scientific reports

Check for updates

OPEN

Artificial intelligence-based methods for fusion of electronic health records and imaging data

Farida Mohsen[1], Hazrat Ali[1], Nady El Hajj[1,2] & Zubair Shah[1]✉

Healthcare data are inherently multimodal, including electronic health records (EHR), medical images, and multi-omics data. Combining these multimodal data sources contributes to a better understanding of human health and provides optimal personalized healthcare. The most important question when using multimodal data is how to fuse them—a field of growing interest among researchers. Advances in artificial intelligence (AI) technologies, particularly machine learning (ML), enable the fusion of these different data modalities to provide multimodal insights. To this end, in this scoping review, we focus on synthesizing and analyzing the literature that uses AI techniques to fuse multimodal medical data for different clinical applications. More specifically, we focus on studies that only fused EHR with medical imaging data to develop various AI methods for clinical applications. We present a comprehensive analysis of the various fusion strategies, the diseases and clinical outcomes for which multimodal fusion was used, the ML algorithms used to perform multimodal fusion for each clinical application, and the available multimodal medical datasets. We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines. We searched Embase, PubMed, Scopus, and Google Scholar to retrieve relevant studies. After pre-processing and screening, we extracted data from 34 studies that fulfilled the inclusion criteria. We found that studies fusing imaging data with EHR are increasing and doubling from 2020 to 2021. In our analysis, a typical workflow was observed: feeding raw data, fusing different data modalities by applying conventional machine learning (ML) or deep learning (DL) algorithms, and finally, evaluating the multimodal fusion through clinical outcome predictions. Specifically, early fusion was the most used technique in most applications for multimodal learning (22 out of 34 studies). We found that multimodality fusion models outperformed traditional single-modality models for the same task. Disease diagnosis and prediction were the most common clinical outcomes (reported in 20 and 10 studies, respectively) from a clinical outcome perspective. Neurological disorders were the dominant category (16 studies). From an AI perspective, conventional ML models were the most used (19 studies), followed by DL models (16 studies). Multimodal data used in the included studies were mostly from private repositories (21 studies). Through this scoping review, we offer new insights for researchers interested in knowing the current state of knowledge within this research field.

Over the past decade, digitization of health data have grown tremendously with increasing data repositories spanning the healthcare sectors[1]. Healthcare data are inherently multimodal, including electronic health records (EHR), medical imaging, multi-omics, and environmental data. In many applications of medicine, the integration (fusion) of different data sources has become necessary for effective prediction, diagnosis, treatment, and planning decisions by combining the complementary power of different modalities, thereby bringing us closer to the goal of precision medicine[2,3].

Data fusion is the process of combining several data modalities, each providing different viewpoints on a common phenomenon to solve an inference problem. The purpose of fusion techniques is to effectively take advantage of cooperative and complementary features of different modalities[4,5]. For example, in interpreting medical images, clinical data is often necessary for making effective diagnostic decisions. Many studies found that missing pertinent clinical and laboratory data during image interpretation decreases the radiologists' ability

[1]College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, 34110 Doha, Qatar. [2]College of Health and Life Sciences, Hamad Bin Khalifa University, Qatar Foundation, 34110 Doha, Qatar. ✉email: zshah@hbku.edu.qa

Previous reviews	Year	Scope and coverage	Comparative contribution of our review
A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics[20]	2022	Their review focused on the fusion of different medical imaging modalities.	Our review focused on the fusion of medical imaging with multimodal EHR data and considered different imaging modalities as a single modality. The two reviews did not share any common studies.
Advances in multimodality data fusion in neuroimaging[21]	2021	Their review focused on the fusion of different imaging modalities, considering neuroimaging applications for brain diseases and neurological disorders.	Our review focused on the fusion of medical imaging with EHR data, considering various diseases, such as neurological disorders, cancer, cardiovascular diseases, psychiatric disorders, eye diseases, and Covid-19. The two reviews did not share any common studies.
An overview of deep learning methods for multimodal medical data mining[22]	2022	Their review focused on the fusion of different types of multi-omics data with EHR and different imaging modalities, only considering DL models for specific diseases (COVID-19, cancer, and Alzheimer's).	Our review focused on the fusion of medical imaging with EHR data, considering all AI models for various diseases, such as neurological disorders, cancer, cardiovascular diseases, psychiatric disorders, eye diseases, and Covid-19. The two reviews did not share any common studies.
Multimodal deep learning for biomedical data fusion: a review[23]	2022	Their review focused on the fusion of different types of multi-omics data with EHR and imaging modalities, considering only DL models. Moreover, they did not provide a summary of the freely accessible multimodal datasets and a summary of evaluation measures used to evaluate the multimodal models.	Our review focused on the fusion of medical imaging with EHR data, considering all AI models. Moreover, o ur study provided a summary of the accessible multimodal datasets and a summary of evaluation measures used to evaluate the multimodal models. The two reviews only shared two common studies.
A comprehensive survey on multimodal medical signals fusion for smart healthcare systems[24]	2021	Their survey did not focus on fusing medical imaging with EHR but rather covered the fusion of IoMTs data for smart healthcare applications and covered studies published untill 2020. Moreover, in their review, multimodality referred to fusing either different 1D medical signals (such as electrocardiogram (ECG) and biosignals), different medical imaging modalities, or 1D medical signals with imaging.	Our review focused on the fusion of medical imaging with EHR (structured and unstructured) for different clinical applications. It included 34 studies, most of them published in 2021 and 2022, with no study common between the two reviews.
Machine learning for multimodal electronic health records-based research: Challenges and perspectives[27]	2021	Their review focused on the fusion of structured and unstructured EHR data and did not consider medical imaging modalities. Moreover, they did not provide a summary of the freely accessible multimodal datasets and a summary of evaluation measures used to evaluate the multimodal models.	Our review focused on the fusion of medical imaging with EHR and considered structured and unstructured data in EHR as a single modality. The two reviews did not share any common studies.
Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines[26]	2020	Their review focused on the fusion of structured EHR data and medical imaging, considering only DL models, and included only 17 studies published until 2019.	Our review focused on the fusion of medical imaging with EHR data, considering all AI models, and included 34 studies, almost more than half published in 2020 and 2021.

Table 1. Comparison with previous reviews.

to accurately make diagnostic decisions[6]. The significance of clinical data to support the accurate interpretation of imaging data is well established in radiology as well as in a wide variety of imaging-based medical specialties such as dermatology, ophthalmology, and pathology that depend on clinical context to interpret imaging data correctly[7–9].

Thanks to the advances of AI and ML models, one can achieve a useful fusion of multimodal data with high-dimensionality[10], various statistical properties, and different missing value patterns[11]. Multimodal ML is the domain that can integrate different data modalities. In recent years, multimodal data fusion has gained much attention for automating clinical outcome prediction and diagnosis. This can be seen in Alzheimer's disease diagnosis and prediction[12–15] when imaging data were combined with specific lab test results and demographic data as inputs to ML models, and better performance was achieved than the single-source models. Similarly, fusing pathological images with patient demographic data observed an increase in performance in comparison with single modality models for breast cancer diagnosis[16]. Several studies found similar advantages in various medical imaging applications, including diabetic retinopathy prediction, COVID-19 detection, and glaucoma diagnosis[17–19].

This scoping review focuses on studies that use AI models to fuse medical images with EHR data for different clinical applications. Modality fusion strategies play a significant role in these studies. In the literature, some other reviews have been published on the use of AI for multimodal medical data fusion[20–26]; however, they differ from our review in terms of their scope and coverage. Some previous studies focused on the fusion of different medical imaging modalities[20,21]; they did not consider the EHR in conjunction with imaging modalities. Other reviews focused on the fusion of omics data with other data modalities using DL models[22,23]. Another study[24] focused on the fusion of various internet of medical things (IoMTs) data for smart healthcare applications. Liu et al.[27] focused exclusively on integrating multimodal EHR data, where multimodality refers to structured data and unstructured free texts in EHR, using conventional ML and DL techniques. Huang et al.[26] discussed fusion strategies of structured EHR data and medical imaging using DL models emphasizing fusion techniques and feature extraction methods. Furthermore, their review covered the research till 2019 and retrieved only 17 studies. In contrast, our review focuses on studies using conventional ML or DL techniques with EHR and medical imaging data, covering 34 recent studies. Table 1 provides a detailed comparison of our review with existing reviews.

The primary purpose of our scoping review is to explore and analyze published scientific literature that fuses EHR and medical imaging using AI models. Therefore, our study aims to answer the following questions:

1. Fusion Strategies: what fusion strategies have been used by researchers to combine medical imaging data with EHR? What is the most used method?
2. Diseases: For what type of diseases are fusion methods implemented?
3. Clinical outcomes and ML methods: What types of clinical outcomes are addressed using the different fusion strategies? What kind of ML algorithms are used for each clinical outcome?
4. Resource: What are the publicly accessible medical multimodal datasets?

We believe that this review will provide a comprehensive overview to the readers on the advancements made in multimodal ML for EHRs and medical imaging data. Furthermore, the reader will develop an understanding of how ML models could be designed to align data from different modalities for various clinical tasks. Besides, we believe that our review will help identify the lack of multimodal data resources for medical imaging and EHR, thus motivating the research community to develop more multimodal medical data.

