A Systematic Collection of Medical Image Datasets for Deep Learning

JOHANN LI, GUANGMING ZHU*, CONG HUA, and MINGTAO FENG, Xidian University, China
BASHEER BENNAMOUN, The University of Notre Dame, Australia
PING LI and XIAOYUAN LU, Shanghai Broadband Network Center, China
JUAN SONG and PEIYI SHEN, Xidian University, China
XU XU and LIN MEI, The Third Research Institute of The Ministry of Public Security, China
LIANG ZHANG*, Xidian University, China
SYED AFAQ ALI SHAH*, Edith Cowan University, Australia
MOHAMMED BENNAMOUN, The University of Western Australia, Australia

The astounding success made by artificial intelligence in healthcare and other fields proves that AI can achieve human-like performance. However, success always comes with challenges. Deep learning algorithms are data-dependent and require large datasets for training. Many junior researchers faced a lack of data, because of a variety of reasons. Medical image acquisition, annotation, and analysis are costly, and their usage is constrained by ethical restrictions. They also require several other resources, such as professional equipment and expertise. That makes it difficult for novice and non-medical researchers to have access to medical data. Thus, as comprehensive as possible, this paper provides a collection of medical image datasets with their associated challenges for deep learning research. We have collected the information of around three hundred datasets and challenges mainly reported between 2007 and 2020 and categorized them into four categories: head & neck, chest & abdomen, pathology & blood, and “others”. The purpose of our paper is to provide a list, as up-to-date and complete as possible, that can be used as a reference to easily find the datasets for medical image analysis and the information related to these datasets.

CCS Concepts: • Social and professional topics → Computing / technology policy; Medical information policy.

Additional Key Words and Phrases: Medical image analysis, Deep learning, Datasets, Challenges, Computer-aided diagnosis

1 Introduction

Medical imaging technology marked a new era in medicine since its inception [30]. The development of medical imaging began with the introduction of X-rays and has since evolved to include various other imaging techniques [72, 119], such as 3D computed tomography (CT) [1], magnetic resonance imaging (MRI) [3, 6], nuclear medicine...
Medical images can provide critical insight into the diagnosis and treatment of many diseases. The human body’s different reactions to imaging modalities are used to produce scans of the body. Reflection and transmission are commonly used in medical imaging because the reflection or transmission ratio of different body tissues and substances are different. Some other methods acquire images by changing the energy transferred to the body, e.g., magnetic field changes or the rays radiated from a chemical agent.

Before modern AI was applied in medical image analysis, radiologists and pathologists needed to manually look for the critical “biomarkers” in the patient’s scans. These “biomarkers”, such as tumors and nodules, are the basis for the medics to diagnose and devise treatment plans. Such a diagnostic process needs to be performed by medics with extensive medical knowledge and clinical experience. However, problems such as diagnostic bias and the lack of medical resources are prevalent and cannot be avoided. With the recent advancements in AI, which have achieved human-like performance in image recognition [33, 85, 164–166, 191, 198] and even won games like Go [190] and real-time strategy games [215], there has been a growing interest in developing AI-based automatic medical image analysis algorithms. As a result, AI has become a major research focus in medical image analysis and has achieved significant success in recent years [65, 98, 149, 163, 168, 172, 176].

Many researchers brought their focus to AI-based medical image analysis methods thinking that it might be one of the solutions to the challenges (e.g., medical resource scarcity) and taking advantage of the technological progress [106, 137, 143, 189, 216]. Traditional medical image analysis focuses on detecting and classifying biomarkers for diagnosis and treatment. AI imitates the medic’s diagnosis through classification, segmentation, detection, regression, and other AI tasks in an automated or semi-automated way.

AI has achieved a significant performance for many computer vision tasks. This success is yet to be translated to the medical image analysis domain. Deep learning (DL), a branch of AI, is a data-dependent method as it needs massive training data. However, when DL is applied to medical image analysis, the paucity of labeled data becomes a major challenge and a bottleneck.

Data scarcity is a common problem when applying DL methods to a specific domain, and this problem becomes more severe in the case of medical image analysis. Researchers, who apply DL methods to medical image analysis research, do not usually have a medical background, commonly computer scientists. They cannot collect data independently because of the lack of access to medical equipment and patients. They cannot also annotate the acquired data either because they lack the relevant medical knowledge. Furthermore, medical data is owned by institutions, which cannot easily make it public due to privacy and ethics restrictions.

To address some of these problems, MICCAI, ISBI, AAPM, and other conferences and institutions have launched many DL-related medical image analysis challenges. These aim to design and develop automatic or semi-automatic algorithms and promote medical image analysis research with computer-aided methods. Concurrently, some researchers and institutions also organize projects to collect medical datasets and publish them for research purposes.

Despite all these works, it is still challenging for novices of medical image analysis to find medical data. Therefore, we present this comprehensive survey of medical datasets and relevant challenges with the aim to help researchers easily find the required datasets for their research.

This paper refers to other research papers with a number between square brackets and refers to the datasets listed in the tables with numbers between parentheses.

The paper is organized as follows. Section 2 summarizes the datasets and challenges at a macro-level and focuses on information relevant to medical data such as the years, body parts, and tasks. Section 3, Section 4, and Section 5 discuss the datasets and challenges of the head and neck, the chest and abdomen organs, and the pathology and...
Fig. 1. An overall taxonomy to outline the organization of the paper.

Fig. 2. Summary of medical image datasets and challenges from 2013 to 2020. Figure 2A shows the number of datasets and challenges published in each year. Figures 2B, C, and D show the year-by-year trends along with the trends in relative numbers for each of the different categories by year. The numbers listed right are the summary count of each category, and the summary counts are not the same with the total numbers, because of (1) some of the categories are not shown, and (2) a dataset counts two times if it includes two of the categories. Figures 2E, F, and G show the most predominant body parts, data modalities, and main tasks with the percentage of their respective dataset.

For other datasets, such as bone, skin, phantom, and animals, respectively, we place them into supplementary materials, because of the space limitation. We have also created a website with a git repo¹, which shows the list of these datasets and their respective challenges.

¹The website with the git repo will be public after the paper is accepted.
2 Medical image datasets

In this section, we provide an overview of the datasets and challenges. Our collection contains over three hundred medical image datasets and challenges organized between 2004 and 2020. This paper focuses mainly on the ones between 2013 and 2020. Subsections 2.1, 2.2, 2.3, and 2.4 provide information about the year, body parts, modalities, and tasks, respectively. In Subsection 2.5, we introduce the sources from where we have collected these datasets and the challenges. Details about the categorization of these datasets and challenges into four groups are provided in the subsequent sections. We provide a taxonomy of our paper in Figure 1 to help the reader navigate through the different sections.

2.1 Dataset Timeline

The timeline of these medical image datasets can be split into two, starting from 2013 as the watershed, since the excellent success of AlexNet[116] in the ILSVRC competition 2012. The success of AlexNet motivated more researchers to solve a variety of problems[60, 66, 115, 132, 139] with different DL-based methods[84, 85, 91, 93, 174, 191]. At the same time, it also motivated many researchers to create and publish medical datasets. Consequently, many researchers focused on analyzing medical images with deep learning, which is one of the driving factors and reasons for the release of medical datasets. Another reason for us to mainly focus on datasets released after 2012 is that some of the datasets developed before 2012 are not aimed at computer-aided diagnosis with DL-based methods, for example, ADNI-1 (55), although those data could be used for DL.

Figure 2A shows the statistics of the datasets and challenges per year between 2013 and 2020. As shown in Figure 2A, the number of related datasets and challenges increased year by year because of the progress and success of DL in computer vision and medical image analysis. That led more and more researchers to focus on medical image analysis with DL-based methods and more and more datasets and challenges with different body parts and tasks started to appear.

Figure 2B shows the number of datasets related to the brain, heart, lung, liver, and eye. As can be noted, the proportion of the brain-related datasets is initially large. Then, the number of datasets related to these body parts (i.e., heart, lung, liver, and eye) is increased, but the proportion of the brain-related datasets starts to decrease.

Figure 2C shows that the number of datasets related to segmentation, classification, detection and other tasks is increased. The research mainly focused on the segmentation(28, 29, 54) task in the early stages. It then diversified to a variety of tasks or combinations of tasks that are also essential for clinical needs, including classification(154, 197, 259), segmentation(1, 2), detection(199, 232), regression(213, 214), generation(14, 19), tracking(220, 245), and registration(24, 25, 166), as time progresses(62).

Figure 2D shows no visible changes in the proportion of different modalities over years, but the number of datasets related to different modalities increased.

2.2 Body parts

As Figure 2E shows, the top-5 focused organs, in these datasets and challenges, include the brain, lung, heart, eye, and liver.

One of reason why datasets and challenges focus on these organs is that the tasks of datasets, such as a classification and segmentation are helpful in the diagnosis and treatment of cancer (a major threat to human life) or other diseases. For example, eye related diseases, which cause blindness, incited the collection of eye related datasets and the release of challenges. Some other datasets and challenges focus on the small organs, such as the prostate, which are challenging to analyze due to the low resolution of images.

As AI has been shown to be more accurate than humans to tackle some complex tasks [79, 86, 188], researchers have attempted to solve the imaging problem of an organ with the AI methods successfully tested on other other organs. That has lead the researchers to focus on more and more organs.
2.3 Modalities
There are several types of medical image modalities. As shown in Figure 2F, the frequently used modalities to acquire medical datasets include Magnetic Resonance Images (MRI), Computed Tomography (CT), Ultrasound (US), Endoscopy, Positron Emission Tomography (PET), Computed Radiography (CR), Electrocardiography, and Optical Coherence Tomography (OCT). Limited by space, we introduce the main modalities in the supplementary materials.

MR, CT, and other modalities are the most commonly used imaging modalities. MR can provide sharp images without harmful radiations of soft tissues. It is therefore widely used in the imaging of brain, heart and many other small organs. CT is an economical and simple imaging approach, and it is widely used for the diagnosis of cancer, e.g., the neck, chest, and abdomen. A pathology image is different from MR and CT, because it is a cell-level imaging method. Pathology is widely used in cancer-related diagnosis.

2.4 Tasks
As Figure 2G shows, we categorize these tasks into seven categories: classification, segmentation, detection, regression, generation, registration, and tracking. Limited by space, we introduce these tasks in the supplementary materials.

2.5 Source and Destination
To conduct a systematic search for relevant datasets and challenges, we first reviewed papers that addressed medical image segmentation and classification. This enabled us to identify several sources for datasets, including Grand Challenge, a website that hosts medical imaging-related competitions, and The Cancer Imaging Archive (TCIA) [56], a website for cancer-related competitions. Additionally, we utilized Google to search for other relevant websites and collections that featured datasets and challenges related to medical image analysis with deep learning. Through this process, we discovered Kaggle, Codalab, OpenNeuro, PhysioNet [77], and scientific journals such as Scientific Data, which were inspired by Grand Challenge and TCIA.

Next, we crawled through these websites to gather information about the datasets/competitions featured on these sites, carefully selecting datasets and challenges based on the available information.

The paper lists datasets and challenges primarily sourced from Grand Challenge and Codalab, which are predominantly related to medical imaging. These were directly included in our work. However, the datasets and challenges from TCIA, Kaggle, PhysioNet, OpenNeuro, and journals were not all related to medical imaging or deep learning. Therefore, we filtered these datasets and challenges using specific keywords such as “machine learning”, “deep learning”, “segmentation”, and “classification”. Some datasets were removed as they were not suitable for DL and AI methods, due to the lack of annotations or other factors. It is worth noting that not all medical image analysis tasks require annotations. For example, the generation task, such as generating synthetic CT images from MR ones [44], does not require annotations. Furthermore, for senior researchers, data in The Cancer Imaging Archive, OpenNeuro, UK Biobank, and other databases can be useful after manual annotations.

We then categorized the remaining datasets and challenges into different groups. Systematically categorizing the datasets and challenges is challenging. Thus we used an asymmetric categorization to group these datasets and challenges into four groups, as shown in Figure 1. This means that we did not use the same sub-taxonomy in each category or subcategory.

First, we simply split the medical datasets and challenges into two groups: body-level (Section 3 & 4) and cell-level (Section 5), according to the imaged body part. Second, we grouped the datasets and challenges of the brain, eye, and neck into one group (Section 3), because these are parts of the head. Third, we organized the datasets and challenges related to the chest and abdomen into the same group (Section 4). Finally, for the datasets and challenges that cannot be categorized into the above groups, we grouped them under “other”, and place them in

ACM Comput. Surv.
the supplementary materials, due to the limitation of space. These datasets and challenges are related to the skin, bone, phantom, and animals.

The introduction of each group and sub-group includes mainly the type of modality, the task, the disease, and the body part. However, not all the groups of datasets can be introduced in that way. For some groups, we introduce the datasets and challenges according to the domain-specific problems. For example, we categorize the pathology datasets into microscopic and macroscopic tasks.

2.6 Notes to the tables

We have compiled tables in this paper to present the datasets and challenges. These tables contain basic information such as the index, dataset/challenge name, release year, and other relevant details. Some information is not available in the table due to either non-provision or unavailability during our data collection process.

For all the tables, the information includes:

| Reference index | Dataset/challenge | Year | Modalities | Focus | Lesion/Type | DISEASES | ORGAN | CATEGORY | Stain | Resolution | Multi-center | Rank | CONF | LICENSE |
|------------------|-------------------|------|------------|-------|-------------|----------|-------|----------|-------|-------------|-------------|------|------|---------|
| The index to refer the dataset or challenge in our paper, and it's numbered in the order of appearing in the table. | The dataset or challenge’s name or abbreviation of the name. The name in the table is the hyperlink to the website or home page of the dataset or challenge. | The year of dataset or challenge was released. For datasets, it is the year of its paper public or they are noted. For challenges, it is the year that their ‘name’ is declared, but for some challenges, of which the conferences will be hosted after 2020, we note 2020. | The modalities of data. For abbreviations, we note the terms in the footnote of the table. | Some words or a sentence to introduce the focus of datasets and challenges. | The task of dataset or challenge designed or focused on. For abbreviations, we note the terms in the footnote of the table. | The diseases that the dataset or challenge focused on. For abbreviations, we note the terms in the footnote of the table. | The organs that the dataset or challenge focused on. For abbreviations, we note the terms in the footnote of the table. | The category of classification or segmentation. | The stain that is used in pathological images. | The resolution of data in the pixel or the voxel. For some datasets and challenges, we note the size of it in mm. | Whether the dataset or challenge includes multi-center data. We will not mark the checkmark(✓) and crossmark(×), if any related information is left in the description of the dataset or challenge. | Whether the challenge includes a rank to compare the participants’ algorithms. | The name of the conference that related to the challenge. | The license of the dataset. We note the abbreviation in the footnote of the tables. |

3 Head and neck related datasets and challenges

The head and neck are ones of significant parts of the human body because many essential organs, glands, and tissues are located there. Several researchers’ image analysis work relate to the head and neck. To make an effective use of computers for research, diagnosis, and treatment, many researchers released datasets and challenges, for examples: 1) the analysis of tissue structure and functions (1, 2, 3, 5) and 2) diseases diagnosis (30, 40, 44).

