Advances in artificial intelligence and machine learning applications for the imaging of bone and soft tissue tumors,

2nd edition

Edited by
Brandon K. K. Fields, George R. Matcuk Jr. and
Bino A. Varghese

Published in
Frontiers in Radiology

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Advances in artificial intelligence and machine learning applications for the imaging of bone and soft tissue tumors, 2nd edition

Topic editors

Brandon K. K. Fields — University of California, San Francisco, United States
George R. Matcuk Jr. — Cedars Sinai Medical Center, United States
Bino A. Varghese — University of Southern California, United States

Citation

Fields, B. K. K., Matcuk Jr., G. R., Varghese, B. A., eds. (2025). *Advances in artificial intelligence and machine learning applications for the imaging of bone and soft tissue tumors, 2nd edition*. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-8325-5863-8

Publisher's note: In this 2nd edition, the following article has been updated:

Fields BKK, Varghese BA and Matcuk GR Jr (2024) Editorial: Advances in artificial intelligence and machine learning applications for the imaging of bone and soft tissue tumors. *Front. Radiol.* 4:1523389. doi: 10.3389/fradi.2024.1523389

Table of
contents

Editorial: Advances in artificial intelligence and machine learning applications for the imaging of bone and soft tissue tumors

Brandon K. K. Fields[1]*, Bino A. Varghese[2] and
George R. Matcuk Jr[3]

[1]Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco,
CA, United States, [2]Department of Radiology, Keck School of Medicine, University of Southern
California, Los Angeles, CA, United States, [3]Department of Imaging, Cedars-Sinai Medical Center, Los
Angeles, CA, United States

Editorial on the Research Topic
Advances in artificial intelligence and machine learning applications for
the imaging of bone and soft tissue tumors

Growing interest in artificial intelligence (AI) applications in the biological and medical sciences has led to a notable surge in related publications in recent years. A PubMed query for "machine learning" yielded 33,855 results published in the year of 2023 alone, which represents a greater than fourfold increase in comparison to 2018 (1). In the realm of musculoskeletal and soft tissue tumor imaging, there has been increasing enthusiasm for a broad range of AI applications ranging from prognostication and risk stratification to lesion classification and treatment response (Varghese et al., 2–4).

This Research Topic, titled "*Advances in Artificial Intelligence and Machine Learning Applications for the Imaging of Bone and Soft Tissue Tumors*," encompasses a selection of manuscripts that primarily focus on recent developments in AI applications targeting the imaging of these neoplasms while also exploring topics related to quantitative image analysis and interpretation. The following representative manuscripts provide insights into current advancements and methodologies within this evolving field:

Jin et al. describe a machine learning algorithm based on Gray Level Co-occurrence Matrix (GLCM) scores to detect pelvic bone metastases in patients with colorectal cancer using a retrospective cohort of 614 patients who underwent MRI of the pelvis over a 7-year period. Diffusion-weighted images were segmented, from which 48 GLCM features were extracted for further analysis. A generalized linear regression model and four machine learning classification models were then constructed from this dataset. Random forest achieved the highest performance, with AUCs of 0.926 and 0.919 in the training and internal validation sets, respectively. Their results suggest a promising role for radiomics-based machine learning models in detecting pelvic bone metastases. By integrating these advanced modeling techniques into clinical workflows, healthcare

Abbreviations
AI, artificial intelligence; GLCM, Gray Level Co-occurrence Matrix.

providers may enhance their ability to identify osseous metastatic disease more reliably, thereby potentially improving prognostic assessments and guiding treatment interventions tailored to individual patient needs.

Rich et al. publish a systematic review and meta-analysis of deep learning image segmentation approaches for the evaluation of primary and secondary malignant bone tumors. Accurate delineation of osseous lesions on imaging is crucial for performing quantitative analyses, yet is often the rate-limiting step in AI and machine learning studies when performed manually. Additionally, precise delineation of the extent of bony involvement of disease may allow for more holistic clinical assessments and prognostication in situations of widespread metastatic disease. In their article, Rich et al. conducted a comprehensive literature search of studies investigating automated deep-learning segmentation methods for malignant bony lesions, identifying 41 studies published between 2010 and 2023 suitable for inclusion in the final analysis. Overall, they found that most studies applied a U-Net convolutional neural network architecture, most often trained on either MRI or CT images. Models trained on PET/MRI and PET/CT had lower median dice similarity coefficients when compared to models trained on MRI and CT, possibly due to increased image noise and degradation. Furthermore, models trained on 2D data appeared to have slightly higher median dice similarity coefficients, though the difference was not statistically significant. Future optimizations and validations could include training of models on larger, multi-institutional datasets to allow for improved generalizability.

Debs and Fayad and Sabeghi et al. both describe emerging applications of AI and machine learning in musculoskeletal imaging. Debs and Fayad review use cases ranging from image protocoling, examination scheduling, and hanging protocol optimization to results reporting, lesion detection, and lesion classification. They also detail approaches for determining tumor of origin and assessing for treatment response of metastatic spinal lesions. Sabeghi et al. comment on methods for stratifying neoplastic lesions with respect to malignant potential, histologic grade, and response to treatment. They further discuss challenges with assembling sufficiently large, diverse, and heterogeneous datasets to train robust, generalizable models. As models become increasingly more complex and advanced, federated learning may offer an elegant solution for leveraging multicentric training sets without the need for centralized aggregation of potentially sensitive and protected patient data (5), though technical limitations and complexities associated with managing heterogeneous systems and data pose major hurdles.

Gadermayr et al. investigate computer vision and deep learning approaches for automated muscle segmentation of MR images. The authors employ an unpaired image-to-image translation approach to leverage "easy" data in order to improve segmentation performance of "hard" data. Using a novel domain specific loss function alongside four segmentation schemes, namely Gaussian mixture, graph-cut, shape-priority graph-cut, and convolutional neural network approaches. Their results suggest performance

benefits for both unsupervised and supervised methods, with the potential to reduce the amount of necessary training data in related studies.

Woznicki et al. developed an open-source framework termed "AutoRadiomics", which seeks to aggregate all common steps of typical radiomics workflows into a single standardized software package. Their framework provides embedded tools for image segmentation and pre-processing alongside standardized radiomics libraries and machine learning models, and provides multiple optimizations for hyperparametric tuning, data splitting, and oversampling. AutoRadiomics seeks to lower the barrier to entry for radiomics and machine learning studies through a modular, user-friendly interface and intuitive design, promising accessibility without the need for a robust coding background. Consistent application of publicly accessible frameworks such as AutoRadiomics in future radiomics studies can serve to enhance transparency by improving workflow standardization and reproducibility.

Finally, Varghese et al. discuss applications of spatial assessments of texture analysis in oncologic imaging. Spatial assessments are particularly effective in capturing areas of intratumoral heterogeneity as they have the ability to quantify subtle voxel-to-voxel variations in the underlying grayscale intensities. This granular approach allows for a more nuanced understanding of tumor characteristics, which is critical for accurate diagnosis and treatment planning. These spatial assessments can be further classified as neighborhood-based methods, which quantify differences in grayscale intensities relative to neighboring voxels, and spatial filters, which are image processing methods that enhance edges and/or textures at regions of rapid grayscale change. Additionally, the authors also underscore the significance of utilizing optimal methodologies and best practices when conducting statistical analyses in high-dimensional radiomics studies. They offer guidance on accurate reporting of machine learning performance outcomes to ensure that findings are both reliable and reproducible.

In conclusion, this Research Topic highlights many innovative developments in AI and machine learning applications for the imaging of bone and soft tissue neoplasms. While radiomics-based studies have for some time shown promise in serving as novel quantitative imaging biomarkers, deep learning approaches have, in more recent years, also gained traction as powerful decision support tools which integrate diverse hierarchical and multimodal inputs. As imaging-based AI algorithms continue to evolve in complexity and rigor, there exists immense potential for creating an armamentarium of tools aimed at augmenting the work of clinical radiologists and enhancing patient care.

Author contributions

BKKF: Conceptualization, Writing – original draft, Writing – review & editing. BAV: Writing – review & editing,

Conceptualization. GRM: Writing – review & editing, Conceptualization.

Conflict of interest

BKKF received prior research grants from the RSNA R&E (2019-2020 RMS #1909; 2018-2019 RMS #1810); consulting fees from Mendaera; honorarium payments from Neurodiem (invited author) and Elsevier (book proposal reviews); RSNA and institutional support for attending meetings (RSNA RFC stipend, institutional support stipend); Vice-Chair of the RSNA Resident and Fellow Committee; member of the American Board of Radiology Initial Certification Advisory Committee for Diagnostic Radiology, of the RSNA Education Council, and of the Radiology: Imaging Cancer Trainee Editorial Board. GRM is a consultant for Canon Medical Systems, USA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The authors BKKF, BAV, and GRM declared that they were editorial board members of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Pubmed website. Available online at: https://pubmed.ncbi.nlm.nih.gov/?term=machine+learning&size=200 (accessed November 5, 2024).

2. Fields BKK, Hwang D, Cen S, Desai B, Gulati M, Hu J, et al. Quantitative magnetic resonance imaging (q-MRI) for the assessment of soft-tissue sarcoma treatment response: a narrative case review of technique development. *Clin Imaging.* (2020) 63:83–93. doi: 10.1016/j.clinimag.2020.02.016

3. Fields BKK, Demirjian NL, Hwang DH, Varghese BA, Cen SY, Lei X, et al. Whole-tumor 3D volumetric MRI-based radiomics approach for distinguishing between benign and malignant soft tissue tumors. *Eur Radiol.* (2021) 31 (11):8522–35. doi: 10.1007/s00330-021-07914-w

4. Fields BKK, Demirjian NL, Cen SY, Varghese BA, Hwang DH, Lei X, et al. Predicting soft tissue sarcoma response to neoadjuvant chemotherapy using an MRI-based delta-radiomics approach. *Mol Imaging Biol.* (2023) 25(4):776–87. doi: 10.1007/s11307-023-01803-y

5. Rehman MHU, Pinaya WHL, Nachev P, Teo JT, Ourselin S, Cardoso MJ. Federated learning for medical imaging radiology. *Br J Radiol.* (2023) 96 (1150):20220890. doi: 10.1259/bjr.20220890

Image-to-Image Translation for Simplified MRI Muscle Segmentation

Michael Gadermayr[1,2]*, Lotte Heckmann[2], Kexin Li[2], Friederike Bähr[3], Madlaine Müller[3,4], Daniel Truhn[5], Dorit Merhof[2,6] and Burkhard Gess[3,7]

[1] Department of Information Technology and Systems Management, Salzburg University of Applied Sciences, Salzburg, Austria, [2] Institute of Imaging & Computer Vision, RWTH Aachen University, Aachen, Germany, [3] Department of Neurology, RWTH Aachen, University Hospital Aachen, Aachen, Germany, [4] Department of Neurology, Inselspital Bern, Bern, Switzerland, [5] Department of Diagnostic and Interventional Radiology, University Hospital Aachen, Aachen, Germany, [6] Fraunhofer Institute for Digital Medicine MEVIS, Bremen, Germany, [7] Department of Neurology, Evangelisches Klinikum Bethel, Universitätsklinikum OWL, Bielefeld, Germany

Deep neural networks recently showed high performance and gained popularity in the field of radiology. However, the fact that large amounts of labeled data are required for training these architectures inhibits practical applications. We take advantage of an unpaired image-to-image translation approach in combination with a novel domain specific loss formulation to create an "easier-to-segment" intermediate image representation without requiring any label data. The requirement here is that the task can be translated from a hard to a related but simplified task for which unlabeled data are available. In the experimental evaluation, we investigate fully automated approaches for segmentation of pathological muscle tissue in T1-weighted magnetic resonance (MR) images of human thighs. The results show clearly improved performance in case of supervised segmentation techniques. Even more impressively, we obtain similar results with a basic completely unsupervised segmentation approach.

Keywords: MRI, muscle, fatty-infiltration, thigh, generative adversarial networks, convolutional neural networks, segmentation, image processing

1. INTRODUCTION

Within the last few years, deep neural networks showed impressive performance and gained popularity in the field of radiology. However, the requirement for large amounts of labeled data for artificial neural network training still inhibits practical applications. Since three-dimensional (3D) data requires complex models, this is particularly challenging in radiology. In addition, voxel-based 3D data annotation is highly time consuming. Another challenging aspect is given by an often high variability within radiological data. Although variability due to the imaging setting can be compensated by methods such as bias field correction (1) and contrast adjustment (2), semantic variability caused by pathological modifications is hard to compensate.

Due to emerging techniques, such as fully convolutional neural networks (3) and adversarial networks (4), image-to-image translation has recently gained popularity (5–7). These methods enable, for example, a translation from one imaging modality to another (such as MRI to CT and vice versa) (8). Conventional approaches require image pairs (e.g., pairs consisting of a CT and an MRI scan of the same subject) for training the translation models (5, 6). To overcome the restriction of training based on image pairs, unpaired approaches were introduced (7, 9, 10) and also applied to radiology (8, 11, 12). These models only require two data sets, one for each of the modalities [e.g., computed tomography (CT) and magnetic resonance imaging (MRI)]. As image pairs are often not achievable or at least very difficult and expensive to collect, this opens up completely new perspectives for many radiological application scenarios.

For example, if trained models (and especially manually annotated training data) are available for one modality only, data collected based on a different imaging setting can be translated to this modality and can be subsequently processed without further annotation effort.

In this paper, we do not consider a translation from one imaging modality to another using cycle-GAN (7). Instead, we consider a scenario where a certain domain (i.e., a subset of the available data; e.g., non-pathological data) is easier to segment than another domain (13). Image-to-image translation can be applied here to translate from a hard-to-segment image domain to an easy-to-segment domain. If translation is performed appropriately, this approach has the potential to facilitate further processing (here segmentation) and thereby enhance accuracy (e.g., segmentation accuracy) to reduce the amount of required annotated training data or even to facilitate fully unsupervised segmentation.

1.1. Thigh Muscle Segmentation

Muscular dystrophy is a class of diseases caused by inherited mutations in genes encoding for proteins that are essential to the health and function of muscles. They are characterized by a degeneration of muscle tissue, which in muscle imaging appears as so-called fatty infiltration (see **Figures 1C,D** for example MR images). A relevant disease marker is especially given by the so-called fat fraction capturing the ratio between fatty-infiltration and original muscle tissue volume. For computation of the fat fraction, it is crucial to segment the overall muscle tissue including fatty infiltrations. Although a segmentation of healthy muscle tissue (see **Figure 1**) can be obtained easily based on thresholding, difficulties arise in case of severely fat-infiltrated muscle as fatty degenerated muscle tissue cannot be distinguished from subcutaneous fat based on the image's gray values (14) (**Figure 1D**). This problem has been recently addressed in a few studies. Origiu et al. (15) developed an active contours model to detect the muscle boundary and a fuzzy c-means method to distinguish muscle from fat. Gadermayr et al. (14) combined graph-cuts and level-set approaches with statistical shape models. Yao et al. (18) made use of two neural networks to first detect the fascia lata and also incorporate region-based information to finally utilize an active contours method. Although showing best segmentation performance, the latter approach as well as further ones (16–18) are optimized and evaluated on an easier scenario, because all tissue inside the fascia lata is labeled as muscle (apart from the bone).

1.2. Contributions

In this work, we make use of a new procedure for facilitating segmentation tasks in order to boost segmentation accuracy. In our approach, a hard segmentation task is mapped to an easier (intermediate) segmentation task by means of unpaired image-to-image translation making use of a cyclic GAN (7). We consider the segmentation of MR images of human thighs showing fatty infiltrations, which are translated to easy-to-segment non-pathological images. For segmentation, we consider methodologies that proved to be effective in previous works (14, 15, 18). Even though we were unable to investigate each individual configuration, we focus on covering a broad range

of techniques, namely a pixel-based unsupervised approach, a region based method, a region-based method using shape prior, and a convolutional neural network.

2. MATERIALS AND METHODS

In this work, we first perform image-to-image translation to convert a hard-to-segment into an easy-to-segment domain (section 2.1). After conversion to the intermediate "easy" representation, only the generated fake image is segmented (section 2.2) and the obtained mask is simply mapped to the original image without making any changes.

