 % accuracy in Alzheimer's detection
Velazquez et al. [35], 2022	383 patients	Mixed: diffusion tensor imaging + electronic health records	concatenation	DL method: ensemble model	98.81 % accuracy in Alzheimer's detection
Golovanevsky et al. [36], 2022	2384 patients	Mixed: clinical, genetic data + MRI	cross-modal attention	DL method: CNN	96.6 % accuracy in Alzheimer's detection
El-Sappagh et al. [37], 2022	1371 subjects	Mixed: MRI + neuropsychological test	information fusion approach	ML method: SVM, random forest	84.95 % accuracy in Alzheimer's disease progression

(continued on next page)

Table A1 (*continued*)

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Papadopoulos et al. [38], 2020	22 subjects	Biosignals: postural accelerations, typing dynamics	Subject embedding	DL method: CNN	92.8 % sensitivity in Parkinson's detection
Makarious et al. [39], 2022	383 patients	Mixed: genetics, transcriptomics, clinico-demographic	concatenation	DL method: ensemble method	98.81 % accuracy in Parkinson's detection
Feis et al.[40], 2018	103 patients	Bioimaging: aMRI, DTI, rs-fMRI	concatenation	ML method: bvFTD	0.68 AUC in presymptomatic dementia prediction
Kassani et al. [41], 2019	900 subjects	Bioimaging: rs-fMRI, nb-fMRI, em-fMRI	concatenation	ML method: ELM	83.05 % accuracy in brain age prediction
Kang et al. [42], 2021	511 patients	Mixed: demographic and clinical data, neuropsychological test, MRI	concatenation	ML method: GBM	0.892 AUC in mild cognitive impairment prediction
Ko et al. [43], 2022	734 subjects	Mixed: Neuroimaging and the genetic data	deep fusion	DL method: generative and discriminative framework	0.92 AUC in classification and regression task

Table A2
. Summary of studies that apply multi-modality approaches for mental disorders.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Patel et al. [44], 2015	68 patients	Mixed: MRI, demographics, cognitive data	concatenation	ML method: alternating decision tree	87.27 % accuracy in depression classification and 89.47 % accuracy in treatment response
Ding et al. [45], 2019	348 patients	Biosignals: EEG, Eye-tracking and galvanic skin response data	concatenation	ML method: Logistic regression	79.63 % accuracy in depression detection
Zhu et al. [46], 2019	39 patients	Biosignals: EEG, Eye-tracking data	feature fusion	ML method: SVM	81.88 % accuracy in depression recognition
Han et al. [47], 2022	90 subjects	Mixed: EEG, Eye-tracking data	concatenation	DL method: MMSDAE	95.56 % accuracy in autism spectrum disorders
Jiang et al. [48], 2022	17 subjects	Biosignals: ECG, EMG, accelerometers, blood volume changes, skin temperatures	decision-level fusion	DL method: Matching network	96.4 % accuracy in stress monitoring
Pan et al. [49], 2022	1645 patients	Mixed: fMRI and phenotype data	Knn graphs	DL method: MAMF-GCN	99.24 % accuracy in mental disorder prediction
Rahaman et al. [50], 2022	437 patients	Mixed: sMRI, fMRI, Genetic	ICA	DL method: Autoencoder, FFN, attention BiLSTM	92 % accuracy in schizophrenia detection
Soundararajan et al. [51], 2022	11 patients	Biosignals: Acceleration + gyro sensors, ECG, EMG and raw voice data	concatenation + PCA	DL method: VAER K. means + LSTM	99.8 % accuracy in the detection of Parkinson's disease

Table A3
. Summary of studies that apply multi-modality approaches for sleep health.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Sano et al. [52], 2019	186 subjects	Biosignals: Acceleration + skin temperature	concatenation	DL method: LSTM network	Sleep/wake classification accuracy of 96.5 %
Jia et al. [54], 2021	62 subjects	Biosignals: EEG, EMG, EOG	feature fusion module	DL method: CNN	88.8 % of sleep stages classification
Zhou et al. [53], 2020	1985 subjects	Biosignals: EEG, EOG, EMG, ECG, SaO2, respiratory	concatenation	DL method: CRPEMA	AUC of 0.844 in sleep arousal detection
Fatimah et al. [55], 2022	153 recordings	Biosignals: EEG, chin EMG, EOG and event markers	concatenation	ML method: ensemble bagged trees algorithm	93.44 % accuracy in 5-class sleep stages classification

Table A4
. Summary of studies that apply multi-modality approaches for cardiac diseases.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Puyol-Anton et al. [56], 2019	69 subjects	Bioimaging: ultrasound, MRI	motion atlas	ML method: MvLapSVM	92.71 % accuracy in the identification of dilated cardiomyopathy patients
Zhang et al. [57], 2020	62 patients	Biosignals: ECG, PCG, Holter, ECHO, BIO	hybrid feature selection	ML method: SVM	96.67 % accuracy in detection of coronary artery disease
Kim et al. [58], 2022	-	Mixed: EHR, ECG, ABP, PPG	concatenation	ML method: random forest	0.82 AUC in prediction of postoperative cardiac events

Table A5
. Summary of studies that apply multi-modality approaches for COVID-19.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Chen et al [59, 123], 2021	362 subjects	Mixed: clinical data, laboratory tests	concatenation	ML method: random forest	97.22 % accuracy in differentiating between severe and non-severe COVID-19
Sait et al. [123], 2021	10 subjects	Mixed: Breathing sounds and Chest X-ray	concatenation	DL method: CovScanNet	98.72 % accuracy in COVID-19 diagnosis
Zheng et al. [60], 2021	2000 subjects	Mixed: X-ray, CT, Ultrasound, patient-doctor text	temporal self-attention mechanism	DL method: knowledge graph attention embedding model	98.10 % accuracy in COVID-19 diagnosis
Manocha et al. [61], 2022	-	Mixed: X-ray, cough	dynamic fusion strategy	DL method: CNN	95.64 % accuracy in COVID-19 prediction