Because the brain controls human emotions and actions/functions of other organs, the datasets and challenges significantly focus on brain. First, we introduce the datasets and challenges related to the analysis of the brain structure, function, imaging, and other basic tasks in Subsection 3.1. Second, we introduce the datasets and challenges related to brain disease diagnosis in Subsection 3.2.

Moreover, since the eyes are crucial to our vision, the computer-aided diagnosis of eye-related diseases is also an important research focus. The eye-related datasets and challenges are covered in Subsection 3.3. Finally, we introduce the remaining datasets and challenges, related to neck and the research of brain behavior and cognition in Subsection 3.4.

3.1 Structural analysis tasks of the brain

The basic analysis and processing of the brain medical images are clinically critical for the diagnosis, treatment, and other brain-related analysis tasks. The datasets and challenges we discuss are mainly for the segmentation tasks and center around the brain structure. In contrast, some datasets focus on imaging, including MR imaging acceleration (e.g. (14, 16)), the non-linear registration of different modalities (e.g. (24, 25)), and tractography (e.g. (28, 29)).
### Table 1. Summary of datasets and challenges for the basic brain image analysis.

| Reference Index | Dataset/Challenge | Year | Modalities 1 | Focus | Tasks 2 | Resolution | Scale 3 | Multi- | Rank | Core 4 | License 5 |
|-----------------|-------------------|------|--------------|-------|---------|------------|---------|--------|-----|--------|----------|
| 1               | iSeg 2019 [196]   | 2019 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | T1: 1 × 1 × 1 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 2               | iSeg 2017 [119]   | 2017 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | T2: 1.5 × 1.5 × 2.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 3               | MIR2013 [13]      | 2013 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 4               | NEALBrain15       | 2015 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 5               | MIR2017 [13]      | 2017 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 6               | Normal-12 MR BrainS 12 [96] | 2012 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 7               | CIRRohoM [15]     | 2019 | ✓ ✓          | WM, GM, CSE other subcortical structures | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 8               | WMH-Sci. Obliq. [117] | 2017 | ✓ ✓          | White Matter Hypointensities | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 9               | NIVDMS Correlation | 2017 | ✓ ✓          | Correlation. | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 10              | CALMR-07 [210]    | 2017 | ✓ ✓          | Cal-main. | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 11              | MindBlind [111]   | 2012 | ✓ ✓          | Brain atlas 11 | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 12              | Labeled Brain Scan | 2012 | ✓ ✓          | Brain atlas 11 | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 13              | 3DREEL [102]      | 2015 | ✓ ✓          | Hippocampus | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 14              | Acceler2020       | 2020 | ✓ ✓          | Non-linear mapping of different resolutions | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 15              | Calgary Comp. Brain Dataset [110] | 2020 | ✓ ✓          | Brain image registration | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 16              | IntraHE [12,20]   | 2015 | ✓ ✓          | Accelarating magnetic resonance imaging | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 17              | Synthetic skull boundary [12,21] | 2020 | ✓ ✓          | Cranioplasty | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 18              | SkullBridge [12,12] | 2020 | ✓ ✓          | Cranioplasty | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 19              | Anatoimplant [131] | 2020 | ✓ ✓          | Cranioplasty | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 20              | HARDI 2013        | 2013 | ✓ ✓          | Diffusion MRI reconstruction | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 21              | HARDI 2012        | 2012 | ✓ ✓          | Diffusion MRI reconstruction | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 22              | BRAINWM Reconstruct. | 2012 | ✓ ✓          | White matter reconstruction | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 23              | Reconstructed col-lection [29] | 2019 | ✓ ✓          | Fiber reconstruction and medical imaging modality in surgery | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 24              | GADROCUS 2019     | 2019 | ✓ ✓          | Image registration | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 25              | OF-REAS 2019      | 2019 | ✓ ✓          | Image registration | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 26              | MUSIBAC 2018      | 2018 | ✓ ✓          | DWI MR registration and enhancement | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 27              | SD-VoTEM          | 2018 | ✓ ✓          | Fiber tractography | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 28              | DTI Tracking Challenge 2012 | 2012 | ✓ ✓          | Fiber tractography | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |
| 29              | DTI Tracking Chal- lenge 2011 | 2011 | ✓ ✓          | Fiber tractography | ✓     | 0.958 × 0.958 × 3.5 mm³ | ✓ ✓     | ✓ ✓    | ✓ ✓ | MICCAI | C/R     |

1. T1w and T1w-IR: T1-weighted images with and without inversion recovery, respectively.
2. T2w and T2w-FLAIR: T2-weighted images with and without fluid-attenuated inversion recovery, respectively.
3. DWI and DTI: Diffusion-weighted images and diffusion tensor imaging, respectively.
4. MT and PD: Magnetic resonance spectroscopy and proton density imaging, respectively.
5. MRS: Magnetic resonance spectroscopy.
6. EPI: Echo planar imaging.
7. SWI: Susceptibility weighted imaging.
8. CTE: Cortical thickness estimation.
9. Other: Additional datasets not categorized under the above categories.
10. For details of labels see: [http://www.neuromorphometrics.com/?page_id=23](http://www.neuromorphometrics.com/?page_id=23)
Table 1 listed the datasets and challenges related to the basic brain image analysis, such as the segmentation of white/gray matter and registration of images. One of the most popular tasks in Table 1 is the segmentation of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF), and their respective datasets and challenges are introduced in Subsection 3.1.1. Meanwhile, other tissues and functional areas’ segmentation are also the focus of research, and their related datasets and challenges are discussed in Subsection 3.1.2. Subsection 3.1.3 describes the other basic tasks. Table 1 presents an overview of the datasets and challenges related to medical image analysis of the head and neck. The majority of these datasets provide MRI images, while some offer CT, PET, or Ultrasound data. Approximately half of the datasets and challenges in Table 1 include multi-center data, and some do not specify whether they contain multi-center data. Most of the challenges listed in Table 1 are affiliated with conferences such as MICCAI, ISBI, or MIDL, while others do not have an affiliation. All of the challenges shown in Table 1 provide a ranking of participants’ methods. Some of the challenges require registration to access the data and submit methods, while others are public under license.

3.1.1 Segmentation of white and gray matter

The segmentation of WM, GM, CSF, and other parts of brain tissues is always a focus for researchers working on brain-related analysis and diagnosis[96, 150]. Similarly, for AI algorithms, it is also of significance to understand the human brain’s structure. Therefore, MICCAI and others have held many challenges(1, 2, 3, 4, 5, 6, 7, 8) with this research focus, and researchers could design automatic algorithms to segment magnetic resonance images into different parts. We introduce these datasets and challenges related to their modalities and tasks in the following.

**Modality:** Since these datasets and challenges focus on the segmentation of brain tissues, the data usually provided is MR image. Challenges (1, 2, 3, 4, 5, 6) provide mainly two modalities: T1(T1-weighted MRI, based on longitudinal relaxation time[8], providing structural information about tissues and is commonly used for anatomical imaging), T2(T2-weighted MRI, based on transversal relaxation time[8], highlighting fluid and helps identify edema or swelling), while datasets (7, 8) only provide T1 for the white matter hyperintensities segmentation task. Note that, MR scans are sensitive to the hydrogen atom, and such a feature can effectively help image analysts to distinguish between different tissues and parts of the image. Moreover, due to the color of the tissue imaged by MR, these scans are named as “white matter” and “gray matter”.

**Task:** The main focus of these datasets and challenges is the segmentation of WM, GM, and CSF. However, they do not only focus on that. Challenges (3, 4, 5, 6, 7) also provide the annotation of other parts of the brain, including basal ganglia, white matter lesions, cerebellum, and infarction. Challenges (1, 2, 4, 6) use MR images of the neonatal brain, and consider tissue volumes as an indicator of long-term neurodevelopmental performance[96]. Challenges (1, 3) increased regions to segment and provided different data compared with the previous challenges (2, 5). Challenge (8) focuses on the segmentation of white matter hyperintensities in the brain.

**Performance metric:** For the segmentation task, the Dice score is one of the most commonly used metrics, and all these datasets and challenges adopt it as a performance measure. Besides the Dice score, datasets (3, 5, 8) also use Hausdorff distance and volumetric similarity as metrics; datasets (1, 2) use the average the Hausdorff distance and the average surface distance as one of their metrics; moreover, dataset (8) also uses sensitivity and F1-score as metrics for performance evaluation.

3.1.2 Segmentation of functional areas & other tissues

The segmentation of functional areas and tissues has also an essential meaning for brain-related research and computer-aided diagnosis. In this subsection, we introduce the datasets and challenges that are related to the segmentation of functional areas and tissues.

**Tissues segmentation:** While, WM, GM, and CSF were introduced in Subsection 3.1.1, the segmentation of other brain tissues is also an active research area. Challenges (3, 5, 6, 7) aim to segment brain images into different tissues, including ventricles, cerebellum, brainstem, and basal ganglia. These challenges provide MR images and the voxel-level annotations of the regions of interest with thirty or forty scans. Because these regions are essential for brain health, researchers need to overcome the challenges related to their size and shape in order to segment...
them. Dataset (9) is designed for the segmentation of the cerebellum from Diffusion-Weighted Images (DWI) a.k.a. dMRI, which are a type of MRI that utilizes the movement of water molecules to generate images [7] and measures the random movement of water molecules within tissues and is particularly sensitive to acute strokes. On the other hand, dataset (10) focuses on the segmentation of the caudate from brain MRI images.

**Functional areas:** The segmentation of the human brain cortex into different functional areas is of great significance in education, clinical research, treatment, and other applications. Datasets (11, 12) provide images and annotations for the design of automatic algorithms to segment the brain cortex into different functional areas, in other words, brain atlas. Brain atlas is a map to annotate different functional areas or tissues. Dataset (11) uses DTK protocol [111], which is modified from DK protocol [62], and the DTK protocol includes 31 labels, details of which are listed in https://mindboggle.readthedocs.io/en/latest/labels.html. Dataset (12) is a commercial dataset for research in the segmentation of functional areas of the brain cortex.

The dataset (13) focuses on the segmentation of the hippocampus, which can be used to detect the progress of Alzheimer’s disease, according to the challenge (13). It provides MRI images from AD patients as the training set and the test set.

### 3.1.3 Imaging-related tasks

In addition to the segmentation tasks of the brain tissues and the functional areas, some of the datasets and challenges also focus on the generation, registration, and tractography.

**Generation:** Datasets and challenges (14, 15, 16) aim to accelerate MR imaging or generate high-resolution MR images from low-resolution ones. Usually, high-resolution imaging requires higher cost, while low-resolution imaging is cheaper but affects the analytical judgment and may lead to an incorrect diagnosis. These challenges provide many scans at low-resolution to allow researchers to design algorithms to convert or map low-resolution images onto higher-resolution ones. The datasets and challenges mainly focus on the generation tasks. Another focus is the cranioplasty (17, 18, 19) to generate a part of broken skull from CT images of the models of the broken skull. At the same time, challenges (20, 21) focus on another type of generation task, the reconstruction of diffusion MR images. Compared with challenge (21), challenge (20) focused on the effect of the local reconstruction accuracy on the quality of connectivity reconstruction. Challenge (22) focuses on the reconstruction of white matter to answer the question such as “what model (either biophysical or signal model) better explains the underlying tissue environment?” and “what dMRI sequence enables the most accurate estimation of these parameters?”. Dataset (23) focuses on face reconstruction from PET and CT images, which aims to the augmented reality in surgery or other tasks.

**Registration:** Registration is a research area that focuses on aligning images from different modalities. Challenges (24, 25) are centered on the registration of ultrasound data and MR images of the brain. Cross-modality registration can be challenging because the subject is not completely stationary, and ultrasound is a 2D imaging modality while MR is a 3D volume imaging modality. Therefore, these challenges aim to establish the topological relationship between the preoperative MR image and intraoperative ultrasound. Challenge (24) provides registration tasks for pre- and post-resection, while challenge (25) does not. Challenge (26) focuses on the registration of diffusion MR images to eliminate variations caused by different vendors’ hardware devices and protocols.

**Tractography:** Tractography is another segmentation task and focuses on the segmentation and imaging of the fiber in the WM. Dataset (27) aims to segment the fiber bundles from brain images, including phantom, squirrel monkey, and macaque, while challenges (28, 29) focus on the tractography with diffusion tensor imaging (DTI, a kind of DWI), which is provides additional information about the direction of water movement within tissues, allowing for the assessment of white matter tracts in the brain. [7]). The details of the challenges’ tasks are different. Challenge (28) focused on the segmentation of the corticospinal tract from a set of neurosurgical cases, and challenge (29) focused on the segmentation and the analysis based on the segmentation.
3.2 Brain diseases related datasets and challenges

Besides the structural analysis and image processing tasks, computer-aided diagnosis is also a research focus in healthcare. Medical image analysis plays a critical role in clinical research, diagnosis, and treatment. The datasets and challenges we have included are for two tasks: 1) the segmentation of lesions and 2) the classification of diseases. For the segmentation task, the respective datasets and challenges focus on the lesion segmentation of the human brain, mark the lesion’s contour for diagnosis and treatment, and the relevant details are shown in Subsection 3.2.1. For classification tasks, the datasets and challenges have been used for the development of automatic algorithms to classify or predict diseases from medical images, and these datasets and challenges are presented in Subsection 3.2.2.