2.1. Image-to-Image Translation

Supposed we have a set of images $\{h_i\}_{j=1}^{N}$ of a "hard" domain ($\mathcal{H}$), which are difficult to segment, as well as a set of images $\{e_i\}_{j=1}^{M}$ of an "easy" domain ($\mathcal{E}$). Although the underlying distributions (based on the empirical ones $e \sim p_{data}(e)$ and $h \sim p_{data}(h)$) are different, we assume that the underlying distribution of the corresponding ground-truth segmentations s ($s_e \sim p_{data}(s(e))$ and $s_h \sim p_{data}(s(h))$) is similar. Then it follows that, based on a segmentation only, the domain of an image ($\mathcal{H}$ vs. $\mathcal{E}$) cannot be predicted with a higher accuracy than chance. Thus, the translated images could also become indistinguishable even if the segmentation mask stays the same, which is the crucial criterion for this approach. Otherwise, in a GAN setting, the generator would be forced by the discriminator to change the object's shape with the implication that the segmentation of the original $\mathcal{H}$ domain image would not be the same as for the fake $\mathcal{E}$ domain image. As we finally directly map the obtained segmentation mask from the fake $\mathcal{E}$ to the real $\mathcal{H}$ domain image without making any changes, the similarity of the object's shapes is a strong requirement. Inspecting the considered MRI data, we notice high variability between patients in general but no systematic differences in the shapes between the datasets.

Now we focus on a domain adaptation from $\mathcal{H}$ to $\mathcal{E}$ by performing image-to-image translation, specifically by means of a cyclic GAN (7). This method requires only one dataset for each domain without corresponding pairs. During GAN training, two mapping functions, $F \colon \mathcal{H} \rightarrow \mathcal{E}$ and $G \colon \mathcal{E} \rightarrow \mathcal{H}$ are trained optimizing a combination of a cycle consistency loss

$$
\begin{aligned}
\mathcal{L}_c = \; & \mathbb{E}_{e \sim p_{data}(e)}[||F(G(e)) - e||_1] + \\
& \mathbb{E}_{h \sim p_{data}(h)}[||G(F(h)) - h||_1]
\end{aligned}
\tag{1}
$$

as well as a discriminator loss

$$
\begin{aligned}
\mathcal{L}_d = \; & \mathbb{E}_{h \sim p_{data}(h)}[\log(D_H(h)) + \log(1 - D_E(F(h)))] + \\
& \mathbb{E}_{e \sim p_{data}(e)}[\log(1 - D_H(G(e))) + \log(D_E(e))]
\end{aligned}
\tag{2}
$$

encouraging indistinguishable outputs (based on the discriminators D_H and D_E). As the underlying distributions of ground-truth segmentations s_h and e_e are similar, and as there is a correlation between image information and the ground-truth segmentation (which is a natural requirement for all segmentation applications), it can be expected that during
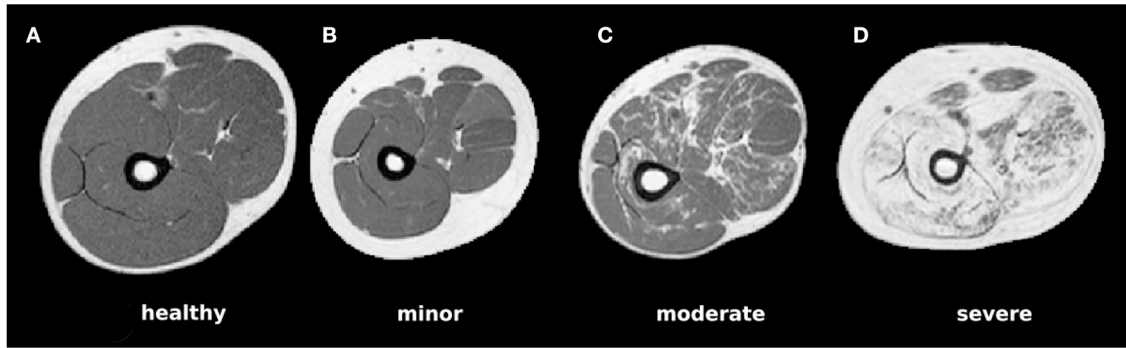
FIGURE 1 | Example MRI slices for each of the four considered pathological categories showing **(A)** healthy muscle only, **(B)** pathological muscle without visible fatty infiltrations, **(C)** moderate infiltrations, and **(D)** largely affected muscle areas.

image-to-image translation using a cyclic GAN (7), the images are translated from domain $\mathcal{H}$ to $\mathcal{E}$ without changing the semantic structure in the image (i.e., the shape of the muscle). To account for the specific application scenario, we introduce a further loss function based on the rectified linear unit (ReLU) r

$$\mathcal{L}_r = \mathbb{E}_{h \sim p_{data}(h)}[r(F(h) - h)] + \\ \mathbb{E}_{e \sim p_{data}(e)}[r(e - G(e))] , \qquad (3)$$

where $r(x) = max(0, x)$. This method is introduced in order to account for the fact that healthy muscle tissue in MR images shows a lower voxel value than pathological muscle tissue. For this purpose, if muscle tissue is translated from $\mathcal{H}$ to $\mathcal{E}$, voxel values should not increase, but only decrease. Vice versa, from $\mathcal{E}$ to $\mathcal{H}$, voxel values should only increase and not decrease. By adding this further constraint, we expect that the overall structure and consequently also the segmentation could be maintained more effectively. This domain specific loss is finally combined with the identity loss

$$\mathcal{L}_i = \mathbb{E}_{e \sim p_{data}(e)}[||F(e) - e||_1] + \\ \mathbb{E}_{h \sim p_{data}(h)}[||G(h) - h||_1] \qquad (4)$$

to focus on maintaining the morphology and to ensure that data from the easy domain $\mathcal{E}$ does not get extremely dark due to $\mathcal{L}_r$. All utilized losses are summarized in **Figure 2**.

2.2. Segmentation

For segmentation, we make use of four methods that were applied to muscle segmentation tasks. Due to the rather small amount of data for training, we focus on the following methods that can be effectively trained with a small amount of data. The first approach is based on the **Gaussian Mixture Model (GMM)**, which is fitted to the data in order to identify clusters of three different classes: muscle, fat, and bone/vessels. Initial cluster centers are fixed to the minimum gray value ($s_{\min}$), maximum gray value ($s_{\max}$), and finally a value in between ($s_{\min} + \frac{s_{\max} - s_{\min}}{6}$). This method is completely unsupervised and does not require any training data. In order to incorporate boundary smoothness constraints, we furthermore investigate

a probabilistic **Graph-Cut (GC)** technique (the initialization is obtained by the GMM and the probabilistic model is trained based on ground-truth annotations). To additionally incorporate a statistical shape model, we make use of the **Shape-Prior Graph-Cut (SPGC)** approach (14). In this case, the shape model (which is optimized for small data sets) is trained by estimating a probability map for each pixel after an initial registration (leading to excellent performance for pathological images). SPGC and GC both require annotated training data as the probabilistic model need to be trained on ground-truth data. Details on these approaches are provided in (14). As reference for a state-of-the-art **convolutional neural network (CNN)** approach, we apply a 2D U-Net (3) including a GAN-Loss, also referred to as Pix2Pix network (5). In this data-driven approach, a segmentation model (implicitly including a shape prior) is automatically learned during optimization of the weights of the convolutional neural networks.

2.3. Experimental Details

The T1-weighted MR images were acquired on a 1.5 Tesla Phillips device with fixed echo time (17 ms), bandwidth (64 kHz) and echo train length (6) and a relaxation time between 721 and 901 ms. The sampling interval was fixed to 1 mm in x-y-direction and 7 mm in z-direction. Bias-field correction was applied to compensate homogeneity (19). Similar to (14, 18), the data are separated into the four categories "healthy," "minor," "moderate," and "severe" corresponding to the degree of fatty infiltration. As the categories "healthy" and "easy" can be rather easily segmented with existing approaches (14), they are not considered during evaluation. Healthy (and easy) scans could also be translated with the proposed pipeline, but remain almost unchanged. Binary ground-truth was acquired to cover muscle volume only, also excluding small fascias (**Figure 4a**). Due to high correlation of consecutive slices and to limit manual effort, each forth slice (transversal plane) was annotated under strong supervision of a medical expert (Madlaine Müller). For parameter optimization of the segmentation stage, grid search combined with leave-one-out cross-validation is applied to determine the best combination individually for both datasets. The parameters of the graph-cut approaches consist of curvature weight $\lambda_s \in$

FIGURE 2 | An illustration depicting the individual losses $\mathcal{L}_d, \mathcal{L}_c, \mathcal{L}_i, \mathcal{L}_r$ contributing to the overall loss on a high level perspective.
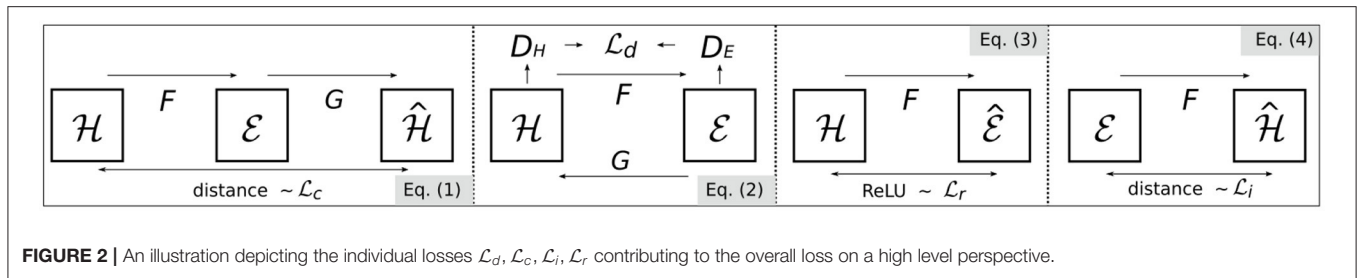
[0.001, 0.002, 0.05, 0.1, 0.2, 0.5], low-pass filtering weight $\sigma \in$ [1, 2], shape prior weight $\lambda_{sp} \in [0.1, 0.2, 0.5, 0.7, 1]$, and neutral probability $p_n \in [0.2, 0.3, 0.4, 0.5]$. The CNN segmentation approach is trained for 200 epochs with learning rate 0.0002 for each setting and each fold. Fourfold cross-validation is conducted. For data augmentation, random cropping (256×256 patches from images padded to 300×300 pixels), rotations with multiples of $90°$ and flipping is applied. For further parameters, we use the defaults from the pytorch reference implementation.

For image translation, a cyclic GAN (based on a ResNet with 9 blocks as generator and the proposed patchwise CNN as discriminator) (7) is trained for 200 epochs with learning rate 0.0002 based on a "hard" and an "easy" dataset. The "easy" dataset contains 2D slices showing "healthy" and "minor" data both showing no visible fatty-infiltrations and the "hard" dataset contains "moderate" and "severe" images. The individual sets are merged to maximize the number of training images (overall, we obtain 649 "hard" (from 19 patients) and 1,124 "easy" 2D images (20 patients) with a size of 256×256 pixels). The losses $\mathcal{L}_d$ and $\mathcal{L}_c$ are equally weighted ($w_d = 1$, $w_c = 1$) (7). For w_i and w_r (corresponding to $\mathcal{L}_i$ and $\mathcal{L}_r$), several relevant parameters are evaluated as shown in **Figure 3**. The standard GAN setting is evaluated with $w_i = 0$ and $w_i = 1$ ($G_{0,0}$, $G_{0,1}$) and three settings for $w_r > 0$ are evaluated with $w_i = 1$ ($G_{.5,1}$, $G_{1,1}$, $G_{2,1}$). In the latter case, the identity loss is required in order to prevent the GAN from generating extremely dark fake-"healthy" MRI scans.

3. RESULTS

Figure 3 shows the segmentation performance individually for the four segmentation methods (GMM, GC, SPGC, CNN) and for the different GAN configurations ($G_{n,m}$, with n and m defining the weights such that $w_r = n$ and $w_i = m$). For completely unsupervised segmentation using GMM, the baseline relying on original images (*OI*) is outperformed clearly. The best median DSCs are obtained with the GAN setting $G_{0.5,1}$ (DSC: 0.82 compared to 0.67 in case of *OI*). A similar effect is observed for GC. The benefit of image translation is clearly smaller in case of SPGC and CNN. For all configurations, $G_{0.5,1}$ exhibits the best DSCs with scores of 0.85/0.82/0.86 compared to 0.83/0.72/0.83 in case of *OI* and 0.83/0.82/0.86 in case of the standard cycle-GAN configuration $G_{0,0}$ for the approaches SPGC/GC/CNN.

Example image translation output and example segmentations for GMM and SPGC are provided in **Figures 4a–g**. Results

show clear improvements for the rather basic methods GMM and GC, which fail in case of original pathological data. For the methods that are capable of learning the shape of the muscles, even with *OI* median scores above 0.84 are achieved. Even for SPGC and CNN further improvements are achieved in case of image translation (CNN: 0.86 compared to 0.83). The bottom row of **Figure 4** additionally shows the impact of different image translation settings for an example image.

4. DISCUSSION

Making use of unpaired image-to-image translation, we propose a methodology to facilitate segmentation tasks for specific scenarios where a hard problem can be mapped to an easier task. The most impressive performance gain is observed in case of fully unsupervised segmentation (GMM) applied to the "severe" data, which was expected due to the high degree of fatty infiltrations complicating a pixel-level classification without contextual knowledge. However, also with probabilistic graph-cuts with (GC) or without a statistical shape model (SPGC) and even for the deep learning based approach (CNN), a slight increase of performance with image translation is observed. For the latter, this is not completely obvious since the segmentation network should be capable of learning the same invariance to pathological data as the translation model. However, for learning the translation model, all available data could be used and not only the annotated data (each forth slice only), which is supposed to be a clear advantage due to the small training data sets. Related work investigating a similar application in digital pathology also suggests that two individual networks performing a task in two steps can be advantageous (20).

Considering the different GAN configuration, we note that especially the introduction of the new loss $\mathcal{L}_r$ leads to best median DSCs and the configuration $G_{.5,1}$ is never outperformed by any other GAN configuration.