Table A6
. Summary of studies that apply multi-modality approaches for oncology.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Kanwal et al. [63], 2022	2797 patients	Mixed: Clinical, RNA, CNA	late fusion features	DL method: CNN-XGBOOST	96.3 % accuracy in multi-cancer prognosis
Karim et al. [62]. 2022	9074 patients	Mixed: Gene and miRNA expression	concatenation	DL method: MCAE model	96.25 % precision in cancer susceptibility prediction
Molina et al. [64], 2014	12 patients	Bioimaging: T2W, DCE-PF, DCE-MTT	interpolation	ML method: SVM	84.4 % sensitivity in prostate cancer classification
Le et al. [65], 2017	364 patients	Bioimaging: ADC, T2WI	concatenation	ML method: SVM	89.85 % sensitivity in prostate cancer diagnosis
Rossi et al. [66], 2021	890 subjects	Bioimaging: axial T2W, HBV, sagittal T2W	merging	DL method: CNN	0.83 AUC in prostate image retrieval
Mokni et al. [67], 2021	286 patients	Bioimaging: DCE MRI and MGs	concatenation	DL method: RBFNN	0.99 AUC in breast cancer detection
Yang et al. [79], 2022	250 subjects	Mixed: H&E images, clinical data	Feature fusion by MCB	DL method: batch normalization layer	0.72 AUC in prediction of breast cancer recurrence
Lu et al. [69], 2018	214 subjects	Mixed: MRI, survival data, histology, IDH, and 1p/19q status	concatenation	ML method: three-level binary classification model	93.2 % accuracy in glioma subtyping
Li et al. [70], 2019	274 patients	Bioimaging; MRI T1, T1c, T2	raw images	DL method: inception-based U-Net	88.5 % sensitivity in glioma segmentation
Wang et al. [71], 2021	121 patients	Bioimaging: MRI T2-Flair, T1Gd, T1, T2	concatenation	DL method: SGPNet	90.7 % sensitivity in glioma genotype predictions
Xiao et al. [72], 2022	24 patients	Bioimaging: White light imaging and fluorescence imaging	fusion-based	DL method: DLS-DARTS	0.843 AUC in glioma grading
Yang et al. [73], 2020	3090 patients	Bioimaging: B-mode US image, elastography	US features	DL method: MCDLM	90.01 % accuracy in thyroid nodule classification
Chen et al. [74], 2021	64 patients	Bioimaging: MRI ADC, DWI, T2w	feature fusion	DL method: CNN	73.6 % accuracy in prediction of pancreatic cancer mutation
Fu et al. [75], 2021	101 patients	Bioimaging: PET, CT	feature embedding	DL method: U-NET+MSAM	81.09 % sensitivity in lung cancer segmentation
Menegotto et al. [76], 2021	46 patients	Mixed: CT, EHR	concatenation	DL method: multimodal xception	86.9 % accuracy in hepatocellular carcinoma diagnosis
Gao et al. [77]. 2022	994 patients	Bioimaging: grayscale and color Doppler US images	features sharing module	DL method: Siamese U-Net	82.54 % accuracy in cervical cancer classification
Hao et al. [78], 2022	1061 patients	Mixed: 3D contrast-enhanced CT, discrete clinical variables	concatenation	DL method: CNN + multi-layer perceptron module	% in gastric cancer survival prediction
Ye et al. [79], 2022	122 patients	Mixed: Tumor (ROI) + T1-T1C-T2 (MRI sequence)	concatenation	ML method: random forest + MLP	0.88 AUC in germinomas classification

Table A7
. Summary of studies that apply multi-modality approaches for ophthalmology.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Chai et al. [80], 2021	4520 patients	Mixed: retinal image, medical indicators	integration module	DL method: Bayesian MLP	92.56 % accuracy and 0.9408 AUC in glaucoma diagnosis
Hervella et al. [81], 2021	1200 images	Bioimaging: retinography–angiography	image alignment	DL method: VGG-Net	0.9528 AUC in pathological myopia
Jin et al. [82], 2021	176 patients	Bioimaging: OCT & OCTA images	feature level fusion	DL method: UFNET	96.77 % accuracy in the detection of age-related macular degeneration
Liu et al [83], 2021	83 patients	Bioimaging: IR, OCT	raw images	DL method: ResNet-152	0.85 AUC in the prediction of visual impairment
Hervella et al. [84], 2022	1200 images	Bioimaging: retinography and fluorescein angiography	self-supervised pre-training approach	DL method: VGG	79.44 % accuracy in diabetic retinopathy grading

Table A8
. Summary of studies that apply multi-modality approaches for pediatric disorders.

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Petrozziello et al. [86], 2019	35000 births	Mixed: Uterine, FHR, Quality	multimodal CNN	DL method: MCNN	0.77 AUC in the detection of fetal compromise
Gao et al. [87], 2021	1823 patients	Mixed: abdominal radiographs, clinical data	concatenation	ML method: LightGBM	0.9414 AUC in the surgical prediction of necrotizing enterocolitis
Salekin et al. [88], 2021	45 neonates	Mixed: Audio, video	decision fusion	DL method: LSTM + spectrograms	79 % accuracy in neonatal pain assessment
Itai Guez et al. [89], 2022	240 children	Mixed: biochemical biomarkers and MRE scans	fusion model	ML method: random forest	0.84 AUC in the assessment of Crohn's disease

Table A9
. Summary of studies that apply multi-modality approaches for other studies (miscellaneous applications).