3.2.1 Datasets for segmentation of lesions

Lesions in the brain affect human’s healthy life and safety, and image analysis is an effective way to diagnose the relevant diseases. In this subsection, related datasets and challenges are introduced, and they are reported in Table 2.

| Reference | Dataset/ Challenge | Year | Modalities | Focus | Tumor/Lesion | Resolution | Scale | Multi-class | Rank | Con1 | License |
|-----------|---------------------|------|------------|-------|--------------|------------|-------|-------------|------|------|---------|
| BraTS 2020 [36], [37] | 2020 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR/T |
| BraTS 2019 [36], [37] | 2019 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR/T |
| BraTS 2018 [36] | 2018 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR/T |
| BraTS 2017 | 2017 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR/T |
| BraTS 2016 | 2016 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR |
| BraTS 2015 | 2015 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR |
| BraTS 2014 | 2014 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR |
| BraTS 2013 | 2013 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR |
| BraTS 2012 | 2012 | ✓ | ✓ | Multi-modalities | ✓ | 240 × 240, 1mm³ | ✓ | ✓ | MICCAI | CR |
| Brain Tumor Pro- | 2018 | ✓ | ✓ | Brain tumor | ✓ | 20 | ✓ | ✓ | MICCAI | CR |
| ICAD 2020 [19] | 2020 | ✓ | ✓ | Cerebral aneurysm | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| CADA-AS | 2020 | ✓ | ✓ | Cerebral aneurysm | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| CADA-RE | 2020 | ✓ | ✓ | Cerebral aneurysm | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| ISLES 2018 | 2018 | ✓ | ✓ | Ischemic stroke | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| ISLES 2017 | 2017 | ✓ | ✓ | Ischemic stroke | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| ISLES 2016 | 2016 | ✓ | ✓ | Ischemic stroke | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| ISLES 2015 | 2015 | ✓ | ✓ | Ischemic stroke | ✓ | 256 × 256 × 220 | ✓ | ✓ | MICCAI | CR |
| MFLAS (131) | 2017 | ✓ | ✓ | Low-grade glioma | ✓ | 159 | ✓ | ✓ | MICCAI | CR |
| CT-ICH (92) | 2020 | ✓ | ✓ | Intracranial hemorrhage | ✓ | 640 × 480 | 82 | ✓ | MICCAI | CR |
| ISBI Challenge (17) | 2019 | ✓ | ✓ | Intracranial hemorrhage | ✓ | 240 × 240 | 220 | ✓ | MICCAI | CR |
| MISeeg Chal- | 2019 | ✓ | ✓ | Multiple sclerosis | ✓ | 240 × 240 | 220 | ✓ | MICCAI | CR |
| 2015 | 2015 | ✓ | ✓ | Multiple sclerosis | ✓ | 240 × 240 | 220 | ✓ | MICCAI | CR |
| MISeg 2018 | 2018 | ✓ | ✓ | Multiple sclerosis | ✓ | 240 × 240 | 220 | ✓ | MICCAI | CR |

1 OF - other modalities, for details, see footnotes.
2 BT for brain tumor or aneurysm, SL for stroke lesion, IL for intracranial hemorrhage, CL for sclerosis lesion.
3 Conference.
4 For BraTS 12 to 16, T1 modality includes T1 image and T1w FLAIR image.
5 For BraTS 17 to 20, T1 modality includes T1 image and T1w FLAIR image.
6 For BraTS 17 to 20, T2 modality includes T2 image and T2w FLAIR image.
The tasks of these datasets mainly include the segmentation of brain tumors, e.g., glioma, the segmentation of cerebral aneurysm, the segmentation of stroke lesion, the segmentation of intracranial hemorrhage, and sclerosis lesion. Most of the datasets and challenges use MR images that include different sub-modalities, while some are using CT. Nearly half of these datasets and challenges listed in Table 2 are reported including multi-center data, while few of them are reported as not included. The rest of them are not reported whether use multi-center data. The challenges in Table 2 provide rankings of the participants’ methods, and most of them are associated with MICCAI. Most of the challenges in Table 2 are required to register when accessing data, while other datasets and challenges are opened under specific licenses.

**Glioma datasets and challenges:** Gliomas are one of the most common and deadly brain malignancies [12]. Therefore, many challenges and datasets focus on the segmentation of glioma for its diagnosis and treatment. BraTS challenge series (30, 31, 32, 33, 34, 35, 36, 37, 38) have been occurring since 2012 to segment the glioma. The challenges of such a segmentation task are caused by the heterogeneous appearance and shape of gliomas. The heterogeneity of glioma reflects its shape, modalities, and many different histological sub-regions, such as the peritumoral edema, the necrotic core, enhancing, and the non-enhancing tumor core. Therefore, these series of challenges provide multi-modal MR scans to help researchers design and train algorithms to segment tumors and their sub-regions. The tasks of this challenge series include 1) low- and high-grade glioma segmentation (37, 38), 2) survival prediction from pre-operative images (32, 33), and 3) the quantification of segmentation uncertainty (30, 31). Besides the BraTS challenge series, dataset (40) is another one for the segmentation of low-grade glioma and provides T1-weight and T2-weight MR images with biopsy-proven gene status of each subject by fluorescence in-situ hybridization, a.k.a. FISH [181]. Dataset (39) focuses on the processing of brain tumor and aims to design and evaluate DL-based automatic algorithms for glioblastoma segmentation and further research.

**Cerebral aneurysm challenges:** Cerebral aneurysm is life-threatening, because the rupture of the cerebral aneurysm is associated with high mortality, and it is cognitive impairment in case of survival. Therefore, challenges (41, 42, 43) focus on the detection of the cerebral aneurysm for the strategy to prevent rupture. These three challenges, respectively, focus on the detection, segmentation, and rupture risk assessment of the cerebral aneurysm.

**Ischemic stroke lesion datasets and challenges:** Similar to tumor segmentation, brain lesion segmentation also focuses on detecting brain abnormalities. However, the difference is that lesion segmentation deals with damaged tissues. Challenges (44, 45, 46, 47, 48) focus on stroke lesion segmentation because stroke is also life-threatening and can disable the surviving patients. Stroke is often associated with high socioeconomic costs and disabilities. Automatic analysis algorithms help to diagnose and treat stroke, since its manifestation is triggered by local thrombosis, hemodynamic factors, or embolic causes. In MR images, the infarct core can be identified with diffusion MR images, while the penumbra (which can be treated) can be characterized by perfusion MR images. The challenge ISLES 2015 (47) focuses on sub-acute ischemic stroke lesion segmentation and acute stroke outcome/penumbra estimation and provides 50 and 60 multi-modalities MR scans of data for training and validation, respectively, for two subtasks, i.e., sub-acute ischemic stroke lesion segmentation and acute stroke outcome/penumbra estimation. The subsequent year’s challenge, ISLES 2016 (46), focuses on the segmentation of lesions and the prediction of the degree of disability. This challenge provides about 70 scans, including clinical parameters and MR modalities, such as DWI, ADC, and perfusion maps. The challenge ISLES 2017 (45) focuses on the segmentation with acute MR images, and ISLES 2018 (44), focuses on the segmentation task based on acute CT perfusion data. Moreover, dataset (48), i.e., ATLAS, focuses on the segmentation of the stroke lesion and the brain itself after stroke for further treatments.

**Intracranial hemorrhage related datasets:** Intracranial hemorrhage (ICH) is a sudden bleeding into the brain tissues or ventricles or both of them, which can be caused by the traumatic brain injuries or other reasons [92]. ICH could lead to disability or even deaths, if it is not treated on time. Normal diagnosis and treatment is based on the analysis of brain CT to localize the regions of ICH. Thus, dataset and challenge (49, 50) focus on the detection and
Alzheimer’s disease (AD) [25, 67, 105, 199] and Parkinson’s disease (PD). Therefore, effective diagnosis and early costs more than two to three hundred billion U.S. dollars.

The disease takes an unbearable toll on the patient and leads to a high cost to their loved everything they know. For example, according to [9], AD became the sixth deadly cause in the U.S. in 2018 and patients not know who they are and what they should do which then progresses to the point until they forget.

AD affects a person’s behavior, cognition, memory, and daily life activities. Such a progressive neurodegenerative disorder affects the normal daily life of patients because suffering from such a disease makes agents to find landmarks associated with AD. For other diseases, MRI or specific modalities are used to better detect diseases. AD has a more detailed division of cognitive impairment, while datasets of other diseases are nearly 50-50, while there is a dataset not report whether used or not. The challenges listed in Table 3, are required to register when accessing, while other datasets are opened or public under specific licenses.

Alzheimer’s disease: AD affects a person’s behavior, cognition, memory, and daily life activities. Such a progressive neurodegenerative disorder affects the normal daily life of patients because suffering from such a disease makes patients not know who they are and what they should do which then progresses to the point until they forget everything they know. The disease takes an unbearable toll on the patient and leads to a high cost to their loved ones and to the society. For example, according to [9], AD became the sixth deadly cause in the U.S. in 2018 and costs more than two to three hundred billion U.S. dollars.

Therefore, researchers are doing everything they could to explore the causes of AD and its treatments. Diagnosis based on medical images has become a research focus because early diagnosis and intervention have significance on the progress of this disease. Hence many researchers work on the classification, i.e., prediction of AD using segmentation of intracranial hemorrhage to help medics locate the hemorrhage regions and decide on a treatment plan. Dataset (51) also provides data for the classification of normal or hemorrhage CT images.

Multiple sclerosis lesion related datasets: Multiple sclerosis lesion is another kind of lesion in the brain which is not life-threatening and deadly but can cause disabilities.Datasets and challenges (52, 53, 54) are related to the multiple sclerosis lesion segmentation with multi-modalities MR data, such as T1w, T2w, and FLAIR (Fluid-attenuated inversion recovery, a kind of MR image, “removing” signal from the cerebrospinal fluid, [5]).

3.2.2 Classification of brain disease Except for the lesion segmentation, brain disease classification also plays an essential role in healthcare. Brain related diseases have a severe effect on patients’ health and their lives, e.g., Alzheimer’s disease (AD) [25, 67, 105, 199] and Parkinson’s disease (PD). Therefore, effective diagnosis and early intervention can effectively reduce the health damage to patients, the effect on the social times of families, and the economical impact on society. In this section, we first introduce the datasets and challenges of AD (55, 56, 57, 58), and then we introduce other diseases (65, 66, 67, 68). Table 3 shows the relevant challenges and datasets.

The datasets listed in Table 3 mainly use MRI and PT. For AD, PET is commonly used with specific contrast agents to find landmarks associated with AD. For other diseases, MRI or specific modalities are used to better detect diseases. AD has a more detailed division of cognitive impairment, while datasets of other diseases are binary classifications. The numbers of datasets and challenges which included or not included multi-center data are 59, 64), and then we introduce other diseases (65, 66, 67, 68). Table 3 shows the relevant challenges and datasets.

Table 3. Summary of datasets and challenges for brain disease classification tasks.

| Reference | Dataset/Challenge | Year  | Modality1 | Datasets2 | Category3 | Scale | Multi-center | License4 |
|-----------|-------------------|-------|-----------|-----------|-----------|-------|--------------|----------|
| 55        | ADNI-1 [156]      | 2004  | ✓          | ✓         | ✓         | NC, MCI, AD | ✓ | O         |
| 56        | ADNI-1 [156]      | 2009  | ✓          | ✓         | ✓         | NC, MCI, AD | ✓ | O         |
| 57        | ADNI-1 [15]       | 2014  | ✓          | ✓         | ✓         | NC, EMIL, MCI, AD | ✓ | O         |
| 58        | ADNI-3 [129]      | 2016  | ✓          | ✓         | ✓         | NC, EMIL, MCI, AD | ✓ | O         |
| 59        | OASIS 1 [145]     | 2007  | ✓          | ✓         | ✓         | NC, AD    | ✓ | O         |
| 60        | OASIS 1 [142]     | 2009  | ✓          | ✓         | ✓         | NC, AD    | ✓ | O         |
| 61        | OASIS 1 [120]     | 2019  | ✓          | ✓         | ✓         | NC, AD    | ✓ | O         |
| 62        | TADPOLE [146]     | 2017  | ✓          | ✓         | ✓         | NC, AD    | ✓ | CR        |
| 63        | MRI and Alzheimer's| 2014  | ✓          | ✓         | ✓         | NC, MCI, AD | ✓ | O         |
| 64        | LADDementia [45]  | 2014  | ✓          | ✓         | ✓         | NC, MCI, AD | ✓ | O         |
| 65        | ANT [40, 140]     | 2019  | ✓          | ✓         | ✓         | NC, AD    | ✓ | O         |
| 66        | PD De Novo [201]  | 2018  | ✓          | ✓         | ✓         | NC, PD    | ✓ | CC0       |
| 67        | SCA2/DTI [147]    | 2018  | ✓          | ✓         | ✓         | NC, SCA2  | ✓ | CC4       |
| 68        | MTDP [10]         | 2016  | ✓          | ✓         | ✓         | Healthy, category I or category II | ✓ | CC4       |

1 DWI: diffusion weighted imaging; PT: positron emission tomography; OT: other modalities, for details, see footnotes.
2 AD: Alzheimer’s disease; PD: Parkinson’s disease; OT: other diseases, for details, see footnotes.
3 NC: normal cognitive; MCI: mild cognitive impairment; EMCI: early MCI; LMCI: late MCI.
4 CC0: Creative Common Zero License; CC4: Creative Commons 4.0 License; O: data is opened under specific license.
5 MRI FLAIR: Fluid-attenuated inversion recovery; T1w: T1 weighted; T2w: T2 weighted; FLAIR: Fluid-attenuated inversion recovery; DWI: diffusion weighted imaging; PT: positron emission tomography; OT: other modalities, for details, see footnotes.
6 NC: normal cognitive; MCI: mild cognitive impairment; EMCI: early MCI; LMCI: late MCI.
brain images. The datasets mainly include “Alzheimer’s Disease Neuroimaging Initiative (ADNI)” and “Open Access Series of Imaging Studies (OASIS)”.

The ADNI is a series of projects that aim to develop clinical, imaging, genetic, and biochemical biomarkers for the early detection and tracking of AD. It includes four stages: ADNI-1 (55), ADNI-GO (56), ADNI-2 (57), and ADNI-3 (58). These projects provide image data of the brain for researchers, and the modalities of images include MR (T1 and T2) and PET (FDG, PIB, Florbetapir, and AV-1451). These four stages consists of 1400 subjects. The subjects can be categorized into normal cognition (NC), mild cognitive impairment (MCI), and AD, where MCI can be split into early mild cognitive impairment (EMCI), later mild cognitive impairment (LMCI).

