PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Acta Informatica Pragensia
2022, Volume 11, Issue 3, pp. 423–457
Special Issue: Sustainable Solutions for Internet of Things Using Artificial Intelligence and Blockchain in Future Networks
https://doi.org/10.18267/j.aip.202

Review Open Access

Comprehensive Review of Multimodal Medical Data Analysis: Open Issues and Future Research Directions

Shashank Shetty [1,2], Ananthanarayana V S [1], Ajit Mahale [3]

[1] Department of Information Technology, National Institute of Technology Karnataka, Mangalore-575025, Karnataka, India
[2] Department of Computer Science and Engineering, Nitte Mahalinga Adyanthaya Memorial Institute of Technology (NMAMIT), NITTE (Deemed to be University), Udupi-574110, India
[3] Department of Radiology, Kasturba Medical College, Mangalore, Manipal Academy of Higher Education, Manipal-575001, India

Corresponding author: Shashank Shetty (shashankshetty@nitte.edu.in)

Abstract

Over the past few decades, the enormous expansion of medical data has led to searching for ways of data analysis in smart healthcare systems. Acquisition of data from pictures, archives, communication systems, electronic health records, online documents, radiology reports and clinical records of different styles with specific numerical information has given rise to the concept of multimodality and the need for machine learning and deep learning techniques in the analysis of the healthcare system. Medical data play a vital role in medical education and diagnosis; determining dependency between distinct modalities is essential. This paper gives a gist of current radiology medical data analysis techniques and their various approaches and frameworks for representation and classification. A brief outline of the existing medical multimodal data processing work is presented. The main objective of this study is to spot gaps in the surveyed area and list future tasks and challenges in radiology. The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (or PRISMA) guidelines were incorporated in this study for effective article search and to investigate several relevant scientific publications. The systematic review was carried out on multimodal medical data analysis and highlighted advantages, limitations and strategies. The inherent benefit of multimodality in the medical domain powered with artificial intelligence has a significant impact on the performance of the disease diagnosis frameworks.

Keywords

AI; Big data analysis; Clinical recommendation system; Multimodality; Structured and unstructured healthcare data; Data extraction; Data classification; Data visualization.

Citation: Shetty, S., Ananthanarayana, V S, & Mahale, A. (2022). Comprehensive Review of Multimodal Medical Data Analysis: Open Issues and Future Research Directions. *Acta Informatica Pragensia*, 11(3), 423–457. https://doi.org/10.18267/j.aip.202

Special Issue Editors: Venkatachalam Kandasamy, University of Hradec Kralove, Czech Republic
Mohamed Abouhawwash, Mansoura University, Egypt
Nebojsa Bacanin, Singidunum University, Serbia

Academic Editor: Zdenek Smutny, Prague University of Economics and Business, Czech Republic

1 Introduction

To begin with, healthcare multimodal knowledge visualization and depiction have a remarkable space in the medical information research outlook. They guide the rich information for doctors and medical research scholars to identify significant new challenges. In general, medical data exist in various formats, mainly structured and unstructured. Medical data are obtained from multiple sources such as electronic health reports, medical reports, signals and 2D and 3D images (Iakovidis & Christos, 2012; Chen et al., 2017; Pai et al., 2021). It is necessary to systematically file and conventionally represent data to extract relevant knowledge for better efficacy in practical applications (Lawonn et al., 2017; Cai et al., 2019; Sultanum et al., 2023).

The first step towards intelligent healthcare system usage of multimodal data was taken in the 1990s, and only a few works were recognized in the initial stage. Moderately, multimodal data became one of the necessary probe cases in the healthcare system. To briefly discuss the history, van der Putten et al. (1995) designed a transparent framework for the physician to access multimodal data from echocardiography, Cathlab databases, hospital information systems and an ECG management system. The workstation was built on UNIX using C language, and an Interbase database and a CD ROM were used as storage modules. In the early stage, the multimodal data analysis concept was a fresh perception, and many hurdles were encountered in integrating and upgrading multimodal data. The multimodal data storage was bulky and expensive, causing significant storage challenges. Wood et al. (1998) proposed a multimodal information system to extract and generate information from the different repositories based on requirements. They aimed to reduce the difficulty in coordinating data between different information sources from a vast pool of domains. To annotate selected data, an object analyser from IntextInc was used. However, the results of the object annotation were not satisfactory. With time, there has been a steady rise in the usage of multimodal data in the healthcare system.

Later, a step ahead from storage to annotation of multimodal data was taken. Generally, medical specialists rely on comparison and interrelation among the data for more accurate clinical diagnosis and prediction. An et al. (2008) published their work on the visualization of multimodal data in the electronic health record. This was a later advance for classifying electronic data into numeric texts and images. Furthermore, the classified data were annotated in the proposed work. In 2010, data fusion, an idea for multimodal data retrieval from electroencephalogram (EEG), magnetic resonance imaging (MRI) and positron emission tomography (PET), was presented (Polikar et al., 2010). This work gave a new dimension to multimodal data processing in the healthcare system with a diverse ensemble classifier solution of 10% to 20% better accuracy than existing work. In 2013, many researchers gave a new insight into multimodal data in healthcare applications. Weibel et al. (2013) proposed an application that was built to analyse multimodal EHR data. The manual coding issues were reduced, and features such as audio track and gaze were augmented for outlined applications. Gradually, in 2016 many researchers started proposing their idea of open-source software for medical imaging to overcome the curse of dimensionality. A leap towards the convolutional neural network (CNN) classifier was observed in (Pinho & Costa, 2016). With the development and increase in storage and processing capacity, there has been a massive increase in healthcare data. Since 2016, significant research has been conducted for analysing big data in the healthcare domain pertaining to data obtained from heterogeneous sources (Amal et al., 2022; Kline et al., 2022; Rehman et al., 2022).

From the 1990s to 2022, a radical change in the growth of multimodal data in the healthcare system is noted. We moved from a technique of storing multimodal data to analysing multimodal data using machine learning, which can be coined as colossal progress (Fleury et al., 2010).

1.1 Big data analysis in healthcare

In recent years, there has been enormous growth in structured, unstructured and semi-structured data in various fields around the globe, including the healthcare industry. Collectively, these heterogeneous multimodal data generated are referred to as big data (Kumar & Singh, 2019; Fei et al., 2021). The notion of "big data" is not new but there has been a constant change in the way it is defined. Big data is an accumulation of data elements whose volume, velocity, variety and complexity necessitate design and development of new hardware or software that can collect, analyse and visualize the data (Lynch et al., 2008; Jacobs, 2009; White et al., 2012; Batko & Ślęzak, 2022). The health sector is the best example of how the characteristics of the data generated are related to the 4 V's of big data, namely volume, velocity, variety and veracity (Raghupathi & Raghupathi, 2014; Jindal et al., 2018; Lv et al., 2020) (see Figure 1).

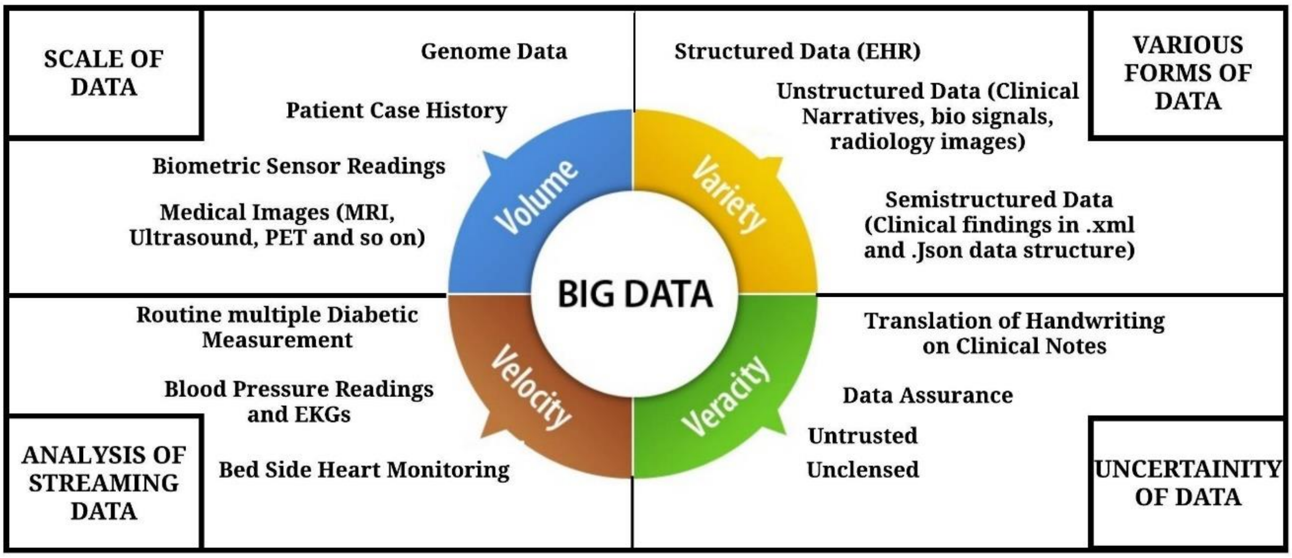

Figure 1. *Four V's of big data in healthcare.*

1.1.1 Big data and their 4 V's in healthcare

1. **Volume:** Over time, a significant amount of healthcare data would be generated and accumulated, producing a massive volume of data.
   Example: Big data in the medical field include patient case history or clinical trial data, medical genetics and genome data, medical images such as X-ray, ultrasound, MRI, etc., and a few modern types of big data such as 3D imaging and biometric sensor readings.
2. **Velocity:** Medical data are accumulated very quickly in real time, and the steady flow of new data stored at an unprecedented rate gives rise to new problems.
   Example: Routine monitoring such as multiple diabetic glucose measurements, blood pressure reading, ECG, bedside heart monitors, etc.
3. **Variety:** Health data are collected in a variety of formats: structured, semi-structured or unstructured (Raghupathi & Raghupathi, 2014; Lv et al., 2020; Batko & Ślęzak, 2022).
   - **Structured data**: Data that are in a standardized format collected in electronic health records (EHR), which contain patient demographics, clinical lab readings, diagnosis list, family list, smoking status, etc., that can be efficiently organized, queried, analysed, recollected and visualized by a machine.
   - **Unstructured data:** Data that do not follow an exact format with no associated data model. Example: Radiology images (MRI, X-ray, PET, etc.), handwritten clinical notes/discharge summaries, etc.

- **Semi-structured data:** Data that consist of semantic tags but do not adhere to the structure of specialized repositories such as a relational database. Example: Clinical narrative contents that use JSON or XML data structure.

4. **Veracity:** The fourth characteristic focuses on data assurance; that is, the outcome of the data obtained should be credible or error-free. Since the medical field is concerned with life and death decisions, accurate, trustworthy information is vital. Notably, in the case of unstructured data, it is challenging to obtain a precise result as they are highly fluctuating. Example: translation of handwriting on clinical notes (Tabassum et al., 2022).

1.1.2 Applications of big data in healthcare

The benefits of big data in healthcare are demonstrated in three focus areas: (1) disease prevention (Razzak et al., 2020), 2) recognition of risk factors of a disease (Nasution et al., 2022), and (3) designing an intervening or recommendation tool for health behaviour change (Miron-Shatz et al., 2014; Dash et al., 2019; Fei et al., 2021). Big data technology has many areas of application in healthcare, such as:

- **Clinical recommendation or decision support systems:** Clinical recommendation systems or decision support systems are mainly used to assist clinicians and hospital administrators to make better decisions from the insights gained from the medical records with volume, velocity (including time notion), variety (different forms of data) and veracity (data correctness) (Comito et al., 2022).
- **Public health:** Assessing the different diseases by analysing disease patterns and predicting the disease outbreaks. This helps in quicker development of resulting vaccines (Castiglione et al., 2021).
- **Evidence-based medicine:** Merging and analysing different forms of structured, unstructured and multimodal data from EHR, financial and operational data to predict the diagnosis for the clinical outcome (Abujaber et al., 2022).
- **Genomic analysis:** The process of studying and examining the different gene sequencing and molecular biology to identify the inherited gene disorders to support medical care decisions (Atta-Ur-Rahman et al., 2022).
- **Privacy and fraud analysis:** Reducing fraud by assessing and analysing the massive set of claim data; velocity of claim requests handled is an issue (Hossain et al., 2021, Haque & Tozal, 2022).
- **Disease surveillance or device monitoring:** Information-based activity involving capturing and analysing large real-time data originating from different origins, e.g., in-hospital and in-home devices for unfavourable event predictions (Taimoor & Rehman, 2022).

1.2 Clinical recommendation systems (CRS)

The Healthcare has massive volumes of data with time and variety attached to them, making it more complicated to handle all the above characteristics of data. There is a necessity to build a CRS that is devised to assist clinicians in making better decisions about a patient by handling various types of data with volume and velocity (Berner, 2010; Sreejith et al., 2022). The necessary characteristic of CRS is to assist in the investigation of diseases as misdiagnosis of conditions may cause adverse and risky effects (Institute of Medicine, 2000; Harada et al., 2021). A mistake in a diagnosis of a medical condition may be due to human error by ignoring minor details or physicians having lesser knowledge of the disease (Schiff et al., 2009; Sutton et al., 2020). Effective decision support systems could significantly resolve this problem (Garg et al., 2005; Hak et al., 2022). CRS should effectively interpret the patient's medical data and assist clinicians in making better decisions by a proper investigation of patient conditions prior the medical care. Clinical recommendation or decision support systems are used in various medical tasks. Some of the common uses are as follows:

- **Prognosis aid:** An expert and knowledge-based recommendation system can assist clinicians or practitioners with no experience to diagnose a patient in intricate cases (Dramburg et al., 2020).
- **Warning or reminder:** The CRS system could be linked to patient monitoring devices and can warn clinicians about any emergency or patient's condition (Chien et al., 2022).
- **Medicine recommendation:** The CRS can advise on prescription recommendations concerning drug-to-drug interaction and medicine overdose faults (Zhou et al., 2022).
- **Knowledge extraction:** The recommendation systems can find and extract suitable and precise data that would be used for the prognosis of specific diseases (Saxena et al., 2021).
- **Medical image identification and analysis**: The CRS can analyse from the available medical images such as X-Rays, ultrasound, PET or MRI and extract the relevant region of interest (RoI) for predictive modelling (Abubaker & Babayigit, 2022).

1.3 Structured and unstructured healthcare data

The rapid growth in healthcare data is with respect to the patient clinical traits, administrative and medical claim data and other various regulatory requirements. In 2009, Health Information Technology for Economic and Clinical Health (HITECH) Act was established to adopt EHR in the United States, which catered the incentives of $30 billion (Petersen, 2022). Adoption of EHR by office-based clinicians has increased dramatically from 21% in 2004 to 87% in 2022, while the adoption of essential EHR tripled from 11% in 2006 to 54% in 2022 (ONC, 2016)

EHR comprises a plethora of structured data such as (1) *numerical quantities:* patient demographics, clinical laboratory results such as height, weight and blood type; 2) *categorical values:* Current Procedural Terminology (CPT) procedures or International Classification of Disease (ICD) codes; (3) *date/time objects:* temporal events of birth or admission; as well as unstructured data such as (4) *natural language free text*, e.g., medical reports containing patient profiles, current health status, patient disease history and discharge summaries; (5) *medical images* such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI), etc. (Gehrmann et al., 2018). In structured EHR, complex processing is not required before statistical or machine learning tasks. However, the majority of data found today are in an unstructured format (Joseph et al., 2021). An EHR consists of extremely large volumes of valuable information and researchers have therefore worked to establish data-driven models (Alqahtani et al., 2022).

The widespread collection of clinical data and variety of data formats thus poses various challenges such as high uncertainty and missing values. This sets up a stage for development of effective and sophisticated techniques such as CRS or predictive analysis frameworks for providing helpful and enhanced information to clinicians. Predictive analysis is a part of data mining, where patterns from historical data are studied to predict an event or outcome (Sundararaman et al., 2018; Ramesh & Santhi, 2020). Predictive analysis has given a boost to many trends in the healthcare sector to help clinicians and patients by supporting better diagnosis or other medical tasks (Ros Maryana Sinaga & Putra 2022). EHR are used to extract disease diagnosis (Comito et al., 2022) and medication (Chen et al., 2020) with higher accuracy and lower expenses.