Preliminaries

We first identify the EHR and medical imaging modalities that are the focus of this review. Then, we present the data fusion strategies that we use to investigate the studies from the perspective of multimodal fusion.

Data modalities. In this review, we focus on studies that use two primary data modalities:

- Medical imaging modality: This includes N-dimensional imaging information acquired in clinical practice, such as X-ray, Magnetic Resonance Imaging (MRI), functional MRI (fMRI), structural MRI (sMRI), Positron Emission Tomography (PET), Computed Tomography (CT), and Ultrasound.
- EHR data: This includes both structured and unstructured free-text data. Structured data include coded data such as diagnosis codes, procedure codes, numerical data such as laboratory test results, and categorical data such as demographic information, family history, vital signs, and medications. Unstructured data include medical reports and clinical notes.

In our review, we consider studies combining the two modalities of EHR and imaging. However, there exist cases where the data could contain only multiple EHR modalities (structured and unstructured) or multiple imaging modalities (e.g., PET and MRI). We consider such data as a single modality, i.e., the EHR modality or imaging modality.

Fusion strategies. As outlined in[26], fusion approaches can be categorized into early, late, and joint fusion. These strategies are classified depending on the stage in which the features are fused in the ML model. Our scoping review follows the definitions in[26] and attempts to match each study to its taxonomy. In this section, we briefly describe each fusion strategy:

- Early fusion: It joins features of multiple input modalities at the input level before being fed into a single ML algorithm for training[26]. The modality features are extracted either manually or by using different methods such as neural networks (NN), software, statistical methods, and word embedding models. When NN are used to extract features, early fusion requires training multiple models: the feature extraction models and the single fusion model. There are two types of joint fusion: type I and type II. Type I fuses the original features without extracting features, while type II fuses extracted features from modalities.
- Late fusion: It trains separate ML models on data of each modality, and the final decision leverages the predictions of each model[26]. Aggregation methods such as weighted average voting, majority voting, or a meta-classifier are used to make the final decision. This type of fusion is often known as decision-level fusion.
- Joint fusion: It combines the learned features from intermediate layers of NN with features from other modalities as inputs to a final model during training[26]. In contrast to early fusion, the loss from the final model is propagated back to the feature extraction model during training so that the learned feature representations are improved through iterative updating of the feature weights. NNs are used for joint fusion since they can propagate loss from the final model to the feature extractor(s). There are two types of joint fusion: type I and type II. The former is when NNs are used to extract features from all modalities. The latter is when not all the input modalities' features are extracted using NNs[26].

Methods

In this scoping review, we followed the guidelines recommended by the PRISMA-ScR[28].

Search strategy. In a structured search, we searched four databases, including Scopus, PubMed, Embase, and Google Scholar, to retrieve the relevant studies. We note here that MEDLINE is covered in PubMed . For Google Scholar search results, we selected the first 110 relevant studies, as, beyond 110 entries, the search results rapidly lost relevancy and were unmatched to our review's topic. Furthermore, we limited our search to English-language articles published in the last seven years between January 1, 2015, and January 6, 2022. The search was based on abstracts and titles and was conducted between January 3 and January 6, 2022.

In this scoping review, we focused on applying AI models to multimodal medical data-based applications. The term multimodal refers to combining medical imaging and EHR, as described in "Preliminaries" section. Therefore, our search string incorporated three major terms connected by AND:(("Artificial Intelligence" OR " machine learning" OR "deep learning") AND "multimodality fusion" AND ("medical imaging" OR "electronic

health records")). We used different forms of each term. We provide the complete search string for all databases in Appendix 1 of the supplementary material.

Inclusion and exclusion criteria. We included all studies that fused EHR with medical imaging modalities using an AI model for any clinical application. As AI models, we considered classical ML models, DL models, transfer learning, ensemble learning, etc as mentioned in the search terms in Appendix 1 of the supplementary material. We did not consider studies that use classical statistical models such as regression in our review. Our definition of imaging modalities is any type of medical imaging used in clinical practice, such as MRI, PET, CT scans, and Ultrasound. We considered both structured and unstructured free-text patients' data for EHR modalities as described in "Preliminaries" section . Only peer-reviewed studies and conference proceedings were included. Moreover, all included studies were limited to English language only. We did not enforce restrictions on types of disorders, diseases or clinical tasks.

We excluded studies that used a single data modality. Also, we excluded studies that used different types of data from the same modality, such as studies that only combined two or more imaging types (e.g. PET and MRI), as we considered this single modality. Moreover, studies that integrated original imaging modalities with extracted imaging features were excluded as this was still considered a single modality. Also, studies that combined multi-omics data modality were excluded. In addition, studies that were unrelated to the medical field or did not use AI-based models were excluded. We excluded reviews, conference abstracts, proposals, editorials, commentaries, letters to editors, preprints, and short letters articles. Non-English publications were also excluded.

Study selection. We used Rayyan web-based review management tool[29] for the first screening and study selection. After removing duplicates, we screened the studies based on title and abstract. Subsequently, full-text of the selected studies from the title and abstract screening were assessed for eligibility using our inclusion and exclusion criteria. Two authors (F.M. and H.A.) conducted the study selection and resolved any conflict through discussion. A third author (Z.S.) was consulted when an agreement could not be reached.

Data extraction. From the final included studies, a data extraction form was designed and piloted on four studies to develop a systematic and accurate data extraction process. The extracted data from the studies are first author's name, year, the country of the first author's institution, disease's name, clinical outcome, imaging modality, EHR modality, fusion strategy, feature extraction methods, data source, AI models, evaluation metrics, and comparison with single modality. In Appendix 2 of the supplementary material, we provide the extracted information description in detail. One author (F.M.) performed the data extraction, and two other authors (Z.S. and H.A.) reviewed and verified the extracted data. Any disagreement was resolved through discussion and consensus between the three authors.

Data synthesis. Following the data extraction, we used a narrative approach to synthesize the data. We analyzed the studies from five perspectives: fusion strategies, diseases, clinical outcomes with ML algorithms, data sources/type, and evaluation mechanism. For fusion strategies, we focused on how the multimodal data was fused. In addition, we recorded implementation details of the model, such as feature extraction and single modality evaluation. We also extracted information on the diseases for which fusion methods were implemented. Furthermore, we analyzed where the data fusion models were applied for clinical outcomes and what ML models were used for each task. Moreover, we focused on the type of imaging and EHR data used by the studies, the source of data, and its availability. Finally, for evaluation, we focused on the evaluation metrics used by each study.

Study quality assessment. In accordance with the guidelines for scoping reviews[30,31], we did not perform quality assessments of the included studies.

Results

Search results. A total of 1158 studies were retrieved from the initial search. After duplicates elimination, 971 studies were retained. Based on our study selection criteria (see Methods), 44 studies remained for full-text review after excluding articles based on their abstract and title. Moreover, 10 studies were removed after the full-text screening. Finally, 34 studies met our inclusion criteria and were selected for data extraction and synthesis. Figure 1 shows a flowchart of the study screening and selection process.

Demographics of the studies. As presented in Table 2, approximately two-thirds of the studies were journal articles ($n = 23, \sim 68\%$)[12–15,17,19,25,32–46], whereas 11 studies were conference proceedings ($\sim 32\%$)[16,47–56]. Most of the studies were published between 2020 and 2022 ($n = 22, \sim 65\%$). Figure 2 shows a visualization of the publication type-wise and year-wise distribution of the studies. The included studies were published in 13 countries; however, most of these studies were from the USA ($n = 10, \sim 30\%$) and China ($n = 8, \sim 24\%$).

Data fusion strategies. We mapped the included studies to the taxonomy of fusion strategies outlined in the "Preliminaries" Section. A primary interest of our review is to identify the fusion strategies that the included studies used to improve the performance of ML models for different clinical outcomes.