The OASIS is a series of projects aiming to provide neuroimaging data of the brain, which researchers can freely access. OASIS released three datasets, which are named OASIS-1, OASIS-2, and OASIS-3. All these three datasets are related to AD, but these datasets are also used in functional areas segmentation and other tasks. The OASIS-1 (59) contains 418 subjects aged from 18 to 96, and for the subjects older than 60, there are 100 subjects diagnosed with AD. The dataset includes 434 MR sessions. The OASIS-2 (60) contains 150 subjects, aged between 60 to 96, and each subject includes three or four MR sessions (T1). About 72 subjects were diagnosed as normal, while 51 subjects were diagnosed with AD. Besides, there are 14 subjects who were diagnosed as normal but were characterized as AD at a later visit. The OASIS-3 (61) includes more than 1000 subjects, more than 2000 MR sessions (T1w, T2w, FLAIR, etc.), and more than 1500 PET sessions (PIB, AV45, and FDG). The dataset includes 609 normal subjects and 489 AD subjects.

Moreover, there are many other challenges based on ADNI and OASIS or independence datasets. Challenge TADPOLE (62) is based on ADNI and aims at the prediction of the longitudinal evolution. Dataset (63) is based on OASIS and it is released on Kaggle for the classification of AD. Challenge CADDementia (64) is an independent AD-related challenge to classify subjects into NC, MCI, and AD.

Other diseases: Similar to AD, other brain diseases are also important from the diagnosis and treatment perspective. However, the number of datasets and challenges of these diseases is not as large as AD. A few datasets focus on Parkinson’s disease (PD) and spinocerebellar ataxia type II (SCA2). Datasets ANT (65) and (66) provide images of PD with MR images and classification labels. Dataset (67) provides images and classification labels of spinocerebellar ataxia-II, i.e., SCA2. Dataset MTOP (68) provides images and annotations for the diagnosis of mild traumatic brain injury.

3.3 Eye related datasets and challenges

As the human’s imaging sensor, the eyes’ health is essential for human beings, and eye diseases may lead to blindness. We introduce the relevant challenges and datasets in this subsection and list them in Table 4. Many of the studies presented the size of their dataset, with approximately half of them separating it into training and testing sets. Table 4 includes a list of challenges that offered rankings for participants’ methods, with many of them being associated with MICCAI and ISBI. It’s worth noting that access to these challenges often requires registration, whereas other datasets may be accessible with specific licenses.

Datasets according to the modality: With regards to the eye-related datasets and challenges, the main used modalities are the fundus photography (69, 70, 71, 72, 74, 75, 76, 77, 78, 79, 84, 86) and OCT (73, 80, 81, 83, 85). The data provided by Challenge (82) consists of surgery videos. The fundus photography can help medics evaluate the eye’s health and locate the retinal lesions because the fundus photography clearly shows the important parts of the eye, such as the blood vessels and the optic disc. OCT is a new imaging approach that is safe for the eye and shows the retinal tissues in details. However, it also has disadvantages — it is not suitable for diagnosing microangioma and the planning of retinal laser for the photocoagulation treatment.

Datasets according to the analysis task: These datasets and challenges can be used for four tasks.
### Table 4. Summary of datasets and challenges of eye-disease related tasks.

| Reference | Dataset/Challenge | Year | Modalities1 | Diseases2 | Tasks3 | Scale | Rank | License5 |
|-----------|-------------------|------|-------------|-----------|--------|-------|------|----------|
| 69        | BIADD [71]        | 2020 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 70        | THE GRAND CHALLENGE [80] | 2020 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 71        | REFUGE [150]      | 2020 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 72        | REFUGE [159]      | 2018 | ✓           | ✓         | ✓      | /     |      | MCCAIM   |
| 73        | AGE [71]          | 2019 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 74        | DRIVE             | 2019 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 75        | ODDE-2019         | 2019 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 76        | PALM              | 2019 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 77        | APTOS-2019        | 2019 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 78        | AAMM [30]         | 2018 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 79        | IRIADD [170]      | 2018 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 80        | Retinal OCT Images [104] | 2018 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 81        | ROCK              | 2017 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 82        | CALENDRS [6, 69]  | 2017 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 83        | RETOUCH [156]     | 2017 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 84        | Diabetic Retinopathy Detection [10] | 2015 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 85        | Seg OCT (DMR) [154] | 2015 | ✓           | ✓         | ✓      | /     |      | ISBI     |
| 86        | RGC [175]         | 2000 | ✓           | ✓         | ✓      | /     |      | ISBI     |

1. OCT: optical coherence tomography; FP: fundus photography; OT: other modalities, for details, see footnotes.
2. AMD: age-related macular degeneration; DR: diabetic retinopathy; G: glaucoma; OT: other diseases, for details, see footnotes.
3. Task: C for Classification, S for Segmentation, D for Detection, O for Other.
4. C for Challenge, D for Dataset, R for Registry. CC0 is Creative Commons 0 License. O means data is opened under specific license.
5. All the diseases of this dataset are listed on the official website. See https://isbi.grand-challenge.org/Data/

1) **Classification** tasks focus on classifying whether the subject has specific diseases or judging whether the subject is abnormal. Datasets and challenges (70, 71, 72, 73, 76, 77, 78, 84) focus on predicting a single disease, while others (69, 75, 79, 80, 81) focus on diagnosing multiple diseases. Challenges (71) is an upgrade of challenge (72) which provided more data for participants.

2) **Segmentation** is another task, which provides more information compared to classification. Datasets and challenges focus on the segmentation of the tissues and lesions, such as disc (72, 76, 79), blood vessels (74), and retinal (83), and DME lesion (85), for further diagnosis and disease analysis.

3) Datasets and challenges (72, 73, 76, 78, 79, 86) focus on the **detection** of lesions or other landmarks. These tasks help medics locate key targets, such as areas and tissues, for effective diagnosis or provide feature details for other automated algorithms.

4) Unlike other tasks, the last one focuses on the **annotation** of the tools used for eye-related surgery (82).

**Datasets according to focused eye diseases:** Researchers mainly focus on these diseases:

- Glaucoma (72, 75)
- Diabetes retinopathy (75, 77, 79, 80, 81, 84, 86)
- Closure glaucoma (73)
- Hypertension (75)
- (Age-related) macular degeneration (75, 78, 80)
- Pathologic myopia (75, 76)
- Cataract (75)
- Diabetic macular edema (79, 85)
- Diabetic macular edema

Besides these diseases, dataset (82) aims at the annotation of images and videos in surgery.

### 3.4 Datasets and challenges of other Subjects

Besides the brain’s structural analysis, the image processing, and the computer-aided diagnosis tasks, another important research focus is the human neck because it holds many essential glands and organs. This subsection discusses the datasets and challenges of the neck and teeth, covered in Subsection 3.4.1 and Subsection 3.4.2, respectively. Moreover, many researchers are working on the analysis of behavior and cognition with DL-based methods. We discuss the details in Subsection 3.4.3. Table 5 includes datasets and challenges that cover various modalities, including CT, PET, MR, and Ultrasound, with a focus on segmentation. Most of the challenges listed in...
when these become abnormal, effective diagnosis and segmentation play an essential role in their treatments. The main tasks of these two challenges include landmark localization and caries segmentation. Challenge (97) focuses on the diagnosis of dental X-Ray images.

3.4.1 Neck related datasets

The neck is also essential for our health. The neck holds many glands and organs, and

3.4.2 Cephalometric and teeth related datasets

Challenges (97, 98) focus on the diagnosis of dental X-Ray images. The neck is also essential for our health. The neck holds many glands and organs, and...
between brain activities and stimulation. They use functional MR images to scan our brain activity, analyze the hemodynamic feedback, and identify the area of the neurons which react. Therefore, the analysis of the reactions of the brain in response to a specific stimulation is an important research focus. Researchers use DL to detect or decode the stimulation of subjects to work out the brain’s functionality. The related datasets are listed in Table 6. These datasets are all opened with specific licenses and all of them are single-center data, which means the publisher collected the data themselves. Due to the difficulty of collecting data, the scale of these datasets is not too large (less than half of a hundred).

Some datasets (99, 100, 101) focus on classifying the stimulations or the subject’s attribution based on the subject’s functional MR images. Dataset (99) aims to identify whether the subject is a beginner or an expert in programming via the reaction of their brain to source codes. Dataset (100) focuses on diagnosing subjects with depression vs. subjects with no-depression using audio stimulations and analyzing the subjects’ brain activity. Dataset (101) focuses on the influence of cannabis on the brain.

Datasets (102, 103, 104, 105, 106, 107, 108, 109) focus on the encoding of the stimulations, i.e., brain activities’ decoding. Datasets (105, 107, 108) aim to rebuild what subjects have seen using DL-based methods from their brain activities using functional MR images. On the other hand, datasets (103, 109) work on the encoding of faces that subjects have seen from functional MR images with similar modalities.

3.5 Summary
Within this section, we have provided an overview of datasets and obstacles concerning the head and neck region. To present this information, we have created six tables detailing each entry within the prior subsections. Table 1 specifically pertains to datasets and challenges centered around brain segmentation, registration, and generation. Table 2 listed datasets and challenges which mainly focused on the segment of brain images, and at the same time, there are many other datasets and challenges focused on the tasks, such as generation and registration. The datasets and challenges, listed in Table 3, mainly focused on the classification of brain disease. Table 4 listed the datasets and challenges related to the classification, segmentation, detection, and other task associated with the diagnosis and surgery of eye disease. Table 5 includes datasets and challenges about diseases of the head and neck, such as the segmentation of the glands or lesions in glands and cephalometric. The datasets listed in Table 6 are related to behavior and perception.

4 Chest and abdomen related datasets and challenges
There are many vital organs in the chest and abdomen. For example, the heart is responsible for the blood supply; the lungs are responsible for breathing; the kidneys are responsible for the production of urine to eliminate toxins from the body. Therefore, the medical image analysis of organs in the chest and abdomen is an important research focus. Most of the tasks are computer-aided diagnosis with classification, detection, and segmentation of lesions being the most targeted tasks.

Many datasets and challenges aim to segment one or more organs in the chest and abdomen for diagnosis or treatment planning. Subsection 4.1 discusses the datasets and challenges related to inter- or intra-organ segmentation. Subsection 4.2 introduces the datasets and challenges that focus on the classification, detection, and segmentation tasks related to the chest and abdomen organs diseases diagnosing. Subsection 4.3 describes the datasets and challenges of the chest and abdomen that are not categorized above, including regression, tracking, registration, and other tasks related to the chest and abdomen organs.

4.1 Datasets for chest & abdomen organ segmentation
This subsection covers the datasets and challenges of the chest and abdomen organs that are used for anatomic segmentation tasks. The anatomic segmentation tasks include the inter-organ segmentation (Subsection 4.1.1) and intra-organ segmentation (Subsection 4.1.2). The inter-organ segmentation aims to separate an organ from the
background or mark the boundaries between multiple organs and the background. The intra-organ segmentation aims to segment the organ into different parts at the anatomical level.

Table 7 showcases the datasets and challenges that are dedicated to the segmentation of organs in the chest and abdomen regions. The majority of these datasets and challenges use CT and MR modalities. Information such as the number of images and division of training/testing sets were reported for most of these datasets and challenges. The challenges listed in Table 7 generally provide rankings for submitted methods, and many of them are affiliated with MICCAI, while a few are associated with ISBI or AAPM. Access to most of these challenges requires registration, while datasets and the remaining challenges are accessible under specific licenses.

4.1.1 Datasets of chest and abdomen organs

The inter-organ segmentation is a necessary information for the preplanning of surgery and diagnosis. A well-segmented contour of the organs provides a precise mask, which helps to produce accurate segmentation results for the diagnosis, treatment, and operation. This subsection introduces datasets and challenges for the inter-organ segmentation of a single organ and of multiple organs.

Chest & abdomen datasets according to the organ:

The inter-organ segmentation is one of the most attractive topics of medical image analysis [97, 103]. It is a basic pre-processing step in diagnosis and treatment, since it provides contour of organs. Some datasets and challenges focus on the larger organs, such as the liver and lungs, while others focus on the multi-organs segmentation or smaller organs. This is because of the challenges, such as imbalanced labels and blurred contour in low-resolution, brought by multi-organs and smaller organs. The datasets and challenges that we have covered involve the following organs and parts:

- Lung (110, 111, 112, 113, 114, 115, 116, 117)
- Liver (110, 111, 115, 116, 118, 119, 120, 121)
- Kidney (110, 115, 116, 118, 120, 122)
- Prostate (111, 123, 124, 125, 126, 127)
- Esophagus (113, 114, 120, 128)
- Heart (111, 113, 128, 129)
- Pancreas (111, 115, 120, 130)
- Aorta (115, 120, 128)
- Trachea (115, 128)
- Colon (111)
- Spleen (115, 118, 120)
- Adrenal glands (115, 120)
- Bladder (110, 115)
- Gall bladder (115, 120)
- Spinal cord (113, 114)
- Breast (131)
- Lymph (132)
- Clavicles (133)
- Stomach (120)

Chest & abdomen datasets according to modality:

The most commonly used image modalities for chest and abdomen organs segmentation are MR and CT. As Table 1 shows, many datasets and challenges use MR images. MR images have higher resolution under certain conditions and have better resolution for soft body tissues and organs, such as the heart and prostate. Meanwhile, CT is the most widely used modality for organ segmentation and other tasks and diagnosis that are related to chest and abdomen, such as the lung and liver, according to our research, because of its convenience, effectiveness, and low cost.

Chest & abdomen datasets according to focus:

The purpose of these datasets and challenges can be categorized into three groups: further analysis, benchmark, and radiotherapy. Most datasets and challenges which provide annotated organs’ contours are provided with the objective to focus on further analysis and treatments. One of the challenges of segmentation is to achieve a robust segmentation of the whole organ and separate it from the background, without omitting the lesions and tumors, and thus, some test benchmarks (111, 115) are provided for researchers to evaluate their algorithms. Another challenge is the imbalance between different organs because of their sizes and shapes, and such an imbalance makes it challenging to segment small organs and provide valuable information for analysis and treatment. Dataset and challenges (113, 128) provide data for researchers to address the problem.

Single chest & abdomen organ segmentation:

The single organ’s contour segmentation tasks usually focus on the segmentation with an anatomical purpose (112, 117, 123, 124, 125, 127) or the research of a region for subsequent tasks (119, 121, 122, 126, 130, 131, 132). The difficulty of the former task is that the lesions and tumors may affect the segmentation by separating the organ from the background, while the latter’s difficulty is for researchers and their algorithms to perform more precise segmentation. Challenge (133) focused on the segmentation of the clavicles.
Chest & abdomen multi-organs segmentation: The chest and abdomen multiple organs segmentation focuses on the inter-organ distinguishing. Some of these datasets and challenges (110, 111, 118) focus on the segmentation of multiple organs, including the relatively larger organs, which are easier to segment, and the relatively smaller organs, which can be more challenging to segment compared to the larger ones, especially when the model is handling the larger and smaller organs at the same time. Similarly, some of these datasets and challenges (113, 128) focus on the “organ at risk” which means that these organs are healthy but might be at risk because of radiation

ACM Comput. Surv.
therapy. Dataset (115) aims to provide a benchmark for the segmentation algorithms. Dataset (120) focuses on multi-atlas-based methods, which are widely used in brain-related research.

