European Journal of Radiology 169 (2023) 111159

Contents lists available at ScienceDirect

European Journal of Radiology

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

A scoping review of interpretability and explainability concerning artificial intelligence methods in medical imaging

Mélanie Champendal [a,b,*], Henning Müller [c,d], John O. Prior [b,e], Cláudia Sá dos Reis [a]

[a] School of Health Sciences HESAV, HES-SO, University of Applied Sciences Western Switzerland, Lausanne, CH, Switzerland
[b] Faculty of Biology and Medicine, University of Lausanne, Lausanne, CH, Switzerland
[c] Informatics Institute, University of Applied Sciences Western Switzerland (HES-SO Valais) Sierre, CH, Switzerland
[d] Medical faculty, University of Geneva, CH, Switzerland
[e] Nuclear Medicine and Molecular Imaging Department, Lausanne University Hospital (CHUV), Lausanne, CH, Switzerland

ARTICLE INFO

Keywords:
Computed Tomography
Magnetic Resonance Imaging (MRI)
Radiography
Machine Learning
Deep Learning
Transparency
Medical Imaging
Explainability

ABSTRACT

Purpose: To review eXplainable Artificial Intelligence/(XAI) methods available for medical imaging/(MI).
Method: A scoping review was conducted following the Joanna Briggs Institute's methodology. The search was performed on Pubmed, Embase, Cinhal, Web of Science, BioRxiv, MedRxiv, and Google Scholar. Studies published in French and English after 2017 were included. Keyword combinations and descriptors related to explainability, and MI modalities were employed. Two independent reviewers screened abstracts, titles and full text, resolving differences through discussion.
Results: 228 studies met the criteria. XAI publications are increasing, targeting MRI (n = 73), radiography (n = 47), CT (n = 46). Lung (n = 82) and brain (n = 74) pathologies, Covid-19 (n = 48), Alzheimer's disease (n = 25), brain tumors (n = 15) are the main pathologies explained. Explanations are presented visually (n = 186), numerically (n = 67), rule-based (n = 11), textually (n = 11), and example-based (n = 6). Commonly explained tasks include classification (n = 89), prediction (n = 47), diagnosis (n = 39), detection (n = 29), segmentation (n = 13), and image quality improvement (n = 6). The most frequently provided explanations were local (78.1 %), 5.7 % were global, and 16.2 % combined both local and global approaches. Post-hoc approaches were predominantly employed. The used terminology varied, sometimes indistinctively using explainable (n = 207), interpretable (n = 187), understandable (n = 112), transparent (n = 61), reliable (n = 31), and intelligible (n = 3).
Conclusion: The number of XAI publications in medical imaging is increasing, primarily focusing on applying XAI techniques to MRI, CT, and radiography for classifying and predicting lung and brain pathologies. Visual and numerical output formats are predominantly used. Terminology standardisation remains a challenge, as terms like "explainable" and "interpretable" are sometimes being used indistinctively. Future XAI development should consider user needs and perspectives.

1. Introduction

Artificial Intelligence (AI) is currently mainly deployed using deep learning (DL). It plays a crucial role in medical image analysis tasks such as detection, classification, diagnosis, segmentation, prediction and image quality enhancement [1,2]. Segmentation, classification, detection and diagnosis have distinct objectives in the analysis of medical images. Segmentation outlines regions of interest by selecting pixels that are part of a specific structure, classification assigns labels to images or image regions, detection locates specific objects or abnormalities and diagnosis proposes a medical pathology (sometimes differential diagnosis with probabilities for each) that can be used by a physician as something similar to a second opinion of a colleague [3]. Machine learning (ML) for these tasks employs trained algorithms or models to make decisions based on data. Deep learning is one the many approaches to machine learning. A distinction is often made between deep learning and classical machine learning, that often relies on handcrafted visual or textual features. These features are created based on expert knowledge

---

* Corresponding author.
E-mail addresses: melanie.champendal@hesav.ch (M. Champendal), henning.mueller@hevs.ch (H. Müller), john.prior@chuv.ch (J.O. Prior), claudia.sadosreis@hesav.ch (C.S. dos Reis).

https://doi.org/10.1016/j.ejrad.2023.111159
Received 25 July 2023; Received in revised form 26 September 2023; Accepted 19 October 2023

using statistical models and are often simple to understand, for example average gray level. In contrast, DL, a branch of machine learning and neural networks, does not require this manual feature definition and extraction step by automatically extracting discriminative features from the multiple layers of interconnected neurons [4]. Their complexity and interpretability remain a challenge, as they are often viewed as a "black box", lacking transparency in the decision-making process [5]. Without clear explanations of how and why a particular decision or output was reached, medical imaging professionals (such as physicians or radiographers), may hesitate to fully rely on the results generated by these algorithms. AI is still struggling to be applied in medical routine because its users do not fully trust it [6]. Therefore, enhancing the understandability of DL models is essential to facilitate the seamless integration into clinical practice [5].

Explainable artificial intelligence (XAI) and interpretable artificial intelligence aim to address this interpretability challenge. XAI techniques provide justifications, explanations and insights into how AI models arrive at specific decisions, with the objective to instill trust, transparency, and confidence in the system. Interpretable AI focuses on designing transparent models that offer human-readable representation of their decision-making process [7].

In the field of medical imaging, there are variations in the terminology used, especially with the terms "explainable" and "interpretable". These terms are often used indistinctively; however some authors may have varying meanings for them depending on the domain and context [8]. Interpretable models are generally understandable by the user without additional tools or methods, while explainable models use supplementary techniques [8].

In addition to terminology, taxonomy also varies slightly from one author to another. To comprehensively explore explainable and interpretable AI in medical imaging, it is important to consider various dimensions: *stage*, *scope*, *input format*, *output format* and the *problem type* that needs to be explained [9–12].

- *Stage*: The term *stage* refers to the specific moment at which a method produces explanations. A distinction is made between *ante-hoc* or *model-based* methods and *post-hoc* methods. *Ante-hoc* methods refer to techniques that aim at designing naturally interpretable and transparent models. These methods are often better suited to simpler AI algorithms such as classical machine learning. On the other hand, *post-hoc* methods focus on explaining the decisions made by often more complex models such as DL after they were trained. Post-hoc models can be either specific or agnostic. Ante-hoc are usually specific by nature.
  - *Model specific* techniques are tailored to a particular AI model or algorithm. These explanations are designed to provide insights into the inner workings of a specific model, making use of its internal structure, parameters and decision-making process.
  - *Model agnostic* techniques aim to explain the decisions of any black-box model, regardless of its specific architecture. These explanations sometimes focus on the inputs and outputs of the model rather than its internal workings but can also use the internal functions for the explications. The aim is to provide a general understanding of the model's behavior and highlight the factors that contribute to its predictions or decisions.
- *Scope*: The models can give either a *local* or a *global* explanation. *Local* focus on understanding the decision-making process of DL models at a case or individual level. While a *global* explanation aims to understand the overall behavior and functioning of DL models across a dataset or a population.
- *Input format:* The input format for medical imaging is mainly images, but these can be accompanied by other data such as text, numerical values, and categorical data. The type of format can have an impact on the choice of explanatory method, for example if DICOM with a full grey-scale resolution can be used or not.
- *Problem type:* AI can be used for a variety of medical imaging tasks, such as detection, classification, diagnosis, segmentation, prediction, and image quality enhancement. Depending on the task, the explanatory model may also change to be the most appropriate.
- *Output format:* There are many different output formats that fall into several categories. These categories vary according to the authors. Van der Velden et al. [9] use three output formats: visual, textual and example based. Vilone and Longo [10] categorize the explanations into visual, textual, numeric, ruled-based and mixes of these 4 possibilities. Adadi and Berrada [12] applied 4 techniques: visual, knowledge extraction, influence methods and example based. Borys et al. [13] employed a different classification scheme that encompasses various categories such as visual, textual, case-based, and auxiliary formats. The case-based category aligns with the example-based approach described by Addi and Berrada, whereas the auxiliary category predominantly encompasses numerical formats, including scores and quantifications of uncertainty. However, there are a few methods that are not classified identically by the authors, such as Testing with Concept Activation Vectors (TCAV), classified into the case-based category by the latter while Bas et al. [9] classify it in the text category. Likewise, T-distributed stochastic neighbour embedding (t-SNE) map is classified in the visual categories by Villone [10] and the t-SNE plot in the auxiliary categories by Borys [13]. These diverse categories represent a comprehensive and evolving taxonomy that, may also undergo variations depending on the specific context of application.

In the literature related to XAI, all articles share a common goal of providing comprehensive insights into the field, along with a definition of taxonomy. However, the published studies differ significantly in both their methodologies and the domains of application they explore. Several researchers, such as Vilone and Longo [10], Adadi and Berrada [12] and Islam [14], have engaged in systematic reviews encompassing a wide range of domains, including healthcare, finance, the military, transportation, law, human–machine interaction, genetics, aviation and many others. They aim to categorize the output formats, including numerical, rule-based, textual, visual, and mixed formats. In contrast, Kök [15] focused specifically on the healthcare domain. Nevertheless, these studies diverge significantly in terms of data sources. Vilone and Longo [10] solely relied on Google Scholar, whereas Adadi and Berrada [12] expanded their scope by incorporating data from databases such as SCOPUS, IEEExplore, ACM Digital Library and others. Adadi and Berrada provided detailed information about their data sources and methodologies, whereas others, such as Yang et al. [16], Kök et al. [15], and Borys et al. [13,17], did not disclose their methodologies. On the other hand, Van der Velden et al. [9] conducted a comprehensive review within the SCOPUS database. Furthermore, each analysis had its unique scope. Groen et al. [18] exclusively focused on computer-assisted diagnosis in radiology, Borys et al. conducted two reviews in XAI methods in medical imaging, initially spanning various techniques [13] and then specifically focusing on saliency-based (visual) methods [17]. Van der Velden concentrated on deep learning-based medical image analysis, while Vilone [10] and Adadi and Berrada [12] explored a broader range of applications beyond medical imaging. These articles contribute valuable insights to the field of XAI, highlighting the diversity of approaches and the adaptability of techniques across multifaceted domains.

This study aligns with the same approach and aims to map the existing literature on explainable and interpretable AI in medical imaging, exploring the techniques, methods and approaches used to improve interpretability. By synthesizing a wide range of studies, this review provides an overview of XAI development, application, and interpretable AI methods, highlighting the progress and challenges related to the transparency and understanding of AI systems as well as the gaps that still need to be further worked on to identify directions for future research.

2. Methodology

A scoping review aims at determining the scope comprehensiveness of existing literature on a given topic in a structured way to address research questions and identify gaps, offering a concise portrayal of the quantity of literature, highlighting its central themes [19].

This scoping review was conducted according to the Joanna Briggs Institute methodology for scoping reviews [20] and the reporting guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [21].

2.1. Eligibility criteria

This scoping review was conducted to summarize the state of evidence of the methods used to explain artificial intelligence to its users (physicians and radiographers considered as end-users). A range of studies was included, examining several designs applied to radiology, nuclear medicine and radio-oncology.

The terminology relating to the explanation of artificial intelligence is relatively broad, with two terms being often used without distinction: "explainable" and "interpretable" [8], as referred before, and therefore both terms were included in the search for this review.

Quantitative, qualitative, and mixed peer-reviewed studies were included, while systematic reviews, guidelines, book sections and editorials were excluded in the search task. Some reviews and their methods are described in the literature review.

2.2. Search strategy

The search strategy (Appendix A) includes both published and unpublished primary studies in seven databases: Pubmed, Embase, Cinhal, Web of Science, BioRxiv, MedRxiv and Google Scholar in October 2022. A combination of keywords and Medical Subject Heading (MeSH) terms related to the terminology of the concept of explainability and modalities or fields from medical imaging were used.

No keywords or MeSH terms related to users were included in the search strategy as they were related to the imaging modalities already included, and it would introduce noise into the results as observed without distinction after a first attempt. Studies in English or French were included.

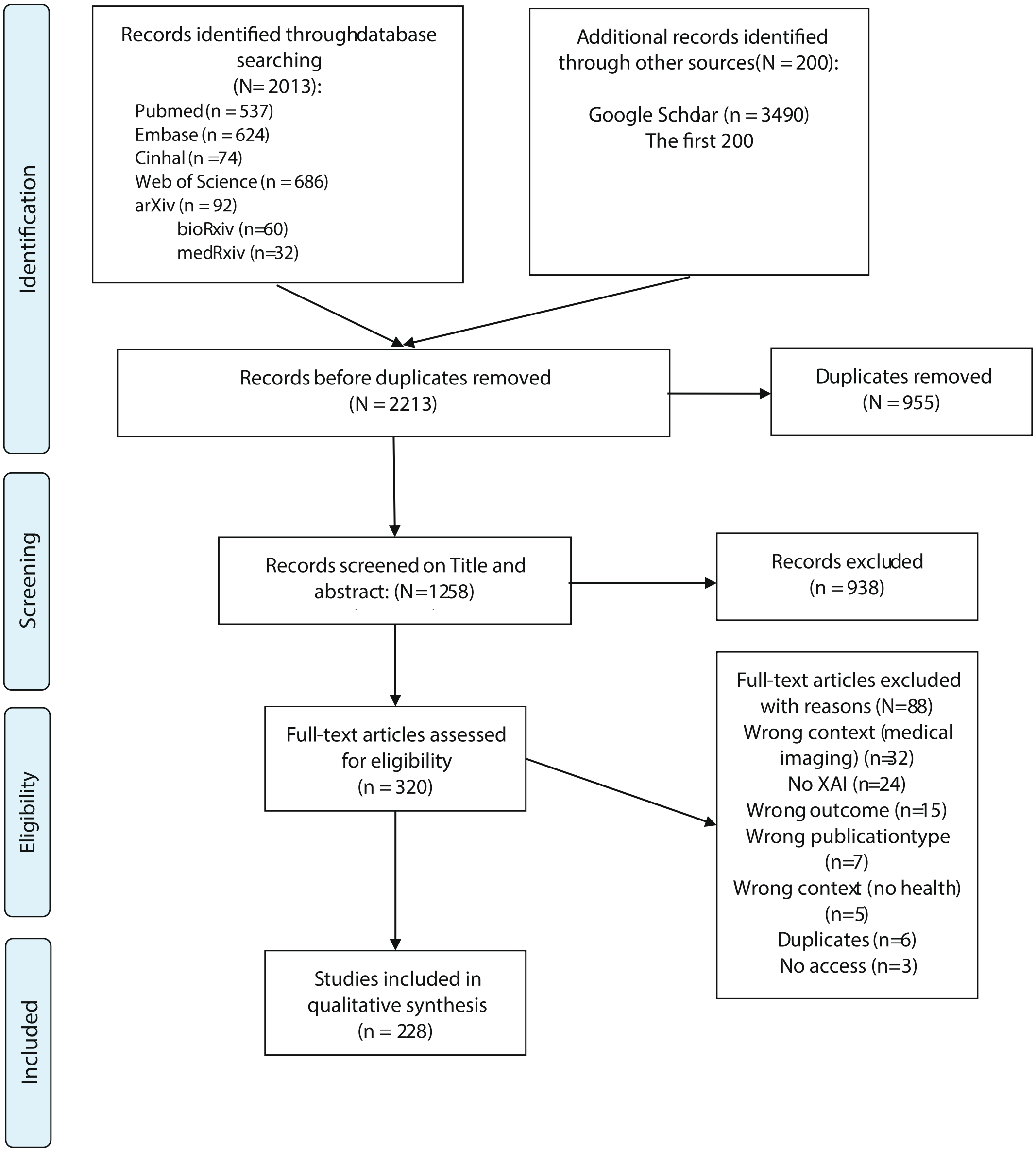


Fig. 1. Search results and study selection and inclusion [23].

2.3. Study selection

All identified studies were uploaded to EndNote 20 and duplicates were eliminated using Bramer's method [22]. Subsequently, the references were imported into Rayyan, a free web-tool, to facilitate study selection. In the first round, titles and abstracts were screened by two independent reviewers to assess their relevance based on the previously described criteria. Full-text articles that met the inclusion criteria were retrieved and reviewed by the same two reviewers in a second round. Full-text studies that did not meet the inclusion criteria were excluded and the reasons for their exclusion were presented (Fig. 1). Any disagreements between the reviewers were resolved through discussion.

2.4. Data extraction and analysis

Data were extracted using an extraction table created by the authors based on the following characteristics: reference/authors, year, country, imaging modality, organs, pathology, sample, AI task, stage (specific or agnostic/ante- or post hoc), scope (local or global), input format, output format, AI terminology used through the article (explainable, interpretable, transparent, understandable, reliable and intelligible).

This scoping review uses the output formats best suited for the medical imaging context namely: visual, numeric, textual, ruled-based, example-based and a mix of these 5 categories.

A descriptive analysis with a narrative summary was performed to present the results.

3. Results

3.1. Search and study selection

After removing duplicates, 1,258 results were identified by the search strategy and 228 studies met all criteria being included for further analysis. The two main reasons for exclusion were wrong context: a medical imaging context not linked to radiology, nuclear medicine or radio-oncology or that no XAI tool was applied (Fig. 1).

3.2. Years

An increase in the XAI-related articles between 2018 and 2022 was observed (Fig. 2). The number of articles has increased each year over the four-year period, with the largest increase (2.3 times more) in 2020 and 2021. In 2018, there were 4 included articles dedicated to XAI, while in 2022, the number of publications reached 94.

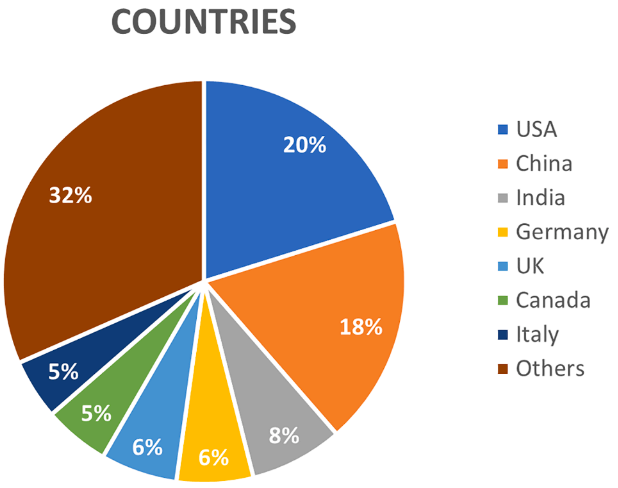


Fig. 3. Countries where studies about XAI were produced (based on first author affiliation).