Early fusion. The majority of the included studies ($n = 22, \sim 65\%$) used early fusion to combine medical imaging and non-imaging data. When the input modalities have different dimensions, such as when combining one-

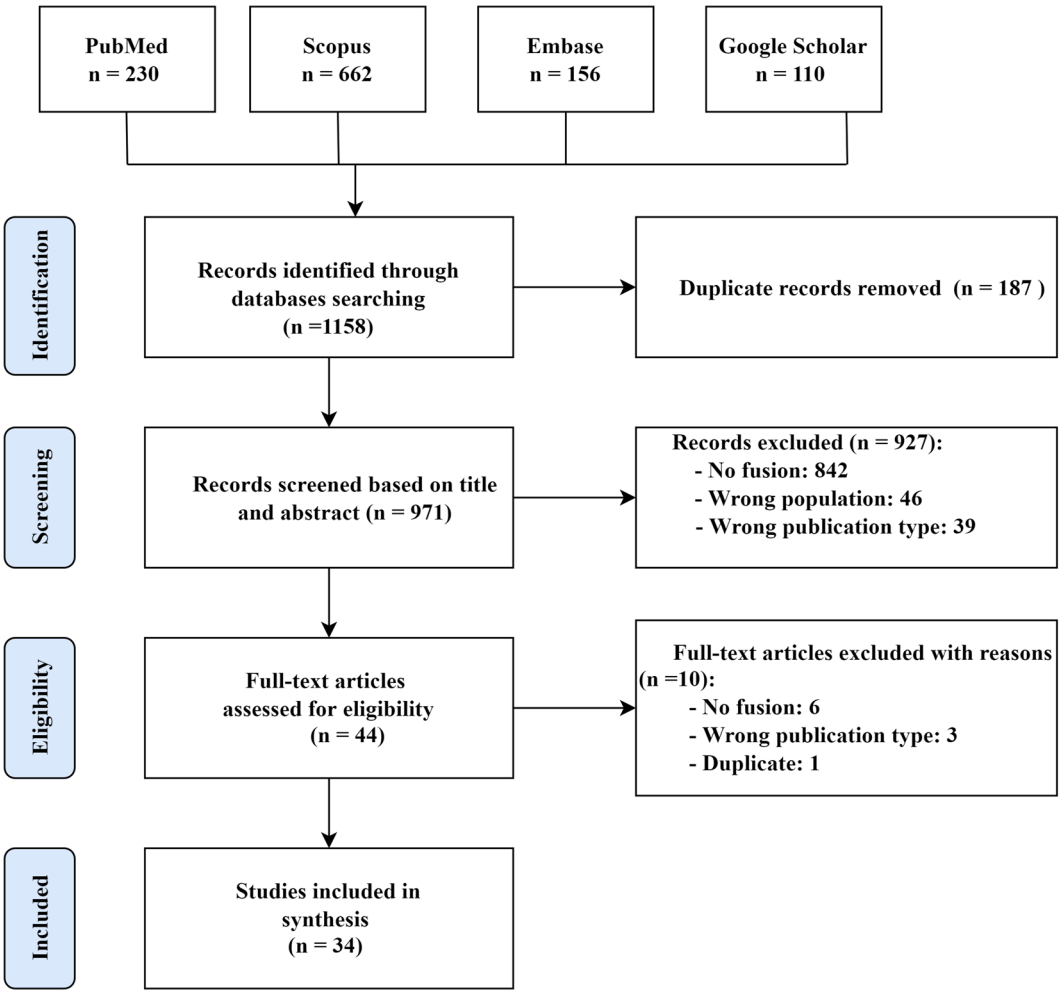

Figure 1. PRISMA flow chart for study identification, screening, and selection.

dimensional (1D) EHR data with 2D or 3D imaging data, it is essential to extract high-level imaging features in 1D before fusing with 1D EHR data. To accomplish this, various methods were used in the studies, including neural network-based features extraction, data generation through software, or manual extraction of features. Out of the 22 early fusion studies, 19 studies[12,13,15,25,33–36,39,41–45,50–53] used manual or software-based imaging features, and 3 studies used neural network-based architectures to extract imaging features before combining with other clinical data modality[16,18,54]. Six out of the 19 studies that used manual or software-based features reduced the feature dimension before concatenating the two modalities' features using different methods[25,36,45,50–52]. Such methods include recursive feature elimination[52], a filter-based method using Pearson correlation coefficient[51], Random Forest feature selection based on Gini importance[50], Relief-based feature selection method[25], a wrapper-based method using backward feature elimination[36], and a rank-based method using Gini coefficients[45]. Moreover, 3 studies[13,15,44] utilized the principal component analysis (PCA) dimensionality reduction technique to reduce the feature dimension.

In the studies that used neural network-based architectures to extract imaging features, CNN architectures were used in three studies[16,18,54]. These studies concatenated the multimodal features (CNN-extracted and EHR features) for their fusion strategy.

Fourteen early fusion studies evaluated their fusion models' performance against that of single modality models[12,13,15,16,18,25,32–34,36,41–44,51] . As a result, 13 of these studies exhibited a better performance for fusion when compared with their imaging-only and clinical-only counterparts[12,13,15,16,18,25,32–34,41–44,51].

Joint fusion. Joint fusion was the second most common fusion strategy used in 10 out of the 34 studies. In these studies, different neural network-based methods were used for processing the imaging and EHR data modalities. Chen et al.[39] used the Visual Geometry Group (VGG-16) architecture to extract features from MRI images, while they used a bidirectional long-short term memory (LSTM) network with an attention layer to learn feature representation from MRI reports. Then, they concatenated the learned features of the two modalities before feeding them into a stacked K-nearest neighbor (KNN) attention pooling layer. Grant et al.[55] used a Residual Network (ResNet50) architecture to extract relevant features from the imaging modality and fully connected NN

Characteristics	Number of studies
Year	
2022	1
2021	14
2020	7
2019	2
2018	5
2016	4
2015	1
Country	
United States of America (USA)	10
China	8
United Kingdom (UK)	4
Germany	2
India	2
Australia	1
Denmark	1
Iran	1
Korea	1
Pakistan	1
Kingdom of Saudi Arabia	1
Singapore	1
Publication type	
Journal	23
Conference	11

Table 2. Demographics of the included studies.

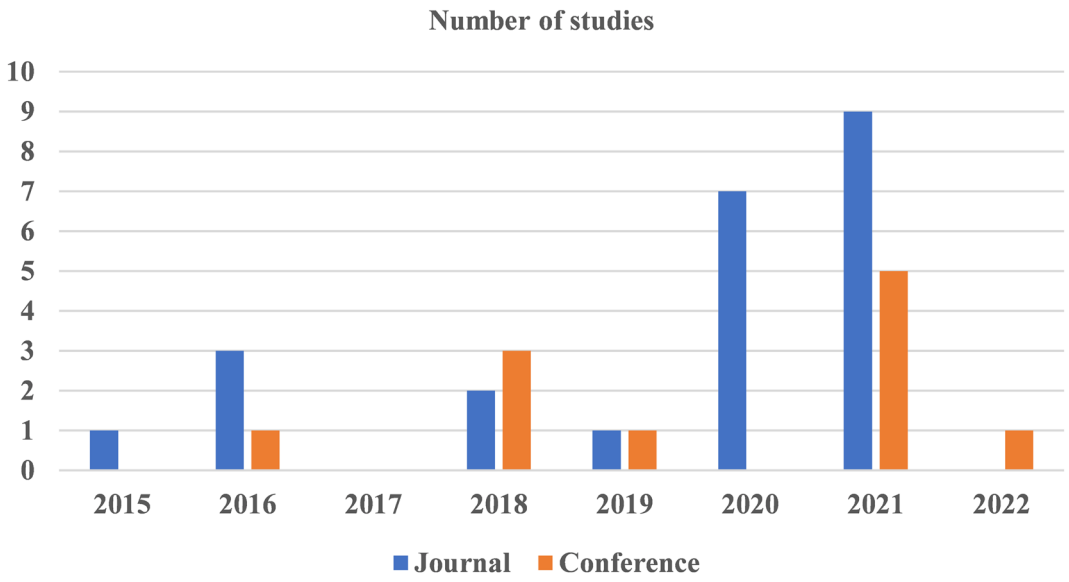

Figure 2. The distribution of studies by the type of publication and the year.

to process the non-imaging data. They directly concatenated the learned feature representation of the imaging and non-imaging data and fed them into two fully connected networks. Yidong et al.[19] used a Bayesian CNN encoder-decoder to extract imaging features and a Bayesian Multilayer perception (MLP) encoder-decoder to process the medical indicators data. The study directly concatenated the two feature vectors and fed the resulted vector into another Bayesian MLP. Samak et al.[47] utilized CNN with a self-attention mechanism to extract the imaging features and fully connected NNs to process the metadata information. Lili et al.[39] used VGG-19 architecture to extract the multimodal MRI features and fully connected networks for clinical data. The study con-

Disease category	Number of studies	Study reference
Neurological disorders	18	
Alzheimer's disease (AD)	7	12–15,44,48,49
Mild cognitive impairment (MCI)	4	37,42,50,51
Ischemic Stroke	2	35,47
Demyelinating diseases	1	32
Neurodevelopmental Deficits	1	39
Epilepsy	1	34
Cancer	5	
Breast Cancer	2	16,41
Glioblastoma	1	43
Lung Cancer	1	55
Upper Gastrointestinal (UGI) Cancer	1	46
Cardiovascular diseases	3	
Aortic stenosis	1	54
Cardiomegaly	1	55
Myocardial Infarction	1	56
Psychiatric disorder	2	
Bipolar disorder	1	33
Schizophrenia	1	36
Eye diseases	2	
Diabetic Retinopathy (DR)	1	17
Glaucoma	1	19
COVID-19	3	18,25,38
Other diseases	3	
Cervical dysplasia	1	53
Pulmonary Embolism (PE)	1	40
Hepatitis B	1	52

Table 3. Disease distribution covered by the 34 studies.

catenated the two feature vectors and fed them into fully connected NN. Another study[46] applied CNN layers for imaging features extraction and word embeddings (Word2vec) with self-attention for textual medical data. In another research[38], Fang et al. applied a ResNet architecture and MLP for imaging and clinical data feature extraction. Then, the authors fused the feature vectors by concatenation and fed them into an LSTM network followed by a fully connected network. Hsu et al.[17] concatenated the imaging features extracted using Inception-V3 model with the clinical data features before feeding them to fully connected NN. In[56], Sharma et al. used CNN to extract image features and then concatenated them directly with the clinical data to feed into a SoftMax classifier. Xu et al.[53] used AlexNet architecture to convert the imaging data into a feature vector fusible with other non-image modalities. Then, they jointly learned the non-linear correlations among all modalities using fully connected NN. Out of 10 joint fusion studies, seven studies evaluated their fusion models' performance against that of a single modality and reported a performance improvement when fusion was used[17,39,46,47,49,53,55].