4.1.2 Chest & abdomen intra-organ segmentation: Different from inter-organ segmentation of the chest and abdomen organs, the intra-organ segmentation aims to segment the organ into different parts. Just as the hand has five fingers, organs are made up of multiple parts, and a typical example is the Couinaud liver segmentation method[58]. This subsection introduces the datasets and challenges for organ segmentation. These datasets and challenges are listed in Table 7.

Heart related datasets and challenges: Most of these datasets and challenges (137, 138, 139, 140, 141, 142, 143) are related to the heart segmentation. The most frequently used modalities are MR and ultrasound, and the aim is to segment the heart into the left atrium, chambers, valves, and other parts. Though MR and ultrasound can effectively image the different tissues of the heart, the heartbeat results in blurred images, which makes the segmentation task more difficult, while for ultrasound, the dynamic nature of ultrasound images is another challenge for the segmentation algorithm.

Others chest & abdomen body parts: Challenge (144) provides 55 CT scans and focuses on the segmentation of the lung with the labeling of its different parts: outside the lungs, the left lung, the upper lobe of the left lung, the lower lobe of the left lung, the upper lobe of the right lung, the middle lobe of the right lung, and the lower lobe of the right lung. The biggest challenge is the effect of the lung lesions and diseases, such as tuberculosis and pulmonary emphysema, on the performance of the segmentation. Moreover, challenges (145, 146) focus on the segmentation of the lung vessels.

---

### Table 8. Summary of datasets and challenges for chest and abdomen organs-related tasks I.

| Reference | Dataset/Challenge | Year | Modalities | Focus | Tasks | Scale | Multi- | Rank | Note | License |
|-----------|-------------------|------|------------|-------|-------|-------|-------|------|------|---------|
| 119 | LITS [34] | 2017 | CT, MR, PET, CR, US | Liver tumor | ✓ | ✓ | ✓ | ISBI | CR |
| 144 | C-MIMI | 2015 | CT, MR, PET, CR, US | Lung cancer, NEURO-IMAGING, OTHER Modalities | ✓ | ✓ | ✓ | ISBI | CR |
| 147 | Soft-tissue-Sarcoma [106] | 2015 | CT, MR, PET | Soft-tissue sarcoma | ✓ | ✓ | ✓ | ISBI | CR |
| 148 | ProstateX | 2015 | CT, MR, PET | Prostate lesion | ✓ | ✓ | ✓ | ISBI | CR |
| 149 | ProstateX | 2015 | CT, MR, PET | Prostate lesion | ✓ | ✓ | ✓ | ISBI | CR |
| 150 | Lung-PET-CT-Atlas [120] | 2015 | Lung cancer | ✓ | ✓ | ✓ | ISBI | CR |
| 151 | C-MIMI | 2015 | CT, MR, PET | Non-small cell lung cancer | ✓ | ✓ | ✓ | ISBI | CR |
| 152 | CheXpert [105] | 2016 | Chest | ✓ | ✓ | ✓ | ISBI | CR |
| 153 | ProstateX | 2016 | CT, MR, PET | Prostate lesion | ✓ | ✓ | ✓ | ISBI | CR |
| 154 | BCS-BCT [46] | 2016 | Breast cancer | ✓ | ✓ | ✓ | ISBI | CR |
| 155 | SIMACX [206] | 2017 | Pneumothorax | ✓ | ✓ | ✓ | ISBI | CR |
| 156 | ACAD [121] | 2017 | Cardiac diseases | ✓ | ✓ | ✓ | ISBI | CR |
| 157 | ProstateX | 2018 | CT, MR, PET | Prostate lesion | ✓ | ✓ | ✓ | ISBI | CR |
| 158 | SIMACX [206] | 2018 | Pneumonia | ✓ | ✓ | ✓ | ISBI | CR |

---

1. CT: computed tomography; MR: magnetic resonance image; PET: positron-emission tomography; CR: computed radiography; US: other modalities, for details, see footnotes.
2. C for classification, S for segmentation, D for detection.
3. Conference.
4. C for Challenge, D for Dataset, B for Benchmark, R for Registry, CC is Creative Commons License. O means data is opened under specific license. T means data is opened under TCIA's license. A means data is opened under Apache License 2.0.
5. E means entry density weighted, Dynamic contrast-enhanced, DE. Digital breast tomosynthesis.
6. Electronic health record and report.
work on this problem by analyzing medical images. Several researchers have designed automatic or semi-automatic
tools, which were associated with various conferences such as MICCAI, ISBI, RSNA, AAPM, ICIAR, among others. Most of
these datasets and challenges are released under specific licenses. Additionally, most of them reported the division of the training/testing set.

The Digital Mammography

Risks from Breast Cancer Screening

Chest & abdomen datasets according to modality: R means data is opened under Apache License 2.0. O means data is opened under specific license. T means data is opened under TCIA’s license. A means data is opened under ACRIN-FLT-Breast registry. CC0 is Creative Commons 0 License.

Table 9. Summary of datasets and challenges for chest and abdomen organs-related tasks II.

| Reference | Dataset/Challenge | Year | Modality | Focus | Tasks | Scale | Multi- | Rank | Cones | License |
|-----------|-------------------|------|----------|-------|-------|-------|-------|------|-------|---------|
| 107       | RIMDR Long CT [234] | 2013 | ✓        | Pulmonary nodules | ✓     | 122/126 | x     | T    |       |         |
| 108       | DREAM Challenge [21, 22] | 2014 | ✓        | Non-small cell lung cancer | ✓     | 648/871 | x     | T    |       |         |
| 109       | The Digital Mammography DREAM Challenge | 2016 | ✓ ✓ ✓ | Breast cancer | ✓ ✓ ✓ | 460/140 | ✓ ✓ ✓ | C/R  |       |         |
| 170       | ACRIN-FLT-Breast [18] | 2015 | ✓        | Pulmonary nodules | ✓     | 10/28 | x     | T    |       |         |
| 171       | ImageCLEF 2018 – Tuberculosis | 2018 | ✓        | Tuberculosis | ✓     | 1080/1081 | x     | T    |       |         |
| 172       | Data Science Bowl 2013 | 2013 | ✓        | Lung cancer | ✓     | 2013 | x     | T    |       |         |
| 173       | NSCLC - Radioprotection: Initial Stanford Study of 26 Cases [14] | 2013 | ✓ ✓ ✓ | Non-small cell lung cancer | ✓ ✓ ✓ | 26 | x     | T    |       |         |
| 174       | LIDC-IDRI [22] | 2013 | ✓ ✓ ✓ | Pulmonary nodules | ✓ ✓ ✓ | 6 | x     | T    |       |         |
| 175       | Lung Force CT-Panology [16] | 2018 | ✓ ✓ ✓ | Pulmonary nodules | ✓ ✓ ✓ | 6 | x     | T    |       |         |
| 176       | Automatic Nodule Detection 2009 [19] | 2009 | ✓        | Pulmonary nodules | ✓ | train/example/test | x ✓ | SPIE C/R |
| 177       | EndoCV 2021 [17] | 2020 | ✓ ✓ ✓ | Colon, rectal cancer, colorectal cancer, and stomach | ✓ ✓ ✓ | ✓ ✓ ✓ | ✓ ✓ ✓ | ISBI C/R |
| 178       | EAD 2020 | 2020 | ✓        | Artifical region | ✓ | ✓ | ✓ | C/R |
| 179       | EDDI 2020 | 2020 | ✓ ✓ ✓ | Colon, rectal cancer, colorectal cancer, and stomach | ✓ ✓ ✓ | ✓ ✓ ✓ | ✓ ✓ ✓ | C/R |
| 180       | EAD 2019 [18] | 2019 | ✓ ✓ ✓ | Artifical region | ✓ ✓ ✓ | ✓ ✓ ✓ | ✓ ✓ ✓ | C/R |
| 181       | SARAS-ESAD [13] | 2020 | ✓ ✓ ✓ | Surgeon action detection | ✓ ✓ ✓ | ✓ ✓ ✓ | ✓ ✓ ✓ | MICCAI CC/AC |
| 182       | RSNA Protonematos Detection Challenge | 2018 | ✓        | Protonematos | ✓ | ✓ | ✓ | RSNA C/R |
| 183       | SMIC Chest X-Ray [12] | 2017 | ✓ ✓ ✓ | Lung Disease | ✓ ✓ ✓ | ✓ | ✓ | O |
| 184       | CT COLONOGRAPHY [10] | 2011 | ✓ ✓ ✓ | Polyp & colorectal cancer | ✓ ✓ ✓ | ✓ | ✓ | T |
| 185       | British 16monthly 1opy | 2020 | ✓ ✓ ✓ | Breast cancer | ✓ ✓ ✓ | ✓ | ✓ | CC |
| 186       | NSCLC Breast (ACRIN 6588) [114] | 2017 | ✓ ✓ ✓ | Breast cancer | ✓ ✓ ✓ | ✓ | ✓ | O |
| 187       | LIDC-IDRI Lung Nodules [16] | 2017 | ✓ ✓ ✓ | Pulmonary nodules | ✓ ✓ ✓ | ✓ | ✓ | C/R |
| 188       | Low Dose CT Challenge | 2015 | ✓ ✓ ✓ | Lung cancer & pulmonary nodules | ✓ ✓ ✓ | ✓ | ✓ | AAPM C/R |
| 189       | Venricular Intact Segmentation | 2012 | ✓ ✓ ✓ | Left ventricular myocardial infarction segmentation | ✓ ✓ ✓ | ✓ | ✓ |         |

1. CT: computed tomography. MR: magnetic resonance imaging. PT: positron emission tomography. DR: digital radiography. Other modalties for details, see footnotes.
2. T for classification. S for segmentation. D for detection.
3. Conference.
4. C for Challenge. D for Dataset. B for Benchmark. R for Registry. CC0 is Creative Commons 0 License. O means data is opened under specific license. T means data is opened under TCIA’s license. A means data is opened under ACRIN-FLT-Breast registry.
5. Mammography. 6. Pathology image.

4.2 Datasets for diagnosis of chest & abdomen diseases

Diseases of organs in the chest and abdomen have a significant impact on human health. Therefore, many researchers work on this problem by analyzing medical images. Several researchers have designed automatic or semi-automatic algorithms for the classification, segmentation, detection, and characterization tasks to help medics diagnose these diseases. In this subsection, we describe the datasets and challenges related to the diagnosis of diseases of the chest and abdomen that are reported in Tables 8, 9, and 10, respectively.

The majority of datasets and challenges listed in Tables 8, 9, and 10 provided information on the scale of the data, including whether they included multi-center data and whether they had rankings. Many of the challenges were associated with various conferences such as MICCAI, ISBI, RSNA, AAPM, ICIAR, among others. Most of these challenges required registration to access the data, while other datasets and challenges were available under specific licenses. Additionally, most of them reported the division of the training/testing set.

Chest & abdomen datasets according to modality: According to the datasets and challenges collected, CT is the most commonly used imaging modality for the chest & abdomen, because of its suitable imaging quality and ability to clearly display tissues and lesions. Some datasets and challenges also provide CT images using contrast agents for clearer images. Besides CT imaging, there are other modalities, including MR, X-Ray digital radiographs, PET,
endoscopy, etc. The MR images are used in breast-related diagnosis, cardiac-related tasks, soft tissue sarcoma detection, and ventilation imaging. Because of the organs’ size and the CT’s resolution, which is limited by the imaging exposure time and radiation dose, MR is a more suitable imaging modality for small or specific organs. The PET is always used with other modalities, such as CT and MRI. The contrast agent’s density is related to the metabolism, which means the density of radiation from contrast agent will be high in the tumor, so PET is always used for tumor related tasks. Endoscopy images are used for medical inspection of the stomach, intestines, and others.

Chest & abdomen datasets according to classification of diseases: The classification of diseases intends to determine whether a subject is healthy or not. A quick and early diagnosis can allow effective interventions to increase the probability of the patient recovery before the condition worsens.

The main focus of these datasets is to judge whether there is any cancer, lesion, or tumor, such as soft tissue sarcoma (147), prostate lesion (148, 149), lung cancer (150, 151, 152, 153), and breast cancer (154).

Another focus is the classification of diseases. These diseases include mainly pneumothorax (155), cardiac diseases (156), tuberculosis (157), pneumonia (158), and COVID-19, which are discussed at the end of this subsection.

The endoscopy related challenges (159, 160, 161) provide data with the aim to classify RGB images and videos into “normal” vs. “abnormal” categories. The “abnormal” in (160) include VA (mucosa showing villous atrophy), CH (mucosa showing crypt hypertrophy), and VACH (mucosa showing both villous atrophy and crypt hypertrophy); the “abnormal” in (161) include GMP (gastric metaplasia), BAR (intestinal metaplasia), NPL (neoplastic mucosa).

Dataset (162) focuses on the classification based on the diagnostic records. These datasets and challenges provide data for researchers to design AI-based algorithms to diagnose common diseases.

ACM Comput. Surv.
Chest & abdomen datasets for attribute classification: The characterization task of the tumor and lesion is also called attribute classification, which focuses on the subsequent characterization analysis of the tumors and lesions following the detection and segmentation tasks using automatic analysis algorithms. A typical example is the attributes classification of pulmonary and lung cancer (164, 165, 166, 167, 168). The datasets and challenges usually provide CT scans with the annotation of different attributes, such as lesion type, spiculation, lesion localization, margin, lobulation, calcification, cavity, etc. Each attribute includes two or more categories. Another focus is the characterization of the breast related lesions and tumors (169, 170).