3.3. Countries

The included studies were produced in 39 countries, with most first authors based in the United States of America (USA) (n = 46; 20 %) and China (n = 42; 18 %). Other countries that contributed with a noticeable number of articles were India (n = 17), Germany (n = 14), the United Kingdom (n = 14), Canada (n = 12) and Italy (n = 11). 32 countries, classified as others, contributed also to this field at a lower level (Fig. 3).

3.4. Imaging modalities

The most frequently (32 %) studied modality with XAI was Magnetic Resonance Imaging (MRI) (n = 73), followed by radiography in 20 % (n = 47) and Computed Tomography (CT) (n = 46). It should be underlined that 12 papers proposed the use of XAI for two imaging modalities [24–35], 4 of which were related to CT and radiography [24,25,27,34]. The area of radiation oncology remains underexplored with XAI, with only one study focusing on radiotherapy cone-beam CT (CBCT) for Image-guided radiotherapy (IGRT) [36] (Fig. 4).

3.5. Anatomical regions/Organs & pathologies

Lungs (n = 82) and brain (n = 74) are the two most frequently studied anatomical regions in the studies included in this scoping review

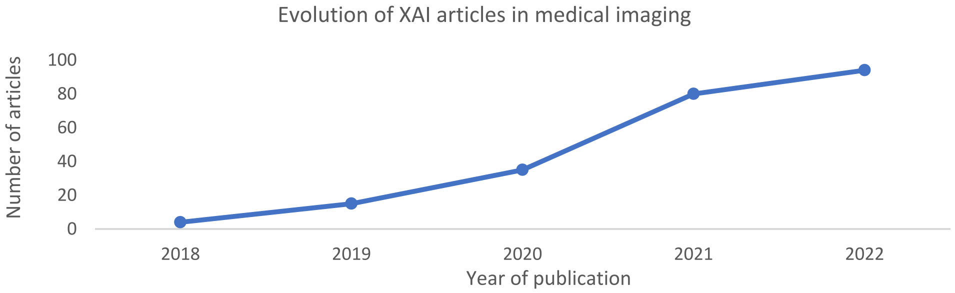


Fig. 2. Number of articles published in the last 5 years using an XAI tool for medical imaging (radiology, nuclear medicine or radio-oncology).

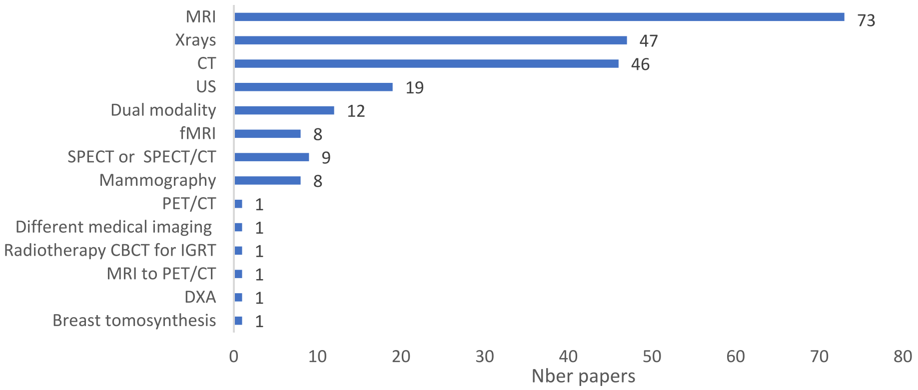


Fig. 4. Number of papers published per imaging modality.

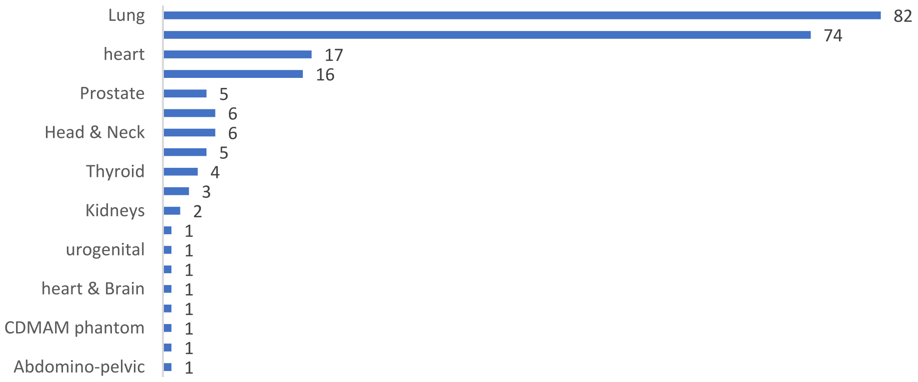


Fig. 5. Number of published papers about XAI per anatomical regions/organs. Of the pathologies studied, Covid-19 was the most frequently analysed (n = 48), followed by Alzheimer's disease (n = 25) and brain tumors (n = 15).

(Fig. 5). The heart (n = 17), breasts (n = 16), prostate (n = 6), liver (n = 6), ear, nose and throat (ENT) regions (n = 6), bone (n = 5) or thyroid (n = 4) are less frequently studied.

3.6. Journals of the articles included in the review

The articles included in this scoping review were classified according to whether they were conference papers (n = 44/228; 19.3 %) or journal articles (n = 182/228; 79.8 %) in 120 different journals. Only 2 articles had not been published at the time of the study but were accepted for publication. The journals were grouped according to their scope, including Computer Science (71/228; 31.1 %) (e. g., Computers in Biology and Medicine or Neural Networks), Medical Imaging (42/228; 18.4 %) (e. g., European Radiology, Neuroradiology or Journal of Magnetic Resonance Imaging), General (39/228; 17.1 %) (e.g. Frontiers in Medicine, IEEE Access Journal or mdpi/diagnostics), Medical specialties other than radiology or pathology (28/228; 12.3 %) (e.g. Frontiers in Neuroscience, Epilepsia Open or Journal of Orthopaedic Research) and 2 in other journals.

3.7. Datasets

The datasets used in the different articles varied in number, provenance and composition. More than half were publicly accessible datasets (58.6 %), 39.7 % were locally acquired and 1.7 % were composed of both public and local data. Table 1 lists the public data by pathology. Local datasets were mainly monocentric data (n = 82/91), 8 were acquired from 2 sites and 1 in 3 sites. Most studies used images only (n = 199), 30 studies included clinical data, laboratory results and text reports.

3.8. AI tasks or problem type

The tasks performed by the algorithms and explained by XAI are illustrated in Fig. 6. The results indicate that classification (n = 89) is the most frequent task performed by the algorithms, followed by prediction (n = 47), diagnosis (n = 39), detection (n = 29), segmentation (n = 13) and image quality improvement (n = 6).

3.9. XAI output format

Five output formats and the possibility of combining two or more of these formats were selected to explain the decision-making patterns of the algorithms (Fig. 7). The visual format was the most frequently applied format (n = 186/228), followed by the numerical format (n = 67), the rule-based (n = 11), the textual (n = 11) and the example-based (n = 6). In 45 papers, the authors combined 2 formats (n = 38), while others combined 3 (n = 6) and 4 (n = 1). When two formats were used, in 66 % (n = 25/38) of the cases, the visual and numerical combination was used. 5 of the 6 articles mixing 3 output formats used the visual,

Table 1
Public datasets used ordered by pathologies.

Pathologies	Public datasets used
Covid-19 (XRays) [30–49]	Kaggle repository (COVIDx, COVID-19 RADIOGRAPHY, covid-19-X-ray-10000-images, covid19-chest-xray-image-dataset, largest-covid19-dataset), Github repository (chestxray-dataset), COVID19 + from the Medical Imaging Databank in Valencia Region Medical Image Bank (BIMCV), RSNA International COVID-19 Open Radiology Database (RICORD), Chest X-ray Images Pneumonia (CXRIP), The MONTGOMERY County CXR dataset, The SHENZHEN Hospital X-ray dataset, the National Institute of Health (NIH) Chest X-ray dataset, COVID-19 Image Data Collection (CIDC), COVID-19-positive radiographs from the GitHub-COVID repository, COVID-19-negative radiographs from the ChestX-ray14 repository of the National Institutes of Health (NIH), COVID dataset by v7-Darwin labs, CheXpert, Italian Radiological Case CASE + Radiopaedia.org COVIDx v3.0, Covid-GAN, Covid-Net mini Chest X-ray, BIMCV, COVID-19-NY-SB, Curated Dataset for COVID-19 Posterior-Anterior Chest Radiography Images (X-Rays) from Mendeley, COVID-19 image data collection from Cohen and COVID-19 dataset originated from the QUIBIM imagingcovid19 platform database, IEEE, RadioGyan and the British Society of Thorathic Imaging
Covid-19 (CT) [17,18,20,27,50–62]	Kaggle repository (SARS-CoV-2 CT Scan, COVID19-CT dataset), Signal Processing Grand Challenge on COVID-19 dataset (SPGC-COVID), COVIDx CT, COVIDx CT-2A & COVIDx CT-2B, CC-CCII, MosMedData, COVID-Ctset, LTRC dataset, CT Chest Images Dataset from Mendeley, COVID Ayademic, iRoads, Caltech-256, Caltech-101
Covid-19 (US) [70]	POCOVID and POCUS Atla platform
Pneumonia (XRays) [71–79]	RSNA Pneumonia Detection Challenge dataset, Chest X-ray Images Pneumonia (CXRIP), MIMIC dataset and a subset of NIH dataset
Lung cancer (CT) [80–83]	Lung Image Database Consortium (LIDC), LUNA 16
Lung abnormalities (XRays) [84–88] (CT) [89]	CheXpert, HUM-CXR, VinDr-CXR, MNIST handwritten digit database and COPDGene dataset
Pneumothorax (XRays) [90]	The CANDID-PTX (Chest X-ray Anonymised Dataset In Dunedin–Pneumothorax) dataset and SIIM-ACR dataset
Pediatric pulmonary health (XRays) [91]	Dataset from Italy with Covid-19 CXR in github, small pediatric dataset with pneumonia, small chest X-ray Tuberculosis image dataset from kaggle
Pulmonary embolism (CT) [92]	Kaggel dataset RSNA STR Pulmonary Embolism
Cardiac health/conditions (MRI) [93,94] (SPECT-Mibi) [95,96]	UK Biobank, REgistry of Fast Myocardial Perfusion Imaging with NExt generation SPECT (REFINE SPECT) registry, Open-source challenge MICCAI 2017 Bernard
Alzheimer's disease (MRI) [97–116] (MRI & PET FDG) [30,117]	Alzheimer's Disease Neuroimaging Initiative dataset (ADNI), Australian dataset (AIBI), Open Access Series of Imaging Studies (OASIS), TADPOLE Challenge organizers, NACC, NIFD, Parkinson's Progression Markers Initiative (PPMI), FHS dataset, T1 weighted MR dataset from Kaggle
Parkinson's Disease (MRI) [118] (SPECT-Datscan) [119,120]	Parkinson's Progression Markers Initiative (PPMI)
Autism Spectrum Disorder (MRI) [121,122]	Autism Brain Imaging Data Exchange (ABIDE)
Cognitive tasks (fMRI) [123]	The Adolescent Brain Cognitive Development (ABCD)
Gliomas / Brain tumours (MRI) [117–128]	Brain Tumor Segmentation (BraTS) challenge 2013, 2018 and 2020 databases, ADNI, Brain Tumor Classification (MRI) from Kaggle, TCGA dataset from The Cancer Imaging Archive repositories, The brain figshare MRI dataset, diffusion datasets (Human connectome project dataset, Lifespan dataset, Prisma dataset and Pathology dataset), IXI dataset, European CyberKnife Center in Munich
Brain diseases (CT) [136,137]	CQ500 and RSNA data sets
Brain cognitive state (fMRI) [138]	The S1200 release of the Human Connectome Project and BOLD5000 dataset
Cerebral haemorrhages (CT) [139]	Felipe Kitamura's CT dataset from the RSNA 2019 brain CT hemorrhage challenge
Prostate lesions (MRI) [140] (MRI & US) [29]	The PROSTATEx dataset and open-source database at the Cancer Imaging Archive
Breast cancer & calcifications (Mammography) [141–147]	CBIS-DDSM dataset, NYU Breast Cancer Screening Dataset v1.0, Database for Screening Mammography (DDSM), Dataset of breast ultrasound image, INbreast dataset, BUSIS and BUSI
Age prediction (XRays-bone) [148] (XRays-chest) [149] (MRI-brain) [150–152]	Cambridge Centre for Aging and Neuroscience (Cam-CAN), NIH Chest X-ray, Consortium for Reliability and Reproducibility (CoRR), Alzheimer's Disease Neuroimaging Initiative (ADNI), Brain Genomics Superstruct Project (GSP), Functional Connectomes Project (FCP), Autism Brain Imaging Data Exchange (ABIDE), Parkinson's Progression Markers Initiative (PPMI), International Consortium for Brain Mapping (ICBM), Australian Imaging, Biomarkers and Lifestyle (AIBL), Southwest University Longitudinal Imaging Multimodal (SLIM), Information extraction from Images (IXI), Open Access Series of Imaging Studies (OASIS), Consortium for Neuropsychiatric Phenomics (CNP), Center for Biomedical Research Excellence (COBRE), Child and Adolescent NeuroDevelopment Initiative (CANDI) and Brainomics, new dataset CVM-900, The 2017 Pediatric Bone Age Challenge dataset from the Radiological Society of North America (RSNA)
Knee injury (MRI) [153]	MRNet data set, Chiba and Stanford dataset, fastMRI dataset
Sex classification (MRI-brain) [154,155]	The Human Connectome Project (HCP), The Brain Genomics Superstruct Project (GSP), The enhanced Nathan Kline Institute-Rockland Sample (NKI-RS), The Consortium for Reliability and Reproducibility (CoRR) and Southwest University Longitudinal Imaging Multimodal dataset, ABIDE, APCI, COBRE-MIND, Tulsa 1000
Blockin for regional anesthesia (US) [156]	Nerve-UTP Nerve segment dataset from the Kaggle Competition repository
Osteoarthritis (MRI & XRays) [31]	MOST study

numerical, and textual formats to explain the algorithm decision. Appendix B contains comprehensive references categorized by imaging modalities and output format.

A visual explanation is mainly used to explain classifications (n = 82, Fig. 8) presenting saliency maps, attention maps or heat maps. The most frequent XAI tool used is class activation mapping (CAM) (n = 96) and its different extensions like Grad-CAM, Grad-CAM++, Guided Grad-CAM, Score-CAM, FasterScore-CAM, Vanilla CAMHR-CAM, HAM or GLAM. A few visual explanation used were Layer-wise Relevance Propagation (LRP) (n = 18), attention-based mechanisms (n = 18), Local Interpretable Model-Agnostic Explanations (LIME) (n = 15) and other perturbation-based or surrogate models (n = 4), Integrated Gradients (e.g., Oriented modified integrated gradients: OMIG) (n = 9), saliency maps (n = 8), SHapley Additive exPlanations (SHAP) (n = 8), attention weights (n = 4), Deep-Taylor Decomposition (n = 4), Randomized Input Sampling for Explanation (RISE) (n = 4), guided backpropagation (n = 3), GSInquire (n = 3), Occlusion Sensitivity (n = 3), Concept activation vectors (n = 3), multilayer perceptrons (MLP) (n = 2), keras-vis (n = 2), t-Distributed Stochastic Neighbor Embedding (t-SNE) (n = 1), Jacobian map (n = 1), Explainable Boosting Machine (EBM) (n = 1), Uniform Manifold Approximation and Projection (UMAP) (n = 1), depth map-based (n = 1), Explainable and Simplified Image Translation (ESIT) (n = 1), pulse-coupled neural network (m-PCNN) (n = 1), genetic

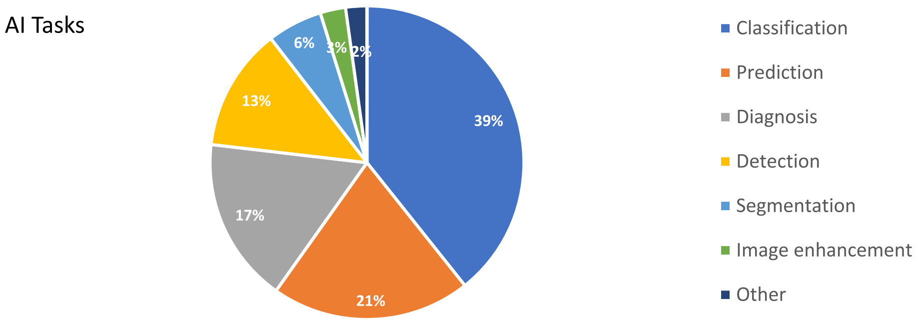
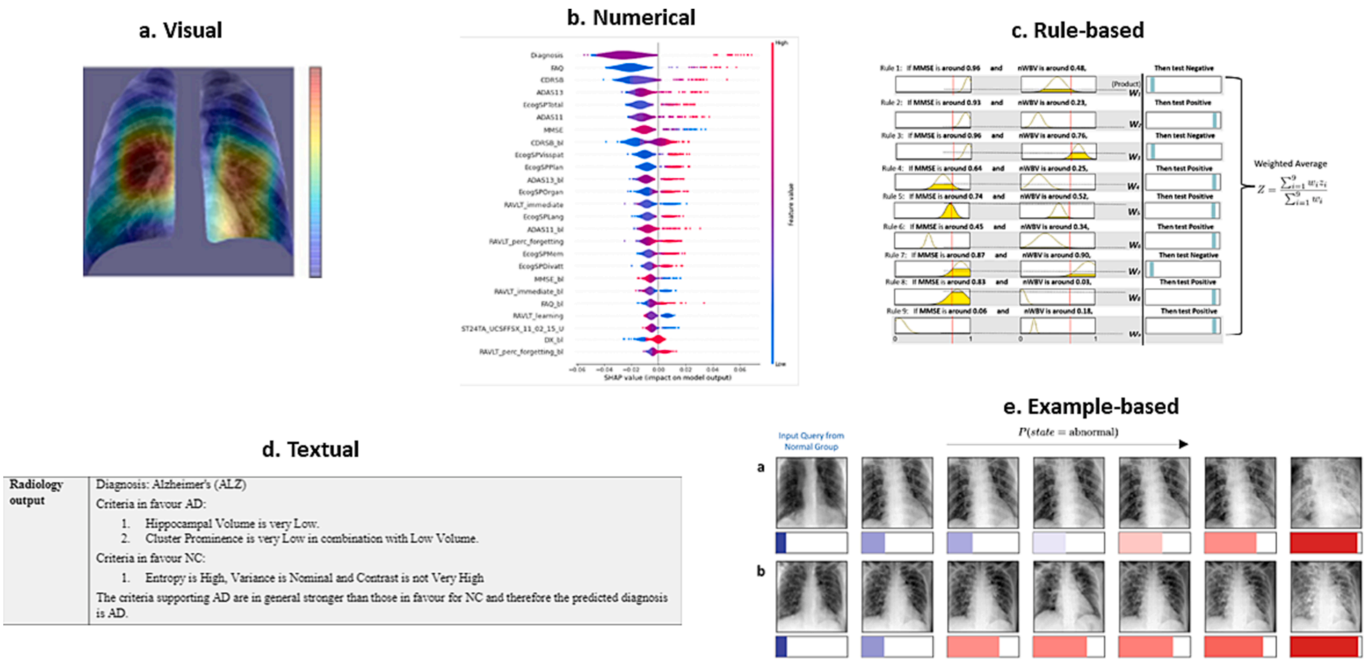


Fig. 6. Number of papers per AI task (total N = 229 papers).