1.4 Multimodal healthcare data

In diverse disciplines, data about the same aspects or subjects can be captured in various modes under different conditions or through numerous experiments. The concept of "modality" refers to these kinds of data acquisition frameworks (Lahat et al., 2014; Acosta et al., 2022). A combined analysis of these fused multimodal data would promise a comprehensive prospect of a particular task and may provide new directions for an unanswered question during an analysis of unimodal data. AI has provided promising results in various applications such as speech recognition, natural image detection and recognition and language translation. However, AI in the healthcare domain is lagging behind due to the high complexity of unique features or signals in the medical cohort (Acosta et al., 2022). There has been an increase in wearable sensor usage, and the ability to capture and aggregate data from multiple sources has been improved due to the wealth of multimodal data. These data can add value to identification, prognosis and prevention of various diseases. Most of the current AI research work focuses on disease discovery, classification and prediction from single modality data. In contrast, clinicians utilize a wide variety of data extracted from multiple sources for diagnostic evaluation and treatment planning. However, the AI models built on multimodal data accessible to clinicians for prognostic assessment have shown promising results in disease detection and prediction compared to unimodal data (Nunes et al., 2019; Soenksen et al., 2022).

There are two basic multimodal fusion approaches for merging text and images: early fusion and late fusion methods (Huang et al., 2020a; Mohsen et al., 2020). Early fusion combines the text and image features into an individual vector, which is then supplied to the final classifier. The important advantage of early fusion is that it can exploit the correlation and communicate between the low-level features. Instead, late fusion utilizes the decision value from each model and then fuses them using different fusing techniques such as averaging, variance, voting schemes, etc., as shown in Figure 2.

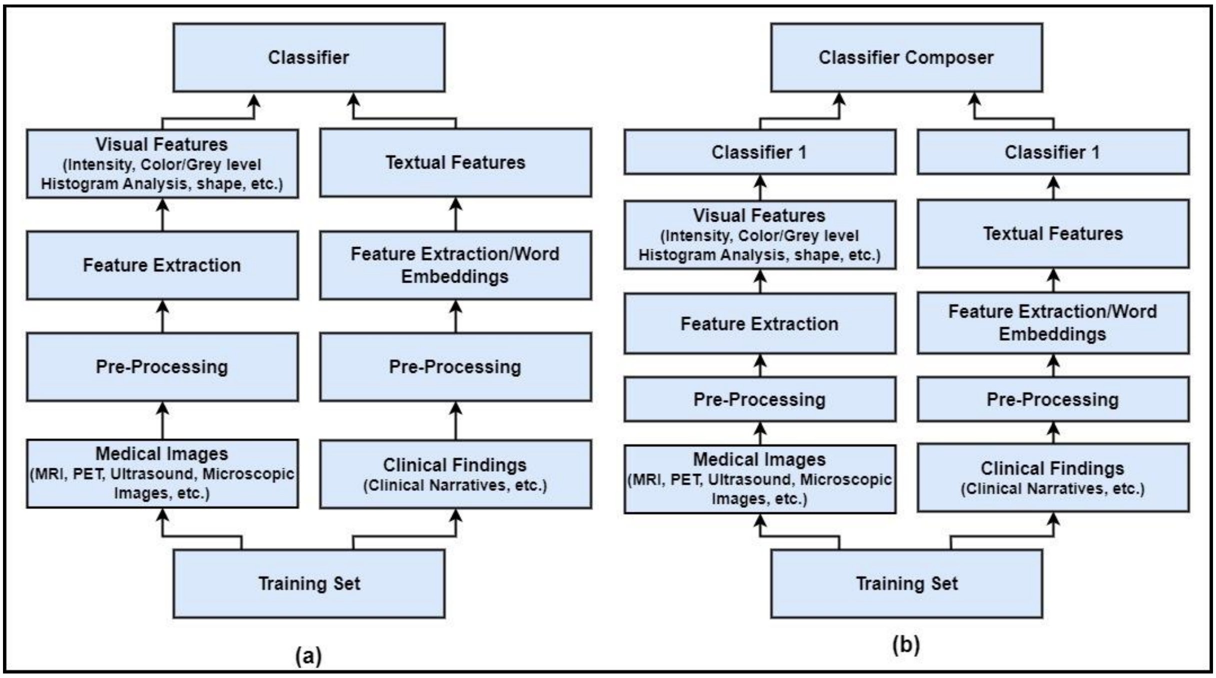

Figure 2. *Early fusion technique (a) and late fusion technique (b).*

After reviewing some of the survey papers on multimodal data in the healthcare system, we noticed that few works reviewed the ideas, guidelines, challenges and different data fusion methodologies. Topics such as advances in multimodal machine learning, early and late data fusion with parallel, non-parallel and hybrid data sets, and data fusion mechanisms on cross-border approach and border association were discussed (Jiang et al., 2020; Zhang et al., 2020).

1.5 Contribution

The contribution of this systematic review is as follows:

- A detailed overview of the multimodal medical data fundamentals is provided and essential characteristics are highlighted.
- A comprehensive review showcasing various multimodal medical data analyses, including feature extraction, data fusion, classification and visualization, is presented.
- The study discusses various data fusion strategies applied to heterogeneous medical data.

- A case study on AI-based radiology text and image analysis is explained with limitations and future research directions.

The paper is organized as follows: Section 2 presents the study methodology, including research questions, search strategy, selection criteria and data extraction. Section 3 mainly focuses on the methodology of multimodal medical data analysis and its current work. The radiology domain deals with a vast amount of multimodal data, including X-ray, CT, PET, unstructured clinical notes, radiology reports, ECG signals, etc. Hence, our survey focuses on data analysis in the radiology domain as a case study. In Section 4, different outcomes of radiology data and their analysis are outlined, and Section 5 presents various challenges and future work in multimodal medical analysis.

2 Methodology

The systematic review was carried out as per the PRISMA guidelines (Page et al., 2021). The main objective is to select eligible articles on multimodal medical data analysis and provide a summary of various open issues and future research directions. Our study mainly deals with outlining the existing research efforts and understanding the ideas put forth by the researchers. Therefore, no quality assessment was conducted on different multimodal medical data analysis techniques. The following subsections present various steps involved in performing the comprehensive review of multimodal medical data analysis.

2.1 Research questions

The various research questions posed during the systematic review are presented in Table 1.

Table 1. *Research questions considered for systematic review.*

No.	Research question (RQ)	Justifications	Sections and subsections dealing with the question
RQ1	What are the various stages, techniques and strategies used to analyse multimodal medical data?	A detailed study is carried out on the general architecture of multimodal medical data analysis comprising various stages, including feature extraction, classification, fusion and visualization.	Section 3 deals with RQ1
RQ2	What are the different data fusion strategies available to integrate multimodal medical data?	We identify and summarize research gaps in the existing literature on different data fusion techniques for integrating multimodal data and provide information to healthcare researchers on the usage of multimodal data fusion.	Section 3.2 deals with RQ2
RQ3	How has AI–based multimodal data analysis impacted on radiology domain?	The radiology domain deals with a vast amount of multimodal data, including X-ray, CT, PET, unstructured clinical notes, radiology reports, ECG signals, etc. Hence, the impact of various AI-based techniques involving radiology cohorts is discussed.	Section 4 deals with RQ3

2.2 Search strategy

Table 2 shows various search strings used for finding the research articles for this study. We have included search strings by selecting one keyword from each column, and multiple combinations were used during the search process. For example: “medical + data fusion + machine learning + radiology data” would be one search string. The search strings were used to find relevant articles from various repositories and publishers such as Web of Science, Google Scholar, Springer Link, Scopus, Elsevier, IEEE digital library, PUBMED and ACM.

Table 2. *Various search string keywords.*

Keyword 1	Keyword 2	Keyword 3	Keyword 4
("medical") OR ("medicine") OR ("health") OR ("healthcare")	("data fusion") OR ("multimodal") OR ("heterogeneous") OR ("multi-view") OR ("cross-modal")	("machine learning") OR ("deep learning") OR ("artificial intelligence")	("radiology data") OR ("medical imaging") OR ("electronic health records") OR ("medical data")

2.3 Selection criteria

The inclusion and exclusion criteria for including the research articles for the final review process are presented in Table 3. The inclusion criteria for selecting relevant articles include AI-based multimodal medical data analysis, feature extraction, classification, visualization and fusion of multimodal medical data. The inclusion criteria also include articles pertaining to radiology data (report + images). The following are the exclusion criteria: articles with no results, partial information, review articles or letters and publications with unimodal medical data analysis are excluded. A detailed overview of selection criteria for the systematic review process is shown in Figure 3.

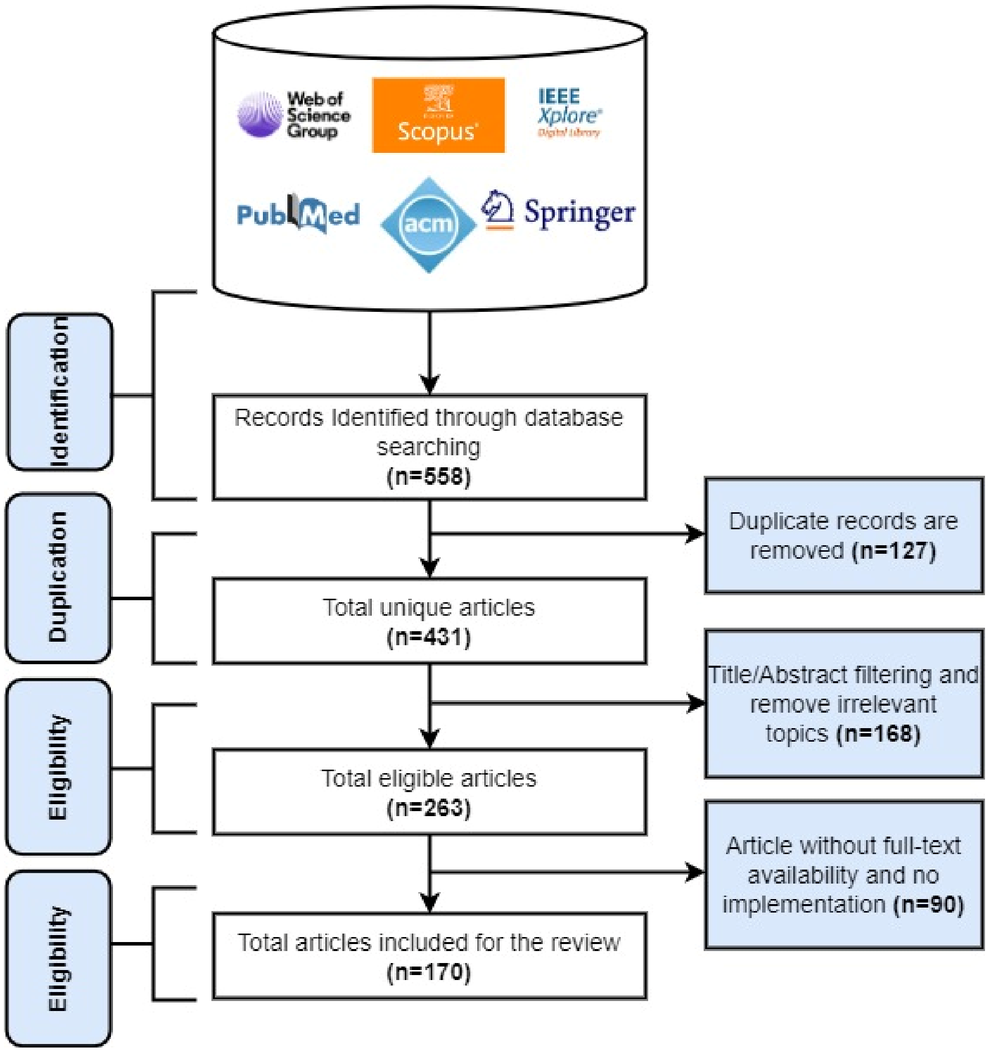

Figure 3. *PRISMA study inclusion process.*

Table 3. *Inclusion and exclusion criteria for review article selection.*

No.	Inclusion criteria	Exclusion criteria
1.	Original research articles published in peer-reviewed journals and conference papers	Articles not published in English.
2.	Articles published within the last 15 years (2007-2022)	Review articles or letters
3.	Articles relevant to AI-based multimodal medical analysis	Articles not dealing with multimodal analysis of medical data
4.	Publications matching the search strings	Duplicate publications and articles with no practical results

2.4 Data extraction

We selected a few design parameters to complete the data extraction process. The data extracted from the included research articles are as follows: first author name, year, methodology for disease diagnosis, modalities, clinical outcome, feature extraction techniques, fusion strategies, classification, visualization methodologies and evaluation metrics. The advantages, drawbacks and future research directions were showcased under particular extracted topics.

3 Multimodal Medical Data Analysis

Data analysis is a systematic process that includes well-designed techniques such as data inspection, cleaning, transformation and modelling. In general, multimodal data analysis uses a well-defined multimodal learning model designed by combining two deep Boltzmann machines with a hidden layer on the top. Trained convolutional neural network models exhibit excellent performance and outcome in analysing the enormous medical datasets (Richard et al., 2022). The general architecture of multimodal medical data analysis is depicted in Figure 4.

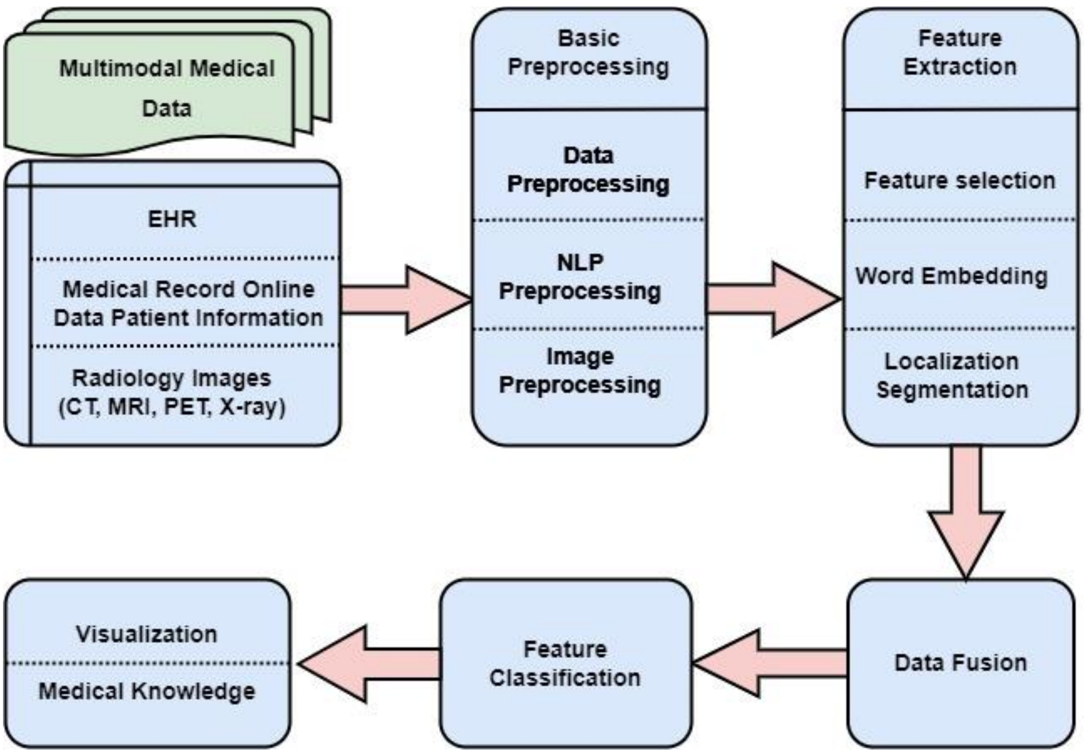

Figure 4. *General architecture of multimodal medical data analysis.*

3.1 Multimodal medical data extraction

With the trend of informatization and mobilization in the healthcare domain, the amount of high-precision multimodal medical data is multiplying. Making effective use of multimodal data helps analyse and solve many healthcare issues. Due to the numerous modalities of data such as text, image and signal, the effectiveness of data extraction algorithms is challenging. Data extraction is a process that includes fetching raw data from various data sources for further data processing and storage (Chaudhury et al., 2017). Existing work presents diverse approaches to extracting application data. In Iakovidis and Christos (2012), an unsupervised data mining approach is practised for mining low-level data with their multiple features, extracted in an extorted multimodal repository, which is represented in a consolidated way. Wang et al. (2018) proposed a novel approach to representing complex medical data into a knowledge-based graph model. Later, graph similarity search is applied to the knowledge graph and lazy learning algorithms, including dynamic time warping, are applied to find similarity between the graphs created. The proposed approach exhibits better accuracy compared to the baseline models.