Chest & abdomen datasets for detection: In most research and clinical situations, classification is not enough. The medics and researchers usually focus on the lesions related to such a disease, and their localization. Further treatment evaluations, plan and interpretability are the specific focus for medics and DL researchers. Thus, detection and segmentation are the tasks which are receiving a lot of attention at present. The detection task aims to find a region of interest and localize its position. The regions of interest usually include:

- Lung cancer and tumor (167, 171, 172, 173, 174)
- Pulmonary nodule (164, 168, 174, 175, 176)
- Celiac-related damage (177, 178, 179, 180)
- Action and artefact of surgeon(180, 181)
- Other lung lesions (182, 183)
- Polyp (179, 184)
- Breast cancer (169, 185, 186)
- Cervical cancer (187)
- Liver cancer (188)

Chest & abdomen datasets for segmentation: Segmentation is a refinement of the detection task because it provides information about the location and the pixel-level labels. Pixel-level annotations can help researchers design pixel-level algorithms for accurate and effective quantification, volume calculations, and other analysis and diagnosis of tumors and lesions at the pixel-level (e.g., monitoring of tumors size). According to the datasets and challenges we have collected, most of them aim at the segmentation of the tumor and lesion from CT of:

- Heart associated conditions (189)
- Kidney tumor (122)
- Liver tumor (119)
- Lung cancer (190, 191, 192)
- Pneumothorax (155)
- Polyp (179)
- Pulmonary nodule (193)

Additionally, challenge (194) centers around the segmentation and ventilation-related tasks of the lung. Meanwhile, challenges (178, 180) concentrate on the segmentation of artifacts (such as polyps) in endoscopic images. The discrepancies between the two challenges are the specific tasks and corresponding data. Challenge (195) encompasses a range of intricate tasks associated with medical images and offers various modalities.

COVID-19: In 2020, COVID-19 became a research focus because it caused more than 100 million infections and two million deaths. Different datasets and challenges focus on this devastating disease and provide data to help researchers develop deep learning models to detect COVID-19 via various medical image modalities.

With respect to modalities, most of these datasets and challenges use either CT or CR images, and some provide both modalities. One exception is dataset (196), which uses ultrasound images. These datasets provide image annotations labeled by radiologists.

Most of these datasets and challenges are related to classification tasks. Datasets (197, 198, 199, 200, 201) directly focus on diagnosing COVID-19 from normal subjects. In contrast, datasets and challenges (196, 202, 203, 204, 205) focus on diagnosing COVID-19 from a few other similar diseases, which can also lead to lung opacity or other symptoms, such as Middle East Respiratory Syndrome (MERS), Severe Acute Respiratory Syndrome (SARS), and Acute Respiratory Distress Syndrome. Moreover, other datasets and challenges (206, 207) focus on the diagnosis task, with natural language processing, or genomics.

Similarly, some other datasets (199, 208, 209) focus on the segmentation or detection of COVID-19 related lesions, such as ground-glass opacity, air-containing space, and pleural effusion.

Additionally, dataset (210) offers images and clinical data to support medical professionals and researchers in their diagnostic efforts and algorithm development.
4.3 Datasets for other chest and abdomen-related tasks

Besides the classification, detection, and segmentation tasks, there are also several other tasks which are the current focus of research. In the following, we present the datasets and challenges related to these tasks, and report them in Table 11.

The datasets and challenges presented in Table 11 primarily consist of MR, CT, and ultrasound data, with some providing information on the size of the dataset. Challenges in this table include rankings of participant methods, with many associated with MICCAI. While most challenges require registration to access data, some datasets and challenges are openly available.

**Chest & abdomen datasets for regression:** Similar to attributes classification, regression is another task which aims to compute or measure the target attributes from given images, but the difference is that the outputs of regression are continuous. A typical example is fetal biometric measurements (211, 212). These challenges provide ultrasound images to help researchers design algorithms to measure such attributes to estimate the gestational age and monitor the fetus’s growth. Besides, another example is cardiac measurements (213, 214, 215, 216, 217, 218, 219). Of these, challenge (214) provided data more clinical, compared with its previous challenge (215). These datasets and challenges provide MR or ultrasound images to analyze the heart’s attributes to detect heart diseases.

**Chest & abdomen datasets for tracking:** Tracking is a crucial task as organs and the body tend to move during imaging. The motion characteristics of organs such as the heart provide valuable information. Challenges (220, 221)...

---

**Table 11. Summary of datasets and challenges of other medical applications in chest and abdomen.**

| Reference Index | Dataset/Challenge | Year | Modalities | Focus | Tasks | Scale | Rank | Cones. | License |
|-----------------|------------------|------|------------|-------|-------|-------|------|--------|---------|
| 211             | HC 18 [210]      | 2018 | ✓          | chest heart circumference | ✓ | random 200/335 | ✓ | CC0    | MICCAI  |
| 212             | Challenge US-IBSI 2012 [109] | 2012 | ✓          | chest heart circumference | ✓ | ISBI |      |        |         |
| 213             | CRE - EPI/GY10    | 2019 | ✓          | heart modeling | ✓ | MICCAI |        |        |         |
| 214             | LVQUAN19          | 2019 | ✓          | heart modeling | ✓ | MICCAI |        |        |         |
| 215             | LVQUAN15          | 2018 | ✓          | heart modeling | ✓ | MICCAI |        |        |         |
| 216             | A-AFMA-Localization 2020 | 2020 | ✓          | heart motion tracking | ✓ | MRA, UM |        |        |         |
| 217             | 7th Annual Mike Sions Research | 2006 | ✓          | cardiac motion fraction | ✓ | ISBI |        |        |         |
| 218             | Coronary Artery Stenosis Detection and Quantification Evaluation [110] | 2012 | ✓          | coronary artery stenosis detection and quantification | ✓ |         |        |        |         |
| 219             | LV Mechanics Challenge [111] | 2017 | ✓          | modeling | ✓ | MICCAI |        |        |         |
| 220             | AUST 2015         | 2017 | ✓          | heart tracking | ✓ | MICCAI |        |        |         |
| 221             | MUSST 2014        | 2014 | ✓          | heart tracking | ✓ | MICCAI |        |        |         |
| 222             | EchoNet-Dynamic   | 2017 | ✓          | heart tracking | ✓ | ISBI |        |        |         |
| 223             | Metrics - Tracking Challenge [202] | 2011 | ✓          | heart motion tracking | ✓ |         |        |        |         |
| 224             | Cardiac Motion Analysis Challenge 2011 | 2011 | ✓          | heart motion tracking | ✓ |         |        |        |         |
| 225             | LUMBC             | 2010 | ✓          | heart registration phantom images | ✓ | ISBI |        |        |         |
| 226             | MMWHS17           | 2017 | ✓          | heart registration phantom images | ✓ | ISBI |        |        |         |
| 227             | Motion Correction Challenge [203] | 2014 | ✓          | heart motion correction | ✓ | ISBI |        |        |         |
| 228             | A-AFMA-Detection 2020 | 2020 | ✓          | cardiac motion tracking | ✓ | MICCAI |        |        |         |
| 229             | A-AFMA-Localization 2020 | 2020 | ✓          | cardiac motion tracking | ✓ | ISBI |        |        |         |
| 230             | Statistical Analysis and Computational Modeling of the Heart - Dataset [204] | 2016 | ✓          | left atrium wall thickness | ✓ | ISBI |        |        |         |
| 231             | A-AFMA-Detection 2020 | 2020 | ✓          | myocardium thickness | ✓ | MICCAI |        |        |         |
| 232             | A-AFMA-Localization 2020 | 2020 | ✓          | myocardium localization | ✓ | ISBI |        |        |         |
| 233             | TeachNDo2020      | 2020 | ✓          | teachNDo2020 | ✓ | MICCAI |        |        |         |

1. CT: computed tomography; MR: magnetic resonance image; US: ultrasound; OT: other modalities, for details, see footnotes.
2. R: reconstruction, S: registration, T: tracking, D: detection, O: other.
3. Cones.
4. CC0 is Creative Commons 0 License. O means data is opened under specific license. T means data is opened under TCIA’s license. A means data is opened under Apache License 2.0.
5. See official description: http://cr.cscs.ch/cr/challenges.html.
6. CT Angiography
7. CT Pulmonary Angiography
8. Endoscopy
9. Classification
10. Generation

---

ACM Comput. Surv.
focus on tracking the liver using ultrasound data to analyze surgery and treatments. The tasks of challenges (220, 221) are similar, but challenge (220) provided more data compared to the previous year. Datasets and challenges (222, 223, 224) focus on the tracking of the heart. They provide ultrasound images to track and analyze the heart.

_Chest & abdomen datasets for registration:_ Challenge (225) focuses on the CT registration of lungs and provides CT scans with and without enhanced and contrast agents. Meanwhile, challenges (226, 227) focus on the registration between different modalities of the heart and provide MR, CT, and other modalities to register images with beating hearts. Challenge (228) focuses on the registration of lungs.

_Datasets for other chest & abdomen related tasks:_ Challenge (229) focuses on the detection of amniotic fluid. Challenges (230, 232) focus on localizing specific landmarks, including the amniotic and the heart, using ultrasound and MR images. Challenge (233) focuses on the classification of surgery videos. Dataset (234) focuses on the reconstruction of the coronary artery.

### 4.4 Summary

In this section, we have provided a summary of the datasets and challenges related to the chest and abdomen. We have organized them into 5 tables and described them in detail in the previous subsections. Table 7 includes datasets and challenges that concentrate on the segmentation of organs in the chest and abdomen, with some focused on single or multiple organ segmentation, while others on intra-organ segmentation. Tables 8, 9, and 10 present datasets and challenges associated with diagnostic tasks, such as disease classification, tumor segmentation, and lesion detection. Table 11 lists other datasets and challenges that focus on tasks like regression, tracking, registration, and generation.

### 5 Datasets and challenges for pathology and blood

Though radiography, MR imaging, and other imaging modalities have been used as the basis for diagnosis, pathology images are also used as a gold standard for diagnosis, particularly for tumors and lesions. Digital pathology images are generally obtained by collecting tissue samples, making slices, staining, and imaging. Therefore, pathology images are also one of the mainstream image modalities that are used for diagnosis.

The focus of these datasets and challenges include 1) the classification and segmentation of basic elements (e.g., cell and nucleus) in pathology images, and 2) blood-based diagnosis from images. In this section, we present datasets and challenges of the pathology images (Subsection 5.1), and at the same time, cover the datasets and challenges of blood images in Subsection 5.2.

#### 5.1 Datasets & challenges for pathology

Pathology images are used as one of the bases for the cancer diagnosis[29]. The pathologists and automatic algorithms analyze images based on specific features, such as cancer cells and cells under mitosis. Many organizations and researchers provide datasets and challenges, which focus on the microscopic pathology and at the whole slide image (WSI) level. The relevant datasets and challenges are listed in Table 12.

Table 12 contains datasets and challenges that primarily use H&E as the staining method. Most of the listed datasets and challenges report the size of the data and whether it includes multi-center data. The division of the training/testing set is also reported for most of them. Multi-center datasets and challenges are more prevalent than single-modalities ones. The challenges provide rankings for submitted methods, and they are mostly associated with MICCAI, ISBI, or other conferences. Some of the challenges require registration to access the data, while others are open with specific licenses.

**Imaging datasets & challenges:** In most situations, WSI is used in pathology diagnosis. Unlike CT or MR images, the pathology image is an optical image similar to the picture photoed by a camera. However, one major difference is that a pathology image is imaged by transillumination, while the usual photo is imaged by reflection. Another difficulty is in the size of the image. WSI is stored in a multi-resolution pyramid structure. A single multi-resolution
WSI is generally achieved by capturing many small high-resolution image patches, and it might contain up to billions of pixels. Thus, WSI is used as a virtual microscope in diagnosis for clinical research, and many challenges use WSI, such as (235, 236, 237, 238, 239, 240).

However, for some machine learning-based algorithms, it is hard to directly process and analyze the whole WSI, for example cell segmentation (241, 242). In some situations, the original image is too large to be analyzed directly by algorithms because of the limited computing resources. In other situations, “end-to-end” deep learning-based algorithms are hard to use for both low- and high-resolution-level features without zoom-in and zoom-out action and intelligent agents. Therefore, pathology image patches are used in several other challenges, such as (243) for visual question answering, (244) for mitosis classification, (241, 242) for multi-organs nucleus detection and segmentation.

**Stains:** Slides made from human tissues are without color, and required to be stained. The commonly used stains include Hematoxylin, Eosin, and Diaminobenzidine. Usually, two or more stains are used in staining the slide, and the most commonly used combinations include Hematoxylin & Eosin (H&E) and Hematoxylin & Diaminobenzidine (H-DAB), as listed in Table 12.

### Table 12. Summary of datasets and challenges for pathology-related image analysis.

| Reference Index | Dataset/Challenge | Year | Focus | Tasks | Stain | Scale | Multi-organ | Rank | Conc. | License |
|-----------------|-------------------|------|-------|-------|-------|-------|------------|------|-------|----------|
| ISBI 2014       | HE2               | 2020 |       | ✓     | H&E   |         |            |      |       |          |
| ISBI 2016       | HE2               | 2016 |       | ✓     | H&E   |         |            |      |       |          |
| ISBI 2017       | HE2               | 2017 |       | ✓     | H&E   |         |            |      |       |          |
| ISBI 2018       | HE2               | 2018 |       | ✓     | H&E   |         |            |      |       |          |

1. C for classification, S for segmentation, D for detection, O for other.
2. Conference.
3. Data is opened under Creative Commons License.
4. Visualization.
5. Tracking.
6. Electron microscope, which is not stained.
7. Cell counting.
8. Immunohistochemistry.
9. Non-linear image registration.

ACM Comput. Surv.
Pathology datasets according to disease: The pathology slides are widely used in the diagnosis of many diseases, especially cancer. The cancer cells and tissues have different shapes compared to their normal counterpart. Thus, the diagnosis via pathology is the gold standard. Many datasets and challenges, such as (241, 242, 245), do not address any specific disease, but the cell-level tasks. At the same time, many datasets and challenges target specific diseases, such as breast cancer (246, 247), cancers in the digestive system (248), cervical cancer (249, 250), lung cancer (251), thyroid cancer (252), and osteosarcoma (253).

Pathology datasets according to task: Generally speaking, the tasks used with these datasets and challenges can be classified into two categories: microscopic task and WSI-level task. The latter targets the diagnosis of diseases, based on a classification task. Expanded from the simple classification tasks, many datasets and research methodologies focus on complex tasks, such as the segmentation of tumor cell areas (241, 242, 254) and the detection of pathological features (239, 248). The microscopic tasks derive from the clinical analysis to identify cells and detect mitosis to extract key features from pathology images to support further disease diagnosis. The following subsections expand on the microscopic tasks and WSI-level tasks, respectively.

5.1.1 Microscopic related datasets: In this subsection, we introduce the microscopic task related datasets and challenges. These tasks focus on microscopic features extraction (e.g., nucleus features), for further diagnosis and WSI-level tasks.