By considering the qualitative results (**Figure 4**), we note that the converted images (in case of $G_{.5,1}$) actually exhibit a high similarity compared to data of healthy subjects and most importantly they finally lead to improved segmentations. Only in some severe cases, it can be observed that the muscle's shape is slightly changed and that small structures are not reconstructed perfectly eventually also affecting the overall segmentation performance. Therefore, we expect that increasing the amount
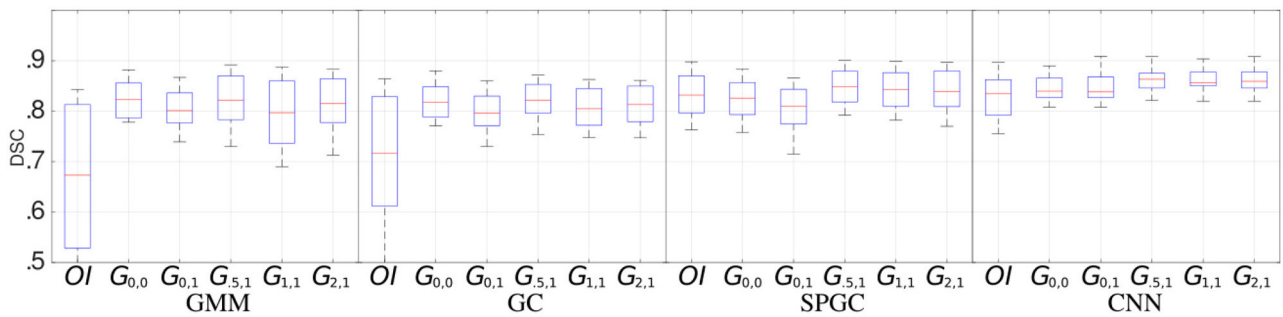
FIGURE 3 | Segmentation performance (median, quartiles, min, and max DSCs) for the four segmentation approaches (GMM, GC, SPGC, CNN) and for individual GAN configurations ($G_{n,m}$) compared to a direct segmentation (i.e., segmentation without image translation) of the original image data (*OI*). The indices of GAN-based methods define the loss weights w_r (first index) and w_i (second index).
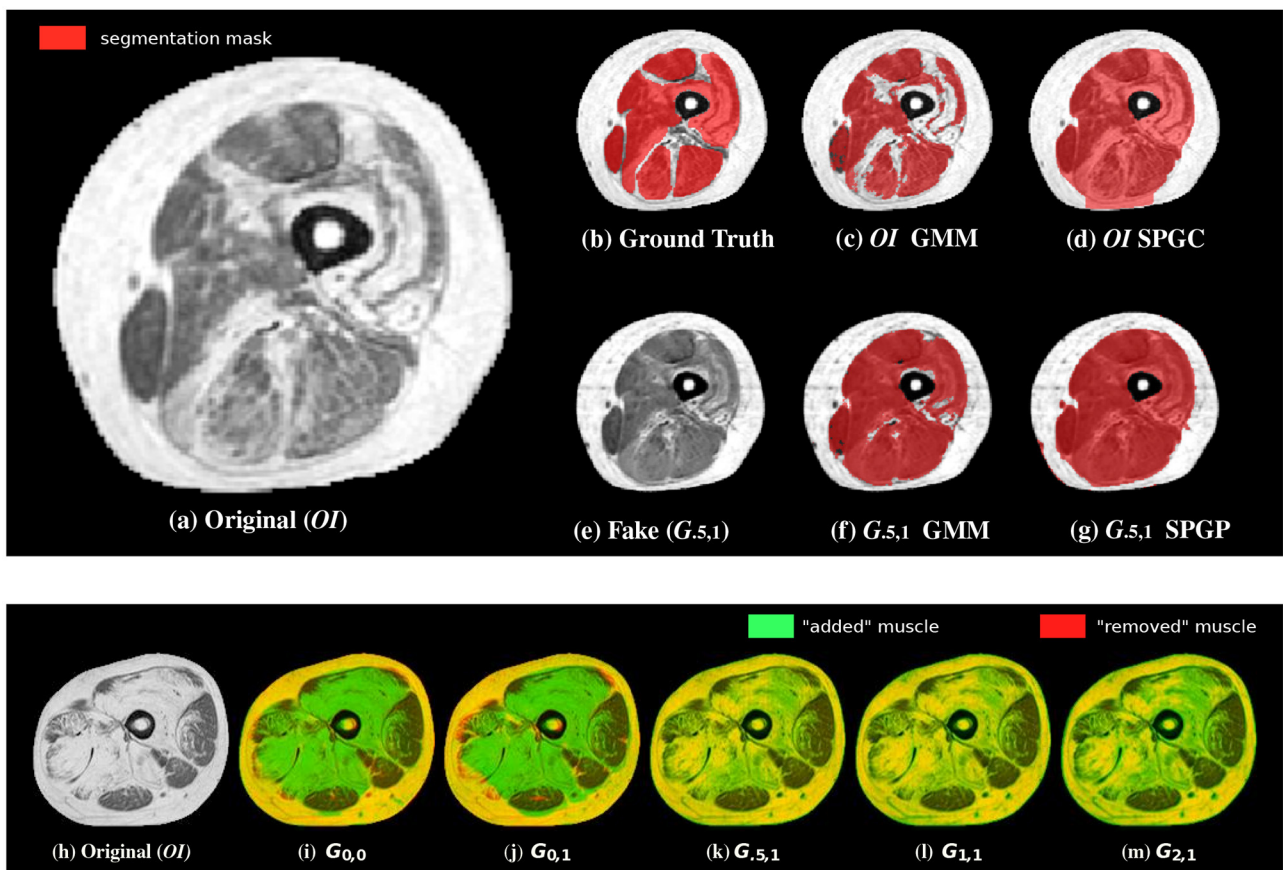
FIGURE 4 | Example segmentations **(c,d)** of the original image **(a)** as well as of the translated images **(e–g)** in comparison to the ground-truth annotations **(b)**. Although small structures often cannot be completely reconstructed (especially SPGC leads to over-smoothed masks), overall segmentation robustness increases in case of the translated image **(f,g)**. The bottom row shows an overlay of an example original image **(h)** with the corresponding translated images. Although green color indicates "added" muscle tissue, red color indicates "removed" muscle. Yellow shows unchanged intensities. The configurations without $\mathcal{L}_r$ show removed muscle tissue and also added muscle in wrong areas **(i,j)**. This is not the case when including the novel domain specific loss **(k–m)**.

of unlabeled training data can help to improve the image-translation process in order to boost the overall performance of (unsupervised) segmentation even further.

For clinical application, we estimate that a DSC of between 0.85 and 0.90 is required for reliable diagnosis. Visual inspection can help to quickly identify scans for which segmentation failed.

After image translation, rates below 0.85 only occurred for severely affected patients.

To conclude, we proposed a methodology to simplify segmentation tasks and thereby boost the segmentation accuracy by mapping a hard segmentation problem to an easier task. For means of enhancing the image-to-image translation approach, we introduced a further domain specific loss function included in GAN training. We considered an application scenario on segmenting MRI scans of human thighs and showed that the proposed approach can be effectively applied to either increase the segmentation performance of supervised segmentation techniques, or even to obtain highly reasonable outcomes with completely unsupervised techniques. We assess the latter case as even more relevant with most significant boosts in DSC (up to 0.15). We are confident that this approach is not limited to the considered application but can be effectively applied to other tasks in radiology as well.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: We are planning to make the data set publicly available either upon request or via a publicly available link. Requests to access these datasets should be directed to michael.gadermayr@fh-salzburg.ac.at.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University Hospitel RWTH Aachen. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

MG and BG primarily designed the study. DM, MM, FB, and DT provided valuable feedback and suggestions for improvements from technical and medical perspective, respectively. KL, LH, and MG were involved in technical implementations. BG, MM, and FB were involved as medical advisors. DM was involved as technical advisor (image analysis). DT was involved as expert radiologist. The paper was mainly written by MG and BG. All co-authors provided feedback and were involved in manuscript revision.

FUNDING

REFERENCES

1. Osadebey M, Bouguila N, Arnold D. Chapter 4 - brain mri intensity inhomogeneity correction using region of interest, anatomic structural map, and outlier detection. In: *Applied Computing in Medicine and Health* (2016). p. 79–98.
2. Sahnoun M, Kallel F, Dammak M, Mhiri C, Ben Mahfoudh K, Ben Hamida A. A comparative study of mri contrast enhancement techniques based on traditional gamma correction and adaptive gamma correction: case of multiple sclerosis pathology. In: *Proceedings of the International Conference on Advanced Technologies for Signal and Image Processing*. Sousse (2018). p. 1–7.
3. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: *Proceedings of the International Conference on Medical Image Computing and Computer Aided Interventions (MICCAI'15)* (2015). p. 234–41.
4. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative adversarial nets. In: *Advances in Neural Information Processing Systems*. Montreal, QC (2014). p. 2672–80.
5. Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'17)*. Honolulu, HI (2017).
6. Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: *Proceedings of the European Conference on Computer Vision (ECCV'16)* (2016).
7. Zhu JY, Park T, Isola P, Efros AA. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: *Proceedings of the International Conference on Computer Vision (ICCV'17)* (2017).
8. Jin CB, Kim H, Liu M, Jung W, Joo S, Park E, et al. Deep CT to MR synthesis using paired and unpaired data. *Sensors.* (2019) 19:2361. doi: 10.3390/s19102361
9. Yi Z, Zhang H, Tan P, Gong M. DualGAN: Unsupervised dual learning for image-to-image translation. In: *Proceedings of the IEEE International Conference on Computer Vision (ICCV'17)*. Venice (2017).
10. Park T, Efros AA, Zhang R, Zhu JY. Contrastive learning for conditional image synthesis. In: *Proceedings of the European Conference on Computer Vision (ECCV'20)*. Virtual (2020).
11. Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, van den Berg CAT, Išgum I. Deep MR to CT synthesis using unpaired data. In: *Proceedings of the International MICCAI Workshop Simulation and Synthesis in Medical Imaging (SASHIMI'17)*. Quebec City, QC (2017). p. 14–23.
12. Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, et al. Magnetic resonance image (MRI) synthesis from brain computed tomography (CT) images based on deep learning methods for magnetic resonance (MR)-guided radiotherapy. *Quant Imaging Med Surg.* (2020) 10:1223–36. doi: 10.21037/qims-19-885
13. Gadermayr M, Tschuchnig M, Merhof D, Krämer N, Truhn D, Gess B. An asymmetric cycle-consistency loss for dealing with many-to-one mappings in image translation: a study on thigh mr scans. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*. Nice (2021).
14. Gadermayr M, Tschuchnig M, Merhof D, Krämer N, Truhn D, Gess B. A comprehensive study on automated muscle segmentation for assessing fat infiltration in neuromuscular diseases. *Magn Recon Imaging.* (2018) 48:20–6. doi: 10.1016/j.mri.2017.12.014
15. Orgiu S, Lafortuna CL, Rastelli F, Cadioli M, Falini A, Rizzo G. Automatic muscle and fat segmentation in the thigh from T1-weighted MRI. *J Magn Reson Imaging.* (2015) 43:601–10. doi: 10.1002/jmri.25031
16. Tan C, Yan Z, Yang D, Li K, Yu HJ, Engelke K, et al. Accurate thigh inter-muscular adipose quantification using a data-driven and sparsity-constrained deformable model. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'15)* (2015).
17. Kovacs W, Liu CY, Summers R, Yao J. Identification of muscle and subcutaneous and intermuscular adipose tissue on thigh MRI of muscular dystrophy. In: *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'16)* (2016).
18. Yao J, Kovacs W, Hsieh N, Liu CY, Summers RM. Holistic segmentation of intermuscular adipose tissues on thigh MRI. In: *Proceedings of the International Conference on Medical Image*

Computing and Computer Assisted Intervention (MICCAI'17) (2017). p. 737–45.

19. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved n3 bias correction. *IEEE Trans Med Imaging.* (2010) 29:1310–20. doi: 10.1109/TMI.2010.20 46908

20. Gadermayr M, Gupta L, Appel V, Boor P, Klinkhammer BM, Merhof D. Generative adversarial networks for facilitating stain-independent supervised and unsupervised segmentation: a study on kidney histology. *IEEE Trans Med Imaging.* (2019) 38:2293–302. doi: 10.1109/TMI.2019.28 99364

frontiers
in Radiology

Review and Prospect: Artificial Intelligence in Advanced Medical Imaging

Shanshan Wang[1,2†], Guohua Cao[3†], Yan Wang[4†], Shu Liao[5†], Qian Wang[3†], Jun Shi[6†], Cheng Li[1] and Dinggang Shen[3,5*]

[1] Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences (CAS), Shenzhen, China, [2] Pengcheng Laboratrory, Shenzhen, China, [3] School of Biomedical Engineering, ShanghaiTech University, Shanghai, China, [4] School of Computer Science, Sichuan University, Chengdu, China, [5] Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China, [6] School of Communication and Information Engineering, Shanghai University, Shanghai, China

Artificial intelligence (AI) as an emerging technology is gaining momentum in medical imaging. Recently, deep learning-based AI techniques have been actively investigated in medical imaging, and its potential applications range from data acquisition and image reconstruction to image analysis and understanding. In this review, we focus on the use of deep learning in image reconstruction for advanced medical imaging modalities including magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET). Particularly, recent deep learning-based methods for image reconstruction will be emphasized, in accordance with their methodology designs and performances in handling volumetric imaging data. It is expected that this review can help relevant researchers understand how to adapt AI for medical imaging and which advantages can be achieved with the assistance of AI.

Keywords: deep learning, magnetic resonance imaging, computed tomography, positron emission tomography, medical imaging reconstruction

INTRODUCTION

Of all the advances in modern medicine, medical imaging is among the most remarkable developments. It allows us to see anatomical structures, organs, and biological processes unreachable by unaided eyes, providing tremendous opportunities for scientific research as well as disease diagnosis and treatment (1, 2). Different modalities such as magnetic resonance imaging (MRI) (3), computational tomography (CT) (4), and positron emission tomography (PET) (5) can provide versatile information, ranging from structure, morphology to physiological function. Specifically, MRI uses powerful magnetic fields, radio waves, and computers to produce details of anatomical structures and functions (6, 7). CT measures the linear attenuation coefficient of tissues inside each voxel element as an X-ray beam transmits through the body. PET measures changes in metabolic processes as well as other physiological activities by counting radioactive emissions of a biochemical metabolite labeled with radioactive material.

To better serve the clinical end-users, abundant studies have been conducted to optimize the scanning process, improve the imaging efficiency, and enhance the image quality of MRI/CT/PET (8, 9). Image reconstruction plays a significant role in this aspect. For MRI, its slow imaging speed has been a long-lasting bottleneck that seriously limits its wider applications in the clinic (10). Among different possible solutions, k-space undersampling has been identified as a highly

effective approach to accelerate the scan (11, 12). Nevertheless, images generated from undersampled k-space data are subject to the low-quality issue, with possible loss of the important information related to disease diagnosis or treatment (13). Thus, high-quality image reconstruction from incomplete k-space data is critical. As for CT and PET, the main focus is to reconstruct high-quality images from deteriorated raw data caused by low-dose imaging demands (14, 15). Many efforts have been devoted to developing image reconstruction methods for MRI/CT/PET, among which deep learning-based methods have shown unprecedented successes (9, 14–16).

During the last decade, deep learning has been extensively applied to medical imaging to handle different problems, such as image reconstruction (17), image registration (18–20), image classification (21, 22), and lesion segmentation (23). Among these applications, image reconstruction is a primary step in the clinical workflow that has a huge impact on the downstream tasks of imaging-based analysis and decision making. Notice that different medical imaging modalities (MRI, CT, and PET) have their own unique imaging physics and principles, and thus numerous deep learning-based methods have been proposed to accomplish respective reconstruction tasks (9, 11, 14, 15). For MRI, existing works have achieved impressive achievement to balance imaging efficiency and imaging quality (9, 11). Similarly, promising results have also been achieved for CT and PET image reconstruction (14, 15). However, current progress is still preliminary for deep learning-based image reconstruction in real applications, and more efforts are needed to make this technology mature enough for wide real-world clinical applications. Thus, it is the right time to review existing works to help beginners as well as non-specialists better understand this relatively new technique and promote more follow-up investigations and applications.

The remainder of this review paper is organized as follows. In section Overall Workflow of Deep Learning-Based Reconstruction, we demonstrate the overall workflow of deep learning-based reconstruction, by briefly introducing the basics of deep learning relevant to the reconstruction task, the purpose of image reconstruction, and the workflow of deep learning-based reconstruction. Detailed technical developments of deep learning-based reconstruction are introduced in section Technical Developments of Deep Learning-Based Reconstruction. Section Clinical Applications and Current Achievements reviews current clinical applications and achievements, followed by descriptions of key challenges and opportunities in section Challenges and Opportunities. Finally, section Conclusion concludes the paper.

OVERALL WORKFLOW OF DEEP LEARNING-BASED RECONSTRUCTION

Basics of Deep Learning

Artificial intelligence (AI) refers to the ability of a machine to simulate human intelligence by thinking and acting like humans (24). Deep learning is a sub-discipline of AI, which specifically addresses various tasks through building deep neural networks (DNNs) (25). Different abstract levels of representations are extracted with multi-layer networks which enable the learning of complex functions. When inputs are images, the low-level features usually represent edges and contours in the images, whereas the high-level features are commonly semantic features (26). One key characteristic for deep learning is that all the parameters for feature extraction are learned automatically with the provided data samples, which can be better self-optimized to specific problems compared to the use of manual feature engineering approaches (26, 27).

Supervised learning, unsupervised learning, and reinforcement learning are the three major paradigms for deep learning (28–31). Supervised learning requires paired data samples for the inputs and the expected outputs (28). Model optimization is performed by minimizing loss functions that are calculated to measure the difference between model outputs and ground truth. In unsupervised learning, only input data samples are provided, and certain assumptions of the data have to be made and then the corresponding model constraints are enforced to facilitate the model learning (29). In reinforcement learning, an algorithm is referred to as an agent. Then, the agent takes an action to change its state, and, at the same time, a reward or penalty is assigned. Different from supervised learning, the training data of reinforcement learning provide only an indication of whether an action is correct or not. The overall goal of reinforcement learning is to achieve the maximum reward over time by learning a policy for the agent to choose proper actions for any given states (31). Most medical image reconstruction models are based on supervised learning or unsupervised learning, while reinforcement learning is less frequently utilized.

Deep Learning-Based Image Reconstruction

MRI

MRI reconstruction aims to generate high-quality images from sampled k-space data. Conventional reconstruction methods (i.e., Fourier transform) require the scanning process to follow the Nyquist sampling theory. Thus, to obtain high-quality images, the sampling frequency should be high enough, which unfortunately makes the scanning process very time-consuming. On the other hand, undersampling, which breaks the Nyquist sampling theory, leads to imperfect MR image reconstruction if using conventional reconstruction methods. To this end, compressed sensing (CS) MRI (CS-MRI) has been proposed by introducing CS theory to reconstruct MR images with significantly fewer measurements than those required by traditional Nyquist sampling theory (32). CS-MRI accomplishes the reconstruction task mainly by exploiting the sparsity of MRI, since most MR images are sparse after transformed into an appropriate domain (32), such as using total variation (33) and wavelet transformation (34).