Today, data are available in various variants from contrasting resources in the healthcare domain (Bleyer, 1997). The extraction and further processing of physiology data such as galvanic skin response, heart rate, facial expression, text and speech are achieved using various techniques such as pattern matching, similarity search, feature extraction, automated annotation, classification and clustering. Kurniawan and Pechenizkiy (2014) proposed a framework for stress analysis from multimodal affective data, such as physiological signals captured from sensors and external user data, including facial expression, speech and text. Pattern mining techniques were applied to extract features from various data models.

Data associated with cardiology can be extracted from different algorithms. The feature extraction can be accomplished by an approach of assigning the same data label for a similar data solution (Syeda-Mahmood et al., 2007). ECG signals can be auto-processed by detecting periodicity from ECG traces as images. The thick waveform peaks are considered the key features of this technique (Wang et al., 2009). Analysis of sub-cancer pixels from MRI and mammography images using a machine learning approach is quite popular. For feature extraction, the decision tree model, chi-square and automatic interaction detection are used (Wu et al., 2019).

The traditional machine learning approaches utilized manual guidance in extracting specific features before passing them to the fusion or classification stage. The advent of the deep learning approach has allowed the automated extraction of features from multimodal medical data. Purwar et al. (2020) utilized an AlexNet CNN model to extract features from red blood cell (RBC) imaging and structured blood reports to detect microcytic hypochromia. The CNN features are concatenated before passing through the classification model. Faris et al. (2021) proposed a multimodal framework for medical diagnosis from 263, 867 unstructured medical questions and structured symptom data. Term frequency (TF) and inverse document frequency (IDF), hashing vectorizer and doc2vec models were utilized to extract features from structured and unstructured data. The data extraction stage was followed by fusion and classification to predict disease diagnosis for telemedicine. Most deep learning feature extraction strategies apply CNN-based models for retrieving features from various data modalities (Hilmizen et al., 2020; Carvalho et al., 2021; Hamidinekoo et al., 2021). Table 4 shows an overview of various multimodal medical data extraction techniques.

Table 4. *Summary of multimodal medical data extraction techniques.*

Paper	Problem addressed	Dataset	Proposed methodology for data extraction	Disease analysed	Outcome
Wood et al. (1998)	Developing a multimodal information managing system	Texts, images, maps related to communicable diseases	Retrieval manager, key matching technique, HINT information extraction method	Communicable disease	Better efficiency compared to existing systems
Syeda-Mahmood et al. (2007)	Designing decision support system for cardiology	(i) ECG (ii) Echo videos (iii) PTB benchmark database	(i) ECG processing: Extraction of ECG waveforms using variants of normalized correlation (ii) Audio mode: envelopes extracted from sound signal, shape features extracted from envelopes (iii) Echo video processing: heart chambers extracted from variation of multiscale normalized graph (iv) Spatio-temporal motion from video: deformable template registration model	Cardiac problem	Not mentioned
Wang et al. (2009)	Heart period predicting medical decision support tool	ECG dataset from hospital	ECG envelope extraction in the form of image segments	Heart diseases	88% accuracy
Iakovidis & Christos (2012)	Model for mining multimodal information from healthcare information systems	Anonymised data from patients admitted in ICU. Ex: blood gasses, body temperature, chest X-ray	(i) Unsupervised data mining. (ii) Clustering: non-negative matrix factorization technique	Pneumonia	Promising results
Kurniawan & Pechenizkiy (2014)	Framework for stress analysis	Physiological signals captured by sensors i. environmental ii. context external user-related data	Pattern mining, feature extraction	Stress	Better results
Wang et al. (2018)	Knowledge graph-based method to connect different types of multimodal data	MIMIC III- public EMR database	Semantic rich knowledge graph. Medledge: Q & A based approach to build medical knowledge graph	Diabetes	62% accuracy
Bu et al. (2017)	3D feature learning framework for multimodal data	(i) SHREC 2007 watertight model (ii) SHREC 2011 non-rigid watertight dataset (iii) McGill shape benchmark	Convolutional neural networks, convolutional deep belief networks	No disease analysed (general usage of 3D hand gesture dataset)	Accuracy up to 94.5%

Paper	Problem addressed	Dataset	Proposed methodology for data extraction	Disease analysed	Outcome
Wu et al. (2019)	Prediction of molecular subtypes of breast cancer	MRI and mammography images	Decision tree model with the chi-squared automatic interaction detector (CHAID) algorithm	Breast cancer	Accuracy of decision tree was 74.1%
Purwar et al. (2020)	To predict microcytic hypochromia detection	(i) RBC images (ii) structured blood test reports	AlexNet CNN model for feature extraction	Hypochromia	Fused features provided better accuracy than the unimodal features
Faris et al. (2021)	Disease diagnosis from patient questions and symptoms	(i) unstructured patient questions (ii) structured symptoms and diagnoses	Tf-Idf, hashing vectorizer and doc2vec	Telemedicine	84.9% accuracy

3.2 Multimodal medical data fusion

Nowadays, in all disciplines, data are captured from distinct sources, which leads to multimodality. Fusion of multimodal data in the medical field can improve decision support systems with impressive results. Data fusion is a study of data sets from various sources communicating with each other (Lahat et al., 2015). The present study suggests that fusion of multimodal data enhances the performance of the methodology or algorithm applied in the data analysis phase. These techniques pose significant challenges in combining and analysing modes of different frequencies and noise. In general, the multimodal data analysis is achieved using a multimodal learning model, which will connect two deep Boltzmann machines with a hidden layer on the top. The model aims to identify missing information from the learnt information (Srivastava & Salakhutdinov, 2014).

Many researchers use data fusion as a proper technique in multimodal data analysis in various applications. Different data fusion approaches such as data fusion for hybrid BCI, rhythm-based BCI and fusion of multiple heartbeat physiological signals (Chandra et al., 2018) have been studied using comparative analysis (Fazli et al., 2015). Most multimodal data contain multiple views, and the multi-view classification of the various subset proves helpful. In Shachor et al. (2020), a novel fusion framework is designed using a neural network where a mixture of views is used for data processing, and proves to perform better. The multimodal medical data fusion also bolsters the 3D neuroanatomical database analysis. Identifying the anatomical structure, feature analysis and labelling approach are used by Barillot et al. (1993). Techniques such as a Markov-Penrose diagram of tensor network notation, Bayesian DAG and coupled matrix tensor factorization are likewise advised as reliable for fusion in the case of neuroimaging. In recent research, even fusion of multi-band images is brought about using a deep-gate convolutional neural network. Fusion of low and high-frequency components gives outstanding results compared to existing systems (Lin et al., 2020). It is shown that MRI, EEG and SMRI can also be fused using joint independent component analysis and transposed independent vector analysis models (Adali et al., 2015). For high spectral information MRI, PET fusion is advised (Abdulkareem, 2018), wherein a low multimodal dataset gives better results. One step ahead, multi-band image fusion offers a wide application in image quality enhancement; filters such as Gaussian (Mohd et al., 2017) and singular value decomposition (Nischitha & Padmavathi, 2017) perform well with fair results.

In general, the multimodal fusion concept is classified into two categories, (i) model agnostic approaches and (ii) model-based approaches. Further classification of multimodal fusion is shown in Figure 5. Model agnostic approaches indicate that any methods/models can be applied and treat the overall approach as a black box, mainly focusing on the fusion stage. Model-based approaches here focus primarily on the type of methods used for fusion. The model-based approaches are further subdivided into early fusion, late fusion and hybrid fusion techniques (Bayoudh et al., 2021). In early fusion or feature-level fusion techniques, the handcrafted features or features obtained from the neural network models are joined before passing on to the classification or prediction model. Li and Fan (2019) applied an early fusion strategy by concatenating the CNN features obtained from MRI and structured clinical test records to predict Alzheimer's disease. Purwar et al. (2020) proposed an early fusion strategy using concatenation to fuse red blood cell (RBC) imaging features and structured lab reports to detect hypochromia. Huang et al. (2020b) presented a neural network-based early fusion strategy by concatenating CNN features extracted from 2500 chest CT images and structured EHR data collected from the Stanford Medical Centre to predict pulmonary embolism. Most early fusion strategies focus on concatenating features obtained from CNN models to form a single vector. The major drawback of concatenation is missing inter-modal interaction between the various features obtained from multiple sources.

The late fusion strategy combines features obtained from multiple classifiers to produce the final prognostic outcome. Reda et al. (2018) presented a late fusion framework to detect early diagnosis of prostate cancer by fusing MRI and clinical biomarkers using the Stacked Nonnegativity Constraint Sparse Autoencoders (SNCSAE) technique. The individual features are passed through two classifiers, and the final fusion technique yields diagnostic probabilities. Faris et al. (2021) proposed a late fusion strategy including summation, ranking and multiplication to fuse unstructured features extracted from patient questionnaires and structured symptom data. Features extracted are separately passed through various machine learning models before being merged for prediction. Hamidinekoo et al. (2021) showcased a fusion of a deep convolutional network (DCN) feature obtained from MRI and whole slide imaging (WSI) pictures using a late fusion technique including majority voting. The existing late fusion strategies focus on combining features from various classifiers using averaging, maximum or majority voting. Similar to early fusion, the main disadvantage of late fusion is minimum intermodal interaction between features from various modalities.

The hybrid fusion strategy combines the early and late fusion approaches to obtain the results. A deep neural network (DNN) is a building block for the hybrid fusion strategy, where the joint representation of the multimodal data is learnt. In hybrid fusion, the loss is propagated back to the feature extraction stage during the training phase. The features from various modalities are learnt at intermediate layers of the neural network, and these learnt features are fused before feeding them into the final model for the prognostic outcome. Hilmizen et al. (2020) presented a joint fusion multimodal framework for predicting COVID-19 from chest X-ray (CXR) and CT images. The imaging features from the CXR and CT images were extracted using the VGG16 and ResNet50 models. The features were further concatenated and passed to the classification model for prediction. Carvalho et al. (2021) proposed a joint fusion strategy for skin cancer detection from dermoscopic images and ABCD pseudo non-imaging features from the ISIC 2017 challenge dataset. The EfficientNet B3 model was used for feature extraction, and concatenation was applied for the fusion. It is seen that the joint fusion strategies provide promising results compared to the early and late fusion techniques in various medical applications. It is also observed that multimodal prediction is superior to unimodal prediction. It is due to the fact that the complementary knowledge gained through the associated alternate features has significantly impacted on the overall performance of the model. Table 5 summarizes the multimodal medical fusion strategies.

Table 5. *Summary of multimodal medical data fusion techniques.*

Paper	Modality	Contribution	Methodology and classifier	Anatomy	Remarks
Barillot et al. (1993)	CT, MRI, DSA, PET, SPECT, MEG	Framework for interpretation of multimodal 3D neuro-anatomical databases	Image analysis methods: stereotactic framework	Brain disease	Final result not claimed
Fazil et al. (2015)	EEG, EMG, NIRS	Framework for sensorimotor rhythm-based BCIs	Novel approach hybrid BCI Data fusion: 1. ECG-EEG 2. EEG-NIRS	Brain disease	Not mentioned
Adali et al. (2015)	fMRI, sMRI, EEG	Application of two novel fusion models are discussed	Joint independent component analysis and transport vector analysis model in fully multivariate and symmetric manner.	Brain disease, schizophrenia	Fusion gave better result compared to fMRI and sMRI data fusion
Karahan et al. (2015)	Functional MRI, EEG, NIRS	Proposed model Markov-Penrose diagram for fusion of multimodal brain images	Coupled matrix-tensor factorization, multiway partial least squares	Brain disease	Not discussed
Bernal et al. (2018)	ISI dataset	Proposed a procedure for multimodal fusion for medical image using wearable sensors	Temporal fusion with multi-layered LSTM	Not mentioned	Achieved 90% accuracy for fused data
Denzil Bosco et al. (2017)	CT, MRI	Proposed a fusion technique for brain tumour images	Enhanced differentiated wavelet co-efficient technique	Brain tumours	The proposed system showcased fusing two medical images, CT and MRI. The combination of the two images provided promising results.
Mohd et al. (2017)	Speech and myoelectric signals	Approach to increase the efficiency of the myo band	Designed a robot with Arduino board	Myo band	95.92% accuracy
Nischitha & Padmavathi (2017)	MRI, PET scan images	Fusion approach for abdominal cancer	(i) Laplacian pyramid fusion rule. (ii) Multi-resolution singular value decomposition	Abdominal cancer	After fusion classification accuracy is 85%
Abdulkareem (2018)	MRI normal axial, MRI normal coronal and MRI Alzheimer	Proposed fusion using discrete wavelet transform	(i) Gaussian filter to enhance quality of fused images (ii) DWT	Alzheimer's brain disease	90% to 95% accuracy without losing spectral and anatomical data

Paper	Modality	Contribution	Methodology and classifier	Anatomy	Remarks
Chandra et al. (2018)	(i) ECG and BP signals of PhysioNet 2014 database (ii) lECG channels of MIT-BIH arrhythmia database	Proposed a heartbeat detector	(i) CNN-based information fusion (CIF) algorithm (ii) Fusing multiple signal without intermediate estimates	Cardiac disease	Achieved 94% to 99% accuracy
Lin et al. (2020)	(i) ImageNet IL SVRC2013 (ii) TNO image fusion dataset	Designed framework for fusion of multi-band images	(i) Deep stack convolutional neural network (ii) Hybrid MDR-DDR image fusion network	Not mentioned	PSNR achieved to 36% to 37%
Shachor et al. (2020)	DDSM dataset	Illustrated a fusion framework for classification of multi-view data	Probabilistic framework: A multi-view data fusion classification	Breast cancer	Cross-validation recall value 0.693
Purwar et al. (2020)	(i) RBC images (ii) Structured blood test reports	To predict microcytic hypochromia detection	AlexNet CNN model for feature extraction and concatenation for fusion	Hypochromia	Fused features provided better accuracy than unimodal features
Huang et al. (2020b)	(i) CT images (ii) Structured EHR	Framework to predict pulmonary embolism	Straightforward concatenation is applied to CNN features	Pulmonary embolism	Comparatively poor performance compared to the late fusion strategy
Faris et al. (2021)	(i) Unstructured patient questions (ii) Structured symptoms and diagnoses	Disease diagnosis from patient questions and symptoms	Late fusion techniques including ranking, summation and multiplication	Telemedicine	Multiplication fusion technique obtained superior performance of 84.9% accuracy
Hamidinekoo et al. (2021)	(i) MRI (ii) WSI	Deep learning framework for glioma detection	Late fusion strategy of majority voting	Brain tumour	Promising results for multimodal fusion
Carvalho et al. (2021)	(i) Dermoscopic images (ii) ABCD pseudo features	Multimodal fusion framework for skin cancer prediction	Joint fusion strategy using concatenating features extracted using EfficientNet B3	Skin cancer	Multi-tasking improved the performance of the prediction outcome
Mohammad et al. (2022)	CT, MRI, PET and SPECT	AI based multimodal medical image fusion model	(i) Modified DWT (ii) CNN network with hybrid optimization dynamic algorithm	Not mentioned	73% fusion factor and 58% standard deviation
Liu et al. (2022a)	CT and MRI	CNN based image fusion technique to preserve functional information	Multiscale mixed attention medical image fusion technique with mixed attention block and multiscale convolution block (encoder)	Brain disease	Achieved 0.8179 linear correlation