Radiology output	Diagnosis: Alzheimer's (ALZ) Criteria in favour AD: 1. Hippocampal Volume is very Low. 2. Cluster Prominence is very Low in combination with Low Volume. Criteria in favour NC: 1. Entropy is High, Variance is Nominal and Contrast is not Very High The criteria supporting AD are in general stronger than those in favour for NC and therefore the predicted diagnosis is AD.

Fig. 7. Output format examples: a. Visual: Heatmap visualisation for pneumonia with GradCAM [52], b. Numerical: The violin plots show the SHAP value impact on the estimation of the probabilities for Alzheimer's Disease [30], c. Rule-based: An illustration in the context of making the rules used for 2 characteristics in a clinical decision-support diagnosis of dementia transparent [157], d. Textual: Gorgias argumentation theory for the assessments of Alzheimer's disease [97], e. Example-based: TraCE generates counterfactual explanations based on diagnoses by incrementally incorporating pertinent patterns into various query images belonging to healthy individuals, thereby enhancing the probability of their classification into the abnormal category. [77].

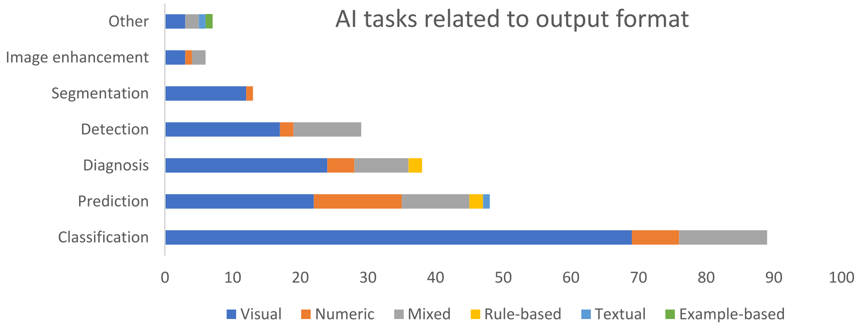


Fig. 8. Type of explanations per AI tasks.

algorithm based brain masking (GABM) (n = 1) or DeepLIFT (n = 1). Ten articles did not describe the model or used a home-made model. In some articles, the authors combined and compared the different methods of visual explanations (n = 31).

Numerical explanations are used for different tasks and mainly for predictions and classifications (Fig. 8). SHAP (n = 31) and, LIME (n = 15) are the two most used frequently techniques to provide numerical explanations through plots & violin plots (SHAP) and pareto chart (LIME). Testing with Concept Activation Vectors (TCAV) (n = 3), semantic interpretability score (SIS) (n = 3), EBM (n = 2), training

Table 2
Scope of the explanation according to the output format.

	Classification (n = 89)	Detection (n = 29)	Segmentation (n = 13)	Diagnosis (n = 38)	Prediction (n = 48)	Image Enhancement (n = 6)	Others (n = 5)
Global (n = 13)	Visual = 2; Numeric = 2; Mix = 1	Numeric = 1	Visual = 1; Numeric = 1	Rule-based = 1	Numeric = 3; Rule-based = 1	–	–
Local (n = 178)	Visual = 65; Numeric = 1; Mix = 7	Visual = 16; Numeric = 1; Mix = 10	Visual = 8	Visual = 24; Numeric = 2; Rule-based = 1; Mix = 5	Visual = 21; Numeric = 1; Rule-based = 1; Textual = 1; Mix = 4	Visual = 2; Numeric = 1; Mix = 2	Visual = 3; Mix = 2
Both (n = 37)	Visual = 2; Numeric = 4; Mix = 5	Visual = 1	Visual = 3	Numeric = 2; Mix = 3	Visual = 1; Numeric = 9; Mix = 6	Visual = 1	–

calibration-based explainers (TraCE) (n = 1), eNetXplorer (n = 1), Accumulated Local Effects (ALE) (n = 1), and Friedman's H-statistic (FHs) (n = 1) are other examples less frequently used. Other human-interpretable explanations including feature importance, similarity score, confidence score or probability score/risk were also employed (n = 18).

Rule-based explanations were employed to explain the diagnosis, the predictions and the classifications made by algorithms (Fig. 8). Decision trees (n = 3), Bayesian Networks (n = 2), Gorgias argumentation theory (n = 1), fuzzy rule-based models (n = 1), Logit Boost Models (LBM) (n = 1), prototypes with logical rules (n = 2) were the rule-based tools used. One article used a home-made approach, meaning one without a specific name.

Textual explanations are mainly useful for prediction tasks (Fig. 8). In 91 % (n = 10/11) of the cases it is used in combination with other types of formats. Concept activation vectors (n = 2), semantic features (n = 4) and different human-interpretable explanation are used for textual explanations (n = 5).

Example-based explanations are the least used in the articles (n = 6) and include prototypes (n = 2), counterfactuals (n = 1), patch similarities (n = 1), content-based image retrieval (CBIR) (n = 1), or TraCE (n = 1). This output format is always combined with other output formats.

Mixed explanations were used in four different tasks: classification, detection, prediction, diagnosis, but not for segmentation and image enhancement (Fig. 8).

3.10. *Scope (global or local)*

The scope of the explanation in the studies was *local* in 215 cases and aimed to explain a given input or sample. The explanation format was typically single (n = 178) with the majority being visual (n = 139) and numeric (n = 6). When the output format was mixed (n = 30), numeric and visual explanations were frequently combined (n = 16) or using triple format of explanations namely visual, numerical and textual types (n = 4).

In 50 articles, the explanation focused on a common pattern across the population and thus had a more global focus. For the article applying a global scope (n = 13), SHAP was the most frequently used (n = 6) with numerical explanations (n = 7), followed by visual (n = 3) and rule-based (n = 2) and one article with a mixed output format (numerical and visual).

Dual scope was used in 37 articles, also with SHAP as the most frequently employed, sometimes (n = 16) alone, other times (n = 7) combined with another approach. When the scope was both global and local, the output format was mainly numeric (n = 28), followed by visual (n = 20), rule-based (n = 3), or textual (n = 3) (Table 2).

Classification was mostly explained locally (82 %), using both scope (12 %) and globally only (6 %). Detection was also mostly explained locally (93 %), as well as diagnosis (84 %), Image enhancement (83 %), segmentation (62 %) and prediction (58 %) (Table 2).

3.11. *Stage (ante-hoc, post-hoc specific or agnostic)*

The models used in the studies were mostly *post-hoc* (n = 222), with

Table 4
Terminology used in the articles.

terminology	Number papers
Explainable (E)	207
INTERPRETABLE (IR)	187
UNDERSTANDABLE (U)	112
TRANSPARENT (T)	61
RELIABLE (R)	31
INTELLIGIBLE (IL)	3

Table 3
Output format and AI task according to Stage Model.

Stage Model		Output	#	AI Task						
				Clas	Det	Seg	Diag	Pred	ImgE	Others
Specific	Ante	Visual	5	2	1	–	2	–	–	–
		Numeric	2	1	1	–	–	–	–	–
		Rule-based	1	–	–	–	–	1	–	–
		Mix	6	2	–	–	1	1	2	–
	Post-hoc	Visual	122	59	12	8	19	19	2	3
		Numeric	2	–	–	–	2	–	–	–
		Rule-based	2	–	–	–	1	1	–	–
		Textual	1	–	–	–	–	1	–	–
		Mix	18	6	5	–	3	2	–	2
Agnostic	Post-hoc	Visual	24	9	4	4	3	3	1	–
		Numeric	24	6	1	1	2	13	1	–
		Rule-based	1	–	–	–	1	–	–	–
		Mix	28	7	9	–	5	7	–	–

Clas - classification; Det - detection; Seg - Segmentation; Diag - Diagnosis; Pred - prediction; ImgE - Image Enhancement; Others

Table 5
Combination of terms used in articles.

Nbr terms used	Nbr papers (combination of terms)
1	***34*** *(E = 23, IR = 10, T = 1)*
2	***70*** *(E + IR = 50, E + U = 10, IR + U = 4, IR + R = 3, IR + T = 2, E + R = 1)*
3	***75*** *(E + IR + U = 49, E + IR + T = 16, E + IR + R = 4, E + T + U = 3, E + T + R = 2, IR + T + U = 1)*
4	***39*** *(E + IR + T + U = 26, E + IR + U + R = 10, E + IR + T + R = 2, E + IR + U + IL = 1))*
5	***9*** *(E + IR + T + U + R = 7, E + IR + U + R + IL = 2)*
6	***1***

77 agnostic, 145 specific and 14 model-based explanations. Eleven papers combined both types of post-hoc models, using several explanatory tools, usually one type of the CAM models (post specific) with another post-agnostic model (LIME, SHAP, RISE, LRP, Occlusion sensitivity). Another paper combined two post-agnostic models (LIME and SAHP) with a CAM modified to be ante-specific (Table 3).

3.12. *Terminology*

The terminology used in the articles was diverse and involved terms such as *explainable* (n = 207), *interpretable* (n = 187), *understandable* (n = 112), *transparent* (n = 61), *reliable* (n = 31) and *intelligible* (n = 3) (Table 4).

Several terms were sometimes combined in the same article. In 34 articles, only one term was used: Explainable (E) (n = 23), Interpretable (IR) (n = 10) and Transparent (T) (n = 1). In 70 occasions, two terms were used throughout the article. While 3 terms were mixed in 75 papers, 4 in 39 articles, 5 in 9 others and in one article all 6 terms were found (Table 5).

The two most indistinctly used terms were explainable and interpretable, being found together in 169 out of the 228 articles.

For ante-hoc models (n = 14), the terms “explainable” (n = 13) and “interpretable” (n = 12) were used frequently, while “understandable” (n = 5) and “transparent” (n = 3) were rarely used. In post hoc models (n = 227), “explainable” (n = 190) and “interpretable” (n = 164) were also the preferred terms, followed by “understandable” (n = 110) and “transparent” (n = 57). Whereas “reliable” (n = 29) and “intelligible” (n = 4) were rare (Fig. 9).

4. Discussion

The aim of this scoping review was to map the existing literature on explainable and interpretable AI in diagnostic/follow up medical imaging modalities, to explore the various techniques, methods and approaches that are employed to improve AI interpretability.

The increased number of XAI-related articles identified in this work can be attributed to several factors, namely the ethical questions or social implications of AI that have gained much attention in recent years [14]. Most notably in 2018, The European Union’s General Data Protection Regulation (GDPR) stipulated that consumers affected by an automatic decision have the right to obtain “meaningful information about the logic involved” - interpreted only as a “right to explanation” [158,159]. In addition, the growing popularity of DL approaches

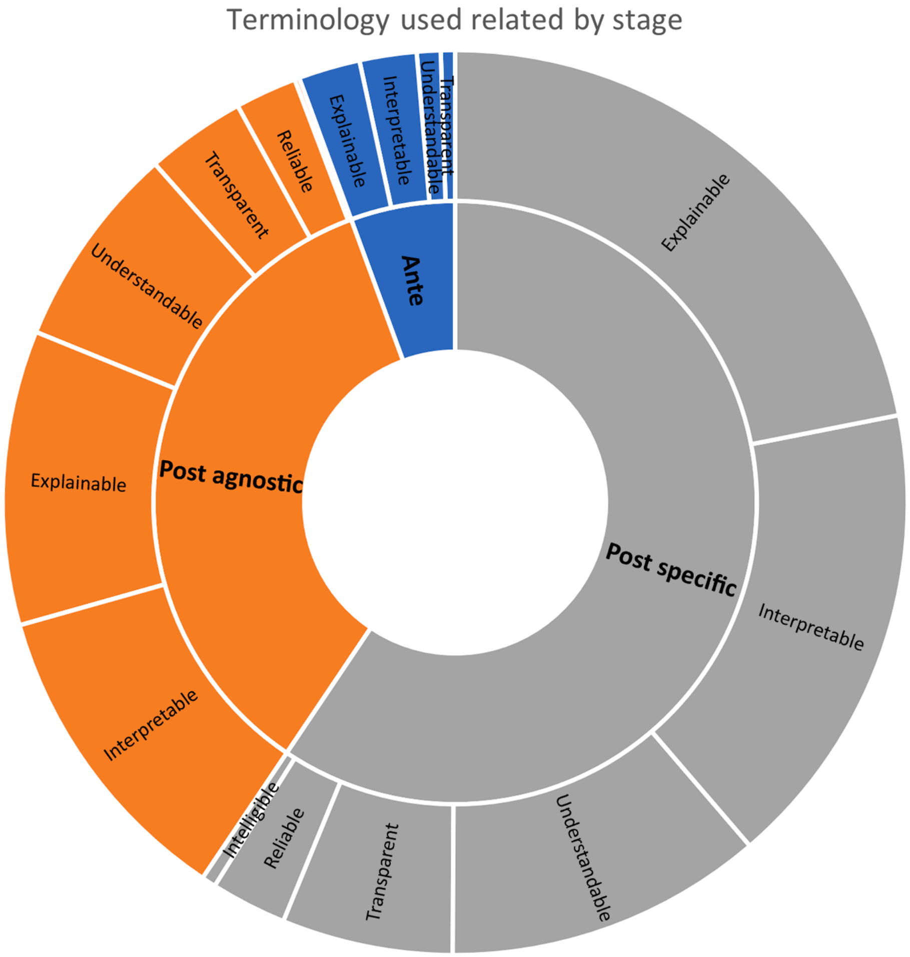


Fig. 9. Terminology used according to stage of deployment.

compared to classical ML, the growth in computing power and processing speed, digitalisation of healthcare and the availability of datasets have also contributed in an important way to the burgeoning field of XAI research [3].

Cross-sectional imaging, such as MRI and CT, and radiography are the main modalities where XAI has been applied. The lungs and the brain with pathologies such as Covid-19 and Alzheimer's disease were well represented in the published literature, as was shown in another similar study [9]. These trends can be attributed to the widespread availability of X-rays and the frequent utilization of cross-sectional images in diagnosis through MRI and CT scans. Furthermore, the worldwide pandemic has accelerated research efforts in understanding the Covid-19 pathology. Lastly, the availability of public datasets required for AI development has also contributed to the advancement of XAI in these areas.

The tasks explained were mainly designed to help physicians to classify (n = 89), predict (n = 47), diagnose (n = 39) and detect (n = 29) pathologies. A few of the more technical tasks, such as improving image quality (n = 6), were more recently explained. This is consistent with the most current use of AI in medical imaging [1,2,160].

In the literature [9–12], authors agree on most terms related to XAI taxonomy. However, the categories of output formats vary, having visual (n = 186), numeric (n = 67), rule-based (n = 11), textual (n = 11) and example-based (n = 6) [9,10,12] as the most frequently applied terms. The visual format is the predominant approach used to illustrate artificial intelligence decision-making processes, and this trend aligns with the dominant model found in existing literature in medical imaging [9,14]. However, visual explanations, such as heatmaps, are criticized by some authors as not being what users expect or need [161]. The same authors draw attention to the importance of distinguishing "user-centric explainable AI from developer-centric XAI" [161]. At present, XAI does not sufficiently take into account the users, such as physicians or radiographers who rely on AI results to take decisions. In addition, the development of XAI should take interdisciplinary approaches into account [162–165]. Wang et al. [166] point out the need to integrate varied and multiple explanations in order to get closer to the users' way of reasoning. In this review, 45 articles combined multiple explanation output formats. Social sciences are a field that can help to develop explanations and social sciences can help to optimize how XAI can incorporate social aspects into explanations to foster interactivity, to priorize user-centered design and to introduce dialogues that mimic human interaction [161,163,166–168]. None of the documents or the XAI frameworks in the field of medical imaging used the oral/dialog/verbal output format for explanations in the identified literature. In this context, explanations must extend beyond simple cognitive and causal aspects, going beyond the probabilities and knowledge possessed by algorithm developers. As Pazzani et al. [161] explain, the aim of explanations is not simply to pass on information, but to enable profane persons to become experts. These explanations should encompass the social process of knowledge integration, creating a symbiotic relationship between the explainer and the user. In addition, it is essential to recognize that users can vary considerably in terms of contexts, backgrounds, levels of knowledge and even modes of reasoning. In particular, users may have distinct mental models and react differently to explanations, which underlines the importance of adapting explanations to these viewpoints and needs [161,163,166–168].