Despite the successes achieved, CS-MRI still has limited performance because of using manually-designed methods to exploit the sparsity in MRI. By contrast, deep learning-based image reconstruction for MRI can automatically and fully exploit the available data information and recover the lost information

under the guidance of certain prior knowledge. Deep learning was first introduced to MR image reconstruction in 2016 by Wang et al. (35). In their work, a three-layer neural network was built to automatically learn the mapping between low-quality and high-quality images (35). Following this work, a series of studies have been published, aiming to build more sophisticated, robust, and optimized deep learning models for MR image reconstruction (36–39).

Existing deep learning-based MR image reconstruction methods can be classified into two major categories, (1) model-based methods and (2) data-driven methods. Model-based methods reconstruct high-quality MR images via solving certain optimization algorithms and utilizing neural network modules to represent the reconstruction steps of the solution. Typical optimization algorithms include alternating direction method of multipliers (ADMM) algorithm (40), iterative shrinkage-thresholding algorithm (ISTA) (41), and primal-dual hybrid gradient (PDHG) algorithm (42). Data-driven methods are the end-to-end approaches that rely on DNNs with large capacities to learn non-linear reconstruction processes. Example models include U-Net (36), residual network (ResNet) (43), and generative adversarial networks (GAN) (44). Model-based methods are more interpretable as the network blocks can correspond to the algorithm solutions, and data-driven methods are more effective in data exploitation. Overall, deep learning-based MR image reconstruction methods have dominated the current research field, with promising performance.

CT

In CT, image reconstruction aims to transform the sensor data, which basically reflects line integrals of the object, to an image representing the object. Until recently, most CT reconstruction methods can be classified as either analytic reconstruction or iterative reconstruction. Analytic reconstruction is based on the mathematical inverse of the forward model of an imaging process, which could either be mathematically derived or numerically modeled after the design of the CT imaging device and the knowledge about how it generates sensor data. A typical example of analytic reconstruction in CT is filtered back-projection (FBP) (45). Iterative reconstruction is based on a numerical forward model combined with a feedback loop (46–51). In the feedback loop, the error between the calculated sensor dataset and the measured sensor dataset is back-transformed to the image domain to update the current image estimation. This process is repeated until the error reaches a small threshold and the optimum image solution is obtained. Iterative reconstruction has been widely used in CT because the measurements are typically noisy or a mathematical inverse is unknown or computationally challenging. Examples of iterative reconstruction in CT include the algebraic reconstruction technique (ART) (52) and the simultaneous algebraic reconstruction technique (SART) (46). Iterative reconstruction usually outperforms analytic reconstruction in terms of the quality of reconstructed images, because iterative reconstruction relies on a more improved forward model and has the ability to bring in various types of

external prior information to expand the information available during reconstruction.

Very recently, a third type of CT reconstruction method – deep learning based reconstruction – was introduced. Deep learning reconstruction was first introduced to CT in 2016, when Kang et al. used a deep learning reconstruction approach at the 2016 Low-Dose X-ray CT Grand Challenge [organized by the American Association of Physicists in Medicine (AAPM)] (53), and, in parallel, when Chen et al. introduced a similar convolutional neural network (CNN) for low-dose CT denoising (54). The successful demonstration of CNN reconstruction in low-dose CT has inspired many other deep learning reconstruction research. For example, a combination of a CNN with the Normalized Metal Artifact Reduction (NMAR) algorithm for CT metal artifact reduction (55), a combination of DenseNet and Deconvolution Network (DD-Net) for sparse-view CT (56), Super-Resolution Convolutional Neural Network (SRCNN) for CT super-resolution (57), and so on.

Deep learning reconstruction does not require an explicit physical imaging model. Instead, deep learning reconstruction can build its own model from a large amount of training data, which becomes more and more readily available due to the wide use of medical imaging in modern healthcare. With larger and more representative training datasets, deep learning reconstruction has the potential to outperform both analytic reconstruction and iterative reconstruction. With unsupervised learning or self-supervised learning, it has been hypothesized that the integration of imaging physics within the machine learning pipeline may further improve the reconstruction quality. For example, a self-supervised and hybrid CT super-resolution model that integrates the advantages of both deep learning network and imaging physics has been just published very recently (51).

PET

Similarly, PET reconstruction aims to generate diagnostic quality images from measurement data. The conventional PET reconstruction methods can be broadly classified into two categories, i.e., (1) analytic (58, 59) and (2) iterative PET reconstruction methods (60, 61). The analytic PET reconstruction methods provide a straightforward mathematical solution for image formation, a typical example of which is the filtered-back projection (FBP). In contrast, based on a more accurate description of the imaging process, iterative methods produce a more complex mathematical solution that requires multiple steps to reach an image. Since it can take into account the noise patterns in the observations and use more realistic models of the system, the iterative methods provide improvements over the analytical methods. The classical iterative methods include Maximum Likelihood-Expectation Maximization (ML-EM) (60) and Ordered Subsets Expectation Maximization (OSEM) (61).

Recently, numerous learning-based methods have also been developed for PET reconstruction, such as random forest (62), sparse representation (SR) (63), and multi-level Canonical Correlation Analysis (mCCA) scheme (64). Yet, these traditional machine learning methods often require complex feature engineering, which largely limits the practicability and

also results in suboptimal reconstruction quality. To address this limitation, deep learning was first introduced to PET reconstruction in 2017 by Xiang et al. (65). The authors proposed a deep CNN model, followed by an auto-context strategy, to estimate standard-dose PET images directly from both the low-dose PET and the corresponding MR images, without the need for handcrafted features. Encouraged by the great success of this work, a series of deep learning-based methods have been developed and successfully applied to various scenarios of PET reconstruction (58–61, 66, 67). In addition, the combination of the conventional iterative reconstruction framework and the deep learning-based method has provided some new approaches for PET reconstruction (14, 68, 69). For instance, Gong et al. (14) used the existing inter-patient information *via* a deep neural network to further improve the quality of the reconstructed PET image. Furthermore, with the introduction and development of new deep learning models such as GAN, more efforts applying new techniques have been continuously conducted for superior PET reconstruction performance (70–72).

Training and Testing Workflow

The image reconstruction framework typically includes an input, a reconstruction model, and an output. Traditionally, the input is a sensor-domain raw data, i.e., sinogram in CT. With deep learning-based reconstruction, the sensor data can be first reconstructed using an analytic reconstruction model to provide a low-quality image, and then this low-quality image is fed into the DNN model to generate the corresponding high-quality image. For MRI, the input and output data pair can be either in k-space or image space. Note, to build a deep learning-based reconstruction framework, two steps, namely model training and model testing, are included, as detailed below.

Model training is performed on the provided training samples to optimize the model parameters. During the model training, the loss between the model-generated outputs and the provided training samples is calculated and back-propagated to optimize the model parameters. The model parameters are updated to minimize this loss. Model training proceeds in a data batch mode. Training is stopped after the model is converged to a certain point, or after reaching a pre-selected number of epochs. To avoid the overfitting issue, data augmentation is commonly utilized. Frequently utilized data augmentation methods include affine transformations and Gaussian noise addition. In a deep learning model, there are usually some hyper-parameters (such as batch size, learning rate, etc.) that need to be adjusted manually or automatically, i.e., using an additional validation set, to improve the model performance.

With the optimized model, testing can be performed. To comprehensively evaluate the model performance, testing with data different from the training/validation data should be conducted. For example, validating and testing data from different centers collected with different machines are often considered to make the model robust enough in real-world applications.

TECHNICAL DEVELOPMENTS OF DEEP LEARNING-BASED RECONSTRUCTION

This section will review various deep learning reconstruction methods developed for MRI, CT, and PET, with typical methods summarized in **Table 1**. We will present technical aspects and performance characterization of deep learning reconstruction. Technical aspects will include data preparation, network architecture design, loss function, and settings or requirements for training.

Data Preparation

When applying deep learning to medical imaging, normally three datasets are in need, namely training, validation, and testing datasets. The training dataset is used to train a neural network that is monitored by the validation dataset to avoid overfitting or underfitting. The testing dataset is to evaluate whether the deep learning models can perform well for the real application scenarios. The datasets should include ground-truth images for supervised learning. While for unsupervised learning, no ground-truth information is needed.

For MRI, different types of datasets have been collected and experimented with for various applications. According to the target region dynamic characteristics, there are static MRI and dynamic MRI. Static MRI is applicable when the imaging target changes slowly with time, such as the knee (36, 38) and the brain (37). Dynamic MRI is often required when the target moves fast, such as cardiac MRI (74, 91). Based on the number of coils utilized to collect the data, MRI datasets can be classified into single-channel MRI (92) and multi-channel MRI (43, 76, 93). When different imaging parameters are used, multi-parametric MRI data are collected to better characterize the physical and physiological properties of the imaging object (94). Besides, quantitative MRI is also available, which can measure tissue-specific parameters (95)[1].

For CT, depending on the goal of network training, various public datasets are available for DNN model training when developing deep learning reconstruction methods. Some datasets are curated for image noise reduction. For example, the Mayo Clinic Low-Dose X-ray CT datasets for the Low Dose CT Grand Challenge organized by the AAPM (54) have clinical CT images acquired at the full-dose level and the corresponding simulated CT images at the quarter-dose level. This Mayo Clinic dataset can be useful for training deep learning models to reduce CT image noise and hence optimize the dose efficiency. Other datasets are curated toward specific diseases or conditions. For example, The Cancer Imaging Archive (TCIA) hosts a large archive of medical CT images of cancer accessible for public download. Noticeably, in the last year, because CT has been successfully proven to be a rapid triaging tool in patients with moderate to severe COVID symptoms in a resource-constrained environment where COVID-19 is highly prevalent (96), we now have abundant publicly-available COVID CT datasets available today. Two particular COVID CT datasets

---

[1]Popular datasets in MRI include fastMRI https://fastmri.org/dataset/ and brainweb https://brainweb.bic.mni.mcgill.ca/.

TABLE 1 | Representative works on deep learning-based MRI/CT/PET image reconstruction.

Modality	Task description	Network architecture	Loss function	Dataset	Evaluation metrics	Reference
MRI	Directly learning the transformation from sensor-space data to image	MLP	Simple squared loss and additional L1-norm penalty	ImageNet database, MGH-USC HCP public database	SNR, RMSE	(37)
MRI	k-space to k-space reconstruction	UNet	L2 loss	Knee k-space dataset, MGH-USC HCP public database	NMSE, PSNR, SSIM	(36)
MRI	Reconstruction with proposed complex convolution operations	ResNet	Mean absolute error (MAE)	Brain dataset, Knee dataset	PSNR, SSIM	(43)
MRI	Reconstructing real-valued and complex-valued MRI data	GAN	Cyclic data consistency loss	IXI database, Data Science Bowl challenge, Knee dataset	PSNR, SSIM, NRMSE	(44)
MRI	Fast and high-quality reconstruction by combining various loss functions	GAN	Content loss, Image domain and frequency domain MSE loss, Perceptual VGG loss	MICCAI 2013 grand challenge dataset, Pathological MRI images	NMSE, PSNR, SSIM	(52)
MRI	Infusing motion information into the modeling process with deep neural networks for enhanced dynamic MRI reconstruction quality	Recurrent neural network (MODRN, Motion-guided Dynamic Reconstruction Network)	L1 loss	Private short-axis cardiac data (21 normal subjects and 3 dyssynchrony disease patients)	NMSE, PSNR, SSIM	(73)
MRI	Reconstruction with both k-space and spatial prior knowledge integrated via multi-supervised network training	CNN	L2 loss	Private cardiac MR data	MSE, PSNR, SSIM	(74)
MRI	Improving MRI reconstruction accuracy and computational speed with a CS-based model	Model-based (Alternating direction method of multipliers algorithm)	NMSE	Brain and chest MR images	NMSE, PSNR, Test time	(40)
MRI	Fast and high-quality reconstruction of clinical accelerated multi-coil MR data	Model-based (Variational network, unrolling iteration)	MSE	Clinical knee dataset	SSIM, NRMSE	(38)
MRI	Deriving deep architectures for inverse problems with the arbitrary structure	Model-based (recursive framework alternating between denoising block and data-consistency layer)	MSE	Brain MR dataset from five volunteers	PSNR, Time	(75)
MRI	Fast parallel MR imaging by exploring both spatial redundancy and multi-coil correlations	Model-based (split Bregman iterative algorithm)	MSE	Private 2D multichannel MR brain dataset	NMSE, PSNR, SSIM	(76)
MRI	Self-supervised deep learning MRI reconstruction by dividing sub-sampled data points into two sets with one for data consistency and another for loss calculation	Model-based (regularized iterative algorithm between data consistency and a regularizer solved by the variable-splitting and quadratic relaxation method)	Normalized L1-L2 loss	Knee MR data from fastMRI initiative database	NMSE, SSIM	(77)
MRI	Accelerate and improve multishot diffusion-weighted MRI reconstruction by combining unrolled network with deep CNNs	Model-based and UNet (recurrences of model-based gradient updates (shotlocally low-rank) and neural networks	L1 loss	Private brain (14 scans from 8 volunteers) and breast (6 scans from 6 volunteers) MR data	NMSE, PSNR, SSIM	(78)
CT	Using U-Net and its variants for recovery of high-frequency edges in sparse-view CT in the image domain	Dual frame and tight frame U-Nets	Pixel-wise soft-max combined with cross entropy function (the original U-net loss function)	10 patient CT scan data from the 2016 AAPM Low Dose CT Grand Challenge Dataset	NMSE, PSNR, SSIM	(56)

(Continued)

TABLE 1 | Continued

Modality	Task description	Network architecture	Loss function	Dataset	Evaluation metrics	Reference
CT	General sparse-view CT image reconstruction	DenseNet combined with deconvolution	Weighted loss between MSE and MS-SSIM	3,059 clinical CT images from the TCIA database	MSE, SSIM, Haralick texture features	(15)
CT	CT super-resolution	Modified U-net	L2 loss	7,670 CT slices	NRMSE, PSNR	(57)
CT	Low-dose CT for mapping low-dose images to normal-dose images; CT image denoising	Residual encoder-decoder CNN (RED-CNN)	MSE loss	7,015 normal-dose CT images from the NBIA dataset and simulated low-dose CT images	RMSE, PSNR, SSIM	(54)
CT	CT image denoising	Framelet-based wavelet residual network	Pixel-wise soft-max combined with cross entropy function (the original U-net loss function)	10 patient CT scan data from the 2016 AAPM Low Dose CT Grand Challenge Dataset	RMSE, PSNR, SSIM	(79)
CT	Sparse-view CT image reconstruction	U-net with skip connection for residual learning	Pixel-wise soft-max combined with cross entropy function (the original U-net loss function)	The 2016 AAPM Low Dose CT Grand Challenge Dataset, plus 500 simulated images and 377 experimental sinograms	SNR	(80)
CT	CT image denoising in low-dose CT	GAN network, consisting of a Generator CNN and a Discriminator CNN	binary cross-entropy, L2 loss	5 low-dose and 5 corresponding routine-dose CT scans of a phantom, and 28 cardiac CT scans from patients	SNR, PSNR	(81)
CT	CT image denoising in low-dose CT	GAN network with Wasserstein distance and perceptual loss (WGAN)	Wasserstein distance based adversarial loss, VGG perceptual loss	10 patient CT scan data from the 2016 AAPM Low Dose CT Grand Challenge Dataset	PSNR, SSIM	(82)
CT	CT super-resolution	GAN-CIRCLE	Adversarial loss, cycle consistency loss, identity loss, joint sparsifying transform loss	Micro-CT dataset from 25 tibia specimen, and the 2016 AAPM Low Dose CT Grand Challenge Dataset	PSNR, SSIM, IFC	(83)
CT	To ensure data consistency even in worst-case scenario, and to guarantee the convergence of a non-convex CT reconstruction problem	Specially designed method that replaces the projector in a projected gradient descent with a CNN, and uses the CNN in the feedback loop to recursively project the result onto the sensor domain	Data consistency loss	500 lower-lung CT images from the 2016 AAPM Low Dose CT Grand Challenge Dataset, and 377 micro-CT slice images of a rat brain	SNR, SSIM	(84)
CT	CT Super-resolution	Self-supervised SADIR-net (super-resolution and deblur based iterative reconstruction), which is a hybrid between deep learning network and imaging physics	Joint loss function combining L2-norm with SSIM	47 clinical CT scans from TCIA database; custom-acquired Catphan^{700} phantom CT sensor data	MTF, RMSE, SSIM, IFC	(51)
PET	Incorporating the neural network into the iterative PET reconstruction framework for PET denoising	UNet with residual learning	Augmented Lagrangian format, L2 loss	19 XCAT phantoms; 6 lung patient data	CR, STD	(14)
PET	Standard-dose PET reconstruction from low-dose PET	Noise-Aware Dual Res-UNet	Dice loss, Binary cross entropy loss, General and adaptive robust loss, SSIM loss	10 subjects referred for whole-body FDG-18 PET/CT scan on a GE Discovery 710 scanner	PSNR, SSIM	(85)
PET	Using patients' own prior information for PET reconstruction	3D UNet	MSE	Phantom and real brain data	CRC, STD	(69)
PET	PET reconstruction from projections data	ANN	MSE	Simulated data	NMSE	(86)