Paper	Modality	Contribution	Methodology and classifier	Anatomy	Remarks
Amal et al. (2022)	EMR and CT	Survey on existing work in multimodal data fusion techniques in cardiovascular medicine field	Survey work	Cardiovascular disease	Comparative analysis of existing work in fusion techniques

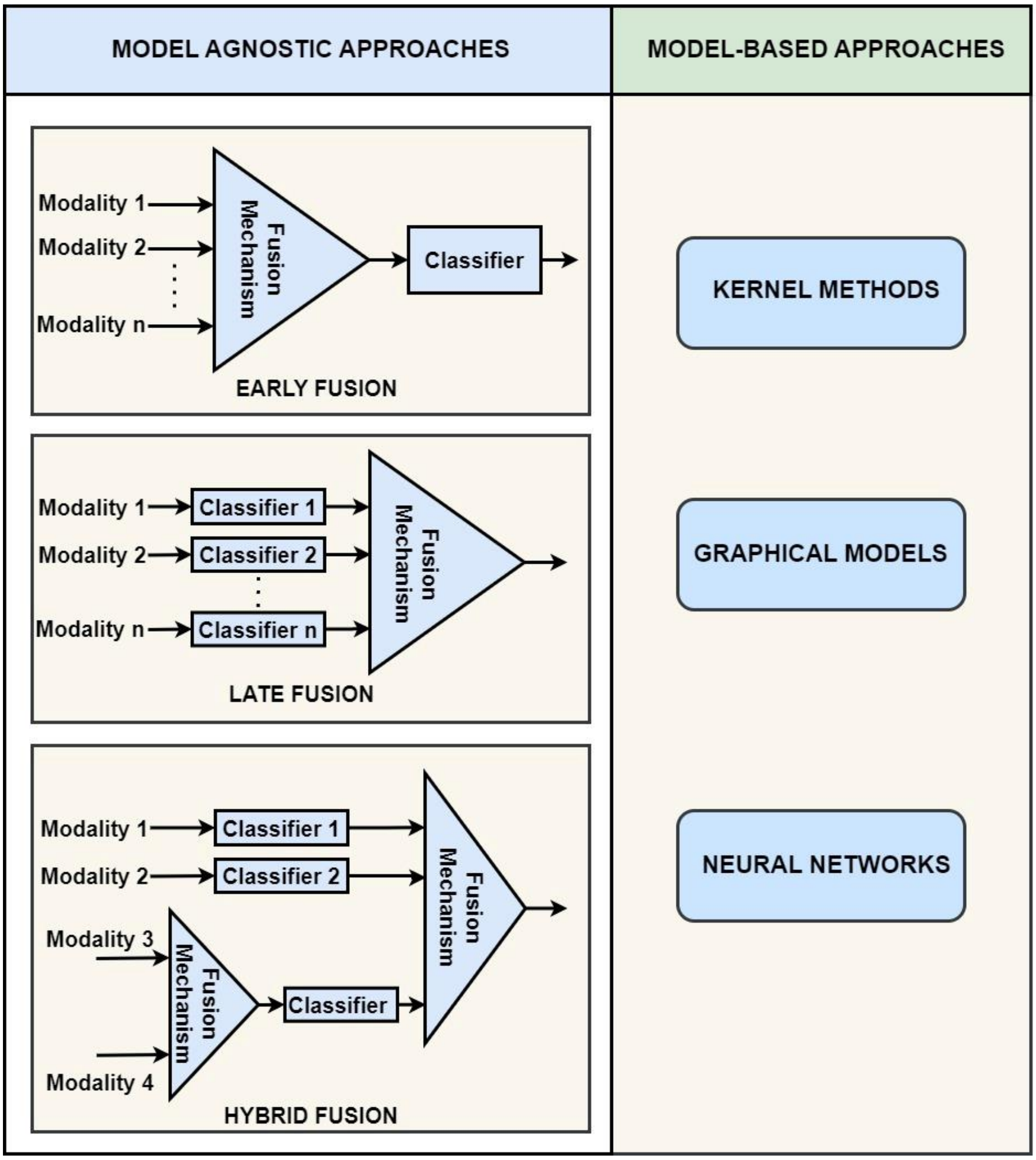

Figure 5. *Overview of multimodal medical data fusion.*

3.3 Multimodal medical data classification

The multimodal data fusion is followed by the classification and visualization task. The classification is a supervised learning approach used to determine the class of new data (Skowron et al., 2005). Data classification is essential in healthcare to analyse different diseases and categorize them accordingly. Multimodal medical data classification can be performed using machine learning and deep learning algorithms. Machine learning algorithms such as Ranklet transform, LSP Ranklet transform and support vector machine (SVM) have been used for breast cancer classification (Xi et al., 2017), whereas ECG, MRI and EEG signal compression classification can be achieved using an encoder-decoder layer followed by a least-square algorithm (Zhang et al., 2017).

Aydin et al. (2019) presented a multimodal deep learning-based binary classification of chest diseases from CXR and associated unstructured radiology reports collected from the Indiana University (IU) dataset. The imaging features were extracted using a pre-trained CNN model, and the textual features were retrieved using a GloVe embedding model. The concatenated features were passed through a fully connected network for classification. Lopez et al. (2020) compared the classification performance of a multimodal model and a unimodal model. The textual and imaging features were extracted using the word2vec and Densenet121 models from CXR and associated radiology reports collected from the IU dataset. The fused features were further passed to a fully connected deep neural network for classification. The outcome of the research also showed a reduction of annotation burden through multimodal learning.

Data classification is not restricted only to visual and textual data. Its application and benefits have been widespread in various fields. Data classification techniques play a significant role in designing a supporting system for Parkinson's patients based on their handwriting. Heidarivincheh et al. (2021) proposed a multimodal classification of Parkinson's disease (PD) in the home environment by extracting features from raw data acquired through a wrist-worn accelerator and RGB-D camera. The silhouette images and the accelerometer signal were passed through data preprocessing stages and classified into PD or healthy using an encoder-decoder CNN model.

Ribeiro et al. (2013) developed an approach to classifying chronic liver disease stages using clinical laboratory and ultrasound data. Techniques such as SVM, Bayes and K-mean clustering were used in the paper. Yi et al. (2022) proposed a multimodal classification framework for categorizing the severity of glaucoma from fundus and grayscale images collected from the Kunming Medical University. CNN-based classifiers were used for the classification task. Hilmizen et al. (2020) used a CNN-based classifier for classifying COVID-19 disease from multimodal CXR and CT features extracted using pre-trained VGG16 and ResNet models.

Multimodal classification has a wide range of applications in various medical domains. Machine learning and deep learning classifiers are applied to categorise data for prognosis outcomes. As we know, machine learning classifiers require supervised learning, which means that human interventions are much needed to manually pick features before passing them through the classifiers. DL classifiers do not require handcrafted features before feeding them into the fully connected layers for classification. Also, ML models do not learn on an incremental basis, and DL classifiers overcome this shortcoming by incrementally learning features. Table 6 presents a systematic review of multimodal medical data classification.

Table 6. *Review of multimodal medical data classification.*

Paper	**Modality**	**Contribution**	**Methodology and classifier**	**Anatomy**	**Remarks**
Bloch (1996)	MR images	Discussed concepts of dumpster-Shafer evidence theory	Dumpster-Shafer evidence theory	Brain diseases	Not mentioned
Valova & Kosugi (1997)	MR images	Classification approach for brain images using a neural network	MMPT neural network adopting depth first searching technique	Brain diseases	Accuracy achieved 89% for grey white block classification and 90.5% for border block classification

Paper	Modality	Contribution	Methodology and classifier	Anatomy	Remarks
Meng et al. (2010)	Three benchmarks from IICBU Biological Image Repository	Designed a framework to solve the challenges in general histology image classification and labelling	C-RSPM classifier	Liver disease	Accuracy achieved 92.70%
Zhang & Shen (2011)	(i) MR brain images (ii) ADNI database	Classification of AD patients data	Multimodal Laplacian regularized least squares technique	Alzheimer's disease (AD)	Accuracy achieved 98.5%
Ribeiro et al. (2013)	Clinical laboratory ultrasound data	Approach to classify and stage chronic liver disease	(i) Bayes Parzen classifier (ii) SVM (iii) K nearest neighbour	Chronic liver disease	Accuracy achieved 98.67% for normal detector
Wu & He (2015)	CT, MR, PET from Journals of Radiology and Radiographics	Tool for automatic classification of modality in medical image	p-norm multiple learning kernel technique	Not mentioned	Accuracy achieved 95.15%
Drotar et al. (2015)	Medical prescriptions	Designed framework for a decision support system for Parkinson's disease	SVM classifier with radial Gaussian kernel	Parkinson's disease	Accuracy achieved 88.13%
Xi et al. (2017)	Breast ultrasound Qianfushan hospital	Robust texture feature analysis medical multimodal data using LSP-Ranklet and multi-task learning	(i) SVM classifier (ii) LSP-Ranklet (iii) Multi-task learning	Breast Tumour	Efficiency achieved 96%
Ben Said et al. (2017)	(i) EMG and ECG signals (ii) DEAP Dataset	Designed compression and classification approach for EMG and ECG signals	Unimodal stacked autoencoder with intra correlation	Not mentioned	Accuracy achieved 78.1%
van der Voort et al. (2017)	MRI	Classification approach for 1p/19q status in presumed low-grade gliomas	Support vector machine	Not mentioned	95% confidence interval for sensitivity and specificity
Aydin et al. (2019)	X-ray and medical report Indiana University and Alfred Hospital	Classifier for multimodal medical data under a small dataset situation	CNN-based classifier	Chest disease	Accuracy increased by 4% to 7% compared to baseline models

Paper	Modality	Contribution	Methodology and classifier	Anatomy	Remarks
Lopez et al. (2020)	X-ray and medical report Indiana University and Alfred Hospital	Reducing data burden using multimodal learning	Densenet121 model for image feature extraction and word2vec for textual features and fully connected DNN for classification	Chest disease	Multimodal classification showed better accuracy than unimodal
Hilmizen et al. (2020)	CXR and CT images	Classification of COVID-19 disease	VGG16 and ResNet model for image feature extraction and CNN-based classifier was used	COVID-19	Significantly higher accuracy compared to baseline models
Heidarivincheh et. al. (2021)	(i) raw accelerometer signal (ii) silhouette images	Classification of PD	Encoder-decoder-based classification model	Parkinson's disease	Precision of 60% achieved by proposed model
Qu & Xiao (2022)	MRI image – RSNA MICCAI dataset	Classification model for multimodal data using deep learning approach	(i) Lite attention mechanism (ii) Recurrent neural network model	Brain tumour	3% better accuracy than existing systems
Yi et al. (2022)	Fundus and grayscale images	Multimodal classification of severity of glaucoma	CNN-based classifier	Glaucoma severity	97% accuracy by proposed model

3.4 Multimodal medical data visualization

Data visualization is the practice of visualizing extracted data acceptably and cleanly with graphical representation or graphs. The main objective of data visualization is to deliver data efficiently without any ambiguity or complexity. Data visualization generally consists of processing, evaluating and communicating the data. This section delineates data visualization for multimodal medical data using various algorithms, software or hardware components. Lavin et al. (2005) designed hardware to visualize 4D cardiac data using motion-based segmentation, which produced significant performance. Data visualization can also be achieved using software such as EHR data multimodal analysis using chromatogram plots. For visualizing the essential data features, OpenCL, C++ and GUI toolkits are adopted; iterative visualization and MVC pattern are some well-noticed algorithms that can be used (Manssour et al., 2000). For MRI, SPECT, CAT and PET data visualization, 2D or 3D image visualization is recommended, and it is tested using inertial moment 3D visualization as it gives a better view. In Joshi et al. (2010) radar plots are used to enhance the saturation and transfer function to visualize multimodal data of image-guided neurosurgery. Trend charts, timelines and data tables can also be practised to visualize EHR and clinical data (An et al., 2008). In Song et al. (2021), a quantitative analysis is performed by integrating the Grad-CAM visualization technique into a multimodal fusion framework applied to MRI and PET images for diagnosing Alzheimer's disease. Table 7 depicts a review of multimodal medical data visualization.

Table 7. *Review of multimodal medical data visualization.*

Paper	Modality	Contribution	Methodology
Mealha et al. (1994)	CT, SPECT	Approach for 3D visualization for multimodal clinical data	Raycaster philosophy
Nunes et al. (1996)	CAT, MRI, SPECT, PET	Presented various technique for multimodal clinical data visualization	2D integrated visualization-slice by slice representation. Semi 3D integrated visualization (one modality represented in 3D and other in 2D displayed as a cut plane).
Manssour et al. (2000)	MRI and PET	Framework for medical data visualization	• Registration • Segmentation • Visualization
Levin et al. (2005)	CT, MRI	3D visualization of multimodal cardiac data	• Volume rendering • Interactive 4D motion segmentation
An et al. (2008)	EHR data	Proposed a procedure for integrated visualization for EHR data	• Numeric EHR data • Data table • Trend chart and timeline.
Cooper et al. (2009)	CT-scan, X-ray, MRI	Designed a client interface for multimodal image analysis system	• DojoFish tree-control • RadLex tree
Joshi et al. (2010)	MRI and CTA data	Improvise multimodal visualizations for image-guided neurosurgery	• Ambient illumination monitoring
Weibel et al. (2013)	EHR Data	Proposed a technique for EHR data visualization	• chromogram plots
Kozlovszky et al. (2016)	Sensor data	Approach for multimodal biophysical data visualization	• For single modality-generic numeric data visualization. • For multimodal data – graphical visualization with continuous modality
Song et al. (2021)	MRI, PET	Proposed 3D Grad-CAM visualization technique	• Class-specific heatmaps generated on disease contours

4 Radiology Data Processing Case Study

Over the years, medical imaging has inclined from clinical routine to progressive human psychology due to its vast application in disease prognosis. The revolution in medical imaging has changed the practice of disease diagnosis through imaging the human body with various electromagnetic waves. The image acquired from shorter and longer wavelengths generates multimodal images with unique characteristics. Radiology is a medical discipline which includes a wide variety of data obtained from heterogeneous sources. As a case study for multimodality in medical AI, we will review various existing works in the area of radiology data processing.

This paper reviews state-of-the-art techniques in medical imaging across various modalities. The general image modalities in radiology include MRI, X-ray, ultrasound and computed tomography. X-ray imaging is a quick, low-cost and popular imaging technique of injecting iodinated contrast agents to the interested region in applications such as cardiovascular, mammography and abdominal imaging. Ultrasound imaging is a fast, non-invasive technique which uses the backscattering effect of acoustic pulses typically used in imaging arteries, blood flow, tissue stiffness, etc. The MR imaging technique applies a magnetic field in a conjunction radio filter, which generates spatial resolution volumetric images. In practice, each of these imaging modalities has varied methodology and characteristics (Panayides et al., 2020). These modalities can be utilized with its associated radiology reports and structured lab data to predict diseases (Shetty & Mahale, 2022). Deshpande et al. (2018) came up with a supporting framework named the "integrated radiology image search engine" to improve search interpretation. The prime requirement of this approach is extracting relevant and meaningful information from the diverse data set, and it is summarised by combining pathology and radiology to give a better diagnosis.

Many researchers have taken surveys on different radiology techniques. Some papers have surveyed CT dose level evaluation using five multi-slice systems (Waidi et al., 2011), and others on enhancing MRI resolution (Akcakaya et al., 2011). Overall, research into medical imaging has recently taken an increasing pace. Radiology has multimodal data consisting of radiology images and accessible text radiology reports. This section discusses different contemporary techniques for radiology data processing. We will delineate various approaches associated with radiology report processing, radiology image processing and data processing of multimodal radiology images with associated reports.

4.1 Radiology report processing

Radiology reports provide information supporting radiologists to envision disease and health conditions. Most radiology reports are free text; therefore, retrieving these unstructured data without a decent text mining technique is troublesome. Radiology free-text processing can be categorised into a rule-based approach (based on pattern matching) and a conventional ML-based method. Our previous work (Shetty et al., 2020) incorporated analysis of radiology reports in low data conditions using an improvised GloVe word embedding technique, a knowledge base and a deep learning framework. Many researchers practise text mining using information retrieval systems such as NeuRadIR, CBIR and MedLEE, which prove to be efficient.