Author, year	Participant/data information	Multi-modality input	Features fusion	Methods	Findings/Results (%)
Milosevic et al. [90], 2016	56 patients	Biosignals: ACM, sEMG	concatenation	ML method: SVM	91 % accuracy in seizure detection
Yao et al. [91], 2016	92 subjects	Biosignals: Heart rate, respiration rate	concatenation	ML method: kNN	93 % sensitivity in infection screening
Tiulpin et al. [92], 2019	2129 subjects	Mixed: radiographic, clinical data	tree gradient boosting	DL method: CNN	0.79 AUC in osteoarthritis progression prediction
Huang et al. [93], 2020	1837 subjects	Mixed: CT, EMR/EHR	late elastic average fusion	DL method: PENet + NN	0.947 AUC in pulmonary embolism classification
Xue at al. [94], 2020	466 patients	Bioimaging: ultrasound image, elastography	combine as a single image top and bottom	DL method: CNN	0.95 AUC in liver fibrosis staging
Subramaniam et al. [95], 2021	67 subjects	Biosignals: ECG, EDA	concatenation	DL method: CNN-LSTM	91.43 % average accuracy in pain monitoring
Tang et al. [97], 2021	1010 patients	Mixed: CT + structured data	aggregation	DL method: CNN + XGBoost	93.6 % accuracy in lung nodule detection
Mattia et al. [96], 2022	63 patients	Bioimaging: sMRI, fMRI	MRI indices	DL method: 3D CNN	96 % accuracy in coma patients
Ming et al. [98], 2022	2020 patients	Bioimaging: CT, PET volumes	adaptive fusion	DL method: DNN	84.3 % accuracy in cervical cancer classification
Wu et al. [99], 2022	1131 patients	Mixed: ultrasound images and clinical records	LSTM	DL method: deep multimodal learning network	0.973 AUC in lymph node metastasis prediction

References

[1] F. Zhang, Z. Li, B. Zhang, H. Du, B. Wang, X. Zhang, Multi-modal deep learning model for auxiliary diagnosis of Alzheimer's disease, Neurocomputing 361 (2019) 185–195, https://doi.org/10.1016/j.neucom.2019.04.093.

[2] O. Faust, Y. Hagiwara, T.J. Hong, O.S. Lih, U.R. Acharya, Deep learning for healthcare applications based on physiological signals: A review, Comput. Methods Programs Biomed. 161 (2018) 1–13, https://doi.org/10.1016/J.CMPB.2018.04.005. Jul.

[3] H.F. Nweke, Y.W. Teh, G. Mujtaba, M.A. Al-garadi, Data fusion and multiple classifier systems for human activity detection and health monitoring: Review and open research directions, Aktuel. Aspekte Kernfusionsforsch., Informationstag. 46 (2019) 147–170, https://doi.org/10.1016/J.INFFUS.2018.06.002. Mar.

[4] K. Doi, Diagnostic imaging over the last 50 years: Research and development in medical imaging science and technology, Phys. Med. Biol. 51 (13) (2006), https://doi.org/10.1088/0031-9155/51/13/R02. Jul. 07.

[5] C.S. Kruse, C. Kristof, B. Jones, E. Mitchell, A. Martinez, Barriers to electronic health record adoption: a systematic literature review, J. Med. Syst. 40 (12) (2016), https://doi.org/10.1007/s10916-016-0628-9. Dec.

[6] S.C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, M.P. Lungren, Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, NPJ Digital Med. 3 (1) (2020), https://doi.org/10.1038/s41746-020-00341-z. Nature Research, Dec. 01.

[7] F. Jiang, et al., Artificial intelligence in healthcare: past, present and future, Stroke Vasc. Neurol. 2 (4) (2017) 230–243, https://doi.org/10.1136/SVN-2017-000101. Dec.

[8] Y. Jiang, W. Li, M.S. Hossain, M. Chen, A. Alelaiwi, M. Al-Hammadi, A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition, Aktuel. Aspekte Kernfusionsforsch., Informationstag. 53 (2020) 209–221, https://doi.org/10.1016/J.INFFUS.2019.06.019. Jan.

[9] T. Searle, Z. Ibrahim, and R. Dobson, "Comparing natural language processing techniques for Alzheimer's dementia prediction in spontaneous speech," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.07358.

[10] N. Bhagwat, J.D. Viviano, A.N. Voineskos, M.M. Chakravarty, Modeling and prediction of clinical symptom trajectories in Alzheimer's disease using longitudinal data, PLoS Comput. Biol. 14 (9) (2018), https://doi.org/10.1371/journal.pcbi.1006376. Sep.

[11] Z.S. Syed, M.S.S. Syed, M. Lech, E. Pirogova, Automated recognition of Alzheimer's dementia using bag-of-deep-features and model ensembling, IEEE Access 9 (2021) 88377–88390, https://doi.org/10.1109/ACCESS.2021.3090321.

[12] D. Lahat, T. Adali, C. Jutten, Multimodal data fusion: an overview of methods, challenges, and prospects, in: Proceedings of the IEEE 103, Institute of Electrical and Electronics Engineers Inc., 2015, pp. 1449–1477, https://doi.org/10.1109/JPROC.2015.2460697. Sep. 01.

[13] D. Ramachandram, G.W. Taylor, Deep multimodal learning: a survey on recent advances and trends, IEEE Signal Process. Mag. 34 (6) (2017) 96–108, https://doi.org/10.1109/MSP.2017.2738401. Nov.

[14] T. Zhou, S. Ruan, S. Canu, A review: deep learning for medical image segmentation using multi-modality fusion, Array 3–4 (2019), https://doi.org/10.1016/j.array.2019.100004. Sep.

[15] B. Huang, F. Yang, M. Yin, X. Mo, C. Zhong, A review of multimodal medical image fusion techniques, in: Computational and Mathematical Methods in Medicine, 2020, Hindawi Limited, 2020, https://doi.org/10.1155/2020/8279342.

[16] C. Cui, et al., Deep multimodal fusion of image and non-image data in disease diagnosis and prognosis: a review, Progress Biomed. Eng. 5 (2) (2023), https://doi.org/10.1088/2516-1091/acc2fe. Apr.

[17] A.S. Kini, et al., Ensemble deep learning and internet of things-based automated COVID-19 diagnosis framework, Contrast Media Mol. Imaging 2022 (2022) 7377502, https://doi.org/10.1155/2022/7377502.