Data: Unlike the WSI-level, the datasets and challenges which focus on microscopic tasks usually provide small size patch-level images with high-resolution. These patches are suitable for the annotation of microscopic-level objects and resource-limited algorithms. The size of images varies depending on the image analysis tasks. For the segmentation and detection of cells and nucleus, the size of images is usually a thousand-pixel square to contain the suitable number of cells or nuclei. For individual cell analysis tasks (e.g., mitosis determination), the size is usually of a single cell. For other tasks (e.g., the patch-level classification), the size varies from dataset to dataset.

Pathology datasets for cell detection & segmentation: The cell is one of the essential parts of the pathology image. The analysis of cells is one of the most effective ways to extract pathology image features for diagnosis. The pathologists analyze the size, shape, pattern, and stained color of the cells with their knowledge and expertise to make judgments about these cells and classify them as normal or abnormal. Thus, many datasets and challenges focus on the segmentation and detection of cells. The cells and nucleus can be placed neatly in the slide. However, during the slide preparation, these cells could overlap or locate randomly on the slide. Aiming at such a problem, challenges (249, 250) focus on the segmentation and detection of overlapping cells and nuclei. The shape and size of cells from different organs might be different and can have different recognition and analysis challenges. Therefore challenges (241, 242) focus on the multi-organ cells or nucleus segmentation.

Pathology datasets for patch-level classification: Generally, the size of WSI is too large to be able to analyze every cell and relationships between cells. DL-based methods can find essential information from the patch-level image to support the diagnosis based on feature learning. The datasets and challenges, which provide patch-level images, mainly focus on the classification, segmentation, or detection tasks. Challenges (252, 255, 256, 257, 258) focus on patch-level image classification to determine whether metastatic or a different tissue is present.

Datasets for other pathology tasks: Besides the detection and segmentation of cells and the patch-level classification, there are other microscopic tasks. Challenge (244) focuses on the mitotic detection for nuclear atypia scoring. The atypical shape, size, and internal organization of cells are related to the progress of cancer. The more advanced the cancer is, the more atypical the cells looks like. Challenge (245) focuses on cell tracking, to know how cells change shapes and move as they interact with their surrounding environment. This is the key to understand cell migration’s mechanobiology and its multiple implications in normal tissue development and many respective diseases. Challenge (243) focuses on the visual question answering task of pathology images using AI where the model is trained to pass the examination of the pathologist. Challenge (263) focuses on the tracking of particles, which mimic moving viruses, vesicles, receptors, and microtubule tips.
5.1.2 Datasets for WSI-level tasks

WSI-level pathology tasks focus on the diagnosis of cancer and pathology image processing. WSI contains all the complete information of a patient to be able to establish an accurate diagnosis. Automatic diagnosis algorithms can quickly analyze the slide. This is useful, especially in developing countries where there is a lack of well-experienced pathologists. However, it is a challenge to directly analyze WSI for both pathologists and algorithms because the size of WSI can be up to 100,000 × 100,000 pixels. Thus, such analysis becomes challenging, and to address this, most of the current datasets and challenges focus on the classification and segmentation of biomarkers, cells, and other regions of interest. At the end of this subsection, we introduce other datasets and challenges that are related to the tasks of regression and localization of tumors and biomarkers.

Datasets for classification of WSI: The prime goal of the examination of pathological images, especially WSI, is to diagnose cancer. Thus, how to classify WSI with large size and limited computing resources becomes a research challenge. Datasets and challenges (235, 236, 259, 260) focus on predicting cancer or evaluating WSI, such as Gleason grade or HER2 evaluation. At the same time, some datasets and challenges (238, 239, 247) focus on the classification of metastasized cancer. The main difference between (239) and (238) is the scale of the dataset, and the latter provided more data for training and testing.

Datasets for segmentation and detection of WSI: DL-based methods are seen as a black box which process pathology images. The performance of these methods has achieved the state-of-the-art performance, but the interpretability of these methods is still difficult. From the pathologists’ point of view, datasets and challenges (237, 238, 239, 248, 261) focus on the segmentation and detection tasks to determine the critical elements which led to a particular diagnosis, such as cancer cell area and signet ring cell.

Datasets for other WSI tasks: Besides classification and detection, there are a few other tasks based on WSI. This includes the registration of pathology images (262) for data pre-processing and the localization of lymphocytes (255).

5.2 Blood-related datasets

Blood image analysis is the basis of the diagnosis of many diseases. In contrast to the pathology images, blood samples’ images mainly contain blood cells, and these datasets and challenges are aimed at blood-related cancer and cell counting. Similar to pathology images, these datasets and challenges also focus on the segmentation, detection, and classification of cells. The relevant datasets and challenges are listed in Table 13. Table 13 includes several datasets and challenges, with most of them reporting the size of data available. Only two of them mention the division of training/testing set. About half of the challenges listed in Table 13 are linked with ISBI, while most of them require registration to access data. Other datasets and challenges are open.

One of the primary tasks of these datasets is the classification of cells, which focuses on identifying the different types of cells. Dataset (264) focuses on classifying red blood cells, white blood cells, platelets, and other cells.
the same time, dataset (265) focuses on the classification of malignant and non-malignant cells. Challenge (266, 267) focuses on the classification of malignant cells from normal cells, in B-ALL white blood cancer microscopic images. Other datasets and challenges (268) (lymphocytes), (269, 270) (multiple myeloma segmentation), (271) (mitochondria segmentation), (272) (malaria detection) focus on the segmentation and detection of blood cells and biomarkers. Dataset (273) focuses on stain normalization for white blood cancer (B-ALL & MM) microscopic images.

5.3 Summary

In this section, we have provided a summary of the datasets and challenges related to pathology and blood. We have organized them into 2 tables to better illustrate and describe them. Table 12 lists the datasets and challenges related to pathology. These include tasks such as WSI-level classification and segmentation related to cancer, cell detection or segmentation, and patch-level tasks. Table 13, on the other hand, lists datasets and challenges related to blood-related microscopic images. Most of the datasets and challenges listed in Table 13 report the scale of data, and some of them provide information on the division of training/testing set. Half of the challenges are associated with ISBI, and most of them require registration when accessing data, while other datasets and challenges are open.

6 Discussions

The success of AI algorithms such as DL has led to their widespread use in several fields, including for medical image analysis. Researchers with different knowledge and background tackle image-based clinical tasks using computer vision tools to design automatic algorithms for different applications [59, 108, 137, 143, 228, 230, 231]. Though AI algorithms can successfully handle many tasks, several unsolved problems and challenges hinder the development of AI-based medical image analysis.

6.1 Problems and challenges

DL-based algorithms learn from input images of real data through gradient descent. Large-scale annotated datasets and a powerful DL model are key to the development of successful DL models. For example, the success of AlexNet [116], GoogleNet [198], ResNet [85] are based on powerful models, which include millions of parameters. At the same time, a large-scale dataset, such as ImageNet [60], is also necessary to train the DL model to be able to tune such a large number of parameters. However, when these methods are applied to medical image analysis, many domain-specific problems and challenges start to appear. This subsection discusses some of these challenges.

6.1.1 Data scarcity

The biggest challenge in the development of DL models is data scarcity. Different from other areas, the scale of the medical image datasets is usually smaller due to many limitations, e.g., the ethical restrictions.

The commonly used datasets for traditional computer vision are in larger scale compared to medical image datasets. For example, the handwritten digits dataset, MNIST [122] includes a training set with 60,000 examples and a testing set with 10,000 examples; the ImageNet dataset [60] includes three million images for training and testing; Microsoft COCO [132] includes more than two million images with annotations. In contrast, many medical image datasets are smaller and only include hundreds or at most thousands of images. For example, the challenge BraTS 2020 (30) includes four hundred subjects and different modalities for each subject; the challenge REFUGE (72) provides about 1200 images of the eye; the challenge LUNA 16 (166) provides 888 CT scans; our recently published dataset of pulmonary lesions [127] just provides 694 scans; the challenge CAMELYON 17 (238) contains only more than 1000 WSI pathology images.

There are multiple reasons for the lack of data. The main cause is due to the restricted access to medical images by non-medical researchers, i.e., barriers between disciplines. The root causes of these barriers relate to the cost and difficulties of annotation and the restricted access due to ethics and privacy.

Access to data: As mentioned in the introduction, the direct cause of the data scarcity is that most non-medical researchers are not allowed to access medical data directly. Though many medical data are generated worldwide
every day, most non-medical researchers have no authorization to access clinical data. The easily accessible data are publicly available datasets, but these datasets are not at a large-scale to be able to properly train a DL model.

Ethical reasons: Ethics of medical data usage is a major bottleneck and a limitation for researchers, particularly, computer scientists, because they do not have resources to collect data. It is also hard for them to access medical data. Medical data stored in databases always contains sensitive or private information, such as name, age, gender, and ID number. In some cases, the data records of medical images can be used to identify a patient. For example, if an MR scan includes the face, an intruder can identify them for a possibly evil purpose. In most countries and regions, it is illegal to distribute such data with private information without the patients’ permission, and nobody would usually consent to such distribution. Therefore, for Deep learning researchers, it is impossible to gain authorization to access these datasets.

Before DL researchers are able to gain authorization even to desensitized data, they still need to pass ethical reviews.

Annotation: Another root cause is the difficulty to annotate medical images. Unlike other computer vision areas, the annotation of medical images requires specialized professionals and knowledge. For example, in autopilot, when annotating objects such as vehicles and pedestrians, there are no specific annotators’ requirements because most of us can easily distinguish a car or a human. However, when annotating medical images, domain-specific knowledge is essential. E.g., few people without professional knowledge would be able to tell the differences between an abnormal and normal tissue. However, it is impossible for a non-specialist to mark the lesion’s contour or diagnose a disease.

This difficulty cannot easily be solved even when professionals are employed to annotate data. First, the cost of annotation of medical data is huge. Once the researcher and their organization have obtained some data, they need then to spend more money to employ few medics for its labeling. Such annotation cost is enormous, particularly where medical resources are scarce or where medical costs are high. For example, the challenge PALM (76) provides about 1,200 images with annotation, but its organizers involved only two clinical medics. Second, the physician who annotates the data is required to have a rich clinical and diagnosis experience, thus reducing the number of people who are suitable for this task even further. Third, to avoid any subjectivity, one image needs to be annotated by two or more physicians. Another problem is what to do if the labels of two annotators are not the same? In many challenges, the organizer employs many junior physicians to annotate and employs a senior physician to decide if the junior physicians’ annotations are not the same. For example, in the challenge AGE (73), each data annotation is determined by the mean of four independent ophthalmologists in a group and it is then manually verified by a senior glaucoma expert.

6.1.2 Challenges of medical data

The characteristics of medical images themselves pose difficulties for the medical image analysis tasks:

There are many types and modalities of images that are used in medical image analysis. Similar to computer vision, the modalities include both 2D and 3D. However, the medical images have several other differences. Though the average scale of a medical image dataset is smaller than computer vision-related field datasets, the size of each sample of data is larger on average than the one of a computer vision-related field.

For 2D images, CR, WSI, and other modalities have large variances in the resolution and color than the other computer vision data. Some modalities might need more bits to encode a pixel, while some modalities are significantly huge. For example, CAMELYON 17 (238) only includes about a thousand of pathology images, but the whole dataset is about three terabytes. Such datasets with few large samples pose a challenge for the AI algorithms, and it requires the researchers to focus on designing algorithms that can learn from limited computational resources and be useful for clinical diagnosis.
For 3D medical images such as CT and MRI, they are dense 3D data, compared with sparse data, such as point cloud, in autopilot. Like the BraTS serial of challenges (30, 31, 32, 33, 34, 35, 36, 37, 38), many researchers face the challenges to design algorithms that can effectively learn from multi-modal dataset.

These characteristics of medical images require well-designed algorithms with a more robust capability to fit the data well and without overfitting. However, that further leads to the need for more data and resources. It is a challenge to learn suitable features from a small sample dataset.

6.2 No silver bullet

The ideal scenario is to find or invent a method or an algorithm to simultaneously solve all of these encountered problems. However, there is no silver bullet. The problems and challenges related to the data and the adopted methods cannot be entirely resolved, or sometimes, a problem arises as another is solved. Nevertheless, many ideas have been introduced to address the current problems, and they are introduced in this subsection.

With respect to the problems and challenges mentioned above, researchers are working on two research directions: 1) a more effective model with less data, and 2) a more practical approach to access data. For the learning methods with small datasets, researchers use approaches such as few-shot learning and transfer learning. In order to access more data, researchers adopt three main approaches, namely federated learning, lifelong learning, and active learning.

6.2.1 Practical learning from small samples

Many medical image datasets have a small number of samples. For example, challenge MRBrains13 (5) only includes 20 subjects for training and testing, while challenge KITS 19 (122) has about two hundred subjects. Therefore, many researchers struggle to find a practical approach to learn from small samples.

Few-shot learning and zero-shot learning

Few-shot learning hits one of the critical spots of DL-based medical image analysis problems, i.e., the development of DL models with fewer data[222]. Humans can effectively learn from few samples. Therefore, different from the standard deep learning-based methods, humans learn to diagnose a disease from images, without the need to view tens of thousands of images (i.e., from only few-shot). Meta-learning, also called learning to learn, is a technique used to address the few-shot learning problem[90]. Meta-learning can learn the meta-features from a small data size. The number of medical images in most datasets and challenges is not as large compared to the regular computer vision-related datasets and challenges. Mondal et al. [155] use few-shot learning and GAN to segment medical images. The GAN is modified for semi-supervised learning with few-shot learning. Similar to few-shot learning, zero-shot learning aims at novel samples. Rezaei et al. [175] cover a review of zero-shot learning from autonomous vehicles to COVID-19 diagnosis. However, zero-shot and few-shot learning have also their disadvantages, such as domain gap, overfitting, and interpretability.

Knowledge transfer:

Transfer learning is another method, which can recognize and apply knowledge and skills learned from a previous task[236]. For example, both white matter and gray matter segmentation and multi-organs segmentation are segmentation tasks[102]. However, the general neural network training is usually independent, which means that almost nobody trains a neural network with two tasks at once. However, it does not mean that these two tasks are unrelated. Besides zero-shot learning and few-shot learning, transfer learning, or say, knowledge transfer, is another method to infer knowledge from a previously learned task. Transfer learning can be applied to two similar tasks and between different domains[107]. The most significant advantage of transfer learning is that they use rich scale datasets to pre-train the neural network and then fine tune and transfer the network to the main task on a few samples dataset.