In our studies, we consider medical images of contrasting modalities such as CT, MRI and X-ray and every modality requires distinct functionalities and algorithms to work upon. Setting up a firm framework is mandatory to envision the narrative medical content. Maghsoodi et al. (2012) considered an automatic sentence classification from radiology reports. Based on three features including modality, literality and recommendation, annotating sentences into seven classes is the core methodology of the proposed work and post tags are used for relevant processing.

The association rule can also be considered a favourable EHR data processing approach. Text weighted schemes such as inverse document frequency prove to be efficient weighted approval for text encoding from radiology reports (Alodadi, 2017). Converting frequent words into a feature vector using the bag-of-word and fuzzy c-mean clustering algorithms is a noted approach. Identifying similar reports would be more accessible using this approach (Turkeli et al., 2017). Adopting a text analysis system is a suitable technique for further improvement in report mining. The MEDAT text analysis system employs annotated system index with propositions. This approach helps identify identical sentences with the same symbols (Friedlin et al., 2011). In recent work, Niu et al. (2021) and Momoki et al. (2022) have illustrated different approaches to labelling radiology reports. In Niu et al. (2021), a labelled dependent attention model is designed with the idea of jointly embedding labels and words, where both the modules will learn from the word weight. In Momoki et al. (2022), an image classifier is built using a pseudo label from a radiology report. Both techniques prove to be efficient compared to the existing methods. Liu et al. (2022b) used the N-gram technique to extract word vectors from liver radiological reports with colorectal cancer and ensemble learning classification algorithms including random forest (RF), logistic regression (LR), etc., were proposed for disease categorization. Table 8 depicts a summary of radiology report data processing.

Table 8. *Summary of radiology report data processing.*

Paper	Modality	Contribution	Methodology	Outcome
Friedlin et al. (2011)	RoentGen corpus dataset (radiology report)	Text analytic system with annotated semantic index	i. Parsing [word→keybased index, sentence→semantic index] ii. Segmentation iii. Semantic annotation	60% of corpus annotated
Maghsoodi et al. (2012)	Breast cancer radiology report	Automated sentence classification for breast cancer report	i. Decide laterality and modality using seven classes ii. Feature extraction iii. Four different feature set extracted [stm, domain knowledge list, MedLEE, stm+DKL+MedLEE] iv. Konstanz information miner for classification and evaluation coreference resolution	92% to 98% accuracy for different classes
Turkeli et al. (2017)	457 MR and CT radiology report from thorax and abdomen	Identify similar radiology report	i. Feature extraction-Bag of words representation of vocabulary by frequency ii. Clustering method [Fuzzy C-mean algorithm, K-mean clustering] iii. Training iv. Evaluation	Similarity rate 77.46%
Alodadi (2017)	Radiology note from EHR	Automated system to read radiology notes	i. Concept-based representation-association rule ii. Extract medical terminology [bag of words concept] iii. Data transformation v. Apply weight schema [TF, IDF] calculate interestingness criteria generate candidate rule from apriori algorithm	Not mentioned
Shetty et al. (2020)	Indiana University chest X-ray reports	Disease classification from radiology reports	Knowledge-based text modelling is applied for text feature extraction followed by DNN for classification	90% accuracy by proposed model and obtained superior results compared to baseline models
Liu et al. (2022b)	Radiology reports	Colorectal cancer classification from liver radiological reports.	(i) N-grams word embedding technique (ii) Ensemble classifier	Proposed model achieved 96% accuracy

4.2 Radiology report data and image processing

We outlined a detailed literature review on radiology report data analysis in the previous section. We observed that researchers practising radiology data analysis have claimed that radiology report analysis with imaging modality produces better accuracy. In the 1990s, radiologists used video for data analysis using the stream of behaviour chronicles technique (Ramey et al., 1991), which was proven to be efficient in the early days. Gradually, radiology data analysis has taken a new direction towards text and image data analysis (Zhu et al., 2022). To handle text and image feature analysis, a match report approach using a knowledge base was developed. The medical finding extractor with SVM classifiers have been utilized for text prediction and standard image feature extraction techniques have been used for image analysis (Bodile & Kshirsagar, 2015). Pre-processing of unstructured reports using basic NLP techniques such as lemmatization, stemming, stop word removal, etc., has significantly improved the performance of the overall disease classification tasks (Gong et al., 2008; Wang et al., 2020). According to the survey, a full 3D image is difficult to process. Authors demonstrated that using a primary form or image slice with text scalar and vector labels simplifies the task. An unsupervised LDA approach was recommended for generating semantic tags in text, and DCNN was considered for mapping generated labels to images (Shin et al., 2015).

There has been significant research in the field of cross-modal retrieval of radiology reports from given input images. Alfarghaly et al. (2021) presented a cross-modal retrieval task to extract radiology reports from CXR images collected from the IU dataset. The CheXNet model was proposed to extract image features and the word2vec model was used for word embedding features. Ramirez-Alonso et al. (2022) reviewed various existing research on radiology report generation techniques using DenseNet, ResNet and VGG models with long and short-term memories (LSTM) and attention models. Sirshar et al. (2022) proposed attention-based report generation, which uses an encoder-decoder model containing VGG 16 as the encoder and LSTM as the decoder. The proposed model has achieved a bilingual evaluation understudy (BLEU) score ranging from 0.155 to 0.58. Table 9 presents a summary of multimodal radiology report data and image processing.

Table 9. *Summary of radiology report data and image processing.*

<table>
<tr><th>Paper</th><th>Modality</th><th>Contribution</th><th colspan="2">Methodology</th><th>Outcome</th></tr>
<tr><td rowspan="3">Shin et al. (2015)</td><td rowspan="3">780k radiology report stored in PACS (CT/MRI)</td><td rowspan="3">Deep learning system for text/images</td><td>Text processing</td><td>Image processing</td><td rowspan="3">Rate of predicted disease is 0.56 for recall at k=1</td></tr>
<tr><td colspan="2">Caffe framework</td></tr>
<tr><td>i. Sentence tokenization
ii. Word/number matching
iii. Stemming
iv. Rule-based information extraction</td><td>i. NLP to extract key image
ii. Image categorization using LDA
iii. Image to text description [vector modelling]
iv. Image to word description [deep CNN regression]</td></tr>
<tr><td rowspan="2">Bodile & Kshirsagar (2015)</td><td rowspan="2">Radiology report [MRI and abdominal]</td><td rowspan="2">Statistical machine translation approach for the radiology report</td><td>Text processing</td><td>Image processing</td><td rowspan="2">Not mentioned</td></tr>
<tr><td>i. Statistical machine translation
ii. Text, hypertext categorization [SVM]
iii. NLP technique [stemming, term mapping, semantic rules]</td><td>i. Image feature extraction (colour feature extraction, edge detection, text extraction)</td></tr>
<tr><td rowspan="2">Gong et al. (2008)</td><td rowspan="2">Brain CT radiology report</td><td rowspan="2">Mining data from the radiology report</td><td>Text processing</td><td>Image processing</td><td rowspan="2">Similarity rate is 77.46%</td></tr>
<tr><td>i. Brain CT radiology
ii. Term mapper
iii. Parser
iv. Finding extractor
v. Report constructor</td><td>i. Image registration
ii. Abnormality area selection
iii. Contour generation and segmentation
iv. Feature extraction</td></tr>
</table>

Paper	Modality	Contribution	Methodology	Outcome
Wang et al. (2020)	Functional MRI	Multisite adaption framework for Austin spectrum	i. Data preprocessing [2 baseline method] ii. Classification using SVM iii. Low-rank representation using SVM/KNN robust domain adaption iv. Geodesic flow kernel	SVM and KNN combining gives better accuracy of 71.88% to 73.44%
Alfarghaly et al. (2021)	Chest X-ray and radiology reports	Radiology report generation	i. Word2Vec for report feature extraction. ii. CheXnet for image feature extraction. iii. LSTM for report generation	Produced significantly higher BLEU score
Sirshar et al. (2022)	Chest X-ray and radiology reports	Radiology report generation	i. Proposed model uses CNN and LSTM for report generation. ii. Encoder used : VGG16 (CNN); decoder used : LSTM, followed by attention	Proposed model achieved BLEU score ranging from 0.155 to 0.58

4.3 Radiology image processing

The current advancement in technology and the digitised world has paved the way for image processing and computer vision in the medical domain. The massive growth of technology and innovation has made data acquisition so easy and cost-reliable that capturing high-resolution images has become a simple step (Khoo et al., 1997). According to a survey of enhancing radiology image quality to 3D resolution (Danzhou et al., 2008), solving integration problems and minimising high-quality issues (Deshpande et al., 2018) has become a practical research interest for radiology image processing. Many researchers have presented various approaches to radiology image processing. In Danzhou et al. (2008), the index assignment approach, enhancing only the ROI slice or bucket are the ideas described to reduce time demands. In Philipsen et al. (2015), processing images from multiple sources, including image enhancement, is presented. Normalising the ideas by decomposing them into different bands has proven to be a better solution and technique; e.g., the average Jaccard index overlap is used for segmentation.

However, dataset complexity and size play a significant role in radiology data processing. Two-class classifiers, known as linear and non-linear classifiers, can simplify the problem, as discussed in Chen et al. (2010). Criteria such as region mean, region variance, weighted least square and inverse filtering are calculated to estimate unbiased parameters in Muzic et al. (1998). Implementation of the Bayesian least square and Gaussian scale mixture approaches to reduce blurring and other acquisition artefacts in MRI modality is discussed (Megalooikonomou & Kontos, 2007; Chen et al., 2010).

Modern machine learning algorithms have a colossal potential for improving the segmentation approach. These approaches consider complete meta-features for evaluation, classification and solution of properties such as confidence connectedness of intra-region intensity to increase the robustness. Setio et al. (2016) proposed a framework for pulmonary nodule detection. Candidate detection, patch extraction and false-positive reaction were the techniques addressed. In Jacenkow et al. (2022), image classification is achieved based on the indication field. The application of self-gating is observed in Rosenzweig et. al. (2020).

In Chun-An et al. (2014), mutual information and practical least square regression are implemented to solve the voxel classification problem for an fMRI dataset, and the accuracy of the results obtained was satisfactory. Using a convolutional neural network to classify radiology images can improve data accuracy up to 89.5%. The ResNet-based neural network model was designed to classify breast cancer effectively. The model utilizes the large data cohort, including 10,00,000 breast-cancer screening mammography exams with breast-level and pixel-level labels. The proposed hybrid model containing both neural networks and radiologists provided superior results compared to the individual outcomes (Wu et al., 2020). Shetty and Mahale (2022) proposed the MS-CheXNet model using a depth-wise convolutional neural network with a multi-channel dilation layer to extract and learn imaging features from radiology CXR. Table 10 presents a summary of radiology image processing.

Table 10. *Summary of radiology report image processing.*

Paper	Modality	Contribution	Methodology	Outcome
Muzic et al. (1998)	PET data	Methodology for quantification of the region to reduce degradation	i. Using Huesman's Region of Interest (ROI) analysis method, image pixels are reconstructed using projection filter ii. Calculate region variance iii. Scaling iv. Inverse filter	Small error of 4% in ROI for myocardium with 10.5 cm thick
Chen et al. (2001)	Ultrasound images	Proposed hybrid two-class classifier suits datasets of all sizes	i. Two-class classification [hybrid linear/ non-linear classifier, binormal ROC theory, Fisher's linear discriminant function]	Not mentioned
Megalooikonomou & Kontos (2007)	fMRI dataset for Alzheimer's disease	Identifies similar radiology report	i. Divide image into smaller sub-region dynamic progressive partition till ROI found ii. test significance such as T-test or rank-sum test iii. P-value threshold to determine discriminative power of statistical image	Accuracy of DDRP test 94%
Danzhou et al. (2008)	MRI	A framework for processing 3D the high-resolution for the internet-based application	i. Partition of high resolution 3D image into bucket ii. Removing duplicate buckets iii. Compress each bucket independently iv. Store the compressed buckets in a Hilbert curve on disk	Processing time of new technique is twice faster than the old
Akcakaya et al. (2011)	MRI	Reconstruction method to improvise blurring and reconstruction artifacts	i. MRI acquired in two different subject cohorts ii. Construction technique [CS threshold, BLS-GSM thresholding, minimization thresholding] iv. Soaptool bubble framework with	At lower RSNR lower MSE

Paper	Modality	Contribution	Methodology	Outcome
			Deriche algorithm on RCA for evaluation	
Chun-An et al. (2014)	fMRI	A framework to select robust feature using partial least square regression and mutual information	i. Extract weight from fMRI using GLM ii. Generate feature importance iii. Index apply voxel select strategy (MI and PLS) iii. Classify testing instance based on voxel selected	Improves overall classification performance
Zwettler & Backfrieder (2015)	CT and MRI	Improvise segmentation in medical dataset using a classification approach	i. Calculate and analyse feature from all region ii. Assign class and label to each feature based on Gaussian shape, normal distribution and similarity iii. Multivariate feature analysis using wavelet transform iv. Local meta feature v. Multivariate similarity calculation	Not mentioned
Philipsen et al. (2015)	Six datasets with 100 posterior-anterior CXR	A framework to normalize the image acquired from a varied source	i. Localised energy-based image normalization using Gaussian kernel ii. CXR normalization (ROI is detected), lung segmentation algorithm	Average area under the receiver operating curve increased from 0.72±0.14 and 0.79±0.06
Setio et al. (2016)	LIDC-IDRI dataset	CAD system for pulmonary module detection using multi-view convolution network	i. Extract 2D patches of the volumetric object using nine views ii. Candidate detection iii. Patch extraction iv. False-positive reduction v. Fusion method (committee fusion, late fusion, mixed fusion) vi. Training vii. Evaluation	Combining multiple candidate detection algorithm boosts sensitivity from 85.7% to 93.3%
Rosenzweig (2020)	Cardiac MRI	Novel self-gating method to acquire data	i. Correction of AC data ii. Dimensionality reduction method (principal component analysis, SSA-FARY iii. Binning	Not mentioned
Wu et al. (2020)	1,000,000 mammogram images	A framework for breast cancer screening exam classification	i. Assign label for images ii. Classification using multitask learning framework iv. CNN based on ResNet architecture	89.5% accuracy in predicting presence of cancer

Paper	Modality	Contribution	Methodology	Outcome
Shetty & Mahale (2022)	Chest X-ray images from IU dataset and data collected from private hospital	Multi-channel framework for predicting pulmonary diseases from CXR	Multi-channel dilation layer with depth-wise separable CNN for image feature extraction	Better accuracy compared to state-of-the-art DL models

5 Discussion and Conclusion

This paper reviewed the existing state-of-the-art methods of medical multimodal data analysis. It discussed various multimodal health analysis techniques and research work on radiology data processing. Research into multimodal data has taken different views. Some applications work on clickstream data, eye tracking, video and EEG to check learning performance (Giannakos et al., 2019) and some on video event detection using temporal analysis and multimodal data mining methods (Min Chen et al., 2006; Richard et al., 2022). Various application-oriented works have also been observed based on multimodal medical data (Mieloszyk et al., 2017) for fixing appointments in hospitals, etc.

In the medical data analysis domain, AI contributes very much to enhance algorithm accuracy (Baltrušaitis et al., 2019). We can say that the use of AI and mainly CNN is changing the research direction of medical data analysis. We outlined a comparative study for each data analysis approach. The main conclusions related to current procedures and algorithms observed during the survey are stated in Table 11.