Late fusion. Late fusion was the least common fusion approach used in the included studies, as only two studies used it. Qiu et al.[37] trained three independent imaging models that took a single MRI slice as input, then aggregated the prediction of these models using maximum, mean, and majority voting. After combining the results of these aggregations by majority vote, the study performed late fusion with the clinical data models. In another study[40], Huang et al. trained four different late fusion models. Three models took the average of the predicted probabilities from the imaging and EHR modality models as the final prediction. The fourth model used an NN classifier as an aggregator, which took as input the single modality models' prediction. The study also created early, joint fusion models and two single modality models to compare with late fusion performance. As a result, the late fusion outperformed both the early and joint fusion models and the single modality models.

Diseases. We categorized the diseases and disorders in the included studies into seven types: neurological disorders, cancer, cardiovascular diseases, Covid-19, psychiatric disorders, eye diseases, and other diseases. The majority of the included studies focused on neurological disorders ($n = 16$). Table 3 shows the distribution of the included studies in terms of the diseases and disorders they covered.

Clinical outcomes and machine learning models. Multimodal ML enables a wide range of clinical applications such as diagnosis, early prediction, patient stratification, phenotyping, biomarkers identification,

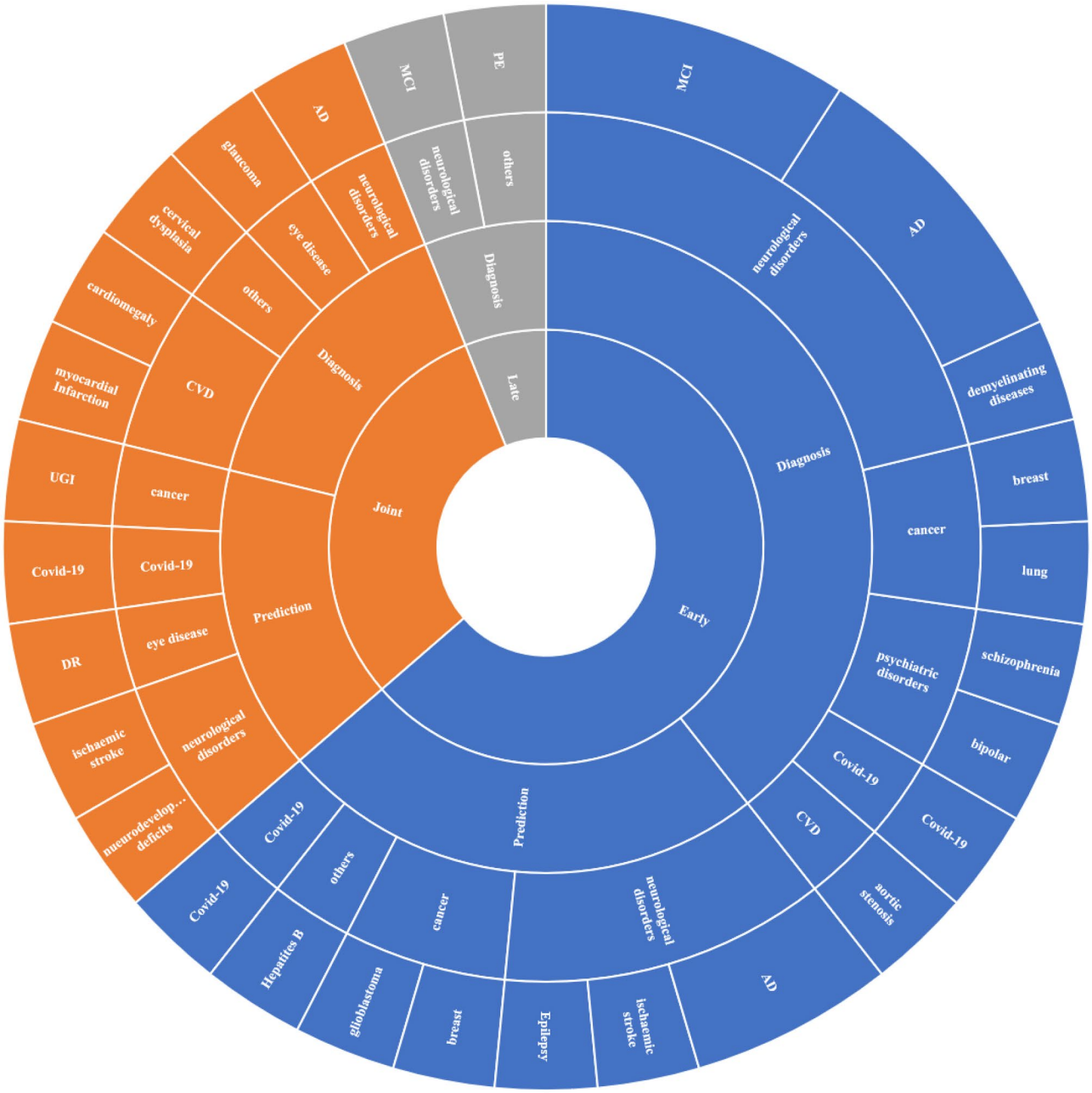

Figure 3. Fusion strategies associated with clinical outcomes for different diseases.

etc. In this review, we labeled each study according to its clinical outcome. We categorized the retrieved clinical tasks into two main categories: diagnosis and prediction. Though some of the studies mentioned detection, classification, diagnosis, and prediction, we categorized them under the diagnosis category. Under the early prediction group, we considered only the studies that predict diseases before onset, identify significant risk factors, predict mortality and overall survival, and predict a treatment outcome. These clinical outcomes were implemented using multimodal ML models. This section summarizes the different clinical tasks of the retrieved studies, the fusion strategy used, and the ML models that were developed for each task. Figure 3 shows the distribution of fusion strategies associated with different diseases' and clinical outcomes.

Diagnosis. The most common applied clinical outcome in the included studies was the diagnosis, reported in 20 ($\sim$ 59%) studies. In these studies, EHRs were combined with medical imaging to diagnose a spectrum of diseases including neurological disorders ($n = 9$)[4,13–15,32,37,42,49,50], psychiatric disorders ($n = 2$)[33,36], CVD ($n = 3$)[54–56], Cancer ($n = 2$)[16,55], and four studies for other different diseases[18,19,40,53]. Specifically, most of the studies that focused on detecting neurological diseases were for AD ($n = 4$)[13–15,49], and MCI ($n = 4$)[37,42,50,51].

Early fusion was the most utilized technique for diagnosis purposes used in 13 studies. These studies employed different ML models on the fused imaging and EHR data for diagnosing different diseases. Most of these studies

were for diagnosing neurological and and psychiatric disorders such as AD[13–15], MCI[42,50,51], demyelinating diseases[32], bipolar disorder[33], and schizophrenia[36] . Parvathy et al.[13] reported diagnosing AD by fusing sMRI and PET imaging features with mini-mental state examination (MMSE) score, clinical dementia rating (CDR), and age of the subjects. They fed the fused features vector to different ML models, including support vector machine (SVM), random forest (RF), and gaussian process (GP) for classification. Niyas et al.[14] classified AD by fusing MRI, PET, demographic data, and lab tests, including cognitive tests and Cerebro-Spinal Fluid (CSF) test. They applied dynamic ensemble of classifiers selection algorithms using a different pool of classifiers on the fused features for classification. Hamid et al.[15] combined MRI and PET imaging features with personal information and neurological data such as MMSE and CRF features for AD early diagnosis. In their study, they fed the fused features into SVM for classification. For MCI diagnosis, Matteo et al.[42] proposed combining MRI imaging with cognitive assessments for MCI diagnosis. They concatenated the features of both modalities and fed them into a linear and quadratic discriminant analysis algorithm for diagnosis. Parisa et al.[50,51] integrated features extracted from MRI and PET images with neuropsychological tests and demographic data (gender, age, and education) to diagnose MCI early. They trained SVM and deep NNs using the fused features for classification in[50,51], respectively. In another study[32], Xin et al. combined MRI imaging with structured data extracted from EHRs to diagnose demyelinating diseases using SVM. For bipolar disorder, Rashmin et al.[33] combined multimodal imaging features with neuropsychological tests and personal information features. They fed them into SVM to differentiate bipolar patients from healthy patients. Ebdrup et al.[36] proposed integrating MRI and diffusion tensor imaging tractography (DTI) imaging with neurocognitive tests and clinical data for schizophrenia classification. Then, they fused the features of the two modalities and fed them to different types of ML classifiers, including SVM, RF, linear regression (LR), decision tree (DT), and Naïve Bayes (NB) for classification.