6.2.2 Effective access to more samples

Besides finding a practical approach to learn from small samples, many researchers have been working on active learning and federated learning (which aims to use data without access to sensitive information). This also reduces annotation costs of deep learning algorithms.
Federated learning: Federated learning provides another way to access data. As discussed previously, the limitation of accessing data is led by privacy and other problems. Instead of directly sharing data, federated learning shares the model to protect privacy from being leaked. With other privacy protection methods, federated learning can effectively use the data from each independent data center or medical center.

However, there are two disadvantages of federated learning: annotation and implementation. The problem of annotation cannot be solved by sharing data but other methods[52, 155, 186]. The main challenges are the design and implementation of a method that allow only feature sharing (needed by algorithms) and not the private information, as only a few institutions have attempted federated learning so far. For example, Intel and other institutions have attempted to apply federated learning for brain tumor-related tasks in their research [187]. The main challenges in their implementation include:

1) The implementation and proof of privacy protection,
2) The methodology for sharing and updating millions of the model’s parameters,
3) Preventing attacks on DL algorithms and leaks of data privacy on the Internet or computing nodes.

Natural language processing: Natural language processing is also a potential tool to automatically or semi-automatically annotate medical image data. It is a standard procedure for a medic to provide a diagnostic report of the patient, particularly after the medical image was taken. Therefore, such large amounts of data (image and text) is useful for medical image analysis after desensitization, and natural language processing can be used for annotation. Several natural language processing-based methods, e.g., [109, 130, 214] have been applied in medical-related research fields.

Active learning: Active learning aims to reduce the annotation cost by indirectly using the unlabeled data to select the “best” samples to annotate. Generally, data annotation for deep learning requires experts to label data so that the neural network can learn from the data. Active learning does not require too many samples at the beginning of training. In other words, active learning can “help” annotators to label their data. Active learning uses the knowledge learned from the labeled data to select and annotate the unlabeled data. The unlabeled data with annotation from algorithms is used to subsequently train the network over the next number of epochs. Active learning [52, 186] is used in the medical image analysis in a loop of 1) algorithm learn from the data annotated by humans, 2) human annotate the unlabeled data selected by the algorithm 3) algorithm add the newly labeled data to the training set. The advantage of active learning is obvious: annotators do not need to annotate all the data they have, and at the same time, the neural network can learn from data faster from such interactive progress.

7 Conclusion
In this work, we have provided a comprehensive survey of the datasets and challenges for medical image analysis, collected between 2013 and 2020. The datasets and challenges were categorized into four themes: head and neck, chest and abdomen, pathology and blood, and others. We provide a summary of all the details about these themes and data. We also discuss the problems and challenges of medical image analysis and the possible solutions to these problems and challenges.

Acknowledgments
We thanks for the projects of National Natural Science Foundation of China (62072358), Zhejiang University special scientific research fund for COVID-19 prevention and control, National Key R&D Program of China under Grant No 2019YFB1311600.

References
[1] 2002. Computed Tomography. John Wiley & Sons, Ltd, Chapter 15, 251–264. https://doi.org/10.1002/0471221155.ch15
[2] 2002. Instrumentation for Nuclear Imaging. John Wiley & Sons, Ltd, Chapter 12, 197–216. https://doi.org/10.1002/0471221155.ch12

ACM Comput. Surv.
[30] S. Bakas et al. 2018. Identification of Alzheimer’s disease via a convolutional neural network model based on T1-weighted magnetic resonance imaging. Scientific Reports 10, 1 (dec 2020), 1–17. https://doi.org/10.1038/s41598-020-70943-9

S. Bakas et al. 2018. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. arXiv (nov 2018). arXiv:1811.02629 http://arxiv.org/abs/1811.02629

S. Balocco et al. 2014. Standardized evaluation methodology and reference database for evaluating IVUS image segmentation. Computerized Medical Imaging and Graphics 38, 2 (mar 2014), 70–90. https://doi.org/10.1016/j.compmedimag.2013.07.001

V. S. Basava et al. 2020. ESAD: Endoscopic surgeon action detection dataset. arXiv (jun 2020). arXiv:2006.07164 http://arxiv.org/abs/2006.07164

B. B. Bejnordi et al. 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. JAMA - Journal of the American Medical Association 318, 22 (dec 2017), 2199–2210. https://doi.org/10.1001/jama.2017.14585

E. Bercovich and M. C. Javitt. 2018. Medical Imaging: From Reimann to the Digital Revolution, and Beyond. Rambam Maimonides Medical Journal 9, 4 (oct 2018). e0034. https://doi.org/10.1016/j.medj.2018.10.003

O. Bernard et al. 2016. Standardized Evaluation System for Left Ventricular Segmentation Algorithms in 3D Echocardiography. IEEE Transactions on Medical Imaging 35, 4 (apr 2016). 915–931. https://doi.org/10.1109/TMI.2015.2493090

O. Bernard et al. 2018. Deep Learning Techniques for Automatic MRI Cardiac Multi-Structures Segmentation and Diagnosis: Is the Problem Solved? IEEE Transactions on Medical Imaging 37, 11 (nov 2018), 2514–2525. https://doi.org/10.1109/TMI.2018.2837502

D. Bharti et al. 2021. CNN Variants for Computer Vision: History, Architecture, Application, Challenges and Future Scope. Electronics 10, 20 (2021). https://doi.org/10.3390/elecronics10204270

P. Blicic et al. 2019. The liver tumor segmentation benchmark (LTS). arXiv abs/1901.01 (2019). arXiv:1901.04056 http://arxiv.org/abs/1901.04056

B. N. Bloch et al. 2018. Large Datasets for Brain Segmentation: A Tutorial. IEEE transactions on medical imaging 37, 10 (2018), 2514–2525. https://doi.org/10.1109/TMI.2018.2837502

P. Bolei et al. 2017. Executing attention networks showed altered relationship with default mode network in PD. NeuroImage: Clinical 13 (2017), 1–8. https://doi.org/10.1016/j.nicl.2016.11.004

F. Brun et al. 2020. POCCOVNet: Automatic detection of COVID-19 from a new lung ultrasound imaging dataset (POCUS). arXiv (apr 2020). arXiv:2004.12084 http://arxiv.org/abs/2004.12084

J. Born et al. 2020. Accelerating COVID-19 differential diagnosis with explainable ultrasound image analysis. arXiv (sep 2020). arXiv:2009.06116

I. Bouzouev et al. 2019. ANHRI: Automatic Non-Rigid Histological Image Registration Challenge. IEEE transactions on medical imaging 39, 10 (2020), 3042–3052. https://doi.org/10.1109/TMI.2019.2986331

B. N. Bloch et al. 2018. Large Datasets for Brain Segmentation: A Tutorial. IEEE transactions on medical imaging 37, 10 (2018), 2514–2525. https://doi.org/10.1109/TMI.2018.2837502

M. Boulanger et al. 2021. Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review. Physica Medica 89 (2021), 265–281. https://doi.org/10.1016/j.ejmp.2021.07.027

E. E. Bron et al. 2015. Standardized evaluation of algorithms for computer-aided diagnosis of dementia based on structural MRI. The CADementia challenge. NeuroImage 111 (may 2015), 562–579. https://doi.org/10.1016/j.neuroimage.2015.01.048

M. Buda et al. 2020. Detection of masses and architectural distortions in digital breast tomosynthesis: a publicly available dataset of 5,060 patients and a deep learning model. (2020). arXiv:2011.07995 http://arxiv.org/abs/2011.07995

G. Campanella et al. 2019. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. Nature Medicine 25, 8 (2019), 1301–1309. https://doi.org/10.1038/s41591-019-0508-1

ACM Comput. Surv.
[179] S. Rueda et al. 2014. Evaluation and comparison of current fetal ultrasound image segmentation methods for biometric measurements: A grand challenge. *IEEE Transactions on Medical Imaging*, 33, 4 (2014), 797–813. https://doi.org/10.1109/TMI.2013.2376641

[180] M. Russe et al. 2017. Co-registration of pre-operative CT with ex vivo surgically excised ground glass nodules to define spatial extent of invasive adenocarcinoma on in vivo imaging: a proof-of-concept study. *European Radiology*, 27, 10 (2017), 4209–4217. https://doi.org/10.1007/s00330-017-4815-0

[181] D. Schiefe et al. 2006. Fluorescence in situ hybridization (FISH) on touch preparations: A reliable method for detecting loss of heterozygosity at 1p and 1q in oligodendrogliomas. *American Journal of Surgical Pathology*, 30, 7 (Jul 2006), 828–837. https://doi.org/10.1097/01.pas.0000213250.44822.2e

[182] K. Schmidmaier and M. Prah. 2018. Brain-Tumor-Progression. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* https://doi.org/10.1007/978-3-319-24571-3_7

[183] A. Seff et al. 2014. 2D view aggregation for lymph node detection using a shallow hierarchy of linear classifiers. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Vol. 8673 LNCS. 544–552. https://doi.org/10.1007/978-3-319-10404-1_68 arXiv:1408.3337

[184] A. A. A. Sertö et al. 2017. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: The LUNA16 challenge. *Medical Image Analysis* 42 (dec 2017), 1–15. https://doi.org/10.1016/j.media.2017.06.015 arXiv:1612.00112

[185] W. Shao et al. 2020. Deep active learning for nucleus classification in pathology images. In *Proceedings - International Symposium on Biomedical Imaging, Vol.* 2018-April. *IEEE*, 199–202. https://doi.org/10.1109/ISBI.2018.8363554

[186] M. J. Sheller et al. 2019. Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Vol. 11383 LNCS. 92–104. https://doi.org/10.1007/978-3-030-11723-8_9 arXiv:1810.04034

[187] Z. Shi et al. 2020. A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nature Communications*, 11 (dec 2020), 6090. https://doi.org/10.1038/s41467-020-19527-w

[188] D. Silver et al. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature*, 3, 1 (jul 2016), 507. https://doi.org/10.1038/nature16961

[189] K. Simonyan and A. Zisserman. 2015. Very deep convolutional networks for large-scale image recognition. In *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings.* arXiv:1409.1556

[190] A. L. Simpson et al. 2019. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv (feb 2019). arXiv:1902.09063 http://arxiv.org/abs/1902.09063

[191] R. Souza et al. 2018. An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage* 170 (apr 2018), 482–484. https://doi.org/10.1016/j.neuroimage.2017.08.021

[192] A. B. Spanier and L. Joskowicz. 2014. Rule-based ventral cavity multi-organ automatic segmentation in CT scans. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* Vol. 8669. 165–170. https://doi.org/10.1007/978-3-319-31972-2_15

[193] S. Steren et al. 2018. A radiomics approach based on the detection of salient features for diagnosis of breast cancer: A proof-of-concept study. *Medical Image Analysis* 42 (dec 2017), 1–15. https://doi.org/10.1016/j.media.2017.06.015 arXiv:1612.00112
A Systematic Collection of Medical Image Datasets for Deep Learning • 37

[221] X. Wang et al. 2017. ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Vol. 2017-Janua. 3462–3471. https://doi.org/10.1109/CVPR.2017.369 arXiv:1705.02315

[222] Y. Wang et al. 2020. Generalizing from a Few Examples: A Survey on Few-shot Learning. Comput. Surveys 53, 3 (apr 2020). https://doi.org/10.1145/3386252 arXiv:1904.05046

[223] D. Wei et al. 2020. MitoEM Dataset: Large-Scale 3D Mitochondria Instance Segmentation from EM Images. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12265 LNCS. 66–76. https://doi.org/10.1007/978-3-030-59722-1_7

[224] M. W. Weiner et al. 2017. The Alzheimer’s Disease Neuroimaging Initiative 3: Continued innovation for clinical trial improvement. Alzheimer’s and Dementia 13, 5 (may 2017). 361–571. https://doi.org/10.1016/j.jalz.2016.10.006

[225] Z. Xiong et al. 2021. A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging. Medical Image Analysis 67 (jan 2021), 101832. https://doi.org/10.1016/j.media.2020.101832

[226] J. Yang et al. 2018. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. Medical Physics 45, 10 (2018), 4568–4581. https://doi.org/10.1002/mp.13141

[227] X. Yang et al. 2014. Automated segmentation of the parotid gland based on atlas registration and machine learning: A longitudinal mri study in head-and-neck radiation therapy. International Journal of Radiation Oncology Biology Physics 90, 5 (dec 2014), 1225–1233. https://doi.org/10.1016/j.ijrobp.2014.08.350

[228] J. Yim et al. 2020. Predicting conversion to wet age-related macular degeneration using deep learning. Nature Medicine 26, 6 (jun 2020), 892–899. https://doi.org/10.1038/s41591-020-0867-7

[229] J. Zhontar et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. arXiv (nov 2018). arXiv:1811.08389 http://arxiv.org/abs/1811.08389

[230] L. Zhang et al. 2020. MeDaS: An open-source platform as service to help break the walls between medicine and informatics. arXiv (jul 2020). arXiv:2007.06013 http://arxiv.org/abs/2007.06013

[231] L. Zhang et al. 2020. Block Level Skip Connections across Cascaded V-Net for Multi-Organ Segmentation. IEEE Transactions on Medical Imaging 39, 9 (2020), 2782–2793. https://doi.org/10.1109/TMI.2020.2975347

[232] Q. Zhang et al. 2016. Deep learning based classification of breast tumors with shear-wave elastography. Ultrasoundics 72 (2016), 150–157. https://doi.org/10.1016/j.ultras.2016.08.004

[233] S. Zhang et al. 2020. Cognitive control of sensory pain encoding in the pregenual anterior cingulate cortex. d1 - decoder construction in day 1, d2 - adaptive control in day 2. https:// doi.org/10.18112/OPENNEURO.DS002596.V1.0.0

[234] B. Zhao et al. 2009. Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology 252, 1 (jul 2009), 263–272. https://doi.org/10.1148/radiol.2522081593

[235] J. Zhao et al. 2020. COVID-CT/Dataset: A CT image dataset about COVID-19. arXiv (mar 2020). arXiv:2003.13865 http://arxiv.org/abs/2003.13865

[236] F. Zhang et al. 2020. A Comprehensive Survey on Transfer Learning. arXiv:1911.02685 [cs.LG]

[237] X. Zhang, 2019. Multivariate Mixture Model for Myocardial Segmentation Combining Multi-Source Images. IEEE Transactions on Pattern Analysis and Machine Intelligence 41, 12 (2019), 2933–2945. https://doi.org/10.1109/TPAMI.2018.2869576 arXiv:1612.08820