Table 11. *Main conclusions related to current procedures and algorithms.*

No.	Conclusions
1.	It is noticed that in multimodal health data classification, data volume is a significant problem. Sometimes data sets are extensive, and features are fewer or a small data group with many elements. The execution speed of the algorithm is affected by this issue.
2.	In most of the approaches, only essential features are picked up for classification, which increases the chance of lowering the accuracy and precision value of the designed technique.
3.	A notable observation is that most of the designed techniques prefer the late fusion method for multimodal classification. This technique reduces the execution time and increases the speed, but the precision value is affected.
4.	Image processing techniques and algorithms are proven to enhance multimodal image quality and features. In current multimodal medical data visualization, significantly lower usage of image processing algorithms or tools is observed.
5.	The current visualization technique is hardware-based with the requirement for a graphics card. Most graphics cards cost more and have higher system specifications, which are challenging to maintain and not cost-efficient.
6.	In multimodal data analysis, multimodal data fusion is a challenging task to achieve. It is observed that most of the fusion process is carried out using a data-driven approach, and handling data of low quality is problematic.
7.	In multimodal fusion, it is observed that most of the fusion techniques such as early, late and joint fusion rely on straightforward concatenation techniques to fuse heterogenous features. A major drawback of this approach is missing intermodal interaction during the fusion process.
8.	It is seen that fusing unstructured clinical notes, reports or structured lab records with radiology images has significantly improved the performance compared to unimodal image analysis.
9.	Very limited work is performed on various medical data such as speech, text, image and video. It is noticed that the algorithm efficiency for speech-text fusion or text-text fusion is lower compared to multimodal image fusion.

No.	Conclusions
10.	Multimodal data benchmarking and publishing of the cohort would significantly improve multimodal medical research as the heterogeneous open source datasets currently available are scarce in numbers or they are restricted to private organizations.

As a case study for multimodal medical AI, we reviewed various research work in the radiology domain. We surveyed papers on unimodal radiology report analysis, image analysis and multimodal radiology image and report analysis. The major limitation observed in radiology data processing is that minimal work on data analysis using radiology text and images is carried out. Usage of unimodal radiology reports or images turned out to be less effective compared to multimodal report and image analysis. Based on the study, a conclusion can be drawn that considering radiology reports and images followed by late or early fusion can give remarkable efficiency. Fusion techniques with intermodal interaction between radiology images and reports can be addressed. The cross-modal retrieval of report generation has achieved a BLEU score up to 0.50–0.60. The unstructured nature of the radiology report poses a severe challenge with respect to the performance of the cross-modal retrieval. Hence, there is a need to develop a practical framework to generate accurate radiology reports.

Additional Information and Declarations

Acknowledgments: We would like to thank the Department of Information Technology, National Institute of Technology Karnataka, Surathkal, for providing us with the resources for carrying out this research. We would like to thank KMC, Mangalore, for technical assistance. We also thank the special issue editors and reviewers for their time and consideration.

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: S.S.: Conceptualization, Methodology, Investigation, Writing – Original draft preparation, Data Curation. A.V.S.: Conceptualization, Methodology, Supervision, Validation, Writing – Reviewing and Editing. A.M.: Validation, Writing – Reviewing and Editing.

References

Abdulkareem, M. B. (2018). Design and Development of Multimodal Medical Image Fusion using Discrete Wavelet Transform. In *2018 Second International Conference on Inventive Communication and Computational Technologies* (ICICCT). IEEE. https://doi.org/10.1109/icicct.2018.8472997

Abubaker, M., & Babayigit, B. (2022). Detection of Cardiovascular Diseases in ECG Images Using Machine Learning and Deep Learning Methods. *IEEE Transactions on Artificial Intelligence*, (in press). https://doi.org/10.1109/tai.2022.3159505

Abujaber, A. A., Nashwan, A. J., & Fadlalla, A. (2022). Harnessing machine learning to support evidence-based medicine: A pragmatic reconciliation framework. *Intelligence-Based Medicine*, 6, 100048. https://doi.org/10.1016/j.ibmed.2022.100048

Acosta, J. N., Falcone, G. J., Rajpurkar, P., & Topol, E. J. (2022). Multimodal biomedical AI. *Nature Medicine*, 28(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2

Adali, T., Levin-Schwartz, Y., & Calhoun, V. D. (2015). Multimodal Data Fusion Using Source Separation: Application to Medical Imaging. *Proceedings of the IEEE*, 103(9), 1494–1506. https://doi.org/10.1109/jproc.2015.2461601

Akcakaya, M., Nam, S., Hu, P., Moghari, M. H., Ngo, L. H., Tarokh, V., Manning, W. J., & Nezafat, R. (2011). Compressed Sensing With Wavelet Domain Dependencies for Coronary MRI: A Retrospective Study. *IEEE Transactions on Medical Imaging*, 30(5), 1090–1099. https://doi.org/10.1109/tmi.2010.2089519

Alfarghaly, O., Khaled, R., Elkorany, A., Helal, M., & Fahmy, A. (2021). Automated radiology report generation using conditioned transformers. *Informatics in Medicine Unlocked*, 24, 100557. https://doi.org/10.1016/j.imu.2021.100557

Almasri, M. M., & Alajlan, A. M. (2022). Artificial Intelligence-Based Multimodal Medical Image Fusion Using Hybrid S2 Optimal CNN. *Electronics*, 11(14), 2124. https://doi.org/10.3390/electronics11142124

Alqahtani, A., Ullah Khan, H., Alsubai, S., Sha, M., Almadhor, A., Iqbal, T., & Abbas, S. (2022). An efficient approach for textual data classification using deep learning. *Frontiers in Computational Neuroscience*, 16. https://doi.org/10.3389/fncom.2022.992296

Alodadi, M. (2017). Radiology Clinical Notes Mining Using Weighted Association Rules. In *2017 IEEE International Conference on Healthcare Informatics* (ICHI). IEEE. https://doi.org/10.1109/ichi.2017.23

Amal, S., Safarnejad, L., Omiye, J. A., Ghanzouri, I., Cabot, J. H., & Ross, E. G. (2022). Use of Multi-Modal Data and Machine Learning to Improve Cardiovascular Disease Care. *Frontiers in Cardiovascular Medicine*, 9. https://doi.org/10.3389/fcvm.2022.840262

An, J., Lu, X., & Duan, H. (2008). Integrated Visualization of Multi-Modal Electronic Health Record Data. In *2008 2nd International Conference on Bioinformatics and Biomedical Engineering*. IEEE. https://doi.org/10.1109/icbbe.2008.156

Atta-Ur-Rahman, Nasir, M. U., Gollapalli, M., Zubair, M., Saleem, M. A., Mehmood, S., Khan, M. A., & Mosavi, A. (2022). Advance Genome Disorder Prediction Model Empowered With Deep Learning. *IEEE Access*, 10, 70317–70328. https://doi.org/10.1109/access.2022.3186998

Aydin, F., Zhang, M., & Haffari, G. (2019). Medical Multimodal Classifiers Under Scarce Data Condition. *arXiv*. https://doi.org/10.48550/arXiv.1902.08888

Barillot, C., Lemoine, D., Le Briquer, L., Lachmann, F., & Gibaud, B. (1993). Data fusion in medical imaging: merging multimodal and multipatient images, identification of structures and 3D display aspects. *European Journal of Radiology*, 17(1), 22–27. https://doi.org/10.1016/0720-048x(93)90024-h

Batko, K., & Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. *Journal of Big Data*, 9(1). https://doi.org/10.1186/s40537-021-00553-4

Bayoudh, K., Knani, R., Hamdaoui, F., & Mtibaa, A. (2022). A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets. *The Visual Computer*, 38, 2939–2970. https://doi.org/10.1007/s00371-021-02166-7

Ben Said, A., Mohamed, A., Elfouly, T., Harras, K., & Wang, Z. J. (2017). Multimodal Deep Learning Approach for Joint EEG-EMG Data Compression and Classification. In *2017 IEEE Wireless Communications and Networking Conference* (WCNC). IEEE. https://doi.org/10.1109/WCNC.2017.7925709

Bernal, E. A., Yang, X., Li, Q., Kumar, J., Madhvanath, S., Ramesh, P., & Bala, R. (2018). Deep Temporal Multimodal Fusion for Medical Procedure Monitoring Using Wearable Sensors. *IEEE Transactions on Multimedia*, 20(1), 107–118. https://doi.org/10.1109/tmm.2017.2726187

Berner, E. S. (2010). *Clinical Decision Support Systems: Theory and Practice*. Springer. https://doi.org/10.1007/978-0-387-38319-4

Bleyer, W. A. (1997). The U.S. pediatric cancer clinical trials programmes: International implications and the way forward. *European Journal of Cancer*, 33(9), 1439–1447. https://doi.org/10.1016/s0959-8049(97)00249-9

Bloch, I. (1996). Some aspects of Dempster-Shafer evidence theory for classification of multi-modality medical images taking partial volume effect into account. *Pattern Recognition Letters*, 17(8), 905–919. https://doi.org/10.1016/0167-8655(96)00039-6

Bu, S., Wang, L., Han, P., Liu, Z., & Li, K. (2017). 3D shape recognition and retrieval based on multi-modality deep learning. *Neurocomputing*, 259, 183–193. https://doi.org/10.1016/j.neucom.2016.06.088

Bodile, A., & Kshirsagar, M. (2015). Text mining in radiology reports by statistical machine translation approach. In *2015 Global Conference on Communication Technologies* (GCCT). IEEE. https://doi.org/10.1109/gcct.2015.7342797

Cai, Q., Wang, H., Li, Z., & Liu, X. (2019). A Survey on Multimodal Data-Driven Smart Healthcare Systems: Approaches and Applications. *IEEE Access*, 7, 133583–133599. https://doi.org/10.1109/ACCESS.2019.2941419

Carvalho, R., Pedrosa, J., & Nedelcu, T. (2021). Multimodal Multi-tasking for Skin Lesion Classification Using Deep Neural Networks. In *Advances in Visual Computing*, (pp. 27–38). Springer. https://doi.org/10.1007/978-3-030-90439-5_3

Castiglione, A., Vijayakumar, P., Nappi, M., Sadiq, S., & Umer, M. (2021). COVID-19: Automatic Detection of the Novel Coronavirus Disease from CT Images Using an Optimized Convolutional Neural Network. *IEEE Transactions on Industrial Informatics*, 17(9), 6480–6488. https://doi.org/10.1109/tii.2021.3057524

Chandra, B. S., Sastry, C. S., & Jana, S. (2019). Robust Heartbeat Detection From Multimodal Data via CNN-Based Generalizable Information Fusion. *IEEE Transactions on Biomedical Engineering*, 66(3), 710–717. https://doi.org/10.1109/tbme.2018.2854899

Chaudhury, S., Dey, L., Verma, I., & Hassan, E. (2017). Mining Multimodal Data. *Pattern Recognition and Big Data*, (pp. 581–604). World Scientific. https://doi.org/10.1142/9789813144552_0017

Chen, R., Ho, J. C., & Lin, J.-M. S. (2020). Extracting medication information from unstructured public health data: a demonstration on data from population-based and tertiary-based samples. *BMC Medical Research Methodology*, 20(1). https://doi.org/10.1186/s12874-020-01131-7

Chen, M., Hao, Y., Hwang, K., Wang, L., & Wang, L. (2017). Disease Prediction by Machine Learning Over Big Data From Healthcare Communities. IEEE Access, 5, 8869–8879. https://doi.org/10.1109/access.2017.2694446

Chien, S.-C., Chen, Y.-L., Chien, C.-H., Chin, Y.-P., Yoon, C. H., Chen, C.-Y., Yang, H.-C., & Li, Y.-C. J. (2022). Alerts in Clinical Decision Support Systems (CDSS): A Bibliometric Review and Content Analysis. *Healthcare* (Basel, Switzerland), 10(4), 601. https://doi.org/10.3390/healthcare10040601

Chun-An Chou, Kampa, K., Mehta, S. H., Tungaraza, R. F., Chaovalitwongse, W. A., & Grabowski, T. J. (2014). Voxel Selection Framework in Multi-Voxel Pattern Analysis of fMRI Data for Prediction of Neural Response to Visual Stimuli. *IEEE Transactions on Medical Imaging*, 33(4), 925–934. https://doi.org/10.1109/tmi.2014.2298856

Comito, C., Falcone, D., & Forestiero, A. (2022). AI-Driven Clinical Decision Support: Enhancing Disease Diagnosis Exploiting Patients Similarity. *IEEE Access*, 10, 6878–6888. https://doi.org/10.1109/access.2022.3142100

Cooper, J. W., Ebadollahi, S., & Eide E. (2009). A Thin Client Interface to a High Performance Multi-modal Image Analytics System. In *2009 42nd Hawaii International Conference on System Sciences*. IEEE. https://doi.org/10.1109/hicss.2009.45

Danzhou L., Hua, K. A., & Sugaya, K. (2008). A Generic Framework for Internet-Based Interactive Applications of High-Resolution 3-D Medical Image Data. *IEEE Transactions on Information Technology in Biomedicine*, 12(5), 618–626. https://doi.org/10.1109/titb.2008.923770

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1). https://doi.org/10.1186/s40537-019-0217-0

Denzil Bosco, T., Narayanan, A. L., & Veeramuthu, A. (2017). A new multimodal brain tumour image fusion for pre-processing of medical image database. In *2017 International Conference on Communication and Signal Processing* (ICCSP). IEEE. https://doi.org/10.1109/iccsp.2017.8286687

Deshpande, P., Rasin, A., Brown, E., Furst, J., Raicu, D. S., Montner, S. M., & Armato, S. G. (2018). Big Data Integration Case Study for Radiology Data Sources. In *2018 IEEE Life Sciences Conference* (LSC). IEEE. https://doi.org/10.1109/lsc.2018.8572185

Dramburg, S., Marchante Fernández, M., Potapova, E., & Matricardi, P. M. (2020). The Potential of Clinical Decision Support Systems for Prevention, Diagnosis, and Monitoring of Allergic Diseases. *Frontiers in Immunology*, 11. https://doi.org/10.3389/fimmu.2020.02116

Drotar, P., Mekyska, J., Rektorova, I., Masarova, L., Smekal, Z., & Faundez-Zanuy, M. (2015). Decision support framework for Parkinson's disease based on novel handwriting markers. *IEEE Transactions on Neural Systems and Rehabilitation* Engineering, 1–1. https://doi.org/10.1109/tnsre.2014.2359997

Faris, H., Habib, M., Faris, M., Elayan, H., & Alomari, A. (2021). An intelligent multimodal medical diagnosis system based on patients' medical questions and structured symptoms for telemedicine. *Informatics in Medicine Unlocked*, 23, 100513. https://doi.org/10.1016/j.imu.2021.100513

Fazli, S., Dahne, S., Samek, W., Bieszmann, F., & Muller, K.-R. (2015). Learning From More Than One Data Source: Data Fusion Techniques for Sensorimotor Rhythm-Based Brain–Computer Interfaces. *Proceedings of the IEEE*, 103(6), 891–906. https://doi.org/10.1109/jproc.2015.2413993

Fei, Z., Ryeznik, Y., Sverdlov, A., Tan, C. W., & Wong, W. K. (2022). An overview of healthcare data analytics with applications to the COVID-19 pandemic. *IEEE Transactions on Big Data*, 8(6), 1463–1480. https://doi.org/10.1109/tbdata.2021.3103458

Fleury, A., Vacher, M., & Noury, N. (2010). SVM-Based Multimodal Classification of Activities of Daily Living in Health Smart Homes: Sensors, Algorithms, and First Experimental Results. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 274–283. https://doi.org/10.1109/titb.2009.2037317

Friedlin, J., Mahoui, M., Jones, J., & Jamieson, P. (2011). Knowledge Discovery and Data Mining of Free Text Radiology Reports. In *2011 IEEE First International Conference on Healthcare Informatics, Imaging and Systems Biology*. IEEE. https://doi.org/10.1109/hisb.2011.31

Garg, A. X., Adhikari, N. K. J., McDonald, H., Rosas-Arellano, M. P., Devereaux, P. J., Beyene, J., Sam, J., & Haynes, R. B. (2005). Effects of Computerized Clinical Decision Support Systems on Practitioner Performance and Patient Outcomes. *JAMA*, 293(10), 1223. https://doi.org/10.1001/jama.293.10.1223