Moreover, two studies implemented multimodality early fusion to diagnose different cancer diseases[16,55]. Yan et al.[16] fused pathological images and structured data extracted from EHRs to classify malignant and benign breast cancer. Then, they fused the features of the two modalities and fed them to two fully connected NN followed by a SoftMax layer for classification. Seung et al.[55] combined PET imaging with clinical and demographic data for differentiating lung adenocarcinoma (ADC) from squamous cell carcinoma. They fed the integrated features into different algorithms such as SVM, RF, LR, NB, and artificial neural network (ANN) for classification. For COVID-19 diagnosis, Ming et al.[18] combined CT images with clinical features and fed them into different ML models, including SVM, RF, and KNN for diagnosis. Finally, Tanveer et al.[54] combined features from echocardiogram reports and images, with diagnosis information for the detection of patients with aortic stenosis CVD. Their study fed the combined features to an RF learning framework to detect patients likely to have the disease.

Joint fusion was used for diagnostic purposes in 5 studies[19,49,53,55,56]. These studies employed different types of DL architectures to learn and fuse the imaging and EHR data for diagnosis purposes. In[19], they proposed a Bayesian deep multisource learning (BDMSL) model that integrated retinal images with medical indicators data to diagnose glaucoma. For this model, they used Bayesian CNN encoder-decoder to extract imaging features and a Bayesian MLP encoder-decoder to process the medical indicators data. The two feature vectors were directly concatenated and fed into Bayesian MLP for classification. Chen et al.[49] used DL for multimodal feature extraction and classification to detect AD; the authors used the VGG-16 model to extract features from MRI images and a bidirectional LSTM network with an attention layer to learn features from MRI reports. Then, they fed the fused features into a stacked KNN pooling layer to classify the patient's diagnosis data. In[53], Xu et al. proposed an end-to-end deep multimodal framework that can learn better complementary features from the image and non-image modalities for cervical dysplasia diagnosis. They used CNN, specifically AlexNet architecture, to convert the cervigram image data into a feature vector fusible with other non-image modalities. After that, they jointly learned the non-linear correlations among all modalities using fully connected NN for cervical dysplasia classification. Another two studies[55,56] also employed DL models to jointly learn multimodal feature representation for diagnosing CVDs. The former[55] proposed a multimodal network for cardiomegaly classification, which simultaneously integrates the non-imaging intensive care unit (ICU) data (laboratory values, vital sign values, and static patient metadata, including demographics) and the imaging data (chest X-ray). They used a ResNet50 architecture to extract features from the X-ray images and fully connected NN to process the ICU data. To join the learned imaging and non-imaging features, they concatenated the learned feature representation and fed them into two fully connected layers to generate a label for cardiomegaly diagnosis. The latter study[56] proposed a stacked multimodal architecture called SM2N2, which integrated clinical information and MRI images. In their research, they used CNN to extract imaging features, and then they concatenated these features with clinical data to feed into a SoftMax classifier for myocardial infarction detection.

Late fusion was implemented in 2 studies[37,40] for disease diagnosis purposes. Fang et al.[37] proposed the fusion of MRI scans, logical memory (LM) tests, and MMSE for MCI classification. Their study utilized VGG-11 architecture for MRI feature extraction and developed two MLP models for MMSE and LM test results. Then, they combined both MRI and MLP models using majority voting. As a result, the fusion model outperformed the individual models. Huang et al.[40] utilized a non-open dataset comprising CT scans and EHR data to train two unimodal and four late fusion models for PE diagnosis. They used their previously implemented architecture (PENet)[57] to encode the CT images and a feedforward network to encode the tabular data. The late fusion approach performed best among the fusion models and outperformed the models trained on the image-only and the tabular-only data.

Early prediction. Prediction tasks were reported in 14 (∼ 41.2%) studies. In these studies, EHRs were fused with medical imaging to predict different outcomes, including disease prediction, mortality prediction, survival prediction, and treatment outcome prediction. Ten studies of the prediction tasks were disease prediction[12,17,34,38,39,41,44,46,48,52], which involved determining whether an individual might develop a given dis-

ease in the future. The second most common prediction task was treatment outcome prediction reported in 2 studies[35,47], followed by one study for mortality prediction and overall survival prediction[25,43], respectively.

The early fusion technique was used in 6 studies[12,34,41,44,48,52] for disease prediction. Minhas et al.[12] proposed an early fusion model to predict which subjects will progress from MCI to AD in the future. The study concatenated MRI extracted features with demographic and neuropsychological biomarkers before feeding them to an SVM model for prediction. Ali et al.[34] proposed a model to predict Epileptogenic-Zone in the Temporal Lobe by feeding MRI extracted features integrated with set-of-semiology features into various ML models such as LR, SVM, and Gradient Boosting. Ma et al.[41] fused MRI and clinicopathological features for predicting metachronous distant metastasis (DM) in breast cancer. They fed the concatenated features to an LR model. Another study[44] combined MRI-derived features and high-throughput brain phenotyping to diagnose and predict the onset of AD. They fed the fused features into different ML classifiers, including RF, SVM, and LR. Ulyana et al.[48] trained a deep, fully connected network as a regressor in a 5-year longitudinal study on AD to predict cognitive test scores at multiple future time points. Their model produced MMSE scores for ten unique future time points at six-month intervals by combing biomarkers from cognitive test scores, PET, and MRI. They early fused imaging features with the cognitive test scores through concatenation before feeding them into the fully connected network. Finally, Bai et al.[52] compared different multimodal biomarkers (clinical data, biochemical and hemologic parameters, and ultrasound elastography parameters) for predicting the assessment of fibrosis in chronic hepatitis B using SVM.

For disease prediction, joint fusion was used in 4 studies[17,38,39,46]. Hsu et al.[17] proposed a deep multimodal fusion model that trained heterogeneous data from fundus images and non-image data for DR screening. They concatenated the imaging extracted features from Inception-V3 with the clinical data features before feeding them to fully connected NN followed by SoftMax layer for classification. Fang et al.[38] developed a prediction system by jointly fusing CT scans and clinical data to predict the progression of COVID-19 malignancy. In their study, the feature extraction part applied a ResNet architecture and MLP for CT and clinical data, respectively. Then, they concatenated the different features and fed them into an LSTM network followed by a fully connected NN for prediction. In[39], the authors proposed a deep multimodal model for predicting neurodevelopmental deficits at 2 years of age. Their model consisted of a feature extractor and fusion classifier. In the feature extractor, they used VGG-19 architecture to extract MRI features and fully connected NN for clinical data. Then, the study combined the extracted features of the two modalities and fed their combination to another fully connected network in the fusion classifier for prediction. To evaluate the performance of the modality fusion, they tested their model using a single modality of MRI and clinical features. The results showed that multimodal fusion outperformed the single modality performance. Another study[46] also used multimodal joint fusion for UGI cancer screening. Their model integrated features extracted from UGI endoscopic images with corresponding textual medical data. They applied CNN for image feature extraction and word embeddings (Word2vec) with self-attention for textual medical data feature extraction. After that, they concatenated the extracted features of the two modalities and fed them into fully connected NN for prediction. Their results showed that multimodal fusion outperformed the single modality performance.

For treatment outcome prediction[35,47], the former[35] implemented early fusion while the latter[47] used joint fusion. For acute ischemic stroke, Gianluca et al.[35] evaluated the predictive power of imaging, clinical, and angiographic features to predict the outcome of acute ischemic stroke using ML. The study early fused all features into gradient boosting classifiers for prediction. In[47], the authors proposed a DL model to directly exploit multimodal data (clinical metadata and non-contrast CT (NCCT) imaging data) to predict the success of endovascular treatment for ischemic stroke. They utilized CNN with a self-attention mechanism to extract the features of images, and then they concatenated them with the metadata information. Then, the classification stage of the proposed model processed the fused features through a fully connected NN, followed by the Softmax function applied to the outputs. Their results showed that multimodal fusion outperformed the single modality performance.

Both the mortality and overall survival prediction studies[25,43] implemented early fusion. In[25], they developed a model to predict COVID-19 ventilatory support and mortality early on to prioritize patients and manage the hospital resources' allocation. They fused patients' CT images and EHR data features by concatenation before feeding them to different ML models, including SVM, RF, LR, and eXtreme gradient boosting. They evaluated the performance against single modality models and observed that the results for multimodal fusion were better. The other study[43] aimed to develop ML models to predict glioblastoma patients' overall survival (OS) and progression-free survival (PFS) based on combining treatment features, pathological, clinical, PET/CT-derived information, and semantic MRI-based features. They concatenated the features of all modalities and fed them to an RF model. The study showed that the model based on multimodal fusion data outperformed the single modality models.

Datasets. *Patient data types.* The included studies reported medical imaging and EHRs (structured and non-structured) patient's data types. In terms of imaging modality, CT, MRI, fMRI, structural MRI (sMRI), PET, Diffusion MRI, DTI, ultrasound, X-ray, fundus images, and PET were used in the studies. MRI and PET images were the most utilized modalities. Out of the included 34 studies, 13 used MRI images, and 8 used PET images mostly for AD diagnosis and prediction. In terms of EHRs, structured data was the most commonly used modality ($n = 32$). Table 4 summarizes the types of imaging and EHR data used in the studies.