Out of a total of 228 explanations, the majority (n = 178/228, 78.1 %) were given at a local level. A smaller proportion (n = 13/228, 5.7 %) exclusively focused on global explanations, while a subset (n = 37/228, 16.2 %) covered both local and global aspects. In their study, Liao et al. [169] conducted interviews with 20 experience practitioners users and designers. The participants acknowledged the importance of offering a global explanation to facilitate user comprehension of the system functionality and enable the formation of a mental model. Additionally, the researchers noted that users with a background in AI-related subjects displayed a higher tendency to actively seek global explanations. This difference in the explanations' scope may be linked to the context of medical imaging and the need to make a diagnosis for each patient, and therefore a trend towards giving local and therefore person-specific explanations. This distinction in explanatory approaches may also be influenced by the choice of output formats. Notably, in this scoping review, the visual format was predominantly used for local explanations (89.3 %, n = 159/178), whereas the numerical format was preferred for global explanations (61.5 %, n = 8/13). This discrepancy may be attributed to the effectiveness of the visual format in aiding decision-making for individual patients, particularly in tasks related to classification and diagnosis. Conversely, the numerical format is preferred for global explanations to explain which features have the biggest impact on the model in general.

Among the models analyzed, the majority (n = 222) were post-hoc, consisting of 77 agnostic models and 145 specific models. Only a small number of articles (14) employed model-based explanations. Furthermore, the terminology used was not specific to the different stages of the models (Fig. 9). The two most commonly used terms in the field of AI model interpretability, "explainable" and "interpretable", were used indistinctly. Whereas, according to their definition, both "interpretable" and "transparent" should be more suited to ante-models or model-based types, as it seeks to make understandable the model by itself. Likewise, "explainable" should be more in line with post-hoc models, since it involves an additional tool to understand the prediction made by the AI algorithm [8]. Thus, in the literature, these terms are not used strictly according to their meaning, but rather to facilitate rhetoric and avoid repetition.

This scoping review has certain limitations. First, the quality assessment of the included studies was not carried out following the specific methodology of a scoping review. Second, the focus was solely on recent XAI developments, leading to the exclusion of studies published before 2017, which may have resulted in missing out other tools that were explored during that period. Finally, efforts have been made to achieve exhaustive coverage of the articles published in this review; however, due to the absence of descriptors (or Mesh terms) related to XAI and their use as keywords in the search equations, it is possible that certain articles were not found using our search methodology. Nevertheless, the equations and search strategy are available in the supplementary material, ensuring transparency and enabling others to replicate this research.

As a scoping review also aims to identify gaps that still require further work, additional research on XAI can be identified. XAI tools related to other tasks can be developed, such as image enhancement and for other imaging modalities, including those specific to the field of radiotherapy and nuclear medicine. XAI developments should integrate user needs and could be more interactive. For example, qualitative or possibly even quantitative studies with physicians and the impact that XAI has on the decision making or patients should be considered, to avoid being purely developer centric XAI. Furthermore, it is essential to analyse the impact that altered decision-making with XAI can have on the patients. This analysis needs to analyse how patients are affected by the use of explainable XAI in medical practice and considers its potential implications for their overall care and well-being, such as improved or faster decision making.

5. Conclusion

XAI techniques are mainly applied in the context of MRI, CT or Radiography for the analysis of lung and brain pathologies, using available datasets. The predominant formats for presenting results are visual and numerical, with the emphasis on explaining classification and prediction tasks. In medical imaging, explanations tend to be more specific to individual samples or populations than to a global application. Meanwhile, there is a lack of attention to other AI tasks, as image enhancement, related to imaging itself and modalities such as PET/CT or SPECT/CT. Terminology in this area is not yet standardized, and terms such as “explainable” and “interpretable” are often used indistinctly in the literature. In the future, XAI developers should take user and patient needs and perspectives into account.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

CRediT authorship contribution statement

Mélanie Champendal: . **Henning Müller:** Writing – review & editing, Validation, Resources. **John O. Prior:** . **Cláudia Sá dos Reis:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to extend our sincere gratitude to Magali Serex for her help and expertise in conducting effective literature searches and formulating judicious recommendations, which greatly contributed to the success of this research.

Appendix A

PubMed

10.10.22

(“explainable deep learning”[tiab] OR “interpretable deep learning”[tiab] OR “XAI”[tiab] OR “explainable machine learning”[tiab] OR “interpretable machine learning”[tiab] OR “Transparent deep-learning”[tiab] OR “Transparent machine learning”[tiab] OR “Interpretable AI”[tiab] OR “Explainable AI”[tiab] OR “Transparent AI”[tiab] OR ((“explainability”[tiab] OR “Interpretability”[tiab] OR “transparency”[tiab] OR “decomposability”[tiab]) AND (“deep learning”[tiab] OR “AI”[tiab] OR “machine learning”[tiab]))).

AND

(“Radiology”[Mesh] OR “Radiology”[tiab] OR “Diagnostic Imaging”[Mesh] OR “Diagnostic Imaging”[tiab] OR “Magnetic Resonance Imaging”[tiab] OR “MRI”[tiab] OR “Computed Tomography”[tiab] OR “CT”[tiab] OR “Mammograph*”[tiab] OR “Ultrasonograph*”[tiab] OR “Radiograph*”[tiab] OR “Radiotherapy”[Mesh] OR “Radiation Oncology”[Mesh] OR “Radiotherap*”[tiab] OR “radiation therap*”[tiab] OR “Radiation Oncology”[tiab] OR “Tomotherapy”[tiab] OR “LINAC”[tiab] OR “linear accelerator”[tiab] OR “nuclear medicine”[tiab] OR “medical imag*”[tiab] OR “PET/CT”[tiab] OR “PET”[tiab] OR “SPECT/CT”[tiab] OR “SPECT”[tiab]).

AND (2017:2022[pdat]).

Number of references: 537.

Embase.com

10.10.22.

(‘explainable deep learning’:ab,ti,kw OR ‘interpretable deep learning’:ab,ti,kw OR ‘XAI’:ab,ti,kw OR ‘explainable machine learning’:ab,ti,kw OR ‘interpretable machine learning’:ab,ti,kw OR ‘Transparent deep-learning’:ab,ti,kw OR ‘Transparent machine learning’:ab,ti,kw OR ‘Interpretable AI’:ab,ti,kw OR ‘Explainable AI’:ab,ti,kw OR ‘Transparent AI’:ab,ti,kw OR ((‘explainability’:ab,ti,kw OR ‘Interpretability’:ab,ti,kw OR ‘transparency’:ab,ti,kw OR ‘decomposability’:ab,ti,kw) AND (‘deep learning’:ab,ti,kw OR ‘AI’:ab,ti,kw OR ‘machine learning’:ab,ti,kw))).

AND.

(‘Radiology’/exp OR ‘Radiology’:ab,ti,kw OR ‘Radiodiagnosis‘/exp OR ‘Diagnostic Imaging’:ab,ti,kw OR ‘Magnetic Resonance Imaging’:ab,ti,kw OR ‘MRI’:ab,ti,kw OR ‘Computed Tomography’:ab,ti,kw OR ‘CT’:ab,ti,kw OR ‘Mammograph*’:ab,ti,kw OR ‘Ultrasonograph*’:ab,ti,kw OR ‘Radiograph*’:ab,ti,kw OR ‘radiotherapy’/exp OR ‘Radiotherap*’:ab,ti,kw OR ‘radiation therap*’:ab,ti,kw OR ‘radiation oncology’/exp OR ‘radiation oncology’:ab,ti,kw OR ‘Tomotherapy’:ab,ti,kw OR ‘LINAC’:ab,ti,kw OR ‘linear accelerator’:ab,ti,kw OR ‘nuclear medicine’/exp OR ‘nuclear medicine’:ab,ti,kw OR ‘medical imag*’:ab,ti,kw OR ‘PET/CT’:ab,ti,kw OR ‘PET’:ab,ti,kw OR ‘SPECT/CT’:ab,ti,kw OR ‘SPECT’:ab,ti,kw).

AND [2017–2022]/py.

Number of references: 624.

CINAHL

10.10.22.

(TI “explainable deep learning” OR TI “interpretable deep learning” OR TI “XAI” OR TI “explainable machine learning” OR TI “interpretable machine learning” OR TI “Transparent deep-learning” OR TI “Transparent machine learning” OR TI “Interpretable AI” OR TI “Explainable AI” OR TI “Transparent AI” OR AB “explainable deep learning” OR AB “interpretable deep learning” OR AB “XAI” OR AB “explainable machine learning” OR AB “interpretable machine learning” OR AB “Transparent deep-learning” OR AB “Transparent machine learning” OR AB “Interpretable AI” OR AB “Explainable AI” OR AB “Transparent AI” OR ((TI “explainability” OR TI “Interpretability” OR TI “transparency” OR TI “decomposability”) AND (TI “deep learning” OR TI“AI” OR TI “machine learning”)) OR ((AB “explainability” OR AB “Interpretability” OR AB “transparency” OR AB “decomposability”) AND (AB “deep learning” OR AB“AI” OR AB “machine learning”))).

AND.

(MH “Diagnostic Imaging+” OR MH “Nuclear Medicine” OR MH “Radiation Oncology” OR “Radiotherapy+” OR TI “Radiology” OR AB “Radiology” OR TI “Diagnostic Imaging” OR AB “Diagnostic Imaging” OR TI “Magnetic Resonance Imaging” OR AB “Magnetic Resonance Imaging” OR TI

“MRI” OR AB “MRI” OR TI “Computed Tomography” OR AB “Computed Tomography” OR TI “CT” OR AB “CT” OR TI “Mammograph*” OR AB “Mammograph*” OR TI “Ultrasonograph*” OR AB “Ultrasonograph*” OR TI “Radiograph*” OR AB “Radiograph*” OR TI “Radiotherap*” OR AB “Radiotherap*” OR TI “Radiation therapy” OR AB “Radiation therapy” OR TI “Radiation oncology” OR AB “Radiation oncology” OR TI “Tomotherapy” OR AB “Tomotherapy” OR TI “LINAC” OR AB “LINAC” OR TI “Linear accelerator” OR AB “Linear accelerator” OR TI “nuclear medicine” OR AB “nuclear medicine” OR TI “medical imag*” OR AB “medical imag*” OR TI “PET/CT” OR AB “PET/CT” OR TI “PET” OR AB “PET” OR TI “SPECT/CT” OR AB “SPECT/CT” OR TI “SPECT” OR AB “SPECT”).

Number of references: 74.

Web of Science

Web of Science Core collecABon.

10.10.22.

TS=(“explainable deep learning” OR “interpretable deep learning” OR “XAI” OR “explainable machine learning” OR “interpretable machine learning” OR “Transparent deep-learning” OR “Transparent machine learning” OR “Interpretable AI” OR “Explainable AI” OR “Transparent AI” OR ((“explainability” OR “Interpretability” OR “transparency” OR “decomposability”) AND (“deep learning” OR “AI” OR “machine learning”))).

AND.

TS=(“Radiology” OR “Diagnostic Imaging” OR “Magnetic Resonance Imaging” OR “MRI” OR “Computed Tomography” OR “CT” OR “Mammograph*” OR “Ultrasonograph*” OR “Radiograph*” OR “Radiotherap*” OR “radiation therap*” OR “radiation oncology” OR “Tomotherapy” OR “LINAC” OR “linear accelerator” OR “nuclear medicine” OR “medical imag*” OR “PET/CT” OR “PET” OR “SPECT/CT” OR “SPECT”).

AND PY = 2017–2100.

Number of references: 686.

Google scholar

10.10.22.

“explainable|interpretable|transparent ”deep learning“|XAI|”machine learning“” Radiology|“Diagnostic Imaging”|“Magnetic Resonance Imaging”|MRI|CT|mammography|ultrasonography|radiography|radiotherapy|“nuclear medicine”|“medical imaging”|PET|“radiation oncology”|“linear accelerator”|LINAC|Tomotherapy.

(With year limit 2017–2022).

Number of references: 3490.

The first 200.

medRxiv and BioRxiv

10.10.2022.

Advanced Search | medRxiv.

(Interpretable AI OR Explainable AI OR XAI) AND (medical imaging OR medical image analysis) Or (explainable deep learning OR interpretable deep learning OR XAI) AND (medical imaging OR medical image analysis).

Same results, Number of references: 92.

Appendix B. Comprehensive references categorized by imaging modalities and output format

MRI

Output format	References
Visual	[43–93]
Numerical	[94,100,110,152,190–195]
Textual	–
Ruled-based	[157]
Example-based	–
Mixed	Rule-based and Textual: [97] Visual & Numerical:[98,120,129,134,151,196,197] Visual & Textual:[93] Visual, Numerical & Textual:[198]

Radiography (Xrays)

Output format	References
Visual	[115–145]
Numerical	[40,46,51,76,204]
Textual	–
Ruled-based	[75,205]
Example-based	–
Mixed	Numerical & Rule-based:[206] Visual & Numerical:[38] Visual & Rule-based:[53] Numerical & Example-based:[77,85]

(*continued on next page*)

(*continued*)

Output format	References
	Visual & Example-based:[207] Visual, Numerical & Textual:[56]

CT

Output format	References
Visual	[160–188]
Numerical	[81,223–227]
Textual	[228]
Ruled-based	–
Example-based	–
Mixed	Rule-based & Example-based: [65] Visual & Numerical:[60,61,63,137,229–231] Numerical & Example-based:[232] Visual, Numerical & Textual:[82] Visual, Numerical, Rule-based & Textual: [80]

US

Output format	References
Visual	[207–219]
Numerical	[147,244]
Textual	–
Ruled-based	–
Example-based	–
Mixed	Numerical & Rule-based:[146] Visual & Numerical:[245] Numerical & Textual:[246]

Double

Output format	References
Visual	- Radiography & CT:[27] PET/CT & MRI:[28] US to MRI:[33] CBCT & Panoramic images:[35]
Numerical	PET/CT & MRI:[30] US & CT:[32] Mammography & US:[26]
Textual	–
Ruled-based	–
Example-based	–
Mixed	Visual & Numerical: Radiography & CT:[24,25,34] MRI & US:[29] Visual, Numerical & Textual: MRI & Radiography: [31]

fMRI

Output format	References
Visual	[122,138,247–249]
Numerical	[250]
Textual	–
Ruled-based	–
Example-based	–
Mixed	Visual & Numerical: [123,154]

SPECT or SPECT/CT

Output format	References
Visual	[251–255]
Numerical	–
Textual	–
Ruled-based	[95]
Example-based	–
Mixed	Visual & Numerical:[96,101,119]

Mammography

Output format	References
Visual	[141–145,256,257]
Numerical	–
Textual	–
Ruled-based	–
Example-based	–
Mixed	Numerical & Rule-based: [258]

PET/CT

Output format	References
Visual	–
Numerical	–
Textual	–
Ruled-based	–
Example-based	–
Mixed	Visual, Numerical & Textual:[259]

Different medical imaging

Output format	References
Visual	DXA: [260] MRI to PET: [117] Radiotherapy: CBCT for IGRT: [261]
Numerical	–
Textual	–
Ruled-based	–
Example-based	–
Mixed	• Image captioning: Visual & Numerical: [262] • Breast Tomosynthesis: Visual & Numerical:[263]

References

[1] F. Nensa, A. Demircioglu, C. Rischpler, Artificial intelligence in nuclear medicine, Journal of Nuclear Medicine 60 (2019) 29S–37S, https://doi.org/10.2967/jnumed.118.220590.

[2] A. Hosny, C. Parmar, J. Quackenbush, L.H. Schwartz, H.J.W.L. Aerts, H.H. Edu, Artificial intelligence in radiology HHS Public Access, Nature Reviews. Cancer 18 (2018) 500–510, https://doi.org/10.1038/s41568-018-0016-5.Artificial.

[3] S.Y. Wang, S. Pershing, A.Y. Lee, Big data requirements for artificial intelligence, Current Opinion in Ophthalmology 31 (2020) 318–323, https://doi.org/10.1097/ICU.0000000000000676.

[4] J. Díaz-Ramírez, Machine Learning and Deep Learning, Ingeniare 29 (2021) 182–183, https://doi.org/10.4067/S0718-33052021000200180.

[5] W.J. von Eschenbach, Transparency and the Black Box Problem: Why We Do Not Trust AI, Philos Technol 34 (2021) 1607–1622, https://doi.org/10.1007/s13347-021-00477-0.

[6] A.I.F. Poon, J.J.Y. Sung, Opening the black box of AI-Medicine, Journal of Gastroenterology and Hepatology 36 (2021) 581–584, https://doi.org/10.1111/jgh.15384.

[7] M. Reyes, R. Meier, S. Pereira, C.A. Silva, F.M. Dahlweid, H. von Tengg-Kobligk, et al., On the interpretability of artificial intelligence in radiology: Challenges and opportunities. Radiol, Artificial Intelligence (2020) 2, https://doi.org/10.1148/ryai.2020190043.

[8] M. Graziani, L. Dutkiewicz, D. Calvaresi, J.P. Amorim, K. Yordanova, M. Vered, et al., A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences, Springer, Netherlands, 2022.

[9] B.H.M. van der Velden, H.J. Kuijf, K.G.A. Gilhuijs, M.A. Viergever, Explainable artificial intelligence (XAI) in deep learning-based medical image analysis, Medical Image Analysis 79 (2022), 102470, https://doi.org/10.1016/j.media.2022.102470.

[10] Vilone G, Longo L. Explainable Artificial Intelligence: a Systematic Review 2020.

[11] W.J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, B. Yu, Definitions, methods, and applications in interpretable machine learning, Proceedings of the National Academy of Sciences of the United States of America 116 (2019) 22071–22080, https://doi.org/10.1073/pnas.1900654116.

[12] A. Adadi, M. Berrada, Peeking inside the black-box: A survey on explainable artificial intelligence (XAI), IEEE Access 6 (2018) 52138–52160, https://doi.org/10.1109/ACCESS.2018.2870052.

[13] K. Borys, Y.A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C.M. Friedrich, et al., Explainable AI in medical imaging: An overview for clinical practitioners – Beyond saliency-based XAI approaches, European Journal of Radiology (2023) 162, https://doi.org/10.1016/j.ejrad.2023.110786.

[14] M.R. Islam, M.U. Ahmed, S. Barua, S. Begum, A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks, Applied Sciences (2022) 12, https://doi.org/10.3390/app12031353.

[15] Kok I, Okay FY, Muyanli O, Ozdemir S. Explainable Artificial Intelligence (XAI) for Internet of Things: A Survey 2022:1–14.