Gehrmann, S., Dernoncourt, F., Li, Y., Carlson, E. T., Wu, J. T., Welt, J., Foote, J., Moseley, E. T., Grant, D. W., Tyler, P. D., & Celi, L. A. (2018). Comparing deep learning and concept extraction based methods for patient phenotyping from clinical narratives. *PLOS ONE*, 13(2), e0192360. https://doi.org/10.1371/journal.pone.0192360

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, 48, 108–119. https://doi.org/10.1016/j.ijinfomgt.2019.02.003

Gong, T., Tan, C. L., Leong, T. Y., Lee, C. K., Pang, B. C., Lim, C. C. T., Tian, Q., Tang, S., & Zhang, Z. (2008). Text Mining in Radiology Reports. In *2008 Eighth IEEE International Conference on Data Mining*. IEEE. https://doi.org/10.1109/icdm.2008.150

Hak, F., Guimarães, T., & Santos, M. (2022). Towards effective clinical decision support systems: A systematic review. *PLOS ONE*, 17(8), e0272846. https://doi.org/10.1371/journal.pone.0272846

Hamidinekoo, A., Pieciak, T., Afzali, M., Akanyeti, O., & Yuan, Y. (2021). Glioma Classification Using Multimodal Radiology and Histology Data. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, (pp. 508–518). Springer. https://doi.org/10.1007/978-3-030-72087-2_45

Harada, T., Miyagami, T., Kunitomo, K., & Shimizu, T. (2021). Clinical Decision Support Systems for Diagnosis in Primary Care: A Scoping Review. *International Journal of Environmental Research and Public Health*, 18(16), 8435. https://doi.org/10.3390/ijerph18168435

Haque, M. E., & Tozal, M. E. (2022). Identifying Health Insurance Claim Frauds Using Mixture of Clinical Concepts. *IEEE Transactions on Services Computing*, 15(4), 2356–2367. https://doi.org/10.1109/tsc.2021.3051165

Heidarivincheh, F., McConville, R., Morgan, C., McNaney, R., Masullo, A., Mirmehdi, M., Whone, A. L., & Craddock, I. (2021). Multimodal Classification of Parkinson's Disease in Home Environments with Resiliency to Missing Modalities. *Sensors*, 21(12), 4133. https://doi.org/10.3390/s21124133

Hilmizen, N., Bustamam, A., & Sarwinda, D. (2020). The multimodal deep learning for diagnosing covid-19 pneumonia from chest ct-scan and x-ray images. In *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems* (ISRITI), (pp. 26–31). IEEE. https://doi.org/10.1109/ISRITI51436.2020.9315478

Hossain, Md. J., Wadud, Md. A. H., Rahman, A., Ferdous, J., Alam, M. S., Amir - Ul - Haque Bhuiyan, T. M., & Mridha, M. F. (2021). A Secured Patient's Online Data Monitoring through Blockchain: An Intelligent way to Store Lifetime Medical Records. In *2021 International Conference on Science & Contemporary Technologies* (ICSCT). IEEE. https://doi.org/10.1109/icsct53883.2021.9642537

Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I., & Lungren, M. P. (2020a). Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. *Npj Digital Medicine*, 3(1). https://doi.org/10.1038/s41746-020-00341-z

Huang, S.-C., Pareek, A., Zamanian, R., Banerjee, I., & Lungren, M. P. (2020b). Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: a case-study in pulmonary embolism detection. *Scientific Reports*, 10(1), 22147. https://doi.org/10.1038/s41598-020-78888-w

Iakovidis, D., & Christos, S. (2012). A semantic model for multimodal data mining in healthcare information systems. *Studies in health technology and informatics*, 180, 574–578.

Institute of Medicine. (2000). *To Err Is Human*. National Academies Press. https://doi.org/10.17226/9728

Jacenkow, G., O'Neil, A. Q., & Tsaftaris, S. A. (2022). Indication as Prior Knowledge for Multimodal Disease Classification in Chest Radiographs with Transformers. In *2022 IEEE 19th International Symposium on Biomedical Imaging* (ISBI). IEEE. https://doi.org/10.1109/isbi52829.2022.9761567

Jacobs, A. (2009). The pathologies of big data. *Communications of the ACM*, 52(8), 36. https://doi.org/10.1145/1536616.1536632

Jindal, A., Dua, A., Kumar, N., Das, A. K., Vasilakos, A. V., & Rodrigues, J. J. P. C. (2018). Providing Healthcare-as-a-Service Using Fuzzy Rule Based Big Data Analytics in Cloud Computing. *IEEE Journal of Biomedical and Health Informatics*, 22(5), 1605–1618. https://doi.org/10.1109/jbhi.2018.2799198

Jiang, Y., Li, W., Hossain, M. S., Chen, M., Alelaiwi, A., & Al-Hammadi, M. (2020). A snapshot research and implementation of multimodal information fusion for data-driven emotion recognition. *Information Fusion*, 53, 209–221. https://doi.org/10.1016/j.inffus.2019.06.019

Joseph, N., Bernadin, S., Hodges, D., & Sekhar, P. (2021). A Case Study on using Unstructured Data Analysis Methods to identify local Covid-19 Hotspots. In *SoutheastCon 2021*. IEEE. https://doi.org/10.1109/southeastcon45413.2021.9401921

Joshi, A., Papanastassiou, A., Vives, K. P., Spencer, D. D., Staib, L. H., & Papademetris, X. (2010). Light-sensitive visualization of multimodal data for neurosurgical applications. In *2010 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE. https://doi.org/10.1109/isbi.2010.5490128

Karahan, E., Rojas-Lopez, P. A., Bringas-Vega, M. L., Valdes-Hernandez, P. A., & Valdes-Sosa, P. A. (2015). Tensor Analysis and Fusion of Multimodal Brain Images. *Proceedings of the IEEE*, 103(9), 1531–1559. https://doi.org/10.1109/jproc.2015.2455028

Khoo, V. S., Dearnaley, D. P., Finnigan, D. J., Padhani, A., Tanner, S. F., & Leach, M. O. (1997). Magnetic resonance imaging (MRI): considerations and applications in radiotherapy treatment planning. *Radiotherapy and Oncology*, 42(1), 1–15. https://doi.org/10.1016/s0167-8140(96)01866-x

Kline, A., Wang, H., Li, Y., Dennis, S., Hutch, M., Xu, Z., Wang, F., Cheng, F., & Luo, Y. (2022). Multimodal machine learning in precision health: A scoping review. *Npj Digital Medicine*, 5(1), 1–14. https://doi.org/10.1038/s41746-022-00712-8

Kozlovszky, M., Batbayar, K., Garaguly, Z., & Karozkai, K. (2016). Multimodal biophysical data visualization for patient monitoring. In *2016 IEEE 11th International Symposium on Applied Computational Intelligence and Informatics* (SACI). IEEE. https://doi.org/10.1109/saci.2016.7507411

Kumar, S., & Singh, M. (2019). Big data analytics for healthcare industry: impact, applications, and tools. *Big Data Mining and Analytics*, 2(1), 48–57. https://doi.org/10.26599/bdma.2018.9020031

Kurniawan, H., & Pechenizkiy, M. (2014). Towards the Stress Analytics Framework: Managing, Mining, and Visualizing Multi-modal Data for Stress Awareness. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*. IEEE. https://doi.org/10.1109/cbms.2014.129

Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects. *Proceedings of the IEEE*, 103(9), 1449–1477. https://doi.org/10.1109/jproc.2015.2460697

Lahat, D., Adaly, T., & Jutten, C. (2014). Challenges in multimodal data fusion. In *2014 22nd European Signal Processing Conference* (EUSIPCO), (pp. 101–105). IEEE.

Lawonn, K., Smit, N. N., Bühler, K., & Preim, B. (2017). A Survey on Multimodal Medical Data Visualization. *Computer Graphics Forum*, 37(1), 413–438. https://doi.org/10.1111/cgf.13306

Levin, D., Aladl, U., Germano, G., & Slomka, P. (2005). Techniques for efficient, real-time, 3D visualization of multi-modality cardiac data using consumer graphics hardware. *Computerized Medical Imaging and Graphics*, 29(6), 463–475. https://doi.org/10.1016/j.compmedimag.2005.02.007

Li, H., & Fan, Y. (2019). Early Prediction Of Alzheimer's Disease Dementia Based On Baseline Hippocampal MRI and 1-Year Follow-Up Cognitive Measures Using Deep Recurrent Neural Networks. 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). https://doi.org/10.1109/isbi.2019.8759397

Lin, S., Han, Z., Li, D., Zeng, J., Yang, X., Liu, X., & Liu, F. (2020). Integrating model- and data-driven methods for synchronous adaptive multi-band image fusion. *Information Fusion*, 54, 145–160. https://doi.org/10.1016/j.inffus.2019.07.009

Liu, Y., Yan, B., Zhang, R., Liu, K., Jeon, G., & Yang, X. (2022a). Multi-Scale Mixed Attention Network for CT and MRI Image Fusion. *Entropy*, 24(6), 843. https://doi.org/10.3390/e24060843

Liu, W., Zhang, X., Lv, H., Li, J., Liu, Y., Yang, Z., Weng, X., Lin, Y., Song, H., & Wang, Z. (2022b). Using a classification model for determining the value of liver radiological reports of patients with colorectal cancer. *Frontiers in Oncology*, 12. https://doi.org/10.3389/fonc.2022.913806

Lopez, K., Fodeh, S. J., Allam, A., Brandt, C. A., & Krauthammer, M. (2020). Reducing Annotation Burden Through Multimodal Learning. *Frontiers in Big Data*, 3, Article 19. https://doi.org/10.3389/fdata.2020.00019

Lv, Z., & Qiao, L. (2020). Analysis of healthcare big data. *Future Generation Computer Systems*, 109, 103–110. https://doi.org/10.1016/j.future.2020.03.039

Lynch, C. (2008). How do your data grow? *Nature*, 455(7209), 28–29. https://doi.org/10.1038/455028a

Maghsoodi, A., Sevenster, M., Scholtes, J., & Nalbantov, G. (2012). Sentence-based classification of free-text breast cancer radiology reports. In *2012 25th IEEE International Symposium on Computer-Based Medical Systems* (CBMS). IEEE. https://doi.org/10.1109/cbms.2012.6266374

Manssour, I. H., Furuie, S. S., Nedel, L. P., & Freitas, C. M. D. S. (2000). A multimodal visualization framework for medical data. In *Proceedings 13th Brazilian Symposium on Computer Graphics and Image Processing*. IEEE. https://doi.org/10.1109/sibgra.2000.895844

Nasution, B. I., Nugraha, Y., Prasetya, N. L., Aminanto, M. E., Sulasikin, A., Kanggrawan, J. I., & Suherman, A. L. (2022). COVID-19 Mortality Risk Factors Using Survival Analysis: A Case Study of Jakarta, Indonesia. *IEEE Transactions on Computational Social Systems*, 1–10. https://doi.org/10.1109/tcss.2022.3163325

Mealha, O. E., Rocha, C., Santos, B. S., Pereira, A. S., Pedroso de Lima, J. J., Botelho, M. F., & Ribeiro, M. J. (1994). 3D registration and integrated visualization of multimodal clinical data. In *Proceedings of 16th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. https://doi.org/10.1109/iembs.1994.411883

Megalooikonomou, V., & Kontos, D. (2007). Medical Data Fusion for Telemedicine. I*EEE Engineering in Medicine and Biology Magazine*, 26(5), 36–42. https://doi.org/10.1109/emb.2007.901790

Meng, T., Lin, L., Shyu, M.-L., & Chen, S.-C. (2010). Histology Image Classification Using Supervised Classification and Multimodal Fusion. In *2010 IEEE International Symposium on Multimedia*. IEEE. https://doi.org/10.1109/ism.2010.29

Mieloszyk, R. J., Rosenbaum, J. I., Bhargava, P., & Hall, C. S. (2017). Predictive modeling to identify scheduled radiology appointments resulting in non-attendance in a hospital setting. In *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (EMBC). IEEE. https://doi.org/10.1109/embc.2017.8037394

Min Chen, Shu-Ching Chen, Mei-Ling Shyu, & Wickramaratna, K. (2006). Semantic event detection via multimodal data mining. *IEEE Signal Processing Magazine*, 23(2), 38–46. https://doi.org/10.1109/msp.2006.1621447

Miron-Shatz, T., Lau, A. Y. S., Paton, C., & Hansen, M. M. (2014). Big Data in Science and Healthcare: A Review of Recent Literature and Perspectives. *Yearbook of Medical Informatics*, 23(01), 21–26. https://doi.org/10.15265/iy-2014-0004

Mohd, T. K., Carvalho, J., & Javaid, A. Y. (2017). Multi-modal data fusion of Voice and EMG data for robotic control. In *2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference* (UEMCON). IEEE. https://doi.org/10.1109/uemcon.2017.8249063

Mohsen, F., Ali, H., El Hajj, N., & Shah, Z. (2022). Artificial intelligence-based methods for fusion of electronic health records and imaging data. *Scientific Reports*, 12(1), 17981. https://doi.org/10.1038/s41598-022-22514-4

Momoki, Y., Ichinose, A., Shigeto, Y., Honda, U., Nakamura, K., & Matsumoto, Y. (2022). Characterization of Pulmonary Nodules in Computed Tomography Images Based on Pseudo-Labeling Using Radiology Reports. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(5), 2582–2591. https://doi.org/10.1109/tcsvt.2021.3073021

Muzic, R. F., Chi-Hsien Chen, & Nelson, A. D. (1998). A method to correct for scatter, spillover, and partial volume effects in region of interest analysis in PET. *IEEE Transactions on Medical Imaging*, 17(2), 202–213. https://doi.org/10.1109/42.700732

Nischitha, & Padmavathi, N. B. (2017). Fusion of multimodal abdominal cancerous images and classification using support vector machine. In *2017 International Conference on Intelligent Sustainable Systems* (ICISS). IEEE. https://doi.org/10.1109/iss1.2017.8389411

Niu, S., Yin, Q., Song, Y., Guo, Y., & Yang, X. (2021). Label Dependent Attention Model for Disease Risk Prediction Using Multimodal Electronic Health Records. In *2021 IEEE International Conference on Data Mining* (ICDM). IEEE. https://doi.org/10.1109/icdm51629.2021.00056

Nunes, J., Mealha, O., Santos, B. S., Pereira, A. S., Botelho, M. F., & Lima, J. P. (1996). Three different approaches in visualization of multimodal clinical data. In *Proceedings of 18th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. https://doi.org/10.1109/iembs.1996.651906

Nunes, N., Martins, B., André da Silva, N., Leite, F., & J. Silva, M. (2019). A Multi-modal Deep Learning Method for Classifying Chest Radiology Exams. In *Progress in Artificial Intelligence*, (pp. 323–335). Springer. https://doi.org/10.1007/978-3-030-30241-2_28

ONC. (2016). Office-based physician electronic health record adoption. Office of the National Coordinator for Health Information Technology. https://dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & McGuinness, L. A. (2021). The PRISMA 2020 statement: an Updated Guideline for Reporting Systematic Reviews. *British Medical Journal*, 372(71), n71. https://doi.org/10.1136/bmj.n71

Pai, M. M. M., Ganiga, R., Pai, R. M., & Sinha, R. K. (2021). Standard electronic health record (EHR) framework for Indian healthcare system. *Health Services and Outcomes Research Methodology*, 21, 339–362. https://doi.org/10.1007/s10742-020-00238-0

Panayides, A. S., Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A., Foran, D., Do, N., Golemati, S., Kurc, T., Huang, K., Nikita, K. S., Veasey, B. P., Zervakis, M., Saltz, J. H., & Pattichis, C. S. (2020). AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE Journal of Biomedical and Health Informatics*, 24(7), 1837–1857. https://doi.org/10.1109/JBHI.2020.2991043

Petersen, S. (2022). HITECH (Health Information Technology for Economic and Clinical Health) Act of 2009. https://www.techtarget.com/searchhealthit/definition/HITECH-Act

Pinho, E., & Costa, C. (2016). Extensible Architecture for Multimodal Information Retrieval in Medical Imaging Archives. In *2016 12th International Conference on Signal-Image Technology & Internet-Based Systems* (SITIS). IEEE. https://doi.org/10.1109/sitis.2016.58