Patient data resources. Almost two-thirds of the studies included in this scoping review used private data sources (clinical data that are not publicly available) ($n = 21, \sim 59\%$). In contrast, publicly accessible datasets were used in only 13 studies. We observed that the most used public dataset was the "Alzheimer's Disease Neuroimaging Initiative" dataset (ADNI)[58] , where 7 out of 13 studies used the dataset. Other publicly available datasets that were used among the included studies were the "National Alzheimer's Coordinating Center" (NACC)

Data Type	Number of studies	Study reference
Imaging data		
MRI imaging		
MRI	13	12,14,15,32,33,37,41–43,48–51
DTI	3	33,36,39
fMRI	2	33,39
sMRI and Diffusion MRI	1	44
PET	8	13–15,43,45,48,50,51
CT	7	18,35,38,40,43,45,47
X-ray	2	25,55
fundus images	2	17,19
Ultrasound	1	52
Echocardiography	1	54
Pathological images	1	16
Cervigram images	1	53
Endoscopy images	1	46
EHR data		
Structured	32	12–19,25,32–45,47,48,50–56
Unstructured	2	46,49

Table 4. Patient data types used in the included studies.

Public dataset	Description	URL	Clinical outcomes	Study reference
ADNI	ADNI represents a series of studies, including ADNI 1, 2, and 3, designed to study MCI and its progression into AD. It has MRI and PET images along with clinical and genetic information[58]	https://adni.loni.usc.edu/data-samples/data-types/	Disease diagnosis (AD) Disease diagnosis (MCI) Disease Prediction (AD)	13,1550,5112,44,48
ADNI TADPOLE	ADNI has a simplified counterpart, TADPOLE, which has a subset of ADNI-3 samples and features. ATDPOLE does not include raw images, but it has processed structural information about the images such as ROI averages, thicknesses of the cortex and volumes of brain sub-regions, etc[61]	https://tadpole.grand-challenge.org/Data/	Disease diagnosis (AD)	14
NACC	The NACC dataset was established to facilitate collaborative AD research. The dataset comprises MRI data, demographic data, neuropsychological testing scores, and clinical diagnosis of patients[59]	https://naccdata.org/requesting-data/nacc-data	Disease Diagnosis (MCI)	37
MIMIC-CXR, MIMIC-IV	MIMIC-CXR is a dataset of patient chest radiographs. It contains X-ray studies for 64,588 patients[63]. MIMIC-IV is a database for patients admitted to critical care units comprising patient stay information, patient's ICU data, and lookup tables to allow linking to MIMIC-CXR[60]	MIMIC-CXR:https://www.nature.com/articles/s41597-019-0322-0 MIMIC-IV:https://physionet.org/content/mimiciv/0.4/	Disease diagnosis (Cardiomegaly)	55
NCI	Data collections produced by major NCI initiatives are listed in the NCI Data Catalog, including Clinical data, Genomics, imaging, and Proteomics	https://datascience.cancer.gov/resources/nci-data-catalog	Disease diagnosis (Cervical dysplasia)	53
MR CLEAN Trial	A longitudinal study of 500 patients treated with endovascular therapy in The Netherlands for acute ischemic stroke comprising NCCT images, CT Angiography (CTA) images, and clinical metadata information on its patients[62]	https://www.mrclean-trial.org/home.html	Treatment outcome prediction (ischemic stroke)	47

Table 5. Multimodal medical datasets and clinical outcome applications.

dataset[59], the "Medical Information Mart for Intensive Care" (MIMIC-IV) dataset[60], the "National Cancer Institute" (NCI) dataset, ADNI TADPOLE dataset[61] , and MR CLEAN Trial dataset[62]. In Table 5, we summarize the public multimodal medical datasets and their clinical applications. Considering these datasets for each clinical task, the most popular is ADNI for AD and MCI disease diagnosis and prediction.

Evaluation metrics	Number of studies	Study reference
Accuracy	31	12,15–19,25,32–42,44–47,49–56
Sensitivity (recall)	20	12–15,17,19,32–34,37–41,45,46,50,51,53,54
AUC	17	12,15–17,19,25,32,35,37–39,41,45,47,52,53,55
Specificity	15	12,14,15,17,32–34,38–41,46,50,51,53
Precision	7	13,19,34,37,45,54
Positive predictive value (PPV) and Negative predictive value(NPV)	3	15,34,40
Matthews correlation coefficient (MCC)	2	34,41
C-index	1	43
Root-Mean Squared Error (RMSE)	1	48

Table 6. The distribution of evaluation metrics in the included studies. The numbers in the second column do not sum up to 34 as many studies used more than a single metric.

Evaluation metrics. Evaluation metrics are mainly dependent on the clinical task. Typically, accuracy, the area under the curve (AUC), sensitivity, specificity, F1- measure, and precision are mostly used for the evaluation of diagnosis and prediction tasks. Table 6 shows the distribution of the evaluation measures used in the included studies

Discussion

This section summarizes our findings and provides future directions for research on the multimodal fusion of medical imaging and EHR.

Principal findings. We found that multimodal models that combined EHR and medical imaging data generally outperformed single modality models for the same task in disease diagnosis or prediction. Since our review shows that the fusion of medical imaging and clinical context data can improve the performance of AI models, we recommend attempting fusion approaches when multimodal data is obtainable. Moreover, through this review, we observed certain trends in the field of multimodality fusion in the medical area, which can be categorized as:

- *Resources*: We observed that multimodal data resources of medical imaging and EHR are limited owing to privacy considerations. The most prominent dataset was the ADNI, containing MRI and PET images collected from about 1700 individuals in addition to clinical and genetic information. Considering ADNI's contributions in advancing the research, similar multimodal datasets should be developed for other medical data sources too.
- *Fusion implementation*: Early fusion was the most commonly used technique in most applications for multimodal learning. Before fusing 1D EHRs data with image data in 2D or 3D, images data was converted to a 1D vector by extracting high-level representations using manual or software-generated features[12,13,15,25,33–36,39,41–45,50–53], or CNN-extracted features[8,16,54]. The learned imaging features from CNN often resulted in better task-specific performance than manually or software-derived features[64]. Based on this reviewed studies, early fusion models performed better than conventional single-modality models on the same task. Researchers can use the early fusion method as a first attempt to learn multimodal representations since it can learn to exploit the interactions and correlations between features of each modality. Furthermore, it only requires one model to be trained, making the pipeline for training easier than that of joint and late fusion. However, if imaging features are extracted with CNN, early fusion requires multiple models to be trained. Joint fusion was the second most commonly used fusion approach. From a modality perspective, CNNs appeared to be the best option for image feature extraction. Tabular data were mainly processed using dense layers when fed into a model, while text data were mostly processed using LSTM layers followed by the attention layer. Most of the current research directly concatenated the feature vectors of the different modalities to combine multimodal data. Using NNs to implement joint fusion can be a limitation when dealing with small datasets, which means that joint fusion is preferred with large datasets. For small datasets, it is preferable to use early or late fusion methods as they can be implemented using classical ML techniques. Nevertheless, we expect and agree with[26] that joint fusion models can provide better results than other fusion strategies because they update their feature representations iteratively by propagating the loss to all the feature extraction models, aiming to learn correlations across modalities. Based on the performance reported in the included studies, it is preferred to try the early and joint fusion when the relation between the two data modalities is complementary. In this review, AD diagnosis is an example in which imaging and EHRs data are dependent as relevant and accurate knowledge of the patient's current symptomatology, personal information and imaging reports can help doctors interpret imaging results in a suitable clinical context, resulting in a more precise diagnosis. Therefore, all AD diagnosis studies in this review implemented either early fusion[13–15] or joint fusion[49] for multimodal learning. On the other hand, it is preferred to try late fusion when input modalities do not complement each other. For example, the brain MRI pixel data and the quantitative result of an MMSE (e.g., Qiu et al.[37]) for diagnosing MCI are independent, making them appropriate candidates for inclusion in the late fusion strategy. Also, late fusion does not impose the

requirement of a huge amount of training data, so it could be used when the modalities data sizes are small. Moreover, late fusion strategy could be attempted when the concatenation of feature vectors from multiple modalities results in high-dimensional vectors that are difficult for ML algorithms to learn without overfitting unless many input samples are available. In late fusion, multiple models are employed, each specialized in a single modality, thereby limiting the size of the input feature vector for each model. Furthermore, late fusion could be used when data is incomplete or missing, i.e., some patients have only imaging data but no clinical data or vice versa. This is because late fusion uses independent models for different modalities, and aggregation methods like averaging and majority voting can be used even when predictions from a modality are not present. Moreover, predictions could be disproportionately influenced by the most feature-rich input modality when the number of features is very different between the input data modalities[65]; in this scenario, late fusion is preferable because it allows training each model using each modality separately.