(Continued)

TABLE 1 | Continued

Modality	Task description	Network architecture	Loss function	Dataset	Evaluation metrics	Reference
PET	Using multilayer perceptron (MLP) to enhance MAP reconstructed PET images	MLP with backpropagation	Least squares loss	PET phantom images, two patient PET imaging datasets	NMSE, NSD, Contrast	(87)
PET	Ultra-low-dose PET reconstruction	ResNet	L1 loss, SSIM, MS-SSIM	9 PET/MRI images from patients with glioblastoma (GBM)	PSNR, SSIM, NRMSE	(66)
PET	Using dilated convolutions for recovering full-count PET images from low-count PET images	UNet with dilated convolution	L1 loss	35 PET data extracted from an IRB approved psychiatric study	MAPE, PSNR, SSIM	(88)
PET	Reconstruction of PET image from sinogram data	CNN	VGG, MAE, MS-SSIM	Whole-body PET studies:40 patients for training, 4 for validation, and 10 for testing	SNR, Bias, MAE, MS-SSIM	(89)
PET	Anatomy-aided PET image reconstruction	3D CNN	L2 loss	Simulation study: 20 XCAT51 phantoms real patients studies: 6 hybrid lesion patients, 6 lung cancer patients	CR, STD	(90)
PET	Using a deep learning prior for iterative PET reconstruction	DnCNN + local linear fitting (LLF)	L2 loss	27 control subjects and clinical patients	Bias and standard deviation; NRMSE; SSIM	(68)
PET	Reconstruction of PET image from sinogram data	GAN	Adversarial loss, L1 loss	Simulated data of the three phantoms using Monte Carlo simulations, including Zubal thorax phantom with 64Cu-ATSM, Hoffman brain phantom with 18F-FDG and Zubal brain phantom with 11C-Acet ate	Bias, Variance	(71)
PET	Reconstruction of PET images from sinogram data	GAN	MSE, Relativistic Average LS adversarial loss	Human brain PET dataset with nine subjects	Bias, Variance, PSNR, SSIM	(70)
PET	Low-dose PET image denoising	CycleWGAN	Adversarial loss, Cycle-consistency loss, Identity loss	Eighteen patients with biopsy-proven primary lung cancer or patients with suspicious radiological abnormalities	NRMSE, PSNR, SSIM, SUV_{mean} and SUV_{max}	(72)

could be useful for training deep learning models. One is the BIMCV-COVID-19+ (97), a large dataset from the Valencian Region Medical Image Bank earlier in the pandemic period, and another is the RSNA International COVID-19 Open Radiology Database (RICORD), which is an ongoing international effort in curating potentially the largest international COVID-19 CT dataset.

For PET, the datasets mainly include static PET (98–100) and dynamic PET (101–104) based on data types. On the other hand, according to the number of tracers imaged in a single scan, the datasets can be classified as single-tracer PET (105), dual-tracer PET (106, 107), and multi-tracer PET (108). When it comes to the injected tracer dose level, the datasets can also be broadly categorized as low-dose PET (L-PET) and full-dose PET (F-PET) (65, 67, 88). Although the use of real PET data in studies is more clinically relevant, these real data are often difficult to obtain due to various factors. Therefore, simulated phantom data is becoming a popular alternative in research works (68, 88, 109, 110).

Network Architecture

The neural network architectures employed for different tomographic imaging tasks share some similar properties. The most frequently used architectures include multilayer perceptron (MLP), U-Net, generative adversarial networks (GAN), ResNet, etc. Here, we introduce these typical network architectures.

MLP

The MLP, which is an artificial neural network (ANN) with all layers fully-connected, can map sets of input data into a set of desired outputs. In the past decades, researchers have worked on exploiting MLP in medical image analysis. For example, a multilayer perceptron was proposed for accelerated parallel MRI (111). Zhu et al. (37) proposed an MLP-based manifold learning framework to emulate the fast-Fourier transform and learn an end-to-end mapping between k-space data and image domains and achieve the purpose of acceleration. For PET, MLP was also employed for simple low-resolution PET reconstruction (86). Furthermore, Yang et al. (87) developed an MLP-based

framework to enhance the maximum a posteriori (MAP) reconstructed PET images, which constructs a highly non-linear and spatial-varying mapping between the MAP reconstructed image patches and the corresponding enhanced image patches.

U-Net

U-net consists of an encoder structure and a decoder structure, which was originally designed for biomedical image segmentation (112, 113). The encoder gradually down samples the input images to extract image features with different levels of semantic information. The decoder receives the features from the encoder and recovers the feature map resolution step-by-step to generate the outputs, which are often the same size as the inputs and can then be treated as the reconstructed images. Skip connections between the encoder and the decoder are introduced to improve the localization accuracy during decoding.

For MRI, Ye et al. (114) used deep residual learning to accelerate MRI. The proposed deep residual learning network is composed of two separately trained amplitude and phase difference networks, which can successfully learn and remove aliasing artifacts. Furthermore, Ye et al. also proposed a U-Net-based domain adaptation architecture for radial k-space undersampled MR (115), and a fully data-driven deep learning algorithm for k-space interpolation (36). These methods have been successfully applied to MR image reconstruction, and have achieved better results than the classic CS method. Duan et al. (116) proposed a fast and accurate deep learning reconstruction method for human lung gas MRI, which consists of coarse-to-fine nets (C-net and F-net) based on U-Net. The proposed deep learning method can better reconstruct the human lung gas MR images acquired from highly undersampled k-space compared with the traditional CS-MRI. Hyun et al. (117) proposed an under sampling MRI reconstruction method using U-Net, which shows excellent performance and can generate high-quality MR images with a small amount of data.

For CT, U-net and its variants have also been successfully applied to solve various problems in CT reconstruction, including sparse-view CT reconstruction, artifact reduction, noise suppression, and CT super-resolution, etc. For sparse-view CT reconstruction, which can reduce radiation dose and accelerate scanning speed, Han et al. (56) achieved better reconstruction performance by framing U-Net via deep convolutional framelets. Also, for sparse-view CT reconstruction, Kofler et al. (118) proposed a cascade of U-nets and data consistency layers, and Zhang et al. (15) developed DD-Net by combining DenseNet and deconvolution and arranging them in a network topology similar to U-Net. For the purpose of CT artifact reduction, Zhang et al. (55) tried U-net and found promising results of U-net in reducing global and local CT artifacts. To reduce noise in low-dose CT images, Liu et al. (119) adopted stacked denoising autoencoders to suppress noise and recover structure details. For CT super-resolution, Park et al. (57) used a modified U-net to learn an end-to-end mapping between low-resolution and high-resolution CT images.

For PET, U-net is also a commonly used framework in many PET reconstruction works (14, 69, 86, 120–122). Gong et al. (14) designed an iterative reconstruction framework that combines

the U-net structure and the residual network for PET denoising by utilizing dynamic data of prior patients. Taking the noise level of low-count PET into account, Xiang et al. (85) developed a noise-aware dual Res-UNet (NADRU) framework for low-dose PET reconstruction. The proposed method first identified an attention map indicating the location of high-intensity noise in the low-dose PET images. Then, the noise attention map was incorporated with the original image for high-quality PET reconstruction. In addition to reconstructing high-quality images within PET, many efforts have also been made to reconstruct PET from other modalities. For example, Sikka et al. (121) adopted a 3D U-Net architecture to estimate PET from MRI images. By considering non-local and non-linear correlations, the proposed method showed a significant improvement in the diagnostic accuracy of Alzheimer's disease. Employing a modified 3D U-net as the network structure, Gong et al. (69) designed an iterative reconstruction framework that incorporates the personalized deep neural network to generate PET data from a patient's own MRI prior image(s). Furthermore, Cui et al. (122) utilized CT/MR prior information to perform PET denoising based on a modified 3D U-net structure in an unsupervised manner.

ResNet

ResNet is proposed to solve the difficulty of training very deep CNNs and avoid model performance degradation (123). The core idea of ResNet lies in residual learning, which is based on the assumption that it is easier to optimize the residual mapping than to optimize the original and unreferenced mapping (123). With the success of residual learning, the ResNet has also been widely used in medical image reconstruction.

For MRI, Shi et al. (124, 125) proposed a residual-learning-based MR image super-resolution reconstruction network. The network can improve image reconstruction performance using both global residual learning (GRL) and local residual learning (LRL). Wang et al. (43) proposed a new framework Deepcomplex MRI using a deep residual CNN for parallel imaging. It considers the correlation between the real and imaginary parts of MR complex images and achieved better results than real-value networks. Li et al. (126) designed a deep ResNet using variable density spiral trajectory to accelerate fMRI reconstruction. The proposed deep ResNet consists of various residual blocks. Du et al. (127) proposed a residual CNN for reconstructing single anisotropic 3D MR images based on residual learning. The residual CNN with long and short skip connections can effectively recover uncollected high-frequency details of MR images.

For CT, ResNet or more generally residual learning has also been demonstrated its effectiveness in CT reconstruction, particularly in noise suppression and artifact reduction. Chen et al. (128) developed a residual encoder-decoder CNN (RED-CNN) for low-dose CT. RED-CNN combines autoencoder, deconvolution network, and shortcut connections. It can effectively suppress noise, preserve structure details, and enhance lesion detection. For CT image denoising, Kang et al. (79) proposed a wavelet residual network based on a deep convolutional framelet and achieved better performance compared to their earlier algorithm using directional deep

convolutional-wavelet neural network (53). To reduce the sparse-view CT artifact, Dong et al. (129) proposed a residual deep learning CNN to interpolate the sinogram of sparse-view micro-CT, and the deep learning interpolated sinogram was FBP-reconstructed into high-quality images. Also for sparse-view CT, Jin et al. (81) proposed FBPConvNet, which first reconstructs sparse-view CT sinogram with FBP and then improves the FBP-reconstructed image using a modified U-net with the addition of residual learning.

For PET, residual learning is also employed in the reconstruction task. In order to effectively restore the low-dose PET images to the standard-dose quality, Xu et al. (66) proposed an encoder-decoder residual deep network, in which residual learning and skip connections were adopted for learning the difference between standard-dose and low-dose PET images. Similarly, Spuhler et al. (88) designed a novel multiscale dilated CNN approach to predict full-count PET images from low-count images. The proposed method integrated the residual learning to capture the difference of low-count and full-count PET images and enhance the convergence of the network. The experiments of these studies showed that residual learning was beneficial for high-quality PET reconstruction. Moreover, in Chen et al. (54), a deep learning-based framework with low-count PET and multimodal MRI as inputs was presented for diagnostic-quality PET image synthesis through residual learning.

GAN

GAN (130), as one of the most popular generative models in deep learning, has demonstrated its superior performance in many computer vision tasks and attracted growing interest in medical image reconstruction.

For MRI, Yang et al. (52) proposed De-Aliasing GAN (DAGAN) for fast compressed sensing MRI reconstruction. The authors designed a refinement learning method to stabilize the U-Net-based generator. In order to better preserve texture and edge information, DAGAN combines adversarial loss and innovative content loss in the image reconstruction process and takes into account the frequency information at the same time. The reconstruction result of DAGAN is better than the traditional CS-MRI algorithm. Quan et al. (44) proposed an improved model, RefineGAN, based on fully residual convolutional autoencoder and GANs for fast and accurate CS-MRI reconstruction. It can perform faithful interpolation for a given undersampled k-space data by employing a deeper generator and discriminator with cyclic data consistency loss. RefineGAN outperforms the state-of-the-art CS-MRI reconstruction algorithms in terms of both image quality and running time. Mardani et al. (131) proposed a novel CS framework based on LSGAN and pixel-wised l1/l2 loss for MRI reconstruction, namely GANCS. GANCS can reconstruct higher quality images with improved fine texture details compared to existing methods.

For CT, Wolterink et al. (81) used a GAN network that consists of a Generator CNN and a Discriminator CNN to reduce the noise level in CT images. They produced better images for more accurate coronary calcium quantification. Similarly, for the purpose of image denoising in low-dose CT, Yang et al. (82) modified the original GAN network by using

the Wasserstein distance, instead of the Jensen-Shannon (JS) divergence, to compare data distributions. The Wasserstein distance is combined with the well-known pre-trained VGG-19 network (132) to build a joint loss function. This modified GAN network also achieved promising results in image denoising. For the purpose of CT super-resolution, You et al. (83) developed a GAN network constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). GAN-CIRCLE incorporates deep CNN, residual learning, and network-in-network techniques for feature extraction and restoration, and employed a cycle Wasserstein regression adversarial training framework. It is noted that many GAN networks also employed the technique of residual learning in their architectures.

For PET, Liu et al. (71) employed a conditional GAN (cGAN) framework to learn the mapping from sinogram data to reconstructed PET images directly. Inspired by the promising results achieved by cGAN, the authors further presented an end-to-end model for PET reconstruction, which adopts two coupled networks to sequentially denoise low dose sinogram and reconstruct activity map (70). Zhou et al. (72) designed a cycle Wasserstein regression adversarial model (CycleWGAN) using Wasserstein distance, instead of JS divergence and cycle-loss, to boost the low-dose PET image quality, which shows the superior performance of Wasserstein distance in effectively preserving the edge information. To reduce the loss of contextual information, Wang et al. (133) developed a concatenated 3D cGAN for high-quality PET image estimation from low count PET. Considering the various contributions of different image locations and the complementary information in different modalities, they further proposed an auto-context-based locality adaptive GANs (LA-GANs) (67) model to reconstruct the full count PET image from both the low count PET and the accompanying multimodal MRI images. Besides, many other works also attempted to reconstruct PET images from other modality information in consideration of the expensive cost of PET imaging and the hazards of radiation exposure. Ben-Cohen et al. (134) proposed to generate simulated PET images from given CT data without manually annotated labels. They first adopted FCN to generate an initial PET-like image and then employed cGAN to refine the FCN output so that the synthesized image could be more realistic. Based on 3D GAN, Yaakub et al. (135) designed a two-stage approach to predict accurate PET images from T1-weighted MRI scans. It is worth noting that many GAN-based models have also introduced residual learning to further improve the reconstruction performance (136, 137).

Modality-Specific Module Design

To improve the reconstruction accuracy or enhance the reliability of the reconstruction results, special network modules are usually designed taking the specific properties of different imaging modalities into consideration.

For MRI, in addition to modules utilized by every model, including the convolutional layers, the normalization layers, and the activation layers, there is commonly a data consistency layer to guarantee that the data on scanned points are correct (138). According to the data acquisition process of MRI, undersampling happens in the k-space by neglecting a certain portion of data

points. Therefore, theoretically, on the scanned data points, the reconstruction results should be consistent with the acquisitions. With the data consistency layer, the reconstruction is forced to be correct on these sampling points and the reconstruction of unscanned data points is accordingly improved. Besides, because the data acquisition of MRI proceeds in a different domain (k-space) from the image domain, reconstruction can be performed in individual domains (38, 139) or cross-domains (74, 140). Furthermore, complex-valued neural networks are proposed to specifically process the complex-valued MR data (43, 141, 142).

For CT, although the reconstruction results from most reported deep learning algorithms are so far remarkable in terms of image quality, there is still some concern about whether those reconstruction results can be trusted, especially in real-world applications of diagnostic imaging. One main limitation of those deep learning algorithms is that they seldom provide guarantees in the worst-case scenario. To address this limitation, Gupta et al. (84) proposed a specially designed CT image reconstruction method that replaces the projector in a projected gradient descent with a CNN and uses the CNN in the feedback loop to recursively project the result onto the sensor domain. This reconstruction method can enforce measurement consistency, is guaranteed to converge, and, under certain conditions, converges to a local minimum of a non-convex inverse problem. On the other hand, while iterative CT reconstruction can yield high-quality images, careful tuning of hyper-parameters in these iterative reconstruction problems is inevitable. To achieve automatic parameter tuning, Shen et al. (143) employed deep reinforcement learning to train a system that can automatically adjust parameters in a human-like manner, and demonstrated that CT images reconstructed from their approach attain quality similar or better than those reconstructed with manually tuned parameters.