[16] G. Yang, Q. Ye, J. Xia, Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond, Inf Fusion 77 (2022) 29–52, https://doi.org/10.1016/j.inffus.2021.07.016.

[17] K. Borys, Y.A. Schmitt, M. Nauta, C. Seifert, N. Krämer, C.M. Friedrich, et al., Explainable AI in medical imaging: An overview for clinical practitioners - Saliency-based XAI approaches, European Journal of Radiology 162 (2023), 110787, https://doi.org/10.1016/j.ejrad.2023.110787.

[18] A.M. Groen, R. Kraan, S.F. Amirkhan, J.G. Daams, M. Maas, A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: Limited use of explainable AI? European Journal of Radiology 157 (2022), 110592 https://doi.org/10.1016/j.ejrad.2022.110592.

[19] H. Arksey, L. O'Malley, Scoping studies: Towards a methodological framework, Int J Soc Res Methodol Theory Pract 8 (2005) 19–32, https://doi.org/10.1080/1364557032000119616.

[20] M.D.J. Peters, C. Marnie, A.C. Tricco, D. Pollock, Z. Munn, L. Alexander, et al., Updated methodological guidance for the conduct of scoping reviews, JBI Evid Synth. 18 (2020) 2119–2126, https://doi.org/10.11124/JBIES-20-00167.

[21] A.C. Tricco, E. Lillie, W. Zarin, K.K. O'Brien, H. Colquhoun, D. Levac, et al., PRISMA extension for scoping reviews (PRISMA-ScR): Checklist and explanation, Annals of Internal Medicine 169 (2018) 467–473, https://doi.org/10.7326/M18-0850.

[22] W.M. Bramer, D. Giustini, G.B. de Jonge, L. Holland, T. Bekhuis, De-duplication of database search results for systematic reviews in EndNote, Journal of the Medical Library Association 104 (2016) 240–243, https://doi.org/10.3163/1536-5050.104.3.014.

[23] M.J. Page, J.E. McKenzie, P. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, et al., The prisma 2020 statement: An updated guideline for reporting systematic reviews, Med Flum. 57 (2021) 444–465, https://doi.org/10.21860/medflum2021_264903.

[24] M.M. Ahsan, R. Nazim, Z. Siddique, P. Huebner, Detection of covid-19 patients from ct scan and chest x-ray data using modified mobilenetv2 and lime, Healthc 9 (2021) 1–12, https://doi.org/10.3390/healthcare9091099.

[25] M.M. Ahsan, K.D. Gupta, M.M. Islam, S. Sen, M.L. Rahman, H.M. Shakhawat, COVID-19 Symptoms Detection Based on NasNetMobile with Explainable AI Using Various Imaging Modalities, Mach Learn Knowl Extr 2 (2020) 490–504, https://doi.org/10.3390/make2040027.

[26] Y. Zou, Y. Shi, F. Sun, J. Liu, Y. Guo, H. Zhang, et al., Extreme gradient boosting model to assess risk of central cervical lymph node metastasis in patients with papillary thyroid carcinoma: Individual prediction using SHapley Additive exPlanations, Computer Methods and Programs in Biomedicine 225 (2022), 107038, https://doi.org/10.1016/j.cmpb.2022.107038.

[27] F. Ullah, J. Moon, H. Naeem, S. Jabbar, Explainable artificial intelligence approach in combating real-time surveillance of COVID19 pandemic from CT scan and X-ray images using ensemble model, The Journal of Supercomputing 78 (2022) 19246–19271, https://doi.org/10.1007/s11227-022-04631-z.

[28] T. Escobar, S. Vauclin, F. Orlhac, C. Nioche, P. Pineau, L. Champion, et al., Voxel-wise supervised analysis of tumors with multimodal engineered features to highlight interpretable biological patterns, Medical Physics 49 (2022) 3816–3829, https://doi.org/10.1002/mp.15603.

[29] M.R. Hassan, M.F. Islam, M.Z. Uddin, G. Ghoshal, M.M. Hassan, S. Huda, et al., Prostate cancer classification from ultrasound and MRI images using deep learning based Explainable Artificial Intelligence, Future Generation Computer Systems 127 (2022) 462–472, https://doi.org/10.1016/j.future.2021.09.030.

[30] Hernandez M, Ramon-Julvez U, Ferraz F. Explainable AI toward understanding the performance of the top three TADPOLE Challenge methods in the forecast of Alzheimer's disease diagnosis. vol. 17. 2022. 10.1371/journal.pone.0264695.

[31] M.R. Karim, J. Jiao, T. Dohmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, et al., DeepKneeExplainer: Explainable Knee Osteoarthritis Diagnosis From Radiographs and Magnetic Resonance Imaging, IEEE Access 9 (2021) 39757–39780, https://doi.org/10.1109/ACCESS.2021.3062493.

[32] M. Ma, R. Liu, C. Wen, W. Xu, Z. Xu, S. Wang, et al., Predicting the molecular subtype of breast cancer and identifying interpretable imaging features using machine learning algorithms, European Radiology 32 (2022) 1652–1662, https://doi.org/10.1007/s00330-021-08271-4.

[33] X. Song, H. Chao, X. Xu, H. Guo, S. Xu, B. Turkbey, et al., Cross-modal attention for multi-modal image registration, Medical Image Analysis 82 (2022), 102612, https://doi.org/10.1016/j.media.2022.102612.

[34] W. Shi, L. Tong, Y. Zhu, M.D. Wang, COVID-19 Automatic Diagnosis with Radiographic Imaging: Explainable Attention Transfer Deep Neural Networks, IEEE J Biomed Heal Informatics 25 (2021) 2376–2387, https://doi.org/10.1109/JBHI.2021.3074893.

[35] S. Yang, H. Lee, B. Jang, K.D. Kim, J. Kim, H. Kim, et al., Development and validation of a visually explainable deep learning model for classification of C-shaped canals of the mandibular second molars in periapical and panoramic dental radiographs, Journal of Endodontia 48 (2022) 914–921, https://doi.org/10.1016/j.joen.2022.04.007.

[36] J.P. Cruz-Bastida, E. Pearson, H. Al-Hallaq, Toward understanding deep learning classification of anatomic sites: lessons from the development of a CBCT projection classifier, Journal of Medical Imaging 9 (2022) 1–12, https://doi.org/10.1117/1.jmi.9.4.045002.

[37] Khobahi S, Agarwal C, Soltanalian M. CoroNet: A Deep Network Architecture for Semi-Supervised Task-Based Identification of COVID-19 from Chest X-ray Images 2020. doi.org/10.1101/2020.04.14.20065722.

[38] M. Bhandari, T.B. Shahi, B. Siku, A. Neupane, Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI, Computers in Biology and Medicine (2022) 150, https://doi.org/10.1016/j.compbiomed.2022.106156.

[39] M.A. Khan, M. Azhar, K. Ibrar, A. Alqahtani, S. Alsubai, A. Binbusayyis, et al., COVID-19 Classification from Chest X-Ray Images: A framework of deep explainable artificial intelligence, Computational Intelligence and Neuroscience (2022), https://doi.org/10.1155/2022/4254631.

[40] X.V. Nguyen, E. Dikici, S. Candemir, R.L. Ball, L.M. Prevedello, Mortality Prediction Analysis among COVID-19 Inpatients Using Clinical Variables and Deep Learning Chest Radiography Imaging Features, Tomography 8 (2022) 1791–1803, https://doi.org/10.3390/tomography8040151.

[41] S.N. Nillmani, L. Saba, N.N. Khanna, M.K. Kalra, M.M. Fouda, et al., Segmentation-Based Classification Deep Learning Model Embedded with Explainable AI for COVID-19 Detection in Chest X-ray Scans, Diagnostics (2022) 12, https://doi.org/10.3390/diagnostics12092132.

[42] A. Sharma, P.K. Mishra, Covid-MANet: Multi-task attention network for explainable diagnosis and severity assessment of COVID-19 from CXR images, Pattern Recognition (2022) 131, https://doi.org/10.1016/j.patcog.2022.108826.

[43] Sharma V, Piyush, Chhatwal S, Singh B. An explainable artificial intelligence based prospective framework for COVID-19 risk prediction. MedRxiv 2021:1–15. 10.1101/2021.03.02.21252269.

[44] D. Shome, T. Kar, S.N. Mohanty, P. Tiwari, K. Muhammad, A. Altameem, et al., Covid-transformer: Interpretable covid-19 detection using vision transformer for healthcare, International Journal of Environmental Research and Public Health (2021) 18, https://doi.org/10.3390/ijerph182111086.

[45] N. Tsiknakis, E. Trivizakis, E. Vassalou, G. Papadakis, D. Spandidos, A. Tsatsakis, et al., Interpretable artificial intelligence framework for COVID-19 screening on chest X-rays, Experimental and Therapeutic Medicine 20 (2020) 727–735, https://doi.org/10.3892/etm.2020.8797.

[46] Y. Wang, C. Jiang, Y. Wu, T. Lv, H. Sun, Y. Liu, et al., Semantic-powered explainable model-free few-shot learning scheme of diagnosing COVID-19 on Chest X-Ray, IEEE J Biomed Heal Informatics 26 (2022) 5870–5882, https://doi.org/10.1109/JBHI.2022.3205167.

[47] L. Wang, Z.Q. Lin, A. Wong, COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images, Scientific Reports 10 (2020) 1–12, https://doi.org/10.1038/s41598-020-76550-z.

[48] Y. Xu, H.K. Lam, G. Jia, MANet: A two-stage deep learning method for classification of COVID-19 from Chest X-ray images, Neurocomputing 443 (2021) 96–105, https://doi.org/10.1016/j.neucom.2021.03.034.

[49] M. Chetoui, M.A. Akhloufi, B. Yousefi, E.M. Bouattane, Explainable COVID-19 detection on chest X-rays using an end-to-end deep convolutional neural network architecture, Big Data Cogn Comput (2021) 5, https://doi.org/10.3390/bdcc5040073.

[50] A.J. DeGrave, J.D. Janizek, S.I. Lee, AI for radiographic COVID-19 detection selects shortcuts over signal, Nat Mach Intell 3 (2021) 610–619, https://doi.org/10.1038/s42256-021-00338-7.

[51] L.V. de Moura, C. Mattjie, C.M. Dartora, R.C. Barros, A.M. Marques da Silva, Explainable machine learning for COVID-19 pneumonia classification with texture-based features extraction in chest radiography, Front Digit Heal 3 (2022) 1–13, https://doi.org/10.3389/fdgth.2021.662343.

[52] A. Dhere, J. Sivaswamy, COVID detection from Chest X-Ray images using multi-scale attention, IEEE J Biomed Heal Informatics 26 (2022) 1496–1505, https://doi.org/10.1109/JBHI.2022.3151171.

[53] P.S. Gidde, S.S. Prasad, A.P. Singh, N. Bhatheja, S. Prakash, P. Singh, et al., Validation of expert system enhanced deep learning algorithm for automated screening for COVID-Pneumonia on chest X-rays, Scientific Reports 11 (2021) 1–12, https://doi.org/10.1038/s41598-021-02003-w.

[54] R. Harkness, G. Hall, A.F. Frangi, N. Ravikumar, K. Zucker, The Pitfalls of Using Open Data to Develop Deep Learning Solutions for COVID-19 Detection in Chest X-Rays, Studies in Health Technology and Informatics 290 (2022) 679–683, https://doi.org/10.3233/SHTI220164.

[55] Q. Hu, F.N.B. Gois, R. Costa, L. Zhang, L. Yin, N. Magaia, et al., Explainable artificial intelligence-based edge fuzzy images for COVID-19 detection and identification, Applied Soft Computing 123 (2022), 108966, https://doi.org/10.1016/j.asoc.2022.108966.

[56] Karim R, Till D, Cochez M, Beyan O, Rebholz-schuhmann D, Decker S. DeepCOVIDExplainer : Explainable COVID-19 Diagnosis from Chest X-ray Images 2021:2020–3.

[57] M. Li, Y. Fang, Z. Tang, C. Onuorah, J. Xia, S.J. Del, et al., Explainable COVID-19 Infections Identification and Delineation Using Calibrated Pseudo Labels, IEEE Trans Emerg Top Comput Intell 7 (2023) 26–35, https://doi.org/10.1109/TETCI.2022.3189054.

[58] I.P. de Sousa, M.M.B.R. Vellasco, E.C. da Silva, Explainable artificial intelligence for bias detection in covid ct-scan classifiers, Sensors 21 (2021), https://doi.org/10.3390/s21165657.

[59] M. Pennisi, I. Kavasidis, C. Spampinato, V. Schinina, S. Palazzo, F.P. Salanitri, et al., An explainable AI system for automated COVID-19 assessment and lesion categorization from CT-scans, Artificial Intelligence in Medicine 118 (2021), 102114, https://doi.org/10.1016/j.artmed.2021.102114.

[60] G. Singh, Think positive: An interpretable neural network for image recognition, Neural Networks 151 (2022) 178–189, https://doi.org/10.1016/j.neunet.2022.03.034.

[61] G. Singh, K.C. Yow, Object or background: An interpretable deep learning model for covid-19 detection from ct-scan images, Diagnostics 11 (2021), https://doi.org/10.3390/diagnostics11091732.

[62] X.J. Yao, Z.Q. Zhu, S.H. Wang, Z.YD. Csgbbnet, An explainable deep learning framework for covid-19 detection, Diagnostics 11 (2021), https://doi.org/10.3390/diagnostics11091712.

[63] Ye Q, Xia J, Yang G. Explainable AI for COVID-19 CT Classifiers: An initial comparison study. Proc - IEEE Symp Comput Med Syst 2021;2021-June:521–6. 10.1109/CBMS52027.2021.00103.

[64] H. Alshazly, C. Linse, E. Barth, T. Martinetz, Explainable COVID-19 detection using chest CT scans and deep learning, Sensors (switzerland) 21 (2021) 1–22, https://doi.org/10.3390/s21020455.

[65] P. Angelov, E. Soares, Towards explainable deep neural networks (xDNN), Neural Networks 130 (2020) 185–194, https://doi.org/10.1016/j.neunet.2020.07.010.

[66] Chetoui M, Akhloufi MA. Automated Detection of COVID-19 Cases using Recent Deep Convolutional Neural Networks and CT images. 2021 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE; 2021, p. 3297–300. 10.1109/EMBC46164.2021.9629689.

[67] H. Gunraj, L. Wang, A. Wong, COVIDNet-CT: A tailored deep convolutional neural network design for detection of COVID-19 Cases From Chest CT Images, Frontiers in Medicine 7 (2020) 1–11, https://doi.org/10.3389/fmed.2020.608525.

[68] H. Gunraj, A. Sabri, D. Koff, A. Wong, COVID-Net CT-2: Enhanced Deep Neural Networks for Detection of COVID-19 From Chest CT Images Through Bigger, More Diverse Learning. Front Med 8 (2022) 1–15, https://doi.org/10.3389/fmed.2021.729287.

[69] He X, Wang S, Chu X, Shi S, Tang J, Liu X, et al. Automated Model Design and Benchmarking of Deep Learning Models for COVID-19 Detection with Chest CT Scans. 35th AAAI Conf Artif Intell AAAI 2021 2021;6A:4821–9. 10.1609/aaai.v35i6.16614.

[70] Maximino J, Coimbra M, Pedrosa J. Detection of COVID-19 in Point of Care Lung Ultrasound. Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS 2022;2022-July:1527–30. 10.1109/EMBC48229.2022.9871235.

[71] M. Charachon, C. Hudelot, P.H. Cournède, C. Ruppli, R. Ardon, Combining similarity and adversarial learning to generate visual explanation: Application to medical image classification, Proc - Int Conf Pattern Recognit (2020) 7188–7195, https://doi.org/10.1109/ICPR48806.2021.9413253.

[72] Han Y, Chen C, Tewfik A, Ding Y, Peng Y. Pneumonia Detection On Chest X-Ray Using Radiomic Features And Contrastive Learning. 2021 IEEE 18th Int. Symp. Biomed. Imaging, IEEE; 2021, p. 247–51. 10.1109/ISBI48211.2021.9433853.

[73] Kashyap S, Karargyris A, Wu J, Gur Y, Sharma A, Wong KCL, et al. Looking in the Right Place for Anomalies: Explainable Ai Through Automatic Location Learning. Proc - Int Symp Biomed Imaging 2020;2020-April:1125–9. 10.1109/ISBI45749.2020.9098370.

[74] Rajaraman S, Thoma G, Antani S., Candemir S. Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs. In: Hahn HK, Mori K, editors. Med. Imaging 2019 Comput. Diagnosis, SPIE; 2019, p. 27. 10.1117/12.2512752.

[75] H. Ren, A.B. Wong, W. Lian, W. Cheng, Y. Zhang, J. He, et al., Interpretable Pneumonia Detection by Combining Deep Learning and Explainable Models with Multisource Data, IEEE Access 9 (2021) 95872–95883, https://doi.org/10.1109/ACCESS.2021.3090215.

[76] R.K. Sheu, L.C. Chen, C.L. Wu, M.S. Pardeshi, K.C. Pai, C.C. Huang, et al., Multi-Modal Data Analysis for Pneumonia Status Prediction Using Deep Learning (MDA-PSP), Diagnostics 12 (2022), https://doi.org/10.3390/diagnostics12071706.

[77] J.J. Thiagarajan, K. Thopalli, D. Rajan, P. Turaga, Training calibration-based counterfactual explainers for deep learning models in medical image analysis, Scientific Reports 12 (2022) 1–15, https://doi.org/10.1038/s41598-021-04529-5.

[78] C.C. Ukwuoma, Z. Qin, M. Belal Bin Heyat, F. Akhtar, O. Bamisile, A.Y. Muaad, et al., A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images, Journal of Advanced Research (2022), https://doi.org/10.1016/j.jare.2022.08.021.

[79] Y. Yang, G. Mei, F. Piccialli, A Deep Learning Approach Considering Image Background for Pneumonia Identification Using Explainable AI (XAI), IEEE/ACM Trans Comput Biol Bioinforma (2022) 1–12, https://doi.org/10.1109/TCBB.2022.3190265.