- *Applications*: In this review, we found that AD diagnosis and prediction[12–15,44,48,49] were the most common applications addressed in a multimodal setting among studies. Using ML fusion techniques consistently demonstrated improved AD diagnosis, while clinicians experience difficulty with accurate and reliable diagnosis even when multimodal data is available[26]. This emphasizes the utility and significance of multimodal fusion approaches in clinical applications.
- *Prospects*: In this review, we noted that multimodal medical data fusion is growing due to its potential in achieving state-of-the-art performance for healthcare applications. Nonetheless, this growth is hampered by the absence of adequate data for benchmarking methods. This is not surprising, given the privacy concerns surrounding revealing healthcare data. Moreover, we observed a lack of complexity in the used non-imaging data, particularly in the context of heavily feature-rich data included in the EHR. For example, the majority of studies focused mostly on basic demographic data like gender and age[12,15,44,51], a limited number of studies also included medical histories such as smoking status and hypertension[18,55] or specific clinical characteristics that are known to be associated with a certain disease, such as an MMSE for diagnosing AD. In addition to selecting the disease-associated features, future research may benefit from using vast amounts of feature-rich data, as demonstrated in domains outside of medicine, such as autonomous driving[66].

Future directions. Although we focus on EHR and medical imaging as multimodal data, other modalities such as multi-omics and environmental data could also be integrated using the aforementioned fusion approaches. As the causes of many diseases are complex, many factors, including inherited genetics, lifestyle, and living environments, contribute to the development of diseases. Therefore, combining multisource data, e.g. EHR, imaging, and multi-omics data, may lead to a holistic view that can improve patient outcomes through personalized medicine.

Although we focus on EHR and medical imaging as multimodal data, other modalities such as multi-omics and environmental data could also be integrated using the aforementioned fusion approaches. As the causes of many diseases are complex, many factors, including inherited genetics, lifestyle, and living environments, contribute to the development of diseases. Therefore, combining multisource data, e.g. EHR, imaging, and multi-omics data, may lead to a holistic view that can improve patient outcomes through personalized medicine.

Moreover, the unavailability of multimodal public data is a limitation that hinders the development of corresponding research. Many factors (e.g., gender, ethnicity, environmental factors) could influence the research directions or even clinical decision, relying on a few publicly available datasets might not be enough for making conclusive clinical claims to the global population[27]. Consequently, it is imperative to encourage the sharing of flexible data among institutions and hospitals in order to facilitate the exploration of a wider range of population data for clinical research. In ML, federated learning (FL)[67,68] provides the ability to collect data safely and securely from multiple centers. It may be used to collect multimodal data from various centers to train a large-scale model without collecting data directly.

Limitations. Our search was limited to studies published within the previous seven years (2015–2022). We only considered studies published in English, which may have led to leaving out some studies published in other languages. We solely included studies fusing EHR with medical imaging. We did not include studies that used other data modalities such as multi-omics data, as they are out of the scope of this work. Because positive results are typically reported disproportionately, publication bias might be another limitation of this review. This bias may result in an overestimation of the benefits associated with multimodal data analysis. The studies included in this review employed various input modalities, investigated various clinical tasks for different diseases, and reported different performance metrics; hence a direct comparison of the results presented in the studies is not always applicable. Furthermore, not all articles provided confidence bounds, making it difficult to compare their results statistically.

Conclusion

Multimodal ML is an area of research that is gaining attention within the medical field. This review surveyed multimodal medical ML literature that combines EHR with medical imaging data. It discussed fusion strategies, the clinical tasks and ML models that implemented data fusion, the type of diseases, and the publicly accessible multimodal data for medical imaging and EHRs. Furthermore, it highlighted some directions to pave the way for future research. Our finding suggests that there is a growing interest in multimodal medical data. Still, most studies combine the modalities with relatively simple strategies, which despite being shown to be effective, might not fully exploit the rich information embedded in these modalities. As this is a fast-growing field and new AI models with multimodal data are constantly being developed, there might exist studies that fall outside our definition of fusion strategies or use a combination of these strategies. We believe that the development of this

field will give rise to more comprehensive multimodal medical data analysis and will be of great support to the clinical decision-making process.

Data availability

The data generated during this scoping review is provided as supplementary materials.

Received: 14 June 2022; Accepted: 17 October 2022
Published online: 26 October 2022

References

1. Murdoch, T. B. & Detsky, A. S. The inevitable application of big data to health care. *JAMA* **309**, 1351–1352 (2013).
2. Obermeyer, Z. & Emanuel, E. J. Predicting the future-big data, machine learning, and clinical medicine. *N. Engl. J. Med.* **375**, 1216 (2016).
3. Roski, J., Bo-Linn, G. W. & Andrews, T. A. Creating value in health care through big data: Opportunities and policy implications. *Health Aff.* **33**, 1115–1122 (2014).
4. Lozano-Perez, T. *Autonomous Robot Vehicles* (Springer, 2012).
5. Castanedo, F. A review of data fusion techniques. *Sci. World J.* **2013**, 704504 (2013).
6. Cohen, M. D. Accuracy of information on imaging requisitions: Does it matter?. *J. Am. Coll. Radiol.* **4**, 617–621 (2007).
7. Comfere, N. I. *et al.* Provider-to-provider communication in dermatology and implications of missing clinical information in skin biopsy requisition forms: a systematic review. *Int. J. Dermatol.* **53**, 549–557 (2014).
8. Jonas, J. B. *et al.* Glaucoma. *The Lancet* **390**, 2183–2193. https://doi.org/10.1016/S0140-6736(17)31469-1 (2017).
9. Comfere, N. I. *et al.* Dermatopathologists' concerns and challenges with clinical information in the skin biopsy requisition form: A mixed-methods study. *J. Cutan. Pathol.* **42**, 333–345 (2015).
10. Li, Y., Wu, F.-X. & Ngom, A. A review on machine learning principles for multi-view biological data integration. *Brief. Bioinform.* **19**, 325–340 (2018).
11. Ramachandram, D. & Taylor, G. W. Deep multimodal learning: A survey on recent advances and trends. *IEEE Signal Process. Mag.* **34**, 96–108 (2017).
12. Minhas, S. *et al.* Early MCI-to-AD conversion prediction using future value forecasting of multimodal features. *Comput. Intell. Neurosci.* **2021**, 6628036 (2021).
13. Pillai, P.S., Leong, T.-Y., Initiative, A.D.N. *et al.* Fusing heterogeneous data for Alzheimer's disease classification. In *MEDINFO 2015: eHealth-enabled Health*, 731–735 (IOS Press, 2015).
14. KP, M. N. & Thiyagarajan, P. Alzheimer's classification using dynamic ensemble of classifiers selection algorithms: A performance analysis. *Biomed. Signal Process. Control* **68**, 102729 (2021).
15. Akramifard, H., Balafar, M. A., Razavi, S. N. & Ramli, A. R. Early detection of Alzheimer's disease based on clinical trials, three-dimensional imaging data, and personal information using autoencoders. *J. Med. Signals Sensors* **11**, 120 (2021).
16. Yan, R. *et al.* Richer fusion network for breast cancer classification based on multimodal data. *BMC Med. Inform. Decis. Mak.* **21**, 1–15 (2021).
17. Hsu, M.-Y. *et al.* Deep learning for automated diabetic retinopathy screening fused with heterogeneous data from EHRs can lead to earlier referral decisions. *Transl. Vis. Sci. Technol.* **10**, 18 (2021).
18. Xu, M. *et al.* Accurately differentiating between patients with COVID-19, patients with other viral infections, and healthy individuals: Multimodal late fusion learning approach. *J. Med. Internet Res.* **23**, e25535 (2021).
19. Chai, Y., Bian, Y., Liu, H., Li, J. & Xu, J. Glaucoma diagnosis in the Chinese context: An uncertainty information-centric Bayesian deep learning model. *Inf. Process. Manag.* **58**, 102454 (2021).
20. Azam, M. A. *et al.* A review on multimodal medical image fusion: Compendious analysis of medical modalities, multimodal databases, fusion techniques and quality metrics. *Comput. Biol. Med.* **144**, 105253. https://doi.org/10.1016/j.compbiomed.2022.105253 (2022).
21. Zhang, Y.-D. *et al.* Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. *Inf. Fusion* **64**, 149–187. https://doi.org/10.1016/j.inffus.2020.07.006 (2020).
22. Behrad, F. & Saniee Abadeh, M. An overview of deep learning methods for multimodal medical data mining. *Expert Syst. Appl.* **200**, 117006. https://doi.org/10.1016/j.eswa.2022.117006 (2022).
23. Stahlschmidt, S. R., Ulfenborg, B. & Synnergren, J. Multimodal deep learning for biomedical data fusion: A review. *Brief. Bioinform.* **23**, bbab569 (2022).
24. Muhammad, G. *et al.* A comprehensive survey on multimodal medical signals fusion for smart healthcare systems. *Inf. Fusion* **76**, 355–375. https://doi.org/10.1016/j.inffus.2021.06.007 (2021).
25. Aljouie, A. F. *et al.* Early prediction of COVID-19 ventilation requirement and mortality from routinely collected baseline chest radiographs, laboratory, and clinical data with machine learning. *J. Multidiscip. Healthc.* **14**, 2017 (2021).
26. Huang, S.-C., Pareek, A., Seyyedi, S., Banerjee, I. & Lungren, M. P. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *NPJ Digit. Med.* **3**, 1–9 (2020).
27. Liu, Z. *et al.* Machine learning for multimodal electronic health records-based research: Challenges and perspectives. arXiv preprint arXiv:2111.04898 (2021).
28. Tricco, A. C. *et al.* Prisma extension for scoping reviews (PRISMA-ScR): Checklist and explanation. *Ann. Intern. Med.* **169**, 467–473 (2018).
29. Ouzzani, M., Hammady, H., Fedorowicz, Z. & Elmagarmid, A. Rayyan—A web and mobile app for systematic reviews. *Syst. Rev.* **5**, 1–10 (2016).
30. Arksey, H. & O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **8**, 19–32 (2005).
31. Grant, M. J. & Booth, A. A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Inf. Libraries J* **26**, 91–108 (2009).
32. Xin, B., Huang, J., Zhou, Y., Lu, J. & Wang, X. Interpretation on deep multimodal fusion for diagnostic classification. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1–8 (IEEE, 2021).
33. Achalia, R. *et al.* A proof of concept machine learning analysis using multimodal neuroimaging and neurocognitive measures as predictive biomarker in bipolar disorder. *Asian J. Psychiatr.* **50**, 101984 (2020).
34. Alim-Marvasti, A. *et al.* Machine learning for localizing epileptogenic-zone in the temporal lobe: Quantifying the value of multimodal clinical-semiology and imaging concordance. *Front. Digit. Health* **3**, 8 (2021).
35. Brugnara, G. *et al.* Multimodal predictive modeling of endovascular treatment outcome for acute ischemic stroke using machine-learning. *Stroke* **51**, 3541–3551 (2020).
36. Ebdrup, B. H. *et al.* Accuracy of diagnostic classification algorithms using cognitive-, electrophysiological-, and neuroanatomical data in antipsychotic-naïve schizophrenia patients. *Psychol. Med.* **49**, 2754–2763 (2019).
37. Qiu, S. *et al.* Fusion of deep learning models of MRI scans, mini-mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimers Dement. Diagn. Assess. Dis. Monit* **10**, 737–749 (2018).