[18] A. Holzinger, B. Malle, A. Saranti, B. Pfeifer, Towards multi-modal causability with graph neural networks enabling information fusion for explainable AI, Aktuel. Aspekte Kernfusionsforsch., Informationstag. 71 (2021) 28–37, https://doi.org/10.1016/j.inffus.2021.01.008. Jul.

[19] P. Fu, J. Wang, X. Zhang, L. Zhang, R.X. Gao, Dynamic routing-based multimodal neural network for multi-sensory fault diagnosis of induction motor, J. Manuf. Syst. 55 (2020) 264–272, https://doi.org/10.1016/j.jmsy.2020.04.009. Apr.
[20] W. Liu, J.-L. Qiu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep canonical correlation analysis," Aug. 2019, [Online]. Available: http://arxiv.org/abs/1908.05349.
[21] D. Zhang, D. Shen, Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in Alzheimer's disease, Neuroimage 59 (2) (2012) 895–907, https://doi.org/10.1016/j.neuroimage.2011.09.069. Jan.
[22] S. Liu, et al., Multimodal neuroimaging feature learning for multiclass diagnosis of Alzheimer's disease, IEEE Trans. Biomed. Eng. 62 (4) (2015) 1132–1140, https://doi.org/10.1109/TBME.2014.2372011. Apr.
[23] J. Shi, X. Zheng, Y. Li, Q. Zhang, S. Ying, Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer's disease, IEEE J. Biomed. Health Inform 22 (1) (2018) 173–183, https://doi.org/10.1109/JBHI.2017.2655720. Jan.
[24] Y. Wang, et al., Diagnosis and prognosis of Alzheimer's disease using brain morphometry and white matter connectomes, Neuroimage Clin. 23 (2019), https://doi.org/10.1016/j.nicl.2019.101859. Jan.
[25] S. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A parameter-efficient deep learning approach to predict conversion from mild cognitive impairment to Alzheimer's disease, Neuroimage 189 (2019) 276–287, https://doi.org/10.1016/j.neuroimage.2019.01.031. Apr.
[26] O. Pelka, et al., Sociodemographic data and APOE-ε4 augmentation for MRI-based detection of amnestic mild cognitive impairment using deep learning systems, PLoS One 15 (no. 9) (2020), https://doi.org/10.1371/journal.pone.0236868. September, Sep.
[27] M.N. Muhammed, P. Thiyagarajan, Alzheimer's classification using dynamic ensemble of classifiers selection algorithms: a performance analysis, Biomed. Signal Process. Control 68 (2021), https://doi.org/10.1016/j.bspc.2021.102729. Jul.
[28] Y. Wang, X. Liu, C. Yu, Assisted diagnosis of alzheimer's disease based on deep learning and multimodal feature fusion, complex. 2021 (2021), https://doi.org/10.1155/2021/6626728.
[29] X. Song, M. Mao, X. Qian, Auto-metric graph neural network based on a meta-learning strategy for the diagnosis of alzheimer's disease, IEEE J. Biomed. Health Inform. 25 (8) (2021) 3141–3152, https://doi.org/10.1109/JBHI.2021.3053568. Aug.
[30] S. El-Sappagh, J.M. Alonso, S.M.R. Islam, A.M. Sultan, K.S. Kwak, A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease, Sci. Rep. 11 (1) (2021), https://doi.org/10.1038/s41598-021-82098-3. Dec.
[31] X. Gao, F. Shi, D. Shen, M. Liu, Task-induced pyramid and attention gan for multimodal brain image imputation and classification in Alzheimer's disease, IEEE J. Biomed. Health Inform 26 (1) (2022) 36–43, https://doi.org/10.1109/JBHI.2021.3097721. Jan.
[32] Y. Zhang, et al., Predicting AT(N) pathologies in Alzheimer's disease from blood-based proteomic data using neural networks, Front. Aging Neurosci. 14 (2022), https://doi.org/10.3389/fnagi.2022.1040001. Nov.
[33] L. Ilias, D. Askounis, Multimodal deep learning models for detecting dementia from speech and transcripts, Front. Aging Neurosci. 14 (2022), https://doi.org/10.3389/fnagi.2022.830943. Mar.
[34] S. Qiu, et al., Multimodal deep learning for Alzheimer's disease dementia assessment, Nat. Commun. 13 (1) (2022), https://doi.org/10.1038/s41467-022-31037-5. Dec.
[35] M. Velazquez, Y. Lee, Multimodal ensemble model for alzheimer's disease conversion prediction from early mild cognitive impairment subjects, Comput. Biol. Med. 151 (2022), https://doi.org/10.1016/j.compbiomed.2022.106201. Dec.
[36] M. Golovanevsky, C. Eickhoff, R. Singh, Multimodal attention-based deep learning for Alzheimer's disease diagnosis, J. Am. Med. Inform. Assoc. 29 (12) (2022) 2014–2022, https://doi.org/10.1093/JAMIA/OCAC168. Nov.
[37] S. El-Sappagh, F. Ali, T. Abuhmed, J. Singh, J.M. Alonso, Automatic detection of Alzheimer's disease progression: an efficient information fusion approach with heterogeneous ensemble classifiers, Neurocomputing 512 (2022) 203–224, https://doi.org/10.1016/j.neucom.2022.09.009. Nov.
[38] A. Papadopoulos, et al., Unobtrusive detection of Parkinson's disease from multi-modal and in-the-wild sensor data using deep learning techniques, Sci. Rep. 10 (1) (2020), https://doi.org/10.1038/s41598-020-78418-8. Dec.
[39] M.B. Makarious, et al., Multi-modality machine learning predicting Parkinson's disease, NPJ Parkinsons Dis. 8 (1) (2022), https://doi.org/10.1038/s41531-022-00288-w. Dec.
[40] R.A. Feis, et al., Single-subject classification of presymptomatic frontotemporal dementia mutation carriers using multimodal MRI, Neuroimage Clin. 20 (2018) 188–196, https://doi.org/10.1016/j.nicl.2018.07.014. Jan.
[41] P.H. Kassani, A. Gossmann, Y.P. Wang, Multimodal sparse classifier for adolescent brain age prediction, IEEE J. Biomed. Health Inform. 24 (2) (2020) 336–344, https://doi.org/10.1109/JBHI.2019.2925710. Feb.
[42] S.H. Kang, et al., Machine learning for the prediction of amyloid positivity in amnestic mild cognitive impairment, J. Alzheimers Dis. 80 (1) (2021) 143–157, https://doi.org/10.3233/JAD-201092. Jan.
[43] W. Ko, W. Jung, E. Jeon, H. Il Suk, A deep generative-discriminative learning for multimodal representation in imaging genetics, IEEE Trans. Med. Imaging 41 (9) (2022) 2348–2359, https://doi.org/10.1109/TMI.2022.3162870. Sep.