[80] N. Alwarasneh, Y.S.S. Chow, S.T.M. Yan, C.H. Lim, Bridging Explainable Machine Vision in CAD Systems for Lung Cancer Detection, The EU-German-Turkish Triangle (2020) 254–269, https://doi.org/10.1007/978-3-030-66645-3_22.

[81] A. Joshi, J. Sivaswamy, G.D. Joshi, Lung nodule malignancy classification with weakly supervised explanation generation, Journal of Medical Imaging 8 (2021), https://doi.org/10.1117/1.jmi.8.4.044502.

[82] A. Meldo, L. Utkin, M. Kovalev, E. Kasimov, The natural language explanation algorithms for the lung cancer computer-aided diagnosis system, Artificial Intelligence in Medicine 108 (2020), 101952, https://doi.org/10.1016/j.artmed.2020.101952.

[83] Qiu B, Furst J, Rasin A, Tchoua R, Raicu D. Learning Latent Spiculated Features for Lung Nodule Characterization. Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS 2020;2020-July:1254–7. 10.1109/EMBC44109.2020.9175720.

[84] N. Gozzi, E. Giacomello, M. Sollini, M. Kirienko, A. Ammirabile, P. Lanzi, et al., Image Embeddings Extracted from CNNs Outperform Other Transfer learning approaches in classification of chest radiographs, Diagnostics 12 (2022), https://doi.org/10.3390/diagnostics12092084.

[85] D. Kim, J. Chung, J. Choi, M.D. Succi, J. Conklin, M.G.F. Longo, et al., Accurate auto-labeling of chest X-ray images based on quantitative similarity to an explainable AI model, Nature Communications 13 (2022) 1–15, https://doi.org/10.1038/s41467-022-29437-8.

[86] D. Mahapatra, A. Poellinger, M. Reyes, Interpretability-Guided Inductive Bias For Deep Learning Based Medical Image, Medical Image Analysis 81 (2022), 102551, https://doi.org/10.1016/j.media.2022.102551.

[87] X. Ouyang, S. Karanam, Z. Wu, T. Chen, J. Huo, X.S. Zhou, et al., Learning Hierarchical Attention for Weakly-Supervised Chest X-Ray Abnormality Localization and Diagnosis, IEEE Transactions on Medical Imaging 40 (2021) 2698–2710, https://doi.org/10.1109/TMI.2020.3042773.

[88] B. Wang, W. Zhang, MARnet: Multi-scale adaptive residual neural network for chest X-ray images recognition of lung diseases, Mathematical Biosciences and Engineering 19 (2022) 331–350, https://doi.org/10.3934/mbe.2022017.

[89] Singla S, Gong M, Ravanbakhsh S, Sciurba F, Poczos B, Batmanghelich KN. Subject2Vec: Generative-discriminative approach from a set of image patches to a vector. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2018;11070 LNCS:502–10. 10.1007/978-3-030-00928-1_57.

[90] S. Feng, Q. Liu, A. Patel, S.U. Bazai, C.K. Jin, J.S. Kim, et al., Automated pneumothorax triaging in chest X-rays in the New Zealand population using deep-learning algorithms, Journal of Medical Imaging and Radiation Oncology 66 (2022) 1035–1043, https://doi.org/10.1111/1754-9485.13393.

[91] Marvin G, Alam MGR. Explainable Augmented Intelligence and Deep Transfer Learning for Pediatric Pulmonary Health Evaluation. 2022 Int Conf Innov Sci Eng Technol ICISET 2022 2022:272–7. 10.1109/ICISET54810.2022.9775845.

[92] X. Ma, E.C. Ferguson, X. Jiang, S.I. Savitz, S. Shams, A multitask deep learning approach for pulmonary embolism detection and identification, Scientific Reports 12 (2022) 1–11, https://doi.org/10.1038/s41598-022-16976-9.

[93] Clough JR, Oksuz I, Puyol-Antón E, Ruijsink B, King AP, Schnabel JA. Global and local interpretability for cardiac MRI classification. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2019;11767 LNCS:656–64. 10.1007/978-3-030-32251-9_72.

[94] Janik A, Dodd J, Ifrim G, Sankaran K, Curran KM. Interpretability of a deep learning model in the application of cardiac MRI segmentation with an ACDC challenge dataset. In: Landman BA, Išgum I, editors. Med. Imaging 2021 Image Process., SPIE; 2021, p. 111. 10.1117/12.2582227.

[95] L.H. Hu, J. Betancur, T. Sharir, A.J. Einstein, S. Bokhari, M.B. Fish, et al., Machine learning predicts per-vessel early coronary revascularization after fast myocardial perfusion SPECT: Results from multicentre REFINE SPECT registry, European Heart Journal Cardiovascular Imaging 21 (2020) 549–559, https://doi.org/10.1093/ehjci/jez177.

[96] Y. Otaki, A. Singh, P. Kavanagh, R.J.H. Miller, T. Parekh, B.K. Tamarappoo, et al., Clinical deployment of explainable artificial intelligence of SPECT for diagnosis of coronary artery disease, JACC: Cardiovascular Imaging 15 (2022) 1091–1102, https://doi.org/10.1016/j.jcmg.2021.04.030.

[97] Achilleos KG, Leandrou S, Prentzas N, Kyriacou PA, Kakas AC, Pattichis CS. Extracting Explainable Assessments of Alzheimer's disease via Machine Learning on brain MRI imaging data. Proc - IEEE 20th Int Conf Bioinforma Bioeng BIBE 2020 2020:1036–41. 10.1109/BIBE50027.2020.00175.

[98] A. Aghaei, M. Ebrahimi Moghaddam, H. Malek, Interpretable ensemble deep learning model for early detection of Alzheimer's disease using local interpretable model-agnostic explanations, International Journal of Imaging Systems and Technology 32 (2022) 1889–1902, https://doi.org/10.1002/ima.22762.

[99] Z. Qin, Z. Liu, Q. Guo, P. Zhu, 3D convolutional neural networks with hybrid attention mechanism for early diagnosis of Alzheimer's disease, Biomedical Signal Processing and Control 77 (2022), 103828, https://doi.org/10.1016/j.bspc.2022.103828.

[100] S. Qiu, M.I. Miller, P.S. Joshi, J.C. Lee, C. Xue, Y. Ni, et al., Multimodal deep learning for Alzheimer's disease dementia assessment, Nature Communications 13 (2022) 1–17, https://doi.org/10.1038/s41467-022-31037-5.

[101] A. Sarica, A. Quattrone, A. Quattrone, Explainable machine learning with pairwise interactions for the classification of Parkinson's disease and SWEDD from clinical and imaging features, Brain Imaging and Behavior 16 (2022) 2188–2198, https://doi.org/10.1007/s11682-022-00688-9.

[102] Shad HA, Rahman QA, Asad NB, Bakshi AZ, Mursalin SMF, Reza MT, et al. Exploring Alzheimer's Disease Prediction with XAI in various Neural Network Models. IEEE Reg 10 Annu Int Conf Proceedings/TENCON 2021;2021-Decem: 720–5. 10.1109/TENCON54134.2021.9707468.

[103] H. Shahamat, A.M. Saniee, Brain MRI analysis using a deep learning based evolutionary approach, Neural Networks 126 (2020) 218–234, https://doi.org/10.1016/j.neunet.2020.03.017.

[104] Sudar KM, Nagaraj P, Nithisaa S, Aishwarya R, Aakash M, Lakshmi SI. Alzheimer's Disease Analysis using Explainable Artificial Intelligence (XAI). Int Conf Sustain Comput Data Commun Syst ICSCDS 2022 - Proc 2022:419–23. 10.1109/ICSCDS53736.2022.9760858.

[105] C. Tinauer, S. Heber, L. Pirpamer, A. Damulina, R. Schmidt, R. Stollberger, et al., Interpretable brain disease classification and relevance-guided deep learning, Scientific Reports 12 (2022) 1–14, https://doi.org/10.1038/s41598-022-24541-7.

[106] L. Yu, W. Xiang, J. Fang, Y.P. Phoebe Chen, R. Zhu, A novel explainable neural network for Alzheimer's disease diagnosis, Pattern Recognition 131 (2022), https://doi.org/10.1016/j.patcog.2022.108876.

[107] F. Zhang, B. Pan, P. Shao, P. Liu, S. Shen, P. Yao, et al., A Single Model Deep Learning Approach for Alzheimer's Disease Diagnosis, Neuroscience 491 (2022) 200–214, https://doi.org/10.1016/j.neuroscience.2022.03.026.

[108] Q. Zhang, Q. Du, G. Liu, A whole-process interpretable and multi-modal deep reinforcement learning for diagnosis and analysis of Alzheimer's disease, Journal of Neural Engineering 18 (2021), https://doi.org/10.1088/1741-2552/ac37cc.

[109] M. Böhle, F. Eitel, M. Weygandt, K. Ritter, Layer-wise relevance propagation for explaining deep neural network decisions in MRI-based Alzheimer's disease classification, Front Aging Neurosci. 10 (2019), https://doi.org/10.3389/fnagi.2019.00194.

[110] L. Bloch, C.M. Friedrich, Data analysis with Shapley values for automatic subject selection in Alzheimer's disease data sets using interpretable machine learning, Alzheimer's Res Ther 13 (2021) 1–30, https://doi.org/10.1186/s13195-021-00879-4.

[111] Bordin V, Coluzzi D, Rivolta MW, Baselli G. Explainable AI Points to White Matter Hyperintensities for Alzheimer's Disease Identification: a Preliminary Study. Proc Annu Int Conf IEEE Eng Med Biol Soc EMBS 2022;2022-July:484–7. 10.1109/EMBC48229.2022.9871306.

[112] H. Guan, C. Wang, J. Cheng, J. Jing, T. Liu, A parallel attention-augmented bilinear network for early magnetic resonance imaging-based diagnosis of Alzheimer's disease, Human Brain Mapping 43 (2022) 760–772, https://doi.org/10.1002/hbm.25685.

[113] Konate S, Lebrat L, Cruz RS, Bourgeat P, Dore V, Fripp J, et al. Smocam: Smooth Conditional Attention Mask For 3d-Regression Models. 2021 IEEE 18th Int. Symp. Biomed. Imaging, IEEE; 2021, p. 362–6. 10.1109/ISBI48211.2021.9433972.

[114] Z. Liu, E. Adeli, K.M. Pohl, Q. Zhao, Going Beyond Saliency Maps: Training Deep Models to Interpret Deep Models, Physiology & Behavior 176 (2021) 71–82, https://doi.org/10.1007/978-3-030-78191-0_6.

[115] E. Nigri, N. Ziviani, F. Cappabianco, A. Antunes, A. Veloso, Explainable Deep CNNs for MRI-Based Diagnosis of Alzheimer's Disease, Proc Int Jt Conf Neural Networks (2020), https://doi.org/10.1109/IJCNN48605.2020.9206837.

[116] M. Odusami, R. Maskeliūnas, R. Damaševičius, An Intelligent System for Early Recognition of Alzheimer's Disease Using Neuroimaging, Sensors 22 (2022), https://doi.org/10.3390/s22030740.

[117] Kao CH, Chen YS, Chen LF, Chiu WC. Demystifying T1-MRI to FDG 18 -PET Image Translation via Representational Similarity. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2021;12903 LNCS: 402–12. 10.1007/978-3-030-87199-4_38.

[118] W.S. Monroe, T. Anthony, M.M. Tanik, F.M. Skidmore, Towards a framework for validating machine learning results in medical imaging opening the black box, ACM Int Conf Proceeding Ser (2019), https://doi.org/10.1145/3332186.3332193.

[119] P.R. Magesh, R.D. Myloth, R.J. Tom, An explainable machine learning model for early detection of parkinson's disease using LIME on DaTSCAN imagery,

Computers in Biology and Medicine 126 (2020), 104041, https://doi.org/10.1016/j.compbiomed.2020.104041.

[120] Sarica A, Quattrone A, Quattrone A. Explainable Boosting Machine for Predicting Alzheimer's Disease from MRI Hippocampal Subfields. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2021;12960 LNAI:341–50. 10.1007/978-3-030-86993-9_31.

[121] Chen Y, Yan J, Jiang M, Zhao Z, Zhao W. Convolutional Networks for Identification of Autism Spectrum Disorder Using Multi-Modal MRI Data n.d.: 374–85.

[122] Y. Liang, G. Xu, S. ur Rehman, Multi-scale attention-based deep neural network for brain disease diagnosis, Comput Mater Contin. 72 (2022) 4545–4661, https://doi.org/10.32604/cmc.2022.026999.

[123] N. Pat, Y. Wang, A. Bartonicek, J. Candia, A. Stringaris, Explainable machine learning approach to predict and explain the relationship between task-based fMRI and individual differences in cognition, Cerebral Cortex 33 (2023) 2682–2703, https://doi.org/10.1093/cercor/bhac235.

[124] M. Svm, S. Maqsood, Multi-Modal Brain Tumor Detection Using Deep Neural, Mdpi (2022).

[125] M. Coupet, T. Urruty, T. Leelanupab, M. Naudin, P. Bourdon, C.F. Maloigne, et al., A multi-sequences MRI deep framework study applied to glioma classfication, Multimedia Tools and Applications 81 (2022) 13563–13591, https://doi.org/10.1007/s11042-022-12316-1.

[126] P. Windisch, P. Weber, C. Fürweger, F. Ehret, M. Kufeld, D. Zwahlen, et al., Implementation of model explainability for a basic brain tumor detection using convolutional neural networks on MRI slices, Neuroradiology 62 (2020) 1515–1518, https://doi.org/10.1007/s00234-020-02465-1.

[127] R.A. Zeineldin, M.E. Karar, Z. Elshaer, J. Coburger, C.R. Wirtz, O. Burgert, et al., Explainability of deep neural networks for MRI analysis of brain tumors, International Journal of Computer Assisted Radiology and Surgery 17 (2022) 1673–1683, https://doi.org/10.1007/s11548-022-02619-x.

[128] M. Esmaeili, R. Vettukattil, H. Banitalebi, N.R. Krogh, J.T. Geitung, Explainable artificial intelligence for human-machine interaction in brain tumor localization, J Pers Med 11 (2021), https://doi.org/10.3390/jpm11111213.

[129] L. Gaur, M. Bhandari, T. Razdan, S. Mallik, Z. Zhao, Explanation-Driven Deep Learning Model for Prediction of Brain Tumour Status Using MRI Image Data, Frontiers in Genetics 13 (2022) 1–9, https://doi.org/10.3389/fgene.2022.822666.

[130] P. Natekar, A. Kori, G. Krishnamurthi, Demystifying Brain Tumor Segmentation Networks: Interpretability and Uncertainty Analysis, Frontiers in Computational Neuroscience 14 (2020) 1–12, https://doi.org/10.3389/fncom.2020.00006.

[131] Niepceron B, Grassia F, Nait Sidi Moh A. Brain Tumor Detection Using Selective Search and Pulse-Coupled Neural Network Feature Extraction. Comput Informatics 2022;41:253–70. 10.31577/cai_2022_1_253.

[132] S. Pereira, R. Meier, R. McKinley, R. Wiest, V. Alves, C.A. Silva, et al., Enhancing interpretability of automatically extracted machine learning features: application to a RBM-Random Forest system on brain lesion segmentation, Medical Image Analysis 44 (2018) 228–244, https://doi.org/10.1016/j.media.2017.12.009.

[133] H. Saleem, A.R. Shahid, B. Raza, Visual interpretability in 3D brain tumor segmentation network, Computers in Biology and Medicine 133 (2021), 104410, https://doi.org/10.1016/j.compbiomed.2021.104410.

[134] C. Severn, K. Suresh, C. Görg, Y.S. Choi, R. Jain, D. Ghosh, A Pipeline for the Implementation and Visualization of Explainable Machine Learning for Medical Imaging Using Radiomics Features, Sensors 22 (2022), https://doi.org/10.3390/s22145205.

[135] R. Tanno, D.E. Worrall, E. Kaden, A. Ghosh, F. Grussu, A. Bizzi, et al., Uncertainty modelling in deep learning for safer neuroimage enhancement: Demonstration in diffusion MRI, NeuroImage 225 (2021), 117366, https://doi.org/10.1016/j.neuroimage.2020.117366.

[136] G. Fu, J. Li, R. Wang, Y. Ma, Y. Chen, Attention-based full slice brain CT image diagnosis with explanations, Neurocomputing 452 (2021) 263–274, https://doi.org/10.1016/j.neucom.2021.04.044.

[137] C. Song, G. Fu, J. Li, Y. Pei, An Explainable Multi-Instance Multi-Label Classification Model for Full Slice Brain CT Images, IFAC-PapersOnLine 53 (2020) 780–785, https://doi.org/10.1016/j.ifacol.2021.05.001.

[138] Z. Jiang, Y. Wang, C.W. Shi, Y. Wu, R. Hu, S. Chen, et al., Attention module improves both performance and interpretability of four-dimensional functional magnetic resonance imaging decoding neural network, Human Brain Mapping 43 (2022) 2683–2692, https://doi.org/10.1002/hbm.25813.

[139] K.H. Kim, H.W. Koo, B.J. Lee, S.W. Yoon, M.J. Sohn, Cerebral hemorrhage detection and localization with medical imaging for cerebrovascular disease diagnosis and treatment using explainable deep learning, Journal of the Korean Physical Society 79 (2021) 321–327, https://doi.org/10.1007/s40042-021-00202-2.

[140] Gulum MA, Trombley CM, Kantardzic M. Improved deep learning explanations for prostate lesion classification through grad-CAM and saliency map fusion. Proc - IEEE Symp Comput Med Syst 2021;2021-June:498–502. 10.1109/CBMS52027.2021.00099.

[141] S.T. Kim, J.H. Lee, H. Lee, Y.M. Ro, Visually interpretable deep network for diagnosis of breast masses on mammograms, Physics in Medicine and Biology 63 (2018), https://doi.org/10.1088/1361-6560/aaef0a.

[142] T. Kobayashi, T. Haraguchi, T. Nagao, Classifying presence or absence of calcifications on mammography using generative contribution mapping, Radiological Physics and Technology 15 (2022) 340–348, https://doi.org/10.1007/s12194-022-00673-3.

[143] La Ferla M. An XAI Approach to Deep Learning Models in the Detection of DCIS 2021:1–9. 10.1007/978-3-031-34171-7_33.