Philipsen, R. H. H. M., Maduskar, P., Hogeweg, L., Melendez, J., Sanchez, C. I., & van Ginneken, B. (2015). Localized Energy-Based Normalization of Medical Images: Application to Chest Radiography. *IEEE Transactions on Medical Imaging*, 34(9), 1965–1975. https://doi.org/10.1109/tmi.2015.2418031

Polikar, R., Tilley, C., Hillis, B., & Clark, C. M. (2010). Multimodal EEG, MRI and PET data fusion for Alzheimer's disease diagnosis. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE. https://doi.org/10.1109/iembs.2010.5627621

Purwar, S., Tripathi, R. K., Ranjan, R., & Saxena, R. (2020). Detection of microcytic hypochromia using cbc and blood film features extracted from convolution neural network by different classifiers. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-019-07927-0

Qu, R., & Xiao, Z. (2022). An Attentive Multi-Modal CNN for Brain Tumor Radiogenomic Classification. *Information*, 13(3), 124. https://doi.org/10.3390/info13030124

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems*, 2(1). https://doi.org/10.1186/2047-2501-2-3

Ramirez-Alonso, G., Prieto-Ordaz, O., Lopez-Santillan, R., & Montes-Y-Gomez, M. (2022). Medical Report Generation through Radiology Images: An Overview. *IEEE Latin America Transactions*, 20(6), 986–999. https://doi.org/10.1109/tla.2022.9757742

Ramesh, T., & Santhi, V. (2020). Exploring big data analytics in health care. *International Journal of Intelligent Networks*, 1, 135–140. https://doi.org/10.1016/j.ijin.2020.11.003

Ramey, J., & Robinson, C. (1991). Video-based task analysis: a tool for understanding your audience (radiology application). In *IPCC 91 Proceedings the Engineered Communication*. IEEE. https://doi.org/10.1109/ipcc.1991.172736

Razzak, M. I., Imran, M., & Xu, G. (2020). Big data analytics for preventive medicine. *Neural Computing and Applications*, 32(9), 4417–4451. https://doi.org/10.1007/s00521-019-04095-y

Reda, I., Khalil, A., Elmogy, M., Abou El-Fetouh, A., Shalaby, A., Abou El-Ghar, M., Elmaghraby, A., Ghazal, M., & El-Baz, A. (2018). Deep Learning Role in Early Diagnosis of Prostate Cancer. *Technology in Cancer Research & Treatment*, 17, 153303461877553. https://doi.org/10.1177/1533034618775530

Rehman, A., Naz, S., & Razzak, I. (2022). Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities. *Multimedia Systems*, 28(4), 1–33. https://doi.org/10.1007/s00530-020-00736-8

Ribeiro, R. T., Marinho, R. T., & Sanches, J. M. (2013). Classification and Staging of Chronic Liver Disease From Multimodal Data. *IEEE Transactions on Biomedical Engineering*, 60(5), 1336–1344. https://doi.org/10.1109/tbme.2012.2235438

Ros Maryana Sinaga, A. S., & Putra, R. E. (2022). Predictive Analytic Healthcare Sector Using Classification Machine Learning Algorithm. In *2022 International Symposium on Information Technology and Digital Innovation* (ISITDI). IEEE. https://doi.org/10.1109/isitdi55734.2022.9944492

Rosenzweig, S., Scholand, N., Holme, H. C. M., & Uecker, M. (2020). Cardiac and Respiratory Self-Gating in Radial MRI Using an Adapted Singular Spectrum Analysis (SSA-FARY). *IEEE Transactions on Medical Imaging*, 39(10), 3029–3041. https://doi.org/10.1109/tmi.2020.2985994

Saxena, A., Singh Tomar, S., Jain, G., & Gupta, R. (2021). Deep learning based Diagnosis of diseases using Image Classification. In *2021 11th International Conference on Cloud Computing, Data Science & Engineering* (Confluence). IEEE. https://doi.org/10.1109/confluence51648.2021.9377154

Schiff, G. D. et al. (2009). Diagnostic Error in Medicine: Analysis of 583 physician-reported errors. *Archives of Internal Medicine*, 169(20), 1881–1887. https://doi.org/10.1001/archinternmed.2009.333

Setio, A. A. A., Ciompi, F., Litjens, G., Gerke, P., Jacobs, C., van Riel, S. J., Wille, M. M. W., Naqibullah, M., Sanchez, C. I., & van Ginneken, B. (2016). Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. *IEEE Transactions on Medical Imaging*, 35(5), 1160–1169. https://doi.org/10.1109/tmi.2016.2536809

Shachor, Y., Greenspan, H., & Goldberger, J. (2020). A mixture of views network with applications to multi-view medical imaging. *Neurocomputing*, 374, 1–9. https://doi.org/10.1016/j.neucom.2019.09.027

Shetty, S., Ananthanarayana, V. S., & Mahale, A. (2020). Medical Knowledge-Based Deep Learning Framework for Disease Prediction on Unstructured Radiology Free-Text Reports Under Low Data Condition. In *Proceedings of the 21st EANN (Engineering Applications of Neural Networks) 2020 Conference*, (pp. 352–364). Springer. https://doi.org/10.1007/978-3-030-48791-1_27

Shetty, S., S., A. V., & Mahale, A. (2022). MS-CheXNet: An Explainable and Lightweight Multi-Scale Dilated Network with Depthwise Separable Convolution for Prediction of Pulmonary Abnormalities in Chest Radiographs. *Mathematics*, 10(19), 3646. https://doi.org/10.3390/math10193646

Shin, H.-C., Le Lu, Kim, L., Seff, A., Yao, J., & Summers, R. M. (2015). Interleaved text/image Deep Mining on a large-scale radiology database. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. https://doi.org/10.1109/cvpr.2015.7298712

Sirshar, M., Paracha, M. F. K., Akram, M. U., Alghamdi, N. S., Zaidi, S. Z. Y., & Fatima, T. (2022). Attention based automated radiology report generation using CNN and LSTM. *PLOS ONE*, 17(1), e0262209. https://doi.org/10.1371/journal.pone.0262209

Skowron, A., Wang, H., Wojna, A., & Bazan, J. (2005). A Hierarchical Approach to Multimodal Classification. In *Lecture Notes in Computer Science*, (pp. 119–127). Springer. https://doi.org/10.1007/11548706_13

Soenksen, L. R., Ma, Y., Zeng, C., Boussioux, L., Villalobos Carballo, K., Na, L., Wiberg, H. M., Li, M. L., Fuentes, I., & Bertsimas, D. (2022). Integrated multimodal artificial intelligence framework for healthcare applications. *Npj Digital Medicine*, 5(1). https://doi.org/10.1038/s41746-022-00689-4

Song, J., Zheng, J., Li, P., Lu, X., Zhu, G., & Shen, P. (2021). An Effective Multimodal Image Fusion Method Using MRI and PET for Alzheimer's Disease Diagnosis. *Frontiers in Digital Health*, 3. https://doi.org/10.3389/fdgth.2021.637386

Sreejith, S., Khanna Nehemiah, H., & Kannan, A. (2022). A clinical decision support system for polycystic ovarian syndrome using red deer algorithm and random forest classifier. *Healthcare Analytics*, 2, 100102. https://doi.org/10.1016/j.health.2022.100102

Srivastava, N., & Salakhutdinov, R. (2014). Multimodal Learning with Deep Boltzmann Machines. *The Journal of Machine Learning Research*, 15(1), 2949–2980.

Sultanum, N., Naeem, F., Brudno, M., & Chevalier, F. (2023). ChartWalk: Navigating large collections of text notes in electronic health records for clinical chart review. *IEEE Transactions on Visualization and Computer Graphics*, 29(1), 1244–1254. https://doi.org/10.1109/tvcg.2022.3209444

Sundararaman, A., Valady Ramanathan, S., & Thati, R. (2018). Novel Approach to Predict Hospital Readmissions Using Feature Selection from Unstructured Data with Class Imbalance. *Big Data Research*, 13, 65–75. https://doi.org/10.1016/j.bdr.2018.05.004

Sutton, R., Pincock, D., Baumgart, D., Sadowski, D., Fedorak, R., & Kroeker, K. (2020). An overview of clinical decision support systems: Benefits, risks, and strategies for success. *NPJ Digital Medicine*, 3(1), 1–10. https://doi.org/10.1038/s41746-020-0221-y

Syeda-Mahmood, T., Wang, F., Beymer, D., Amir, A., Richmond, M., & Hashmi, S. N. (2007). AALIM: Multimodal mining for cardiac decision support. In *2007 Computers in Cardiology*, (pp. 209–212). IEEE. https://doi.org/10.1109/cic.2007.4745458

Tabassum, S., Abedin, N., Rahman, M. M., Rahman, M. M., Ahmed, M. T., Islam, R., & Ahmed, A. (2022). An online cursive handwritten medical words recognition system for busy doctors in developing countries for ensuring efficient healthcare service delivery. *Scientific Reports*, 12(1), 3601. https://doi.org/10.1038/s41598-022-07571-z

Taimoor, N., & Rehman, S. (2022). Reliable and Resilient AI and IoT-Based Personalised Healthcare Services: A Survey. *IEEE Access*, 10, 535–563. https://doi.org/10.1109/access.2021.3137364

Turkeli, S., Akkoca Gazioglu, B. S., Kurt, K. K., Atay, H. T., & Gorur, Y. (2017). Mining similar radiology reports using BoW and Fuzzy C-means clustering. In *2017 International Artificial Intelligence and Data Processing Symposium* (IDAP). IEEE. https://doi.org/10.1109/idap.2017.8090213

Valova, I., & Kosugi, Y. (1997). MR brain image classification by multimodal perceptron tree neural network. Neural Networks for Signal Processing VII. In *Proceedings of the 1997 IEEE Signal Processing Society Workshop*. IEEE. https://doi.org/10.1109/nnsp.1997.622398

van der Putten, N., Vinke, R., Citroen, M., Cornet, R., van Mulligen, E., & den Boer, A. (1995). Integrating medical images, biosignals, and alphanumeric data in a cardiological department. *Computers in Cardiology,* 1995. https://doi.org/10.1109/cic.1995.482748

van der Voort, S. R., Gahrmann, R., van den Bent, M. J., Vincent, A. J. P. E., Niessen, W. J., Smits, M., & Klein, S. (2017). Radiogenomic classification of the 1p/19q status in presumed low-grade gliomas. In 2017 *IEEE 14th International Symposium on Biomedical Imaging* (ISBI 2017). IEEE. https://doi.org/10.1109/isbi.2017.7950601

Wang, F., Syeda-Mahmood, T., & Beymer, D. (2009). Information Extraction from Multimodal ECG Documents. In *2009 10th International Conference on Document Analysis and Recognition*. IEEE. https://doi.org/10.1109/icdar.2009.189

Wang, X., Wang, Y., Gao, C., Lin, K., & Li, Y. (2018). Automatic Diagnosis With Efficient Medical Case Searching Based on Evolving Graphs. *IEEE Access*, 6, 53307–53318. https://doi.org/10.1109/access.2018.2871769

Wang, M., Zhang, D., Huang, J., Yap, P.-T., Shen, D., & Liu, M. (2020). Identifying Autism Spectrum Disorder With Multi-Site fMRI via Low-Rank Domain Adaptation. *IEEE Transactions on Medical Imaging*, 39(3), 644–655. https://doi.org/10.1109/tmi.2019.2933160

Weijie Chen, Metz, C. E., Giger, M. L., & Drukker, K. (2010). A Novel Hybrid Linear/Nonlinear Classifier for Two-Class Classification: Theory, Algorithm, and Applications. *IEEE Transactions on Medical Imaging*, 29(2), 428–441. https://doi.org/10.1109/tmi.2009.2033596

Weibel, N., Ashfaq, S., Calvitti, A., Hollan, J., & Agha, Z. (2013). Multimodal Data Analysis and Visualization to Study the Usage of Electronic Health Records. In *Proceedings of the ICTs for Improving Patients Rehabilitation Research Techniques*. EUDL. https://doi.org/10.4108/icst.pervasivehealth.2013.252025

White, T. (2012). *Hadoop: The Definitive Guide*. O'Reilly Media.

Wood, C., Cross, M., & Phuong La. (1998). Multimodal information retrieval, extraction and generation for use in the health domain. In *Proceedings of the Second International Conference, Knowledge-Based Intelligent Electronic Systems*. IEEE. https://doi.org/10.1109/kes.1998.725988

Wu, H., & He, L. (2015). Combining visual and textual features for medical image modality classification with ℓp−norm multiple kernel learning. *Neurocomputing*, 147, 387–394. https://doi.org/10.1016/j.neucom.2014.06.046

Wu, N., Phang, J., Park, J., Shen, Y., Huang, Z., Zorin, M., Jastrzebski, S., Fevry, T., Katsnelson, J., Kim, E., Wolfson, S., Parikh, U., Gaddam, S., Lin, L. L. Y., Ho, K., Weinstein, J. D., Reig, B., Gao, Y., Toth, H., & Pysarenko, K. (2020). Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening. *IEEE Transactions on Medical Imaging*, 39(4), 1184–1194. https://doi.org/10.1109/tmi.2019.2945514

Wu, M., Zhong, X., Peng, Q., Xu, M., Huang, S., Yuan, J., Ma, J., & Tan, T. (2019). Prediction of molecular subtypes of breast cancer using BI-RADS features based on a "white box" machine learning approach in a multi-modal imaging setting. *European Journal of Radiology*, 114, 175–184. https://doi.org/10.1016/j.ejrad.2019.03.015

Xi, X., Xu, H., Shi, H., Zhang, C., Ding, H. Y., Zhang, G., Tang, Y., & Yin, Y. (2017). Robust texture analysis of multi-modal images using Local Structure Preserving Ranklet and multi-task learning for breast tumor diagnosis. *Neurocomputing*, 259, 210–218. https://doi.org/10.1016/j.neucom.2016.06.082

Yi, S., Zhang, G., Qian, C., Lu, Y., Zhong, H., & He, J. (2022). A Multimodal Classification Architecture for the Severity Diagnosis of Glaucoma Based on Deep Learning. *Frontiers in Neuroscience*, 16, 939472. https://doi.org/10.3389/fnins.2022.939472

Zhang, D., & Shen, D. (2011). Semi-supervised multimodal classification of alzheimer's disease. In *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro* (pp. 1628–1631). IEEE. https://doi.org/10.1109/isbi.2011.5872715

Zhang, Y.-D., Dong, Z., Wang, S.-H., Yu, X., Yao, X., Zhou, Q., Hu, H., Li, M., Jiménez-Mesa, C., Ramirez, J., Martinez, F. J., & Gorriz, J. M. (2020). Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. *Information Fusion*, 64, 149–187. https://doi.org/10.1016/j.inffus.2020.07.006

Zhou, X., Li, Y., & Liang, W. (2021). CNN-RNN Based Intelligent Recommendation for Online Medical Pre-Diagnosis Support. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 18(3), 912–921. https://doi.org/10.1109/tcbb.2020.2994780

Zhu, T., Li, L., Yang, J., Zhao, S., Liu, H., & Qian, J. (2022). Multimodal Sentiment Analysis With Image-Text Interaction Network. *IEEE Transactions on Multimedia*, (in press). https://doi.org/10.1109/tmm.2022.3160060

Zwettler, G., & Backfrieder, W. (2015). Evolution strategy classification utilizing meta features and domain-specific statistical a priori models for fully-automated and entire segmentation of medical datasets in 3D radiology. In *2015 International Conference on Computing and Communications Technologies* (ICCCT). IEEE. https://doi.org/10.1109/iccct2.2015.7292712

Editorial record: The article has been peer-reviewed. First submission received on 14 September 2022. Revisions received on 8 December 2022 and 18 December 2022. Accepted for publication on 24 December 2022. The editors coordinating the peer-review of this manuscript were Venkatachalam Kandasamy, Mohamed Abouhawwash, Nebojsa Bacanin. The editor in charge of approving this manuscript for publication was Zdenek Smutny.

Special Issue: Sustainable Solutions for Internet of Things Using Artificial Intelligence and Blockchain in Future Networks.

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951