[44] M.J. Patel, C. Andreescu, J.C. Price, K.L. Edelman, C.F. Reynolds, H.J. Aizenstein, Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction, Int. J. Geriatr. Psychiatry 30 (10) (2015) 1056–1067, https://doi.org/10.1002/gps.4262. Oct.
[45] X. Ding, X. Yue, R. Zheng, C. Bi, D. Li, G. Yao, Classifying major depression patients and healthy controls using EEG, eye tracking and galvanic skin response data, J. Affect. Disord. 251 (2019) 156–161, https://doi.org/10.1016/j.jad.2019.03.058. May.
[46] J. Zhu, et al., Multimodal mild depression recognition based on EEG-EM synchronization acquisition network, IEEE Access 7 (2019) 28196–28210, https://doi.org/10.1109/ACCESS.2019.2901950.
[47] J. Han, G. Jiang, G. Ouyang, X. Li, A multimodal approach for identifying autism spectrum disorders in children, IEEE Trans. Neural Syst. Rehabil. Eng. 30 (2022) 2003–2011, https://doi.org/10.1109/TNSRE.2022.3192431.
[48] S. Jiang, F. Firouzi, K. Chakrabarty, E.B. Elbogen, A resilient and hierarchical iot-based solution for stress monitoring in everyday settings, IEEE Internet Things J. 9 (12) (2022) 10224–10243, https://doi.org/10.1109/JIOT.2021.3122015. Jun.
[49] J. Pan, H. Lin, Y. Dong, Y. Wang, Y. Ji, MAMF-GCN: multi-scale adaptive multi-channel fusion deep graph convolutional network for predicting mental disorder, Comput. Biol. Med. 148 (2022), https://doi.org/10.1016/j.compbiomed.2022.105823. Sep.
[50] M.A. Rahaman, et al., Deep multimodal predictome for studying mental disorders, Hum. Brain Mapp. 44 (2) (2023) 509–522, https://doi.org/10.1002/hbm.26077. Feb.
[51] R. Soundararajan, et al., Deeply trained real-time body sensor networks for analyzing the symptoms of Parkinson's disease, IEEE Access 10 (2022) 63403–63421, https://doi.org/10.1109/ACCESS.2022.3181985.
[52] A. Sano, W. Chen, D. Lopez-Martinez, S. Taylor, R.W. Picard, Multimodal ambulatory sleep detection using LSTM recurrent neural networks, IEEE J. Biomed. Health Inform. 23 (4) (2019) 1607–1617, https://doi.org/10.1109/JBHI.2018.2867619. Jul.
[53] G. Zhou, R. Li, S. Zhang, J. Wang, J. Ma, Multimodal sleep signals-based automated sleep arousal detection, IEEE Access 8 (2020) 106157–106164, https://doi.org/10.1109/ACCESS.2020.3000272.
[54] Z. Jia, X. Cai, G. Zheng, J. Wang, Y. Lin, SleepPrintNet: a multivariate multimodal neural network based on physiological time-series for automatic sleep staging, IEEE Trans. Artif. Intell. 1 (3) (2020) 248–257, https://doi.org/10.1109/TAI.2021.3060350. Dec.
[55] B. Fatimah, A. Singhal, P. Singh, A multi-modal assessment of sleep stages using adaptive fourier decomposition and machine learning, Comput. Biol. Med. 148 (2022), https://doi.org/10.1016/j.compbiomed.2022.105877. Sep.
[56] E. Puyol-Anton, et al., Regional multi-view learning for cardiac motion analysis: application to identification of dilated cardiomyopathy patients, IEEE Trans. Biomed. Eng. 66 (4) (2019) 956–966, https://doi.org/10.1109/TBME.2018.2865669. Apr.
[57] H. Zhang, et al., Detection of coronary artery disease using multi-modal feature fusion and hybrid feature selection, Physiol. Meas. 41 (11) (2020), https://doi.org/10.1088/1361-6579/abc323. Dec.
[58] R.B. Kim, et al., Prediction of postoperative cardiac events in multiple surgical cohorts using a multimodal and integrative decision support system, Sci. Rep. 12 (1) (2022), https://doi.org/10.1038/s41598-022-15496-w. Dec.
[59] Y. Chen, et al., A multimodality machine learning approach to differentiate severe and nonsevere COVID-19: model development and validation, J. Med. Internet Res. 23 (4) (2021), https://doi.org/10.2196/23948. Apr.
[60] W. Zheng, et al., Pay attention to doctor–patient dialogues: multi-modal knowledge graph attention image-text embedding for COVID-19 diagnosis, Aktuel. Aspekte Kernfusionsforsch., Informationstag. 75 (2021) 168–185, https://doi.org/10.1016/j.inffus.2021.05.015. Nov.
[61] A. Manocha, M. Bhatia, A novel deep fusion strategy for COVID-19 prediction using multimodality approach, Comput. Electr. Eng. 103 (2022), https://doi.org/10.1016/j.compeleceng.2022.108274. Oct.
[62] M.R. Karim, T. Islam, C. Lange, D. Rebholz-Schuhmann, S. Decker, Adversary-aware multimodal neural networks for cancer susceptibility prediction from multiomics data, IEEE Access 10 (2022) 54386–54409, https://doi.org/10.1109/ACCESS.2022.3175816.
[63] S. Kanwal, F. Khan, S. Alamri, A multimodal deep learning infused with artificial algae algorithm – An architecture of advanced E-health system for cancer prognosis prediction, J. King Saud. Univ. Comput. Inf. Sci. 34 (6) (2022) 2707–2719, https://doi.org/10.1016/j.jksuci.2022.03.011. Jun.
[64] J.F. García Molina, L. Zheng, M. Sertdemir, D.J. Dinter, S. Schönberg, M. Rädle, Incremental learning with SVM for multimodal classification of prostatic adenocarcinoma, PLoS One 9 (4) (2014), https://doi.org/10.1371/journal.pone.0093600. Apr.
[65] M.H. Le, et al., Automated diagnosis of prostate cancer in multi-parametric MRI based on multimodal convolutional neural networks, Phys. Med. Biol. 62 (16) (2017) 6497–6514, https://doi.org/10.1088/1361-6560/aa7731. Jul.
[66] A. Rossi, M. Hosseinzadeh, M. Bianchini, F. Scarselli, H. Huisman, Multi-modal siamese network for diagnostically similar lesion retrieval in prostate MRI, IEEE Trans. Med. Imaging 40 (3) (2021) 986–995, https://doi.org/10.1109/TMI.2020.3043641. Mar.
[67] R. Mokni, N. Gargouri, A. Damak, D. Sellami, W. Feki, Z. Mnif, An automatic computer-aided diagnosis system based on the multimodal fusion of breast cancer (MF-CAD, Biomed. Signal Process Control. 69 (2021), https://doi.org/10.1016/j.bspc.2021.102914. Aug.
[68] J. Yang, et al., Prediction of HER2-positive breast cancer recurrence and metastasis risk from histopathological images and clinical information via