[144] K. Liu, Y. Shen, N. Wu, J. Chłędowski, C. Fernandez-Granda, K.J. Geras, Weakly-supervised high-resolution segmentation of mammography images for breast cancer diagnosis, Proc Mach Learn Res 143 (2021) 268–285.

[145] Major D, Lenis D, Wimmer M, Sluiter G, Berg A, Bühler K. INTERPRETING MEDICAL IMAGE CLASSIFIERS BY OPTIMIZATION BASED COUNTERFACTUAL IMPACT ANALYSIS 2020:1096–100. 10.48550/arXiv.2004.01610.

[146] A. Rezazadeh, Y. Jafarian, A. Kord, Explainable ensemble machine learning for breast cancer diagnosis based on ultrasound image texture features, Forecasting 4 (2022) 262–274, https://doi.org/10.3390/forecast4010015.

[147] Zhang B, Vakanski A, Xian M. Bi-Rads-Net: An Explainable Multitask Learning Approach for Cancer Diagnosis in Breast Ultrasound Images. 2021 IEEE 31st Int. Work. Mach. Learn. Signal Process., IEEE; 2021, p. 1–6. 10.1109/MLSP52302.2021.9596314.

[148] C. Wang, Y. Wu, C. Wang, X. Zhou, Y. Niu, Y. Zhu, et al., Attention-based multiple-instance learning for Pediatric bone age assessment with efficient and interpretable, Biomedical Signal Processing and Control 79 (2022), https://doi.org/10.1016/j.bspc.2022.104028.

[149] Karargyris A, Kashyap S, Wu JT, Sharma A, Moradi M, Syeda-Mahmood T. Age prediction using a large chest x-ray dataset. In: Hahn HK, Mori K, editors. Med. Imaging 2019 Comput. Diagnosis, SPIE; 2019, p. 66. 10.1117/12.2512922.

[150] G. Levakov, G. Rosenthal, I. Shelef, T.R. Raviv, G. Avidan, From a deep learning model back to the brain—Identifying regional predictors and their relation to aging, Human Brain Mapping 41 (2020) 3235–3252, https://doi.org/10.1002/hbm.25011.

[151] A. Lombardi, D. Diacono, N. Amoroso, A. Monaco, J.M.R.S. Tavares, R. Bellotti, et al., Explainable Deep Learning for Personalized Age Prediction With Brain Morphology, Frontiers in Neuroscience 15 (2021) 1–17, https://doi.org/10.3389/fnins.2021.674055.

[152] R. Scheda, S. Diciotti, Explanations of Machine Learning Models in Repeated Nested Cross-Validation: An Application in Age Prediction Using Brain Complexity Features, Applied Sciences 12 (2022), https://doi.org/10.3390/app12136681.

[153] Belton N, Welaratne I, Dahlan A, Hearne RT, Hagos MT, Lawlor A, et al. Optimising Knee Injury Detection with Spatial Attention and Validating Localisation Ability. vol. 12722 LNCS. 2021. 10.1007/978-3-030-80432-9_6.

[154] O. Al Zoubi, M. Misaki, A. Tsuchiyagaito, V. Zotev, E. White, M. Paulus, et al., Machine learning evidence for sex differences consistently influences resting-state functional magnetic resonance imaging fluctuations across multiple independently acquired data sets, Brain Connectivity 12 (2022) 348–361, https://doi.org/10.1089/brain.2020.0878.

[155] Gao K, Shen H, Liu Y, Zeng L, Hu D. Dense-CAM: Visualize the Gender of Brains with MRI Images. Proc Int Jt Conf Neural Networks 2019;2019-July:1–7. 10.1109/IJCNN.2019.8852260.

[156] C.A. Jimenez-Castaño, A.M. Álvarez-Meza, O.D. Aguirre-Ospina, D.A. Cárdenas-Peña, Á.A. Orozco-Gutiérrez, Random fourier features-based deep learning improvement with class activation interpretability for nerve structure segmentation, Sensors 21 (2021), https://doi.org/10.3390/s21227741.

[157] T. Chen, P. Su, Y. Shen, L. Chen, M. Mahmud, Y. Zhao, et al., A dominant set-informed interpretable fuzzy system for automated diagnosis of dementia, Frontiers in Neuroscience 16 (2022), https://doi.org/10.3389/fnins.2022.867664.

[158] R. Binns, Algorithmic Accountability and Public Reason, Philos Technol 31 (2018) 543–556, https://doi.org/10.1007/s13347-017-0263-5.

[159] Agarwal S, Kirrane S, Scharf J. Modelling the general data protection regulation. Jusletter IT 2017;2014.

[160] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, et al., A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017) 60–88, https://doi.org/10.1016/j.media.2017.07.005.

[161] Pazzani M, Soltani S, Kaufman R, Qian S, Hsiao A. Expert-Informed, User-Centric Explanations for Machine Learning. Proc 36th AAAI Conf Artif Intell AAAI 2022 2022;36:12280–6. 10.1609/aaai.v36i11.21491.

[162] Brandão R, Carbonera J, de Souza C, Ferreira J, Gonçalves B, Leitão C. Mediation Challenges and Socio-Technical Gaps for Explainable Deep Learning Applications 2019:1–39.

[163] Gerlings J, Shollo A, Constantiou I. Reviewing the need for explainable artificial intelligence (XAI). Proc Annu Hawaii Int Conf Syst Sci 2021;2020-Janua: 1284–93. 10.24251/hicss.2021.156.

[164] T.J. Bradshaw, M.D. McCradden, A.K. Jha, J. Dutta, B. Saboury, E.L. Siegel, et al., Artificial Intelligence Algorithms Need to Be Explainable-or Do They? Journal of Nuclear Medicine 64 (2023) 976–977, https://doi.org/10.2967/jnumed.122.264949.

[165] Abdul A, Vermeulen J, Wang D, Lim BY, Kankanhalli M. Trends and trajectories for explainable, accountable and intelligible systems: An HCI research agenda. Conf Hum Factors Comput Syst - Proc 2018;2018-April. 10.1145/3173574.3174156.

[166] D. Wang, Q. Yang, A. Abdul, B.Y. Lim, Designing theory-driven user-centric explainable AI, Conf Hum Factors Comput Syst - Proc (2019), https://doi.org/10.1145/3290605.3300831.

[167] T. Miller, Explanation in artificial intelligence: Insights from the social sciences, Artificial Intelligence 267 (2019) 1–38, https://doi.org/10.1016/j.artint.2018.07.007.

[168] Eiband M, Schneider H, Bilandzic M, Fazekas-Con J, Haug M, Hussmann H. Bringing transparency design into practice. Int Conf Intell User Interfaces, Proc IUI 2018:211–23. 10.1145/3172944.3172961.

[169] Q.V. Liao, D. Gruen, S. Miller, Questioning the AI: Informing Design Practices for Explainable AI User Experiences, Conf Hum Factors Comput Syst - Proc (2020), https://doi.org/10.1145/3313831.3376590.

[170] N. Bahrami, T. Retson, K. Blansit, K. Wang, A. Hsiao, Automated selection of myocardial inversion time with a convolutional neural network: Spatial temporal ensemble myocardium inversion network (STEMI-NET), Magnetic Resonance in Medicine 81 (2019) 3283–3291, https://doi.org/10.1002/mrm.27680.

[171] C. Biffi, J.J. Cerrolaza, G. Tarroni, W. Bai, A. De Marvao, O. Oktay, et al., Explainable Anatomical Shape Analysis through Deep Hierarchical Generative Models, IEEE Transactions on Medical Imaging 39 (2020) 2088–2099, https://doi.org/10.1109/TMI.2020.2964499.

[172] K.S. Choi, S.H. Choi, B. Jeong, Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network, Neuro-Oncology 21 (2019) 1197–1209, https://doi.org/10.1093/neuonc/noz095.

[173] F. Eitel, E. Soehler, J. Bellmann-Strobl, A.U. Brandt, K. Ruprecht, R.M. Giess, et al., Uncovering convolutional neural network decisions for diagnosing multiple sclerosis on conventional MRI using layer-wise relevance propagation, NeuroImage Clin 24 (2019), 102003, https://doi.org/10.1016/j.nicl.2019.102003.

[174] D.D. Gunashekar, L. Bielak, L. Hägele, B. Oerther, M. Benndorf, A.L. Grosu, et al., Explainable AI for CNN-based prostate tumor segmentation in multi-parametric MRI correlated to whole mount histopathology, Radiation Oncology 17 (2022) 1–10, https://doi.org/10.1186/s13014-022-02035-0.

[175] T. Hepp, D. Blum, K. Armanious, B. Schölkopf, D. Stern, B. Yang, et al., Uncertainty estimation and explainability in deep learning-based age estimation of the human brain: Results from the German National Cohort MRI study, Computerized Medical Imaging and Graphics 92 (2021), https://doi.org/10.1016/j.compmedimag.2021.101967.

[176] V. Jain, O. Nankar, D.J. Jerrish, S. Gite, S. Patil, K. Kotecha, A Novel AI-Based System for Detection and Severity Prediction of Dementia Using MRI, IEEE Access 9 (2021) 154324–154346, https://doi.org/10.1109/ACCESS.2021.3127394.

[177] Y.S. Jeon, K. Yoshino, S. Hagiwara, A. Watanabe, S.T. Quek, H. Yoshioka, et al., Interpretable and Lightweight 3-D Deep Learning Model for Automated ACL Diagnosis, IEEE J Biomed Heal Informatics 25 (2021) 2388–2397, https://doi.org/10.1109/JBHI.2021.3081355.

[178] J. Huang, R. He, J. Chen, S. Li, Y. Deng, X. Wu, Boosting Advanced Nasopharyngeal Carcinoma Stage Prediction Using a Two-Stage Classification Framework Based on Deep Learning, Int J Comput Intell Syst 14 (2021), https://doi.org/10.1007/s44196-021-00026-9.

[179] D. Kim, J. Lee, J. Moon, T. Moon, Interpretable deep learning-based hippocampal sclerosis classification, Epilepsia Open 7 (2022) 747–757, https://doi.org/10.1002/epi4.12655.

[180] A. Kumar, R. Manikandan, U. Kose, D. Gupta, S.C. Satapathy, Doctor's dilemma: Evaluating an explainable subtractive spatial lightweight convolutional neural network for brain tumor diagnosis, ACM Transactions on Multimedia Computing, Communications, and Applications 17 (2021), https://doi.org/10.1145/3457187.

[181] A. Lopatina, S. Ropele, R. Sibgatulin, J.R. Reichenbach, D. Güllmar, Investigation of Deep-Learning-Driven Identification of Multiple Sclerosis Patients Based on Susceptibility-Weighted Images Using Relevance Analysis, Frontiers in Neuroscience 14 (2020) 1–12, https://doi.org/10.3389/fnins.2020.609468.

[182] P. Lu, L. Hu, N. Zhang, H. Liang, T. Tian, L. Lu, A Two-Stage Model for Predicting Mild Cognitive Impairment to Alzheimer's Disease Conversion, Frontiers in Aging Neuroscience 14 (2022) 1–13, https://doi.org/10.3389/fnagi.2022.826622.

[183] G. Maicas, G. Snaauw, A.P. Bradley, I. Reid, G. Carneiro, Model agnostic saliency for weakly supervised lesion detection from breast DCE-MRI australian institute for machine learning, school of computer science, the university of adelaide faculty of applied sciences, Delft University of Technology Science and Isbi (2019) 1057–1060.

[184] E. Moulton, R. Valabregue, M. Piotin, G. Marnat, S. Saleme, B. Lapergue, et al., Interpretable deep learning for the prognosis of long-term functional outcome post-stroke using acute diffusion weighted imaging, Journal of Cerebral Blood Flow and Metabolism 43 (2023) 198–209, https://doi.org/10.1177/0271678X221129230.

[185] M. Pérez-Pelegrí, J.V. Monmeneu, M.P. López-Lereu, L. Pérez-Pelegrí, A. M. Maceira, V. Bodí, et al., Automatic left ventricle volume calculation with explainability through a deep learning weak-supervision methodology, Computer Methods and Programs in Biomedicine 208 (2021), https://doi.org/10.1016/j.cmpb.2021.106275.

[186] A. Termine, C. Fabrizio, C. Caltagirone, L. Petrosini, A Reproducible Deep-Learning-Based Computer-Aided Diagnosis Tool for Frontotemporal Dementia Using MONAI and Clinica Frameworks, Life 12 (2022), https://doi.org/10.3390/life12070947.

[187] P. Tupe-Waghmare, P. Malpure, K. Kotecha, M. Beniwal, V. Santosh, J. Saini, et al., Comprehensive Genomic Subtyping of Glioma Using Semi-Supervised Multi-Task Deep Learning on Multimodal MRI, IEEE Access 9 (2021) 167900–167910, https://doi.org/10.1109/ACCESS.2021.3136293.

[188] S.H. Wang, X.J. Han, J. Du, Z.C. Wang, C. Yuan, Y. Chen, et al., Saliency-based 3D convolutional neural network for categorising common focal liver lesions on multisequence MRI, Insights Imaging. 12 (2021), https://doi.org/10.1186/s13244-021-01117-z.

[190] M. Bang, Y.W. Park, J. Eom, S.S. Ahn, J. Kim, S.K. Lee, et al., An interpretable radiomics model for the diagnosis of panic disorder with or without agoraphobia using magnetic resonance imaging, Journal of Affective Disorders 305 (2022) 47–54, https://doi.org/10.1016/j.jad.2022.02.072.

[191] X. Chen, Y. Li, X. Li, X. Cao, Y. Xiang, W. Xia, et al., An interpretable machine learning prognostic system for locoregionally advanced nasopharyngeal carcinoma based on tumor burden features, Oral Oncology 118 (2021), 105335, https://doi.org/10.1016/j.oraloncology.2021.105335.

[192] A. Conti, C.A. Treaba, A. Mehndiratta, V.T. Barletta, C. Mainero, N. Toschi, An interpretable machine learning model to predict cortical atrophy in multiple sclerosis, Brain Sciences 13 (2023) 3757–3760, https://doi.org/10.3390/brainsci13020198.

[193] S. Kucukseymen, A. Arafati, T. Al-Otaibi, H. El-Rewaidy, A.S. Fahmy, L.H. Ngo, et al., Noncontrast cardiac magnetic resonance imaging predictors of heart failure hospitalization in heart failure with preserved ejection fraction, Journal of Magnetic Resonance Imaging 55 (2022) 1812–1825, https://doi.org/10.1002/jmri.27932.

[194] Seitzer M, Yang G, Schlemper J, Oktay O, Würfl T, Christlein V, et al. Adversarial and perceptual refinement for compressed sensing MRI reconstruction. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2018;11070 LNCS:232–40. 10.1007/978-3-030-00928-1_27.

[195] Y.W. Park, J. Eom, D. Kim, S.S. Ahn, E.H. Kim, S.G. Kang, et al., A fully automatic multiparametric radiomics model for differentiation of adult pilocytic astrocytomas from high-grade gliomas, European Radiology 32 (2022) 4500–4509, https://doi.org/10.1007/s00330-022-08575-z.

[196] Li W, Feng X, An H, Ng XY, Zhang YJ. MRI reconstruction with interpretable pixel-wise operations using reinforcement learning. AAAI 2020 - 34th AAAI Conf Artif Intell 2020:792–9. 10.1609/aaai.v34i01.5423.

[197] B.H.M. van der Velden, M.H.A. Janse, M.A.A. Ragusi, C.E. Loo, K.G.A. Gilhuijs, Volumetric breast density estimation on MRI using explainable deep learning regression, Scientific Reports 10 (2020) 1–9, https://doi.org/10.1038/s41598-020-75167-6.

[198] C.J. Wang, C.A. Hamm, L.J. Savic, M. Ferrante, I. Schobert, T. Schlachter, et al., Deep learning for liver tumor diagnosis part II: convolutional neural network interpretation using radiologic imaging features, European Radiology 29 (2019) 3348–3357, https://doi.org/10.1007/s00330-019-06214-8.

[204] M.S. Kamal, L. Chowdhury, N. Dey, S.J. Fong, K. Santosh, Explainable ai to analyze outcomes of spike neural network in Covid-19 Chest X-rays, Conf Proc - IEEE Int Conf Syst Man Cybern (2021) 3408–3415, https://doi.org/10.1109/SMC52423.2021.9658745.

[205] E. Casiraghi, D. Malchiodi, G. Trucco, M. Frasca, L. Cappelletti, T. Fontana, et al., Explainable machine learning for early assessment of COVID-19 risk prediction in emergency departments, IEEE Access 8 (2020) 196299–196325, https://doi.org/10.1109/ACCESS.2020.3034032.

[206] N. Aslam, Explainable artificial intelligence approach for the early prediction of ventilator support and mortality in COVID-19 patients, Computation 10 (2022), https://doi.org/10.3390/computation10030036.

[207] Kim J, Kim M, Ro YM. Interpretation of Lesional Detection Via Counterfactual Generation. Proc - Int Conf Image Process ICIP 2021;2021-Septe:96–100. 10.1109/ICIP42928.2021.9506282.

[208] A.F. Bayram, C. Gurkan, A. Budak, H. Karataş, Böbrek hastaliklari için açiklanabilir yapay zeka destekli derin öğrenmeye dayali bir tespit ve tahmin modeli, Eur J Sci Technol. (2022) 67–74, https://doi.org/10.31590/ejosat.1171777.

[209] Z. Han, B. Wei, Y. Hong, T. Li, J. Cong, X. Zhu, et al., Accurate Screening of COVID-19 Using Attention-Based Deep 3D Multiple Instance Learning, IEEE Transactions on Medical Imaging 39 (2020) 2584–2594, https://doi.org/10.1109/TMI.2020.2996256.

[210] J. Hao, J. Xie, R. Liu, H. Hao, Y. Ma, K. Yan, et al., Automatic Sequence-Based Network for Lung Diseases Detection in Chest CT, Frontiers in Oncology 11 (2021) 1–14, https://doi.org/10.3389/fonc.2021.781798.

[211] B. Jiang, Y. Zhang, L. Zhang, G. H. de Bock, R. Vliegenthart, X. Xie, Human-recognizable CT image features of subsolid lung nodules associated with diagnosis and classification by convolutional neural networks, Eur Radiol. 31 (2021) 7303–7315, https://doi.org/10.1007/s00330-021-07901-1.