For PET, some studies have incorporated specially designed modules to improve the PET image quality. For instance, taking the location-varying contributions from different imaging modalities into account, Wang et al. (67) proposed a locality adaptive fusion module to automatically fuse local patches from multimodal MRI for high-quality PET image synthesis. In Samuel Matej et al. (58), the authors devised a novel Radon inversion layer to address the computational challenges in multi-slice PET image reconstruction. This specially designed layer was demonstrated to be efficient in performing domain transformation from sinogram to image space. Moreover, to encourage feature reuse and prevent resolution degradation, Du et al. (144) designed residual dense connections followed with pixel shuffle operations (RDPS blocks) in the generator network, achieving promising reconstruction results.

Loss Function
As the task is to restore the quality of the output images in all locations, for the fully supervised learning, the most frequently used loss for the network training is the mean squared error (MSE) between the network prediction and the ground truth. MSE is also known as the L2 loss. Based on MSE, there are also some extended loss functions such as root mean squared

errors (RMSE), normalized mean squared errors (NMSE), and normalized root mean squared errors (NRMSE).

There are alternative losses, such as the mean-absolute-error cost function (MAE), which is also known as the L1 loss. Compared with MSE, MAE is used relatively less, but there are still studies showing that using MAE can preserve better results than MSE.

One common choice of loss function for reconstruction problem is L2, but the reconstructed image obtained is of low quality and lacks high-frequency detail. Therefore, in order to offset the shortcoming of L2 loss, structural similarity index (SSIM), signal to noise ratio (SNR), peak SNR (PSNR), or perceptual loss is used as an additional loss to constrain the prediction results in some literatures. These additional loss functions or the combined loss between them have been shown to improve the reconstruction performance of the model.

Modality-Specific Loss
In MRI, there are also specially designed losses. In Quan et al. (44), the authors proposed a cyclic data consistency loss, which combines the undersampled frequency loss and the fully reconstructed image loss. In practice, MSE, MAE or other functions can be used as the basic function to achieve cyclic loss. Some studies (52) combine MSE and perceptual loss to form a novel content loss to achieve better reconstruction details. There are also studies that combine MAE with perceptual loss (145), or MSE with TV loss (146), for MR image reconstruction.

In CT, Yang et al. (84) employed for their modified GAN network a joint loss function that combines the Wasserstein distance-based adversarial loss with the well-known pre-trained VGG−19 loss (134). Those two loss terms in the joint loss function are balanced with a hyperparameter to control the trade-off between the GAN adversarial loss and the VGG perceptual loss. When comparing the performance of a modularized deep neural network to commercial algorithms for low-dose CT image reconstruction, Shan et al. (147) chose a composite loss function that includes three components: adversarial loss, MSE, and edge incoherence. The adversarial loss is used to train the generator in their GAN network to produce images as close to the reference high-dose images as possible, the MSE is used to reduce image noise, and the edge incoherence is used to enhance the edge information in the denoised image.

In PET, Kim et al. (68) proposed a novel 3D local linear fitting (LLF) function and incorporated it into the cost function, combining the input image with the DnCNN correcting the unwanted bias and finally enhance the image quality. Similarly, Ouyang et al. (105) designed a GAN model with feature matching technique and task-specific perceptual loss to ensure that the synthesized standard-dose amyloid PET images include the correct features.

Requirement for Network Training
The fundamental parameter learning schemes are back-propagation algorithms. Adam optimization with variable parameter momentum is often used in neural network optimization. As for hardware, the graphics card for deep learning network training is essential. According to the literature

we searched and referenced, the types of graphics cards generally used are NVIDIA K80, NVIDIA K40c, GTX 1080Ti, RTX 2080, RTX 2080Ti, Titan X, Titan Xp, Titan V, etc. As for software, TensorFlow, PyTorch, Keras, Caffe, etc. are several commonly-used DNN training frameworks. In addition, Matlab is also used to process data or perform tests in some studies. The system used is generally a Linux system.

CLINICAL APPLICATIONS AND CURRENT ACHIEVEMENTS

Deep learning-based medical imaging techniques have played more and more important roles in today's clinical applications, and have achieved significant progress in solving various major pain points in different imaging modalities.

MRI

For MRI, it has superior soft-tissue contrast and it is radiation-free. However, the major limitation of MRI is its slow acquisition speed. Although lots of acceleration strategies were proposed in the literature, such as parallel imaging and compressed sensing, they have their own limitations such as amplification of Gibbs artifacts and long iterative reconstruction time. Deep learning-based techniques offer a feasible solution to robustly and efficiently reconstruct the MRI images from subsampled K-space data even under high down sampling factors. Moreover, deep learning-based reconstruction techniques can be integrated with conventional acceleration techniques to reach even higher reconstruction quality. For instance, the AI-assisted compressed sensing (ACS) technique developed by United Imaging Intelligence (UII) and United Imaging Healthcare (UIH) integrates the advantages of four acceleration techniques, i.e., (1) deep learning-based reconstruction, (2) partial Fourier transform, (3) parallel imaging, and (4) compressed sensing, into a unified framework, and achieves great success in real-world clinical applications for fast MRI imaging. ACS is able to reduce around 80% scan time on average for most of the FSE sequences, and it supports the scan of different body parts such as head, cervical spine, lumbar spine, hip, pelvis, ankle, and knee. For each body part, ACS normally can achieve a scan time of fewer than 100 s for all the sequences as shown in **Figure 1**.

ACS has received FDA 510K clearance and has also been deployed in different hospitals. Another example is the SubtleMR techniques developed by Subtle Medical, which also adopts deep learning-based techniques for fast MR imaging and received FDA 510K clearance. SubtleMR is able to reduce around 60% scan time and has also been deployed in many hospitals and applied in real-world clinical workflow in the US.

CT

For modalities of CT (as well as PET as introduced below), the radiation dose delivered to the patient must be strictly controlled, because radiation is harmful to the patient and an excessive dose may lead to the result of secondary cancer. However, a lower dose normally leads to inferior image quality, and it may affect the diagnosis accuracy. Therefore, how to obtain high-quality images

under the low-dose condition for CT is essential in real world clinical applications.

Deep learning-based denoising techniques provide a good solution to obtain high-quality CT images under low-dose conditions. The basic principle is to train a deep learning network that learns the mapping between the low dose CT image and the corresponding standard-dose CT image. Once the network is trained, the image quality can be significantly improved by passing the low dose CT image through the network. This strategy has been adopted by many industries and turned into products in real world applications. For instance, the DELTA (i.e., DEep Learning Trained Algorithms), a deep learning-based denoising technique developed by UII and UIH, can reduce the dose up to 80% while the low contrast detectability (LCD) of CT images can be improved up to 157%; some typical examples are shown in **Figure 2**. Canon developed the Advanced intelligent Clear-IQ Engine (AiCE) which can reduce the noise and boost signal in CT images based on deep learning. GE developed the TrueFidelity CT imaging platform, which adopts deep learning-based techniques to improve the image quality of low-dose CT images. DELTA, AiCE, and TrueFidelity all received FDA 510K clearance.

PET

For PET, besides the concern of dose, another pain point is the relatively longer imaging time than other image modalities such as CT and DR, and some patients such as children and patients with bone cancer may not be able to hold their positions during the imaging process. Therefore, how to obtain high-quality images under low-dose conditions and how to accelerate the imaging is essential for real-world clinical applications of PET.

So far, various deep learning-based techniques have been applied to accelerate the acquisition speed of PET imaging and also maintain the high quality of PET images. For instance, the HYPER DLR (Deep-Learning Reconstruction) product developed by UII and UIH can significantly reduce the scanning time of PET imaging from 3 min/bed to < 1 min/bed. In addition, it can effectively reduce the noise level of PET images under low count rate conditions and significantly improve image quality. Specifically, the SNR (Signal-to-Noise Ratios) of PET images can be improved by 42% with an accelerated imaging speed. **Figure 3** shows some typical examples of HYPER DLR. Similarly, Subtle Medical developed the SubtlePET product which also adopts deep learning-based techniques and can denoise the low-count PET images obtained in 25% of the original scan duration, improving patient comfort during PET scans. Both HYPER DLR and SubtlePET received FDA 510K clearance.

PET-MRI

In some applications, cross-modality synthesis techniques are also required. For instance, the PET-MR imaging equipment normally needs to synthesize the CT image from the acquired MR image in order to perform attenuation correction (AC) for the PET image (148). This process is illustrated in **Figure 4**.

There are lots of synthesis strategies. The most simple and straightforward strategy is to segment the MR image into several tissue types and fill the corresponding regions with fixed CT HU
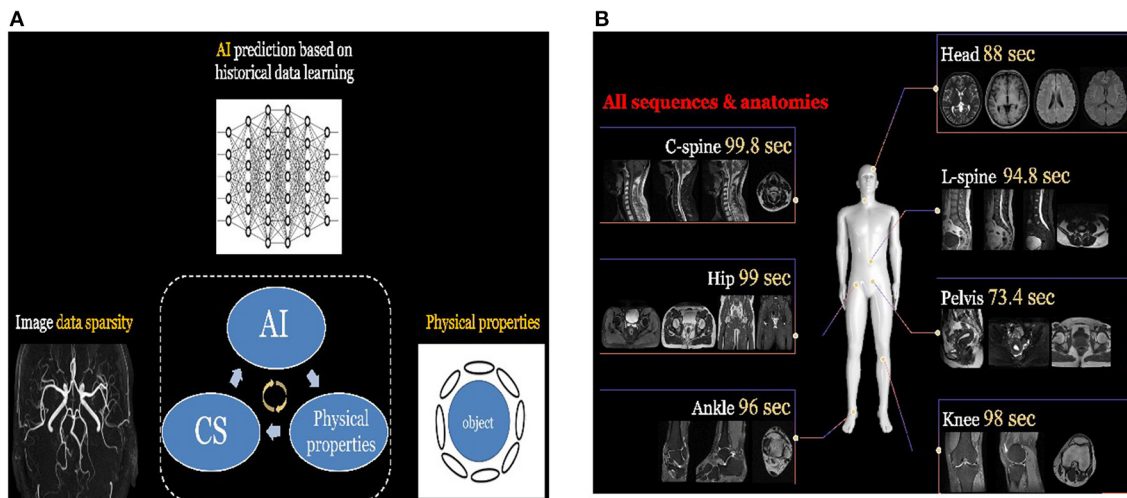
FIGURE 1 | (A) The principle of ACS. It integrates the advantages of different acceleration techniques such as deep learning-based reconstruction, parallel imaging, and compressed sensing. **(B)** ACS normally can achieve great scan speed (i.e., <100 s) for different body parts and sequences.

FIGURE 2 | (A) The low dose abdominal CT image. **(B)** The resulting image after applying DELTA to the low dose image in **(A)**. **(C)** The corresponding standard dose abdominal CT image.
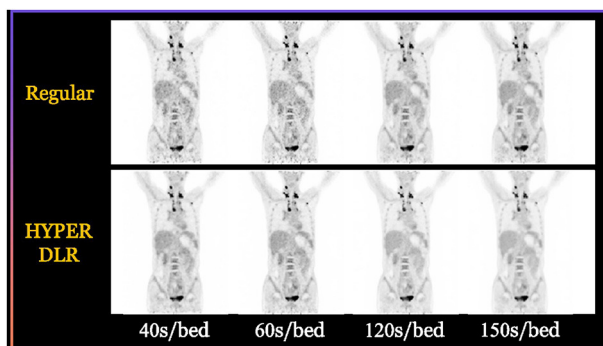
FIGURE 3 | Typical examples of the HYPER DLR PET denoising product developed by UII and UIH. The first row shows PET images obtained by using different acquisition times per bed without HYPER DLR, where the image quality degrades significantly when fast acquisition time. The second row shows the resulting images by applying the HYPER DLR technique, where obvious image quality improvement can be observed.

values. This strategy has been widely adopted in many companies such as Siemens and GE. With the aid of deep learning-based cross-modality synthesis techniques, it is possible to obtain more precise synthesized CT images from the MR images with unsupervised learning techniques and therefore to produce more accurate AC operation. For instance, UII and UIH proposed an unsupervised deep learning-based technique (149) that can effectively synthesize the CT images from the MR sequences. Typical examples are shown in **Figure 5**.

CHALLENGES AND OPPORTUNITIES

The success of deep learning-based methods on image reconstruction for medical imaging has been extensively validated. However, the wide applications in clinical practices are not yet realized. One key limiting issue is the model interpretability. Due to the nature of DNN, the entire non-linear mapping process is a "black box," meaning that no direct physical
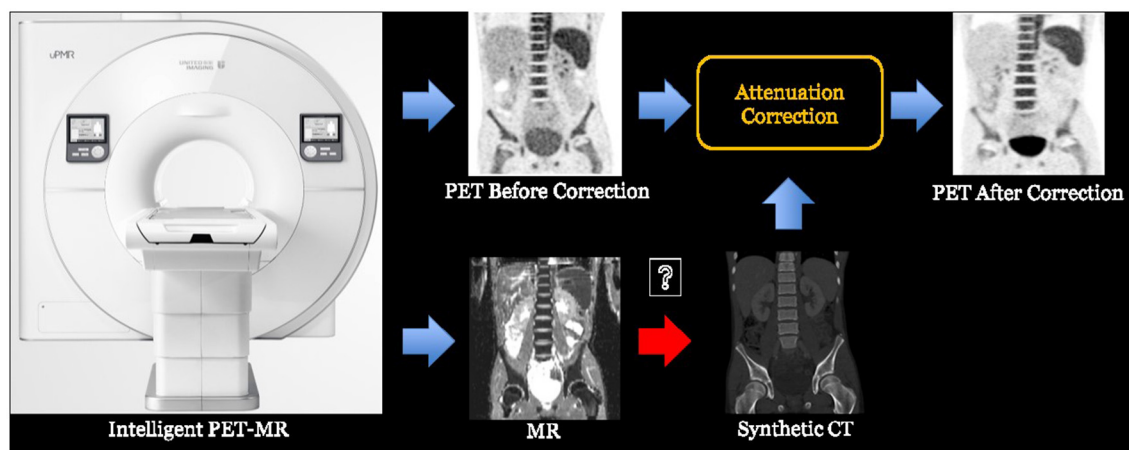
FIGURE 4 | The process of PET-MR attenuation correction, where a synthetic CT image is obtained from the MR image to help the attenuation correction of PET image.
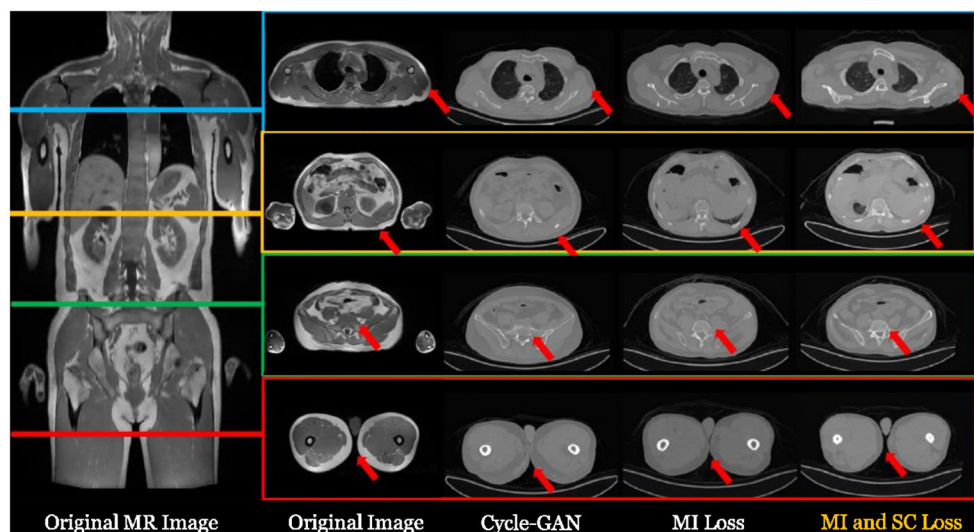
FIGURE 5 | Typical examples of synthesizing CT images from a whole body MR image with deep learning-based techniques. For more details, please refer to Ge et al. (149).

or theoretical mechanism is provided to explain how the inputs are transformed to the outputs (150). Consequently, deep learning reconstruction models find difficulties to get accepted by clinicians. Recently, enhancing model interpretability through building interpretable neural networks or utilizing various visualization techniques becomes a hot topic in deep learning-based natural image analysis (151–154). Similarly, more efforts should be devoted to building both interpretable and high-performance deep learning reconstruction models.