multimodal deep learning, Comput. Struct. Biotechnol. J. 20 (2022) 333–342, https://doi.org/10.1016/j.csbj.2021.12.028. Jan.
[69] C.F. Lu, et al., Machine learning–based radiomics for molecular subtyping of gliomas, Clin. Cancer Res. 24 (18) (2018) 4429–4436, https://doi.org/10.1158/1078-0432.CCR-17-3445. Sep.
[70] H. Li, A. Li, M. Wang, A novel end-to-end brain tumor segmentation method using improved fully convolutional networks, Comput. Biol. Med. 108 (2019) 150–160, https://doi.org/10.1016/j.compbiomed.2019.03.014. May.
[71] Y. Wang, Y. Wang, C. Guo, S. Zhang, L. Yang, SGPNet: a three-dimensional multitask residual framework for segmentation and IDH genotype prediction of gliomas, Comput. Intell. Neurosci. 2021 (2021) 1–9, https://doi.org/10.1155/2021/5520281.
[72] A. Xiao, et al., Intraoperative glioma grading using neural architecture search and multi-modal imaging, IEEE Trans. Med. Imaging 41 (10) (2022) 2570–2581, https://doi.org/10.1109/TMI.2022.3166129. Oct.
[73] W. Yang, et al., Integrate domain knowledge in training multi-task cascade deep learning model for benign–malignant thyroid nodule classification on ultrasound images, Eng. Appl. Artif. Intell. 98 (2021), https://doi.org/10.1016/j.engappai.2020.104064. Feb.
[74] X. Chen, X. Lin, Q. Shen, X. Qian, Combined spiral transformation and model-driven multi-modal deep learning scheme for automatic prediction of TP53 mutation in pancreatic cancer, IEEE Trans. Med. Imaging 40 (2) (2021) 735–747, https://doi.org/10.1109/TMI.2020.3035789. Feb.
[75] X. Fu, L. Bi, A. Kumar, M. Fulham, J. Kim, Multimodal spatial attention module for targeting multimodal PET-CT lung tumor segmentation, IEEE J. Biomed. Health Inform. 25 (9) (2021) 3507–3516, https://doi.org/10.1109/jbhi.2021.3059453, sep.
[76] A.B. Menegotto, C.D.L. Becker, S.C. Cazella, Computer-aided diagnosis of hepatocellular carcinoma fusing imaging and structured health data, Health Inf. Sci. Syst. 9 (1) (2021), https://doi.org/10.1007/s13755-021-00151-x. Dec.
[77] Y. Gao, X. Fu, Y. Chen, C. Guo, J. Wu, Post-pandemic healthcare for COVID-19 vaccine: tissue-aware diagnosis of cervical lymphadenopathy via multi-modal ultrasound semantic segmentation, Appl. Soft. Comput. 133 (2023), 109947, https://doi.org/10.1016/j.asoc.2022.109947. Jan.
[78] D. Hao, et al., SurvivalCNN: a deep learning-based method for gastric cancer survival prediction using radiological imaging data and clinicopathological variables, Artif. Intell. Med. 134 (2022), 102424, https://doi.org/10.1016/j.artmed.2022.102424. Dec.
[79] N. Ye, Q. Yang, P. Liu, Z. Chen, X. Li, A comprehensive machine-learning model applied to MRI to classify germinomas of the pineal region, Comput. Biol. Med. 152 (2023), 106366, https://doi.org/10.1016/j.compbiomed.2022.106366. Jan.
[80] Y. Chai, Y. Bian, H. Liu, J. Li, J. Xu, Glaucoma diagnosis in the Chinese context: an uncertainty information-centric Bayesian deep learning model, Inf. Process Manag. 58 (2) (2021) 102454, https://doi.org/10.1016/j.ipm.2020.102454.
[81] Á.S. Hervella, J. Rouco, J. Novo, M. Ortega, Self-supervised multimodal reconstruction pre-training for retinal computer-aided diagnosis, Expert Syst. Appl. 185 (115598) (2021), https://doi.org/10.1016/j.eswa.2021.115598.
[82] K. Jin, et al., Multimodal deep learning with feature level fusion for identification of choroidal neovascularization activity in age-related macular degeneration, Acta Ophthalmol. 100 (2) (2022) e512–e520, https://doi.org/10.1111/aos.14928. Mar.
[83] T.Y.A. Liu, C. Ling, L. Hahn, C.K. Jones, C.J.F. Boon, M.S. Singh, Prediction of visual impairment in retinitis pigmentosa using deep learning and multimodal fundus images, Br. J. Ophthalmol. (2023) 1–6, https://doi.org/10.1136/BJO-2021-320897, 0May.
[84] Á.S. Hervella, J. Rouco, J. Novo, M. Ortega, Multimodal image encoding pre-training for diabetic retinopathy grading, Comput. Biol. Med. 143 (2022), https://doi.org/10.1016/j.compbiomed.2022.105302. Apr.
[85] S.K. Khare, S. March, P.D. Barua, V.M. Gadre, U.R. Acharya, Application of data fusion for automated detection of children with developmental and mental disorders: a systematic review of the last decade, Aktuel. Aspekte Kernfusionsforsch., Informationstag. (2023), 101898, https://doi.org/10.1016/j.inffus.2023.101898. Nov.
[86] A. Petrozziello, C.W.G. Redman, A.T. Papageorghiou, I. Jordanov, A. Georgieva, Multimodal convolutional neural networks to detect fetal compromise during labor and delivery, IEEE Access 7 (2019) 112026–112036, https://doi.org/10.1109/ACCESS.2019.2933368.
[87] W. Gao, Y. Pei, H. Liang, J. Lv, J. Chen, W. Zhong, Multimodal AI system for the rapid diagnosis and surgical prediction of necrotizing enterocolitis, IEEE Access 9 (2021) 51050–51064, https://doi.org/10.1109/ACCESS.2021.3069191.
[88] M.S. Salekin, G. Zamzmi, D. Goldgof, R. Kasturi, T. Ho, Y. Sun, Multimodal spatio-temporal deep learning approach for neonatal postoperative pain assessment, Comput. Biol. Med. 129 (2021), 104150, https://doi.org/10.1016/j.compbiomed.2020.104150. Feb.
[89] I. Guez, et al., Development of a multimodal machine-learning fusion model to non-invasively assess ileal Crohn's disease endoscopic activity, Comput. Methods Programs Biomed. 227 (2022), https://doi.org/10.1016/j.cmpb.2022.107207. Dec.
[90] M. Milosevic, et al., Automated detection of tonic-clonic seizures using 3-d accelerometry and surface electromyography in pediatric patients, IEEE J. Biomed. Health Inform. 20 (5) (2016) 1333–1341, https://doi.org/10.1109/JBHI.2015.2462079. Sep.