38. Fang, C. *et al.* Deep learning for predicting COVID-19 malignant progression. *Med. Image Anal.* **72**, 102096 (2021).
39. He, L. *et al.* Deep multimodal learning from MRI and clinical data for early prediction of neurodevelopmental deficits in very preterm infants. *Front. Neurosci.* **15**, 753033 (2021).
40. Huang, S.-C., Pareek, A., Zamanian, R., Banerjee, I. & Lungren, M. P. Multimodal fusion with deep neural networks for leveraging CT imaging and electronic health record: A case-study in pulmonary embolism detection. *Sci. Rep.* **10**, 1–9 (2020).
41. Ma, W. *et al.* Distant metastasis prediction via a multi-feature fusion model in breast cancer. *Aging (Albany NY)* **12**, 18151 (2020).
42. De Marco, M., Beltrachini, L., Biancardi, A., Frangi, A. F. & Venneri, A. Machine-learning support to individual diagnosis of mild cognitive impairment using multimodal MRI and cognitive assessments. *Alzheimer Dis. Assoc. Disord.* **31**, 278–286 (2017).
43. Peeken, J. C. *et al.* Combining multimodal imaging and treatment features improves machine learning-based prognostic assessment in patients with glioblastoma multiforme. *Cancer Med.* **8**, 128–136 (2019).
44. Wang, Y. *et al.* Diagnosis and prognosis of Alzheimer's disease using brain morphometry and white matter connectomes. *NeuroImage Clin.* **23**, 101859 (2019).
45. Hyun, S. H., Ahn, M. S., Koh, Y. W. & Lee, S. J. A machine-learning approach using pet-based radiomics to predict the histological subtypes of lung cancer. *Clin. Nucl. Med.* **44**, 956–960 (2019).
46. Ding, S., Huang, H., Li, Z., Liu, X. & Yang, S. SCNET: A novel UGI cancer screening framework based on semantic-level multimodal data fusion. *IEEE J. Biomed. Health Inform.* **25**, 143–151 (2020).
47. Samak, Z. A., Clatworthy, P. & Mirmehdi, M. Prediction of thrombectomy functional outcomes using multimodal data. In *Annual Conference on Medical Image Understanding and Analysis*, 267–279 (Springer, 2020).
48. Morar, U. *et al.* A deep-learning approach for the prediction of mini-mental state examination scores in a multimodal longitudinal study. In *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, 761–766 (IEEE, 2020).
49. Chen, D., Zhang, L. & Ma, C. A multimodal diagnosis predictive model of Alzheimer's disease with few-shot learning. In *2020 International Conference on Public Health and Data Science (ICPHDS)*, 273–277, https://doi.org/10.1109/ICPHDS51617.2020.00060 (2020).
50. Forouzannezhad, P., Abbaspour, A., Cabrerizo, M. & Adjouadi, M. Early diagnosis of mild cognitive impairment using random forest feature selection. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 1–4, https://doi.org/10.1109/BIOCAS.2018.8584773 (2018).
51. Forouzannezhad, P., Abbaspour, A., Li, C., Cabrerizo, M. & Adjouadi, M. A deep neural network approach for early diagnosis of mild cognitive impairment using multiple features. In *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1341–1346, https://doi.org/10.1109/ICMLA.2018.00218 (2018).
52. Bai, Y., Chen, X., Dong, C., Liu, Y. & 0001, Z. Z. A comparison of multimodal biomarkers for chronic hepatitis b assessment using recursive feature elimination. In *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2016*, Orlando, FL, USA, August 16-20, 2016, 2448–2451, https://doi.org/10.1109/EMBC.2016.7591225 (IEEE, 2016).
53. Xu, T., Zhang, H., Huang, X., Zhang, S. & Metaxas, D. N. Multimodal deep learning for cervical dysplasia diagnosis. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016* (eds Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W.), 115–123 (Springer International Publishing, Cham, 2016).
54. Syeda-Mahmood, T. et al. Identifying patients at risk for aortic stenosis through learning from multimodal data. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016* (eds Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W.), 238–245 (Springer International Publishing, Cham, 2016).
55. Grant, D., Papież, B. W., Parsons, G., Tarassenko, L. & Mahdi, A. Deep learning classification of cardiomegaly using combined imaging and non-imaging icu data. In *Medical Image Understanding and Analysis* (eds Papież, B. W., Yaqub, M., Jiao, J., Namburete, A. I. L. & Noble, J. A.),547–558 (Springer International Publishing, Cham, 2021).
56. Sharma, R., Eick, C. F. & Tsekos, N. V. Sm2n2: A stacked architecture for multimodal data and its application to myocardial infarction detection. In *Statistical Atlases and Computational Models of the Heart. M &Ms and EMIDEC Challenges* (eds Puyol Anton, E. *et al.*) 342–350 (Springer International Publishing, Cham, 2021).
57. Huang, S.-C. *et al.* PENet—A scalable deep-learning model for automated diagnosis of pulmonary embolism using volumetric CT imaging. *NPJ Digit. Med.* **3**, 1–9 (2020).
58. Mueller, S. *et al.* The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clin. N. Am.* **15**, 869–877. https://doi.org/10.1016/j.nic.2005.09.008 (2005).
59. Beekly, D. *et al.* The National Alzheimer's Coordinating Center (NACC) database: An Alzheimer disease database. *Alzheimer Dis. Assoc. Disord.* **18**, 270–277 (2004).
60. Alistair, J. et al. Mimic-iv (version 0.4). PhysioNet https://doi.org/10.13026/a3wn-hq05 (2020).
61. Marinescu, R. V. *et al.* Tadpole challenge: prediction of longitudinal evolution in Alzheimer's disease. arXiv preprint arXiv:1805.03909 (2018).
62. Fransen, P. S. *et al.* MR CLEAN, a multicenter randomized clinical trial of endovascular treatment for acute ischemic stroke in the netherlands: Study protocol for a randomized controlled trial. *Trials* **15**, 1–11 (2014).
63. Johnson, A. E. *et al.* MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Sci. Data* **6**, 1–8 (2019).
64. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT press, 2016).
65. Reda, I. *et al.* Deep learning role in early diagnosis of prostate cancer. *Technol. Cancer Res. Treat.* **17**, 1533034618775530 (2018).
66. Hecker, S., Dai, D. & Van Gool, L. End-to-end learning of driving models with surround-view cameras and route planners. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 435–453 (2018).
67. Li, T., Sahu, A. K., Talwalkar, A. & Smith, V. Federated learning: Challenges, methods, and future directions. *IEEE Signal Process. Mag.* **37**, 50–60 (2020).
68. Ali, H., Alam, T., Househ, M. & Shah Z. Federated learning and internet of medical things–opportunities and challenges. in *Advances in Informatics, Management and Technology in Healthcare*. 201–204. https://doi.org/10.3233/SHTI220697 (2022).

Acknowledgements

Open Access funding provided by Qatar National Library.

Author contributions

F.M., H.A., Z.S. contributed to conceptualization. F.M. and H.A. administered the project. F.M. curated the data, performed data synthesis, and contributed to writing-original draft. H.A and N.E performed writing-review and editing. Z.S. and H.A. supervised the study. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-22514-4.

Correspondence and requests for materials should be addressed to Z.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.