Another challenge is the generalization capability of deep learning-based methods. It is known that deep learning is a data-driven method, and the performance of deep learning models depends heavily on the training data (25, 26, 155). Thus, constructing a comprehensive training dataset is critical. Different from natural images, the distributions of medical

images can be quite different if different scanning protocols or scanning machines are utilized. Moreover, due to ethical issues, building large medical image datasets by collecting images from different resources is difficult. As a result, the performance of most existing deep learning models might be over-claimed, and a performance drop can be observed when applying the reported models to the data of end-users. Building robust models that can maintain performance during implementation is important to promote wide applications.

At the same time, the increasing demand for automated image analysis in the clinic to help achieve efficient and accurate imaging-based diagnosis and decision making is providing various opportunities for the introduction of deep learning-based methods. With the rapid development of computing power and optimization of deep learning models, deep learning is expected

to play a significant role in achieving fast, portable, safe, and cheap medical imaging. For instance, the transformer (156) framework proposed in 2017 for NLP has demonstrated inspiring performance in capturing global information and has also shown great potential for applications in many image processing tasks recently. The development of the transformer also provides opportunities for the enhancement of current medical imaging models. Besides, multi-modal imaging and autonomous imaging are also promising directions for future studies.

CONCLUSION

Deep learning has presented inspiring performances in image reconstruction for different medical imaging modalities, including MRI, CT, and PET. In this review paper, we focus on the applications in MRI, CT, and PET. A detailed survey is conducted in the following aspects and sequence: the overall deep learning reconstruction workflow, the technological development of deep learning reconstruction, the clinical applications and current achievements, and a discussion of the challenges and opportunities. In summary, deep learning-based medical image reconstruction presents a great potential to promote a wide spectrum of applications in the clinic, if the remaining issues, such as interpretability and generalizability, can be properly addressed in the future.

AUTHOR CONTRIBUTIONS

SW: MRI review, manuscript preparation, and editing. GC: CT review, manuscript preparation, and editing. YW: PET review, manuscript preparation, and editing. SL, QW, and JS: manuscript preparation and editing. CL: manuscript preparation. DS: topic creating and manuscript editing. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Shung KK, Smith M, Tsui BMW. *Principles of Medical Imaging*. Academic Press. (2012).
2. Doi K. Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Phys Med Biol*. (2006) 51:5–27. doi: 10.1088/0031-9155/51/13/R02
3. Westbrook C, Talbot J. *MRI in Practice*. John Wiley & Sons. (2018).
4. Buzug TM. *Computed Tomography*. Springer handbook of medical technology. (2011):311–42. doi: 10.1007/978-3-540-74658-4_16
5. Bénard F, Romsa J, Hustinx R. Imaging gliomas with positron emission tomography and single-photon emission computed tomography. *Semin Nucl Med*. (2003) 33:148–62. doi: 10.1053/snuc.2003.127304
6. Edelman RR, Warach S. Magnetic resonance imaging. *N Engl J Med*. (1993) 328:785–91. doi: 10.1056/NEJM199303183281109
7. Hoult DI, Bhakar B. NMR signal reception: Virtual photons and coherent spontaneous emission. *Concepts Magn Reson*. (1997) 9:277–91. doi: 10.1002/(SICI)1099-0534(1997)9:5<277::AID-CMR1>3.0.CO;2-W
8. Manduca A, Yu L, Trzasko JD, Khaylova N, Kofler JM, McCollough CM, et al. Projection space denoising with bilateral filtering and CT noise modeling for dose reduction in CT. *Med Phys*. (2009) 36:4911–9. doi: 10.1118/1.3232004
9. Gong Y, Shan H, Teng Y, Tu N, Liang G, Wang G, et al. Parameter-transferred Wasserstein generative adversarial network (PT-WGAN) for low-dose PET image denoising. *IEEE Trans Radiat Plasma Med Sci*. (2020) 5:213–23. doi: 10.1109/TRPMS.2020.3025071
10. Macovski A, Conolly S. Novel approaches to low-cost MRI. *Mag Reson Med*. (1993) 30:221–30. doi: 10.1002/mrm.1910300211
11. Wang S, Tan S, Gao Y, Liu Q, Ying L, Xiao T, et al. Learning joint-sparse codes for calibration-free parallel MR imaging. *IEEE Trans Med Imaging*. (2018) 37:251–61. doi: 10.1109/TMI.2017.2746086
12. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Mag Reson Med*. (2007) 57:1086–98. doi: 10.1002/mrm.21236

13. Adluru G, DiBella EVR. Reordering for improved constrained reconstruction from undersampled k-space data. *Int J Biomed Imaging*. (2008) 2008:341684. doi: 10.1155/2008/341684
14. Gong K, Guan J, Kim K, Zhang X, Yang J, Seo Y, et al. Iterative PET image reconstruction using convolutional neural network representation. *IEEE Trans Med Imaging*. (2018) 38:675–85. doi: 10.1109/TMI.2018.2869871
15. Zhang Z, Liang X, Dong X, Xie Y, Cao G A. sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE Trans Med Imaging*. (2018) 37:1407–17. doi: 10.1109/TMI.2018.2823338
16. Chartrand R. Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data. In: *Proceedings of the International Symposium on Biomedical Imaging: From Nano to Macro*. Boston (2009). p. 262–5. doi: 10.1109/ISBI.2009.5193034
17. Lin F, Man DB, A. hierarchical approach to deep learning and its application to tomographic reconstruction. In: *Proceedings of the International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine Philadelphia*. Niskayuna (2019).
18. Fan J, Cao X, Wang Q, Yapa PT, Shen D. Adversarial learning for mono-or multi-modal registration. *Med Image Anal*. (2019) 58:101545. doi: 10.1016/j.media.2019.101545
19. Cao X, Yang J, Zhang J, Dong N, Kim M, Wang Q, et al. Deformable image registration based on similarity-steered CNN regression. *In: Proceedings of the International Symposium on Medical Image Computing and Computer Assisted Intervention (MICCAI'17)*. Cham: Springer (2017). doi: 10.1007/978-3-319-66182-7_35
20. Luan H, Qi F, Xue Z, Chen L, Shen D. Multimodality image registration by maximization of quantitative-qualitative measure of mutual information. *Pattern Recognition*, (2008) 41, 285–98. doi: 10.1016/j.patcog.2007.04.002
21. Zhang J, Gao Y, Gao Y, Munsell BC, Shen D. Detecting anatomical landmarks for fast Alzheimer's disease diagnosis. *IEEE Trans Med Imaging*. (2016) 35:2524–33. doi: 10.1109/TMI.2016.2582386
22. Mohsen H, El-Dahshan ESA, El-Horbaty ESM, Salem ABM. Classification using deep learning neural networks for brain tumors. *Future Comput Inform J*. (2018) 3:68–71. doi: 10.1016/j.fcij.2017.12.001

23. Abraham N, Khan NM. A novel focal tversky loss function with improved attention u-net for lesion segmentation. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'19) Venice*. (2019). doi: 10.1109/ISBI.2019.8759329

24. Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*. 3rd edn. Applied Mechanics & Materials (2009).

25. Webb S. Deep learning for biology. *Nature*. (2018) 554:555–7. doi: 10.1038/d41586-018-02174-z

26. Lecun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521:436–44. doi: 10.1038/nature14539

27. Shen D, Wu G, Suk HI. Deep learning in medical image analysis. *Annu Rev Biomed Eng*. (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442

28. Murphy KP. Machine learning: a probabilistic perspective. *Chance*. (2014) 27:62–3. doi: 10.1080/09332480.2014.914768

29. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining. Springer Series in Statistics*. (2009). doi: 10.1007/978-0-387-84858-7

30. Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature*. (2014) 518:529–33. doi: 10.1038/nature14236

31. Sutton RS, Barto AG. Reinforcement learning: an introduction. *Kybernetes*. (1998) 27:1093–6. doi: 10.1108/k.1998.27.9.1093.3

32. Lustig M, Donoho DL, Santos JM, Pauly JM. Compressed sensing MRI. *Signal Process Mag*. (2008) 25:72–82. doi: 10.1109/MSP.2007.914728

33. Liang D, Liu B, Wang J, Ying L. Accelerating SENSE using compressed sensing. *Magn Reson Med*. (2009) 62:1574–84. doi: 10.1002/mrm.22161

34. Qu X, Guo D, Ning B, Hou Y, Lin Y, Cai S, et al. Undersampled MRI reconstruction with patch-based directional wavelets. *Magn Reson Imaging*. (2012) 30:964–77. doi: 10.1016/j.mri.2012.02.019

35. Wang S, Su Z, Ling L, Peng X, Zhu S, Liang F, et al. Accelerating magnetic resonance imaging via deep learning. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'16)*. Prague (2016). doi: 10.1109/ISBI.2016.7493320

36. Han Y, Sunwoo L, Ye JC. k-space deep learning for accelerated MRI. *IEEE Trans Med Imaging*. (2020) 39:377–86. doi: 10.1109/TMI.2019.2927101

37. Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. *Nature*. (2018) 555:87–492. doi: 10.1038/nature25988

38. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. (2018) 79:3055–71. doi: 10.1002/mrm.26977

39. Wang S, Xiao T, Liu Q, Zheng H. Deep learning for fast MR imaging: a review for learning reconstruction from incomplete k-space data. *Biomed Signal Process Control*. (2021) 68:102579. doi: 10.1016/j.bspc.2021.102579

40. Yang Y, Sun J, Li H, Xu Z. Deep ADMM-Net for compressive sensing MRI. *NIPS*. (2017) 10–18.

41. Zhang J, Ghanem B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'18)*. Salt Lake (2018). p. 1828–37. doi: 10.1109/CVPR.2018.00196

42. Adler J, Öktem O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Probl*. (2017) 33:124007. doi: 10.1088/1361-6420/aa9581

43. Wang S, Cheng H, Ying L, Xiao T, Ke Z, Zheng H, et al. Deep complex MRI: exploiting deep residual network for fast parallel MR imaging with complex convolution. *Magn Reson Imaging*. (2020) 68:136–47. doi: 10.1016/j.mri.2020.02.002

44. Quan TM, Nguyen-Duc T, Jeong WK. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Trans Med Imaging*. (2018) 37:1488–97. doi: 10.1109/TMI.2018.2820120

45. Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *J Opt Soc Am*. (1984) 1:612–9. doi: 10.1364/JOSAA.1.000612

46. Gordon R, Bender R, Herman GT. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J Theor Biol*. (1970) 29:477–81. doi: 10.1016/0022-5193(70)90109-8

47. Andersen AH, Kak AC. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrason Imaging*. (1984) 6:81–94. doi: 10.1177/016173468400600107

48. Erdogan H, Fessler JA. Ordered subsets algorithms for transmission tomography. *Phys Med Biol*. (1999) 44:2835–51. doi: 10.1088/0031-9155/44/11/311

49. Lange K, Carson R. EM reconstruction algorithms for emission and transmission tomography. *J Comput Assist Tomogr*. (1984) 8:306–16. doi: 10.1097/00004728-198404000-00002

50. Thibault JB, Sauer KD, Bouman CA, Hsieh J. A three-dimensional statistical approach to improved image quality for multislice helical CT. *Med Phys*. (2007) 34:4526–44. doi: 10.1118/1.2789499

51. Zhang Z, Yu S, Qin W, Liang X, Xie Y, Cao G. Self-supervised CT super-resolution with hybrid model. *Comput Biol Med*. (2021) 138:104775. doi: 10.1016/j.compbiomed.2021.104775

52. Yang G, Yu S, Dong H, Slabaugh G, Dragotti PL, Ye X, et al. DAGAN: deep De-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans Med Imaging*. (2018) 37:1310–21. doi: 10.1109/TMI.2017.2785879

53. Kang E, Min J, Ye JC. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med Phys*. (2016) 44:e360–75. doi: 10.1002/mp.12344

54. Chen H, Zhang Y, Kalra NK, Lin F, Chen Y, Liao P, et al. Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE Trans Med Imaging*. (2017) 36:2524–35. doi: 10.1109/TMI.2017.2715284

55. Zhang C, Xing Y. CT artifact reduction via U-net CNN. In: *Proceedings of the Medical Imaging 2018: Image Processing* (2018). doi: 10.1117/12.2293903

56. Han Y, Ye JC. Framing U-Net via deep convolutional framelets: application to sparse-view CT. *IEEE Trans Med Imaging*. (2018) 37:1418–29. doi: 10.1109/TMI.2018.2823768

57. Park J, Hwang D, YunKim K, Kang S, Kim YK, Lee JS. Computed tomography super-resolution using deep convolutional neural network. *Phys Med Biol*. (2018) 63:145011. doi: 10.1088/1361-6560/aacdd4

58. Samuel Matej S, Daube-Witherspoon ME, Karp JS. Analytic TOF PET reconstruction algorithm within DIRECT data partitioning framework. *Phys Med Biol*. (2016) 61:3365–86. doi: 10.1088/0031-9155/61/9/3365

59. Defrise M, Kinahan P. *Data Acquisition and Image Reconstruction for 3d33d Pet*. Springer (1998). p. 11–53. doi: 10.1007/978-94-017-3475-2_2

60. Shepp LA, Vardi Y. Maximum likelihood reconstruction for emission tomography. *IEEE Trans Med Imaging*. (1982) 1:113–22. doi: 10.1109/TMI.1982.4307558

61. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Trans Med Imaging*. (1994) 13:601–9. doi: 10.1109/42.363108

62. Kang J, Gao Y, Shi F, Lalush DS, Lin W, Shen D. Prediction of standard-dose brain PET image by using MRI and low-dose brain [18F]FDG PET images. *Med Phys*. (2015) 42:5301–9. doi: 10.1118/1.4928400

63. Wang Y, Zhang P, An L, Ma G, Kang J, Shi F, et al. Predicting standard-dose PET image from low-dose PET and multimodal MR images using mapping-based sparse representation. *Phys Med Biol*. (2016) 61:791–812. doi: 10.1088/0031-9155/61/2/791

64. An L, Zhang P, Adeli E, Wang Y, Ma G, Shi F, et al. Multi-level canonical correlation analysis for standard-dose PET image estimation. *IEEE Trans Image Process*. (2016) 25:3303–15. doi: 10.1109/TIP.2016.2567072

65. Xiang L, Qiao Y, Nie D, An L, Lin W, Wang Q, et al. Deep auto-context convolutional neural networks for standard-dose PET image estimation from low-dose PET/MRI. *Neurocomputing*. (2017) 267:406–16. doi: 10.1016/j.neucom.2017.06.048

66. Xu J, Gong E, Pauly J, Zaharchuk G. 200x Low-dose PET reconstruction using deep learning. (2017). *arXiv:1712.04119*.

67. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, et al. 3D auto-context-based locality adaptive multi-modality GANs for PET synthesis. *IEEE Trans Med Imaging*. (2019) 38:1328–39. doi: 10.1109/TMI.2018.2884053

68. Kim K, Wu D, Gong K, Dutta J, Kim JH, Son Y, et al. Penalized PET reconstruction using deep learning prior and local linear fitting. *IEEE Trans Med Imaging*. (2018) 37:1478–87. doi: 10.1109/TMI.2018.2832613

69. Gong K, Catana C, Qi J, Li Q. PET image reconstruction using deep image prior. *IEEE Trans Med Imaging*. (2019) 38:1655–65. doi: 10.1109/TMI.2018.2888491

70. Feng Q, Liu H. Rethinking PET image reconstruction: ultra-low-dose, sinogram and deep learning. In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'20)* (2020). p. 783–92. doi: 10.1007/978-3-030-59728-3_76

71. Liu Z, Chen H, Liu H. Deep learning based framework for direct reconstruction of PET images. In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'19)*. Springer (2019). p. 48–56. doi: 10.1007/978-3-030-32248-9_6

72. Zhou L, Schaefferkoetter JD, Tham IWK, Gang H, Yan J. Supervised learning with CycleGAN for low-dose FDG PET image denoising. *Med Image Anal.* (2020) 65:101770. doi: 10.1016/j.media.2020.101770

73. Huang Q, Xian Y, Yang D, Qu H, Yi J, Wu P, et al. Dynamic MRI reconstruction with end-to-end motion-guided network. *Med Image Anal.* (2020) 68:1010901. doi: 10.1016/j.media.2020.101901

74. Wang S, Ke Z, Cheng H, Jia S, Ying L, Zheng H, et al. DIMENSION: dynamic MR imaging with both k-space and spatial prior knowledge obtained via multi-supervised network training. *NMR Biomed.* (2019) 1–16. doi: 10.1002/nbm.4131

75. Aggarwal HK, Mani MP, Jacob M. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Trans Med Imaging.* (2019) 38:394–405. doi: 10.1109/TMI.2018.2865356

76. Chen Y, Xiao T, Li C, Liu Q, Wang S. Model-based convolutional de-aliasing network learning for parallel MR imaging. In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'19)*. Springer (2019). p. 30–8. doi: 10.1007/978-3-030-32248-9_4

77. Yaman B, Hosseini SAH, Moeller S, Ellermann J, Ugurbil K, Akçakaya M. Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'20)*. Iowa City (2020). doi: 10.1109/ISBI45749.2020.9098514

78. Hu Y, Xu Y, Tian Q, Chen F, Shi X, Moran CJ, et al. RUN-UP: accelerated multishot diffusion-weighted MRI reconstruction using an unrolled network with U-net as priors. *Magn Reson Med.* (2020) 85:709–20. doi: 10.1002/mrm.28446

79. Kang E, Cahng W, Yoo J, Ye JC. Deep convolutional framelet denosing for low-dose CT via wavelet residual network. *IEEE Trans Med Imaging.* (2018) 37:1358–69. doi: 10.1109/TMI.2018.2823756

80. Hwan K, McCann MT, Froustey E, Unser M. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans Image Process.* (2017) 26:4509–22. doi: 10.1109/TIP.2017.2713099

81. Wolterink JM, Leiner T, Viergever MA, Išgum I. Generative adversarial networks for noise reduction in low-dose CT. *IEEE Trans Med Imaging.* (2017) 36:2536–45. doi: 10.1109/TMI.2017.2708987

82. Yang Q, Yan P, Zhang Y, Yu H, Shi Y, Mou X, et al. Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE Trans Med Imaging.* (2018) 37:1348–57. doi: 10.1109/TMI.2018.2827462

83. You C, Li G, Zhang Y, Zhang X, Shan H, Li M, et al. CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Trans Med Imaging.* (2020) 39:188–203. doi: 10.1109/TMI.2019.2922960

84. Gupta H, Jin KH, Nguyen HQ, McCann MT, Unser M. CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE Trans Med Imaging.* (2018) 37:1440–53. doi: 10.1109/TMI.2018.2832656

85. Xiang L, Wang L, Gong E, Zaharchuk G, Zhang T. Noise-aware standard-dose PET reconstruction using general and adaptive robust loss. In: *Proceedings of the International Conference on Machine Learning in Medical Imaging (MLMI'20)*. (2020). p. 654–62. doi: 10.1007/978-3-030-59861-7_66

86. Bevilacqua A, Bollini D, Campanini R, Lanconelli N, Galli M. A new approach to image reconstruction in positron emission tomography using artificial neural networks. *Int J Mod Phys C.* (1998) 9:71–85. doi: 10.1142/S0129183198000078

87. Yang B, Ying L, Tang J. Artificial neural network enhanced Bayesian PET image reconstruction. *IEEE Trans Med Imaging.* (2018) 37:1297–309. doi: 10.1109/TMI.2018.2803681

88. Spuhler K, Serrano-Sosa1a M, Cattell R, DeLorenzo C, Huang C. Full-count PET recovery from low-countimage using a dilated convolutional neural network. *Med Phys.* (2020) 47:4928–38. doi: 10.1002/mp.14402

89. Whiteley W, Luk WK, Gregor J, DirectPET: full-size neural network PET reconstruction from sinogram data. *J Med Imag.* (2020) 7: 032503. doi: 10.1117/1.JMI.7.3.032503

90. Xie Z, Li T, Zhang X, Qi W, Asma E, Qi J. Anatomically aided PET image reconstruction using deep neural networks. *Med Phys.* (2021) 48:5244–58. doi: 10.1002/mp.15051

91. Sandino CM, Lai P, Vasanawala SS, Cheng JY. Accelerating cardiac cine MRI using a deep learning-based ESPIRiT reconstruction. *Magn Reson Med.* (2021) 85:152–67. doi: 10.1002/mrm.28420

92. Lee D, Yoo J, Ye JC. Deep residual learning for compressed sensing MRI. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'17)*. Melbourne (2017). p. 15–8. doi: 10.1109/ISBI.2017.7950457

93. Ke Z, Cheng J, Ying L, Zheng H, Zhu Y, Liang D. An unsupervised deep learning method for multi-coil cine *MRI. Phys Med Biol.* (2020) 65:235041. doi: 10.1088/1361-6560/abaffa

94. Liu X, Zhang M, Liu Q, Xiao T, Zheng H, Ying L, et al. Multi-contrast MR reconstruction with enhanced denoising autoencoder prior learning. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'20)*. Iowa (2020). p. 1432–6. doi: 10.1109/ISBI45749.2020.9098334

95. Mills AF, Sakai O, Anderson SW, Jara H. Principles of quantitative MR imaging with illustrated review of applicable modular pulse diagrams. *Radiographics.* (2017) 37:2083–105. doi: 10.1148/rg.2017160099

96. Kwee TC, Kwee RM. Chest CT in COVID-19: What the radiologist needs to know. *Radiographics.* (2020) 40:1848–65. doi: 10.1148/rg.2020200159

97. Vayá MDLI, Saborit JM, Montell JA, Pertusa A, Bustos A, Cazorla M, et al. BIMCV COVID-19+: a large annotated dataset of RX and CT images from COVID-19 patients. (2020). *arXiv:2006.01174*.

98. Wang H, Xu Y, Zhao Y, Zhao Y. A novel static PET image reconstruction method. In: *Proceedings of the Chinese Automation Congress (CAC'17)*. Jinan (2017). p. 4537–41. doi: 10.1109/CAC.2017.8243580

99. Serrano-Sosa M, Spuhler K, DeLorenzo C, Huang C. PET image denoising using structural MRI with a novel dilated convolutional neural network. *J Nucl Med.* (2020) 61:434.

100. Ladefoged CN, Hasbak P, Hornnes C, Højgaard L, Andersen FL. Low-dose PET image noise reduction using deep learning: application to cardiac viability FDG imaging in patients with ischemic heart disease. *Phys Med Biol.* (2021) 66:054003. doi: 10.1088/1361-6560/abe225

101. Novosad P, Reader AJ. MR-guided dynamic PET reconstruction with the kernel method and spectral temporal basis functions. *Phys Med Biol.* (2016) 61:4624–45. doi: 10.1088/0031-9155/61/12/4624

102. Cui J, Liu X, Wang Y, Liu H. Deep reconstruction model for dynamic PET images. *PLoS ONE.* (2017) 12:e0184667. doi: 10.1371/journal.pone.0184667

103. Gong K, Guan J, Liu CC, Qi J. PET image denoising using a deep neural network through fine tuning. *IEEE Trans Radiat Plasma Med Sci.* (2018) 3:153–61. doi: 10.1109/TRPMS.2018.2877644

104. Sun H, Peng L, Zhang H, He Y, Cao S, Lu L. Dynamic PET image denoising using deep image prior combined with regularization by denoising. *IEEE Access.* (2021) 9:52378–92. doi: 10.1109/ACCESS.2021.3069236

105. Ouyang J, Chen KT, Gong E, Pauly J, Zaharchuk G. Ultra-low-dose PET reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Med Phys.* (2019) 46:3555–64. doi: 10.1002/mp.13626

106. Xu J, Liu H. Three-dimensional convolutional neural networks for simultaneous dual-tracer PET imaging. *Phys Med Biol.* (2019) 64:185016. doi: 10.1088/1361-6560/ab3103

107. Xu J, Liu H. Hybrid-loss guided 3D CNN for dynamic dual-tracer PET reconstruction. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'19)*. Venice (2019).

108. Ellis S, Mallia A, McGinnity CJ, Cook GJR, reader AJ. Guided image reconstruction for multi-tracer PET. In: *Proceedings of the Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC'17)*. Atlanta (2017). doi: 10.1109/NSSMIC.2017.8533140

109. Hong X, Zan Y, Weng F, Tao W, Peng Q, Huang Q. Enhancing the image quality via transferred deep residual learning of coarse PET sinograms. *IEEE Trans Med Imaging.* (2018) 37:2322–32. doi: 10.1109/TMI.2018.2830381

110. Häggström I, Schmidtlein C R, Campanella G, Fuchs TJ. DeepPET: a deep encoder–decoder network for directly solving the PET image

reconstruction inverse problem. *Med Image Anal.* (2019) 54:253–62. doi: 10.1016/j.media.2019.03.013

111. Kwon K, Kim D, Park H. A parallel MR imaging method using multilayer perceptron. *Med Phys.* (2017) 44:6209–24. doi: 10.1002/mp.12600

112. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI'15).* Springer (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28

113. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat Methods.* (2019) 16:67–70. doi: 10.1038/s41592-018-0261-2

114. Lee D, Yoo J, Tak S, Ye JC. Deep residual learning for accelerated MRI using magnitude and phase networks. *IEEE Trans Biomed Eng.* (2018) 65:1985–95. doi: 10.1109/TBME.2018.2821699

115. Han Y, Yoo J, Kim HH, Shin HJ, Sung K, Ye JC. Deep learning with domain adaptation for accelerated projection-reconstruction MR. *Magn Reson Med.* (2018) 80:1189–205. doi: 10.1002/mrm.27106

116. Duan C, Deng H, Xiao S, Xie J, Li H, Sun X, et al. Fast and accurate reconstruction of human lung gas MRI with deep learning. *Magn Reson Med.* (2019) 82:2273–85. doi: 10.1002/mrm.27889

117. Hyun CM, Kim HP, Lee SM, Lee S, Seo JK. Deep learning for undersampled MRI reconstruction. *Phys Med Biol.* (2018) 63:135007. doi: 10.1088/1361-6560/aac71a

118. Kofler A, Haltmeier M, Kolbitsch C, Kachelrieß M, Dewey M. A U-Nets cascade for sparse view computed tomography. In: *Proceedings of the Machine Learning for Medical Image Reconstruction (MLMIR'18).* Springer (2018). p. 91–9. doi: 10.1007/978-3-030-00129-2_11

119. Liu Y, Zhang Y. Low-dose CT restoration via stacked sparse denoising autoencoders. *Neurocomputing.* (2018) 284:80–9. doi: 10.1016/j.neucom.2018.01.015

120. Schaefferkoetter J, Yan, J, Ortega C, Sertic A, Lechtman E, Eshet Y, et al. Convolutional neural networks for improving image quality with noisy PET data. *EJNMMI Res.* (2020) 10:105. doi: 10.1186/s13550-020-00695-1

121. Sikka A, Peri SV, Bathula DR. MRI to FDG-PET: cross-modal synthesis using 3D U-Net for multi-modal Alzheimer's classification. In: *Proceedings of the International Conference on Simulation and Synthesis in Medical Imaging (SASHIMI'18).* Springer (2018). doi: 10.1007/978-3-030-00536-8_9

122. Cui J, Gong K, Guo N, Wu C, Meng X, Kim K, et al. PET image denoising using unsupervised deep learning. *Eur J Nucl Med Mol Imaging.* (2019) 46:2780–9. doi: 10.1007/s00259-019-04468-4

123. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'16).* Las Vegas (2016). p. 770–8.

124. Shi J, Liu Q, Wang C, Zhang Q, Ying S, Xu H. Super-resolution reconstruction of MR image with a novel residual learning network algorithm. *Phys Med Biol.* (2018) 63:085011. doi: 10.1088/1361-6560/aab9e9

125. Shi J, Li Z, Ying S, Wang C, Zhang Q, Yan P, et al. image super-resolution via wide residual networks with fixed skip connection. *IEEE J Biomed Health Inform.* (2019) 23:1129–40. doi: 10.1109/JBHI.2018.2843819

126. Li X, Cao T, Tong Y, Ma X, Niu Z, Guo H. Deep residual network for highly accelerated fMRI reconstruction using variable density spiral trajectory. *Neurocomputing.* (2020) 398:338–46. doi: 10.1016/j.neucom.2019.02.070

127. Du J, He Z, Wang L, Gholipour A, Zhou Z, Chen D, et al. Super-resolution reconstruction of single anisotropic 3D MR images using residual convolutional neural network. *Neurocomputing.* (2020) 392:209–20. doi: 10.1016/j.neucom.2018.10.102

128. Chen KT, Gong E, de Carvalho Macruz FB, Xu J, Boumis A, Khalighi M, et al. Ultra–low-dose 18F-florbetaben amyloid PET imaging using deep learning with multi-contrast MRI inputs. *Radiology.* (2019) 290:649–56. doi: 10.1148/radiol.2018180940

129. Dong X, Vekhande S, Cao G. Sinogram interpolation for sparse-view micro-CT with deep learning neural network. In: *Proceedings of the Medical Imaging 2019: Physics of Medical Imaging* (2019).

130. Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair, S, et al. Generative adversarial nets. *Adv Neural Inform Processing Syst.* (2014) 2672–80.

131. Mardani M, Gong E, Cheng JY, Vasanawala SS, Zaharchuk G, Xing L, et al. Deep generative adversarial neural networks for compressive sensing MRI. *IEEE Trans Med Imaging.* (2019) 38:167–79. doi: 10.1109/TMI.2018.2858752

132. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. (2014). *arXiv:1409.1556.*

133. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D conditional generative adversarial networks for high-quality PET image estimation at low dose. *Neuroimage.* (2018) 174:550–62. doi: 10.1016/j.neuroimage.2018.03.045

134. Ben-Cohen A, Klang E, Raskin SP, Soffer S, Ben-Haim S, Kone E. et al. Cross-modality synthesis from CT to PET using FCN and GAN networks for improved automated lesion detection. *Eng Appl Artif Intell.* (2019) 78:186–94. doi: 10.1016/j.engappai.2018.11.013

135. Yaakub SN, McGinnity CJ, Clough JR, Kerfoot E, Girard N, Guedj E, et al. Pseudo-normal PET synthesis with generative adversarial networks for localising hypometabolism in epilepsies. In: *Proceedings of the International Workshop on Simulation and Synthesis in Medical Imaging.* (2019). doi: 10.1007/978-3-030-32778-1_5

136. Kaplan S, Zhu Y-M. Full-dose PET image estimation from low-dose PET image using deep learning: a pilot study. *J Digit Imaging.* (2019) 32:773–8. doi: 10.1007/s10278-018-0150-3

137. Yang L, Dong X, Wang T, Higgins K, Liu T, Curran WJ, et al. Whole-body PET estimation from low count statistics using cycle-consistent generative adversarial networks. *Phys Med Biol.* (2019) 64:215017. doi: 10.1088/1361-6560/ab4891

138. Schlemper J, Caballero J, Hajnal JV, Price AN, Rueckert D, A. deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging.* (2018) 37:491–503. doi: 10.1109/TMI.2017.2760978

139. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, Rueckert D. Convolutional recurrent neural networks for dynamic MR image reconstruction. *IEEE Trans Med Imaging.* (2019) 38:280–90. doi: 10.1109/TMI.2018.2863670

140. Ran M, Xia W, Huang Y, Lu Z, Bao P, Liu Y, et al. MD-Recon-Net: a parallel dual-domain convolutional neural network for compressed sensing MRI. *IEEE Trans Radiat Plasma Med Sci.* (2020) 5:120–35. doi: 10.1109/TRPMS.2020.2991877

141. El-Rewaidy H, Neisius U, Mancio J, Kucukseymen S, Rodriguez J, Paskavitz A, et al. Deep complex convolutional network for fast reconstruction of 3D late gadolinium enhancement cardiac MRI. *NMR Biomed.* (2020) 33:e4312. doi: 10.1002/nbm.4312

142. Cole E, Cheng J, Pauly J, Vasanawala S. Analysis of deep complex-valued convolutional neural networks for MRI reconstruction and phase-focused applications. *Magn Reson Med.* (2021) 86:1093–109. doi: 10.1002/mrm.28733

143. Shen C, Gonzalez Y, Chen L, Jiang SB, Jia X. Intelligent parameter tuning in optimization-based iterative CT reconstruction via deep reinforcement learning. *IEEE