[91] Y. Yao, G. Sun, T. Matsui, Y. Hakozaki, S. Van Waasen, M. Schiek, Multiple vital-sign-based infection screening outperforms thermography independent of the classification algorithm, IEEE Trans. Biomed. Eng. 63 (5) (2016) 1025–1033, https://doi.org/10.1109/TBME.2015.2479716. May.
[92] A. Tiulpin, et al., Multimodal machine learning-based knee osteoarthritis progression prediction from plain radiographs and clinical data, Sci. Rep. 9 (1) (2019) 20038, https://doi.org/10.1038/s41598-019-56527-3.
[93] S.C. Huang, A. Pareek, R. Zamanian, I. Banerjee, M.P. Lungren, Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection, Sci. Rep. 10 (1) (2020) 22147, https://doi.org/10.1038/s41598-020-78888-w.
[94] L.Y. Xue, et al., Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis, Eur. Radiol. 30 (5) (2020) 2973–2983, https://doi.org/10.1007/s00330-019-06595-w. May.
[95] S.D. Subramaniam, B. Dass, Automated nociceptive pain assessment using physiological signals and a hybrid deep learning network, IEEE Sens. J. 21 (3) (2021) 3335–3343, https://doi.org/10.1109/JSEN.2020.3023656. Feb.
[96] G.M. Mattia, et al., Multimodal MRI-based whole-brain assessment in patients in anoxoischemic coma by using 3D convolutional neural networks, Neurocrit. Care 37 (2022) 303–312, https://doi.org/10.1007/s12028-022-01525-z. Aug.
[97] N. Tang, et al., Improving the performance of lung nodule classification by fusing structured and unstructured data, Aktuel. Aspekte Kernfusionsforsch., Informationstag. 88 (2022) 161–174, https://doi.org/10.1016/j.inffus.2022.07.019. Dec.
[98] Y. Ming, X. Dong, J. Zhao, Z. Chen, H. Wang, N. Wu, Deep learning-based multimodal image analysis for cervical cancer detection, Methods 205 (2022) 46–52, https://doi.org/10.1016/j.ymeth.2022.05.004. Sep.
[99] X. Wu, M. Li, X.W. Cui, G. Xu, Deep multimodal learning for lymph node metastasis prediction of primary thyroid cancer, Phys. Med. Biol. 67 (3) (2022), https://doi.org/10.1088/1361-6560/ac4c47. Feb.
[100] T. Khan Mohd, N. Nguyen, A. Y. Javaid, and S. Masood, "Multi-modal data fusion in enhancing human-machine interaction for industry 4.0 robotic applications: a survey".
[101] R. V. Marinescu et al., "TADPOLE challenge: prediction of longitudinal evolution in alzheimer's disease," May 2018, [Online]. Available: http://arxiv.org/abs/1805.03909.
[102] I. Bos, et al., The EMIF-AD Multimodal Biomarker Discovery study: design, methods and cohort characteristics, Alzheimers Res. Ther. 10 (1) (2018), https://doi.org/10.1186/S13195-018-0396-5. Jul.
[103] C.J. Aine, et al., Multimodal neuroimaging in schizophrenia: description and dissemination, Neuroinformatics 15 (4) (2017) 343–364, https://doi.org/10.1007/S12021-017-9338-9. Oct.
[104] B.M. Adhikari, et al., Functional network connectivity impairments and core cognitive deficits in schizophrenia, Hum. Brain Mapp. 40 (16) (2019) 4593–4605, https://doi.org/10.1002/HBM.24723. Nov.
[105] A Di Martino, et al., The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism, Mol. Psychiatr. (2014), https://doi.org/10.1038/mp.2013.78.
[106] B. Kemp, A.H. Zwinderman, B. Tuk, H.A.C Kamphuisen, J.J.L Oberyé, Analysis of a sleep-dependent neuronal feedback loop: the slow-wave microcontinuity of the EEG, ,vol 47 (9) (2000) 1185–1194, https://doi.org/10.1109/10.867928.
[107] C. O'Reilly, N. Gosselin, J. Carrier, T. Nielsen, Montreal archive of sleep studies: an open-access resource for instrument benchmarking and exploratory research, J. Sleep Res. 23 (6) (2014) 628–635, https://doi.org/10.1111/JSR.12169. Dec.
[108] C. Tobon-Gomez, et al., A multimodal database for the 1st cardiac motion analysis challenge, in: Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges, 2012, https://doi.org/10.1007/978-3-642-28326-0_4.
[109] D. Bhattacharya, et al., Coswara: a respiratory sounds and symptoms dataset for remote screening of SARS-CoV-2 infection, Sci. Data 10 (1) (2023), https://doi.org/10.1038/s41597-023-02266-0. Dec.
[110] C. Curtis, et al., The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups, Nature 486 (7403) (2012) 346–352, https://doi.org/10.1038/nature10983. *2012 486:7403*Apr.
[111] M. Vallières, C.R. Freeman, S.R. Skamene, I. El Naqa, A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities, Phys. Med. Biol. 60 (2015) 5471–5496, https://doi.org/10.1088/0031-9155/60/14/5471.
[112] B.H. Menze, et al., The multimodal brain tumor image segmentation benchmark (BRATS), IEEE Trans. Med. Imaging 34 (10) (2015) 1993–2024, https://doi.org/10.1109/TMI.2014.2377694. Oct.
[113] S. Hajeb Mohammad Alipour, H. Rabbani, M.R. Akhlaghi, Diabetic retinopathy grading by digital curvelet transform, Comput. Math. Methods Med. 2012 (2012), 761901, https://doi.org/10.1155/2012/761901.
[114] Walter Steffen et al., "The BioVid heat pain database," 2013. doi: 10.1109/CYBConf.2013.6617456.
[115] X. Gao, F. Shi, D. Shen, M. Liu, T. Alzheimer's disease, "task-induced pyramid and attention gan for multimodal brain image imputation and classification in alzheimer's disease, IEEE J. Biomed. Health Inform. 26 (1) (2022), https://doi.org/10.1109/JBHI.2020.3097721.
[116] A. Serra, P. Galdi, R. Tagliaferri, Multiview learning in biomedical applications, Artif. Intel. Age Neural Networks Brain Comput. (2024) 307–324, https://doi.org/10.1016/B978-0-323-96104-2.00010-5. Jan.
[117] Bharath Shankar and, Carol Anne Hargreave, Adversarial Learning for Improved Patient Representations, 13891, Springer Nature Switzerland, Cham, 2023, https://doi.org/10.1007/978-3-031-32883-1 vol. 13891. in Lecture Notes in Computer Science.
[118] J.-W. Baek, K. Chung. Accident Risk Prediction Model Based on Attention-Mechanism LSTM Using Modality Convergence in Multimodal 27, Springer-Verlag, 2021, pp. 1179–1189.