[212] S.C. Liu, J. Lai, J.Y. Huang, C.F. Cho, P.H. Lee, M.H. Lu, et al., Predicting microvascular invasion in hepatocellular carcinoma: a deep learning model validated across hospitals, Cancer Imaging 21 (2021) 1–16, https://doi.org/10.1186/s40644-021-00425-3.

[213] S. Mohagheghi, A.H. Foruzan, Developing an explainable deep learning boundary correction method by incorporating cascaded x-Dim models to improve segmentation defects in liver CT images, Computers in Biology and Medicine 140 (2022), 105106, https://doi.org/10.1016/j.compbiomed.2021.105106.

[214] O'Shea RJ, Horst C, Manickavasagar T, Hughes D, Cusack J, Tsoka S, et al. Weakly supervised Unet: an image classifier which learns to explain itself. BioRxiv 2022:2022.09.09.507144. 10.1101/2022.09.09.507144.

[215] D.E. Heron, R.S. Andrade, S. Beriwal, R.P. Smith, PET-CT in radiation oncology: the impact on diagnosis, treatment planning, and assessment of treatment response, American Journal of Clinical Oncology 31 (2008) 352–362, https://doi.org/10.1097/COC.0b013e318162f150.

[216] S.M. Ryan, N.E. Carlson, H. Butler, T.E. Fingerlin, L.A. Maier, F. Xing, Cluster activation mapping with application to computed tomography scans of the lung, Journal of Medical Imaging 9 (2022), https://doi.org/10.1117/1.JMI.9.2.026001.

[217] J.S. Suri, S. Agarwal, G.L. Chabert, A. Carriero, A. Paschè, P.S.C. Danna, et al., COVLIAS 2.0-cXAI: Cloud-Based Explainable Deep Learning System for COVID-19 Lesion Localization in Computed Tomography Scans, Diagnostics 12 (2022), https://doi.org/10.3390/diagnostics12061482.

[218] W. Tan, P. Guan, L. Wu, H. Chen, J. Li, Y. Ling, et al., The use of explainable artificial intelligence to explore types of fenestral otosclerosis misdiagnosed when

using temporal bone high-resolution computed tomography, Ann Transl Med. 9 (2021), https://doi.org/10.21037/atm-21-1171.

[219] S.H. Wang, K. Wu, T. Chu, S.L. Fernandes, Q. Zhou, Y.D. Zhang, et al., SOSPCNN: Structurally optimized stochastic pooling convolutional neural network for tetralogy of fallot recognition, Wireless Communications and Mobile Computing 2021 (2021), https://doi.org/10.1155/2021/5792975.

[223] P. Giraud, P. Giraud, E. Nicolas, P. Boisselier, M. Alfonsi, M. Rives, et al., Interpretable machine learning model for locoregional relapse prediction in oropharyngeal cancers, Cancers (basel) 13 (2021) 1–13, https://doi.org/10.3390/cancers13010057.

[224] Han F, Liao S, Yuan S, Wu R, Zhao Y, Xie Y. Explainable Prediction of Renal Cell Carcinoma From Contrast-Enhanced Ct Images Using Deep Convolutional Transfer Learning and the Shapley Additive Explanations Approach. Proc - Int Conf Image Process ICIP 2021;2021-Septe:3802–6. 10.1109/ICIP42928.2021.9506144.

[225] W. He, B. Li, R. Liao, H. Mo, L. Tian, An ISHAP-based interpretation-model-guided classification method for malignant pulmonary nodule, Knowledge-Based Syst 237 (2022), 107778, https://doi.org/10.1016/j.knosys.2021.107778.

[226] M.S. Jabal, O. Joly, D. Kallmes, G. Harston, A. Rabinstein, T. Huynh, et al., Interpretable Machine Learning Modeling for Ischemic Stroke Outcome Prediction, Frontiers in Neurology 13 (2022) 1–8, https://doi.org/10.3389/fneur.2022.884693.

[227] Wang Y, Zamiela C, Thomas T V., Duggar WN, Roberts PR, Bian L, et al. 3D Texture Feature-Based Lymph Node Automated Detection in Head and Neck Cancer Analysis. Proc - 2020 IEEE Int Conf Bioinforma Biomed BIBM 2020 2020: 2113–9. 10.1109/BIBM49941.2020.9313482.

[228] S. Shen, S.X. Han, D.R. Aberle, A.A. Bui, W. Hsu, An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification, Expert Systems with Applications 128 (2019) 84–95, https://doi.org/10.1016/j.eswa.2019.01.048.

[229] K. Goel, R. Sindhgatta, S. Kalra, R. Goel, P. Mutreja, The effect of machine learning explanations on user trust for automated diagnosis of COVID-19, Computers in Biology and Medicine 146 (2022), 105587, https://doi.org/10.1016/j.compbiomed.2022.105587.

[230] J. Liu, Y. Kang, Z. Xia, J. Qiang, J.F. Zhang, Y. Zhang, et al., MRCON-Net: Multiscale reweighted convolutional coding neural network for low-dose CT imaging, Computer Methods and Programs in Biomedicine 221 (2022), 106851, https://doi.org/10.1016/j.cmpb.2022.106851.

[231] T. Refaee, Z. Salahuddin, A.N. Frix, C. Yan, G. Wu, H.C. Woodruff, et al., Diagnosis of Idiopathic Pulmonary Fibrosis in High-Resolution Computed Tomography Scans Using a Combination of Handcrafted Radiomics and Deep Learning, Frontiers in Medicine 9 (2022) 1–10, https://doi.org/10.3389/fmed.2022.915243.

[232] K. Zhang, S. Qi, J. Cai, D. Zhao, T. Yu, Y. Yue, et al., Content-based image retrieval with a Convolutional Siamese Neural Network: Distinguishing lung cancer and tuberculosis in CT images, Computers in Biology and Medicine 140 (2022), 105096, https://doi.org/10.1016/j.compbiomed.2021.105096.

[244] Z. Jin, S. Pei, L. Ouyang, L. Zhang, X. Mo, Q. Chen, et al., Thy-Wise: An interpretable machine learning model for the evaluation of thyroid nodules, International Journal of Cancer 151 (2022) 2229–2243, https://doi.org/10.1002/ijc.34248.

[245] S. Turco, T. Tiyarattanachai, K. Ebrahimkheil, J. Eisenbrey, A. Kamaya, M. Mischi, et al., Interpretable Machine Learning for Characterization of Focal Liver Lesions by Contrast-Enhanced Ultrasound, IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 69 (2022) 1670–1681, https://doi.org/10.1109/TUFFC.2022.3161719.

[246] J. Wang, J. Jiang, D. Zhang, Y.z. Zhang, L. Guo, Y. Jiang, et al., An integrated AI model to improve diagnostic accuracy of ultrasound and output known risk features in suspicious thyroid nodules, Eur Radiol. 32 (2022) 2120–2129, https://doi.org/10.1007/s00330-021-08298-7.

[247] K. Borkar, A. Chaturvedi, P.K. Vinod, R.S. Bapi, Ayu-Characterization of healthy aging from neuroimaging data with deep learning and rsfMRI, Front Comput Neurosci. 16 (2022), https://doi.org/10.3389/fncom.2022.940922.

[248] A. Gotsopoulos, H. Saarimäki, E. Glerean, I.P. Jääskeläinen, M. Sams, L. Nummenmaa, et al., Reproducibility of importance extraction methods in neural network based fMRI classification, NeuroImage 181 (2018) 44–54, https://doi.org/10.1016/j.neuroimage.2018.06.076.

[249] A. Safai, N. Vakharia, S. Prasad, J. Saini, A. Shah, A. Lenka, et al., Multimodal Brain Connectomics-Based Prediction of Parkinson's Disease Using Graph Attention Networks, Frontiers in Neuroscience 15 (2022) 1–13, https://doi.org/10.3389/fnins.2021.741489.

[250] Chun JY, Sendi MSE, Sui J, Zhi D, Calhoun VD. Visualizing Functional Network Connectivity Difference between Healthy Control and Major Depressive Disorder Using an Explainable Machine-learning Method. 2020 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., IEEE; 2020, p. 1424–7. 10.1109/EMBC44109.2020.9175685.

[251] R.J.H. Miller, K. Kuronuma, A. Singh, Y. Otaki, S. Hayes, P. Chareonthaitawee, et al., Explainable deep learning improves physician interpretation of myocardial perfusion imaging, Journal of Nuclear Medicine 63 (2022) 1768–1774, https://doi.org/10.2967/jnumed.121.263686.

[252] M. Nazari, A. Kluge, I. Apostolova, S. Klutmann, S. Kimiaei, M. Schroeder, et al., Explainable AI to improve acceptance of convolutional neural networks for automatic classification of dopamine transporter SPECT in the diagnosis of clinically uncertain parkinsonian syndromes, European Journal of Nuclear Medicine and Molecular Imaging 49 (2022) 1176–1186, https://doi.org/10.1007/s00259-021-05569-9.

[253] M. Nazari, A. Kluge, I. Apostolova, S. Klutmann, S. Kimiaei, M. Schroeder, et al., Data-driven identification of diagnostically useful extrastriatal signal in dopamine transporter SPECT using explainable AI, Scientific Reports 11 (2021) 1–13, https://doi.org/10.1038/s41598-021-02385-x.

[254] N.I. Papandrianos, A. Feleki, S. Moustakidis, E.I. Papageorgiou, I. D. Apostolopoulos, D.J. Apostolopoulos, An explainable classification method of SPECT myocardial perfusion images in nuclear cardiology using deep learning and grad-CAM, Appl Sci. 12 (2022), https://doi.org/10.3390/app12157592.

[255] A. Singh, R.J.H. Miller, Y. Otaki, P. Kavanagh, M.T. Hauser, E. Tzolos, et al., Direct risk assessment from myocardial perfusion imaging using explainable deep learning, JACC: Cardiovascular Imaging 16 (2023) 209–220, https://doi.org/10.1016/j.jcmg.2022.07.017.

[256] N. Amanova, J. Martin, C. Elster, Explainability for deep learning in mammography image quality assessment, Mach Learn Science and Technology 3 (2022), https://doi.org/10.1088/2632-2153/ac7a03.

[257] F. Lizzi, C. Scapicchio, F. Laruina, A. Retico, M.E. Fantacci, Convolutional neural networks for breast density classification: Performance and explanation insights, Applied Sciences 12 (2022), https://doi.org/10.3390/app12010148.

[258] A.J. Barnett, F.R. Schwartz, C. Tao, C. Chen, Y. Ren, J.Y. Lo, et al., A case-based interpretable deep learning model for classification of mass lesions in digital mammography, Nat Mach Intell 3 (2021) 1061–1070, https://doi.org/10.1038/s42256-021-00423-x.

[259] Kraaijveld RCJ, Philippens MEP, Eppinga WSC, Jürgenliemk-Schulz IM, Gilhuijs KGA, Kroon PS, et al. Multi-modal Volumetric Concept Activation to Explain Detection and Classification of Metastatic Prostate Cancer on PSMA-PET/CT. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2022;13611 LNCS:82–92. 10.1007/978-3-031-17976-1_8.

[260] Y. Glaser, J. Shepherd, L. Leong, T. Wolfgruber, L.-Y. Lui, P. Sadowski, et al., Deep learning predicts all-cause mortality from longitudinal total-body DXA imaging, Communication & Medicine 2 (2022) 1–12, https://doi.org/10.1038/s43856-022-00166-9.

[261] J.P. Cruz-Bastida, E. Pearson, H. Al-Hallaq, Toward understanding deep learning classification of anatomic sites: lessons from the development of a CBCT projection classifier, Journal of Medical Imaging 9 (2022), https://doi.org/10.1117/1.JMI.9.4.045002.

[262] Beddiar D, Oussalah M, Tapio S. Explainability for Medical Image Captioning. 2022 11th Int Conf Image Process Theory, Tools Appl IPTA 2022 2022. 10.1109/IPTA54936.2022.9784146.

[263] S.M. Hussain, D. Buongiorno, N. Altini, F. Berloco, B. Prencipe, M. Moschetta, et al., Shape-based breast lesion classification using digital tomosynthesis images: The role of explainable artificial intelligence, Appl Sci. 12 (2022), https://doi.org/10.3390/app12126230.

Further reading

[189] Y. Zhang, D. Hong, D. McClement, O. Oladosu, G. Pridham, G. Slaney, Grad-CAM helps interpret the deep learning models trained to classify multiple sclerosis types using clinical brain magnetic resonance imaging, Journal of Neuroscience Methods 353 (2021), 109098, https://doi.org/10.1016/j.jneumeth.2021.109098.

[199] Corizzo R, Dauphin Y, Bellinger C, Zdravevski E, Japkowicz N. Explainable image analysis for decision support in medical healthcare. Proc - 2021 IEEE Int Conf Big Data, Big Data 2021 2021:4667–74. 10.1109/BigData52589.2021.9671335.

[200] J. Huynh, S. Masoudi, A. Noorbakhsh, A. Mahmoodi, S. Kligerman, A. Yen, et al., Deep Learning Radiographic Assessment of Pulmonary Edema: Optimizing Clinical Performance, Training with Serum Biomarkers. IEEE Access 10 (2022) 48577–48588, https://doi.org/10.1109/ACCESS.2022.3172706.

[201] N. Liao, J. Dai, Y. Tang, Q. Zhong, S. Mo, iCVM: An Interpretable Deep Learning Model for CVM Assessment Under Label Uncertainty, IEEE J Biomed Heal Informatics 26 (2022) 4325–4334, https://doi.org/10.1109/JBHI.2022.3179619.

[202] Mondal AK. Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19 . The COVID-19 resource centre is hosted on Elsevier Connect , the company ' s public news and information 2020.

[203] C.T. Wang, B. Huang, N. Thogiti, W.X. Zhu, F. Lai, Successful real-world application of an osteoarthritis classification deep-learning model using 9210 knees—An orthopedic surgeon's view, Journal of Orthopaedic Research (2022), https://doi.org/10.1002/jor.25415.

[220] Y. Wan, H. Zhou, X. Zhang, An interpretation architecture for deep learning models with the application of COVID-19 diagnosis, Entropy 23 (2021) 1–19, https://doi.org/10.3390/e23020204.

[221] Y. Yin, D. Yakar, R.A.J.O. Dierckx, K.B. Mouridsen, T.C. Kwee, R.J. de Haas, Liver fibrosis staging by deep learning: a visual-based explanation of diagnostic decisions of the model, European Radiology 31 (2021) 9620–9627, https://doi.org/10.1007/s00330-021-08046-x.

[222] Yu W, Zhou H, Choi Y, G.goldin J, Kim GHJ. Mga-Net: Multi-scale guided attention models for an automated diagnosis of idiopathic pulmonary fibrosis (IPF). Proc. - Int. Symp. Biomed. Imaging, vol. 2021- April, 2021, p. 1777–80. 10.1109/ISBI48211.2021.9433956.

[233] M.C. Walker, I. Willner, O.X. Miguel, M.S.Q. Murphy, D. El-Chaâr, F. Moretti, et al., Using deep-learning in fetal ultrasound analysis for diagnosis of cystic hygroma in the first trimester, PLoS One1 17 (2022) 88–98, https://doi.org/10.1371/journal.pone.0269323.

[234] X. Wei, J. Zhu, H. Zhang, H. Gao, R. Yu, Z. Liu, et al., Visual interpretability in computer-assisted diagnosis of thyroid nodules using ultrasound images, Med Sci Monit. 26 (2020), https://doi.org/10.12659/MSM.927007.

[235] S. Wu, Y. Ren, X. Lin, Z. Huang, Z. Zheng, X. Zhang, Development and validation of a composite AI model for the diagnosis of levator ani muscle avulsion, European Radiology 32 (2022) 5898–5906, https://doi.org/10.1007/s00330-022-08754-y.

[236] M. Byra, K. Dobruch-Sobczak, H. Piotrzkowska-Wroblewska, Z. Klimonda, J. Litniewski, Explaining a deep learning based breast ultrasound image classifier with saliency maps, J Ultrason. 22 (2022), https://doi.org/10.15557/JoU.2022.0013.

[237] F. Dong, R. She, C. Cui, S. Shi, X. Hu, J. Zeng, et al., One step further into the blackbox: a pilot study of how to build more confidence around an AI-based decision system of breast nodule assessment in 2D ultrasound, European Radiology 31 (2021) 4991–5000, https://doi.org/10.1007/s00330-020-07561-7.

[238] G. Dong, Y. Ma, A. Basu, Feature-Guided CNN for Denoising Images from Portable Ultrasound Devices, IEEE Access 9 (2021) 28272–28281, https://doi.org/10.1109/ACCESS.2021.3059003.

[239] G. Duffy, I. Jain, B. He, D. Ouyang, Interpretable deep learning prediction of 3d assessment of cardiac function, Pacific Symposium on Biocomputing 27 (2022) 231–241, https://doi.org/10.1142/9789811250477_0022.

[240] A. Patra, J.A. Noble, Incremental Learning of Fetal Heart Anatomies Using Interpretable Saliency Maps, Hospitals (lond) (2020) 129–141, https://doi.org/10.1007/978-3-030-39343-4_11.

[241] X. Qian, J. Pei, H. Zheng, X. Xie, L. Yan, H. Zhang, et al., Prospective assessment of breast cancer risk from multimodal multiview ultrasound images via clinically applicable deep learning, Nature Biomedical Engineering 5 (2021) 522–532, https://doi.org/10.1038/s41551-021-00711-2.

[242] M. Vafaeezadeh, H. Behnam, A. Hosseinsabet, P. Gifani, Automatic morphological classification of mitral valve diseases in echocardiographic images based on explainable deep learning methods, International Journal of Computer Assisted Radiology and Surgery 17 (2022) 413–425, https://doi.org/10.1007/s11548-021-02542-7.

[243] B. VanBerlo, D. Wu, B. Li, M.A. Rahman, G. Hogg, B. VanBerlo, et al., Accurate assessment of the lung sliding artefact on lung ultrasonography using a deep learning approach, Computers in Biology and Medicine 148 (2022), https://doi.org/10.1016/j.compbiomed.2022.105953.