[119] S. Seoni, V. Jahmunah, M. Salvi, P.D. Barua, F. Molinari, U.R. Acharya, Application of uncertainty quantification to artificial intelligence in healthcare: a review of last decade (2013–2023), Comput. Biol. Med. (2023), 107441, https://doi.org/10.1016/J.COMPBIOMED.2023.107441. Sep.

[120] M. Salvi, et al., Automated assessment of glomerulosclerosis and tubular atrophy using deep learning, Comput. Med. Imaging Graph. 90 (2021), 101930, https://doi.org/10.1016/J.COMPMEDIMAG.2021.101930. Jun.

[121] M. Salvi, F. Molinari, N. Dogliani, M. Bosco, Automatic discrimination of neoplastic epithelium and stromal response in breast carcinoma, Comput. Biol. Med. 110 (2019) 8–14, https://doi.org/10.1016/J.COMPBIOMED.2019.05.009. Jul.

[122] F. Veronese, et al., The role in teledermoscopy of an inexpensive and easy-to-use smartphone device for the classification of three types of skin lesions using convolutional neural networks, Diagnostics 11 (3) (2021) 451, https://doi.org/10.3390/DIAGNOSTICS11030451. *2021, Vol. 11, Page 451*Mar.

[123] U. Sait, et al., A deep-learning based multimodal system for Covid-19 diagnosis using breathing sounds and chest X-ray images, Appl. Soft. Comput. 109 (2021), https://doi.org/10.1016/j.asoc.2021.107522. Sep.

[124] S. Seoni, et al., Application of uncertainty quantification to artificial intelligence in healthcare: a review of last decade (2013–2023), Comput. Biol. Med. (2023), https://doi.org/10.1016/j.compbiomed.2023.107441. October.

[125] V. Jahmunah, et al., Uncertainty quantification in DenseNet model using myocardial infarction ECG signals, Comput. Methods Programs Biomed. 23 (2023), https://doi.org/10.1016/j.cmpb.2022.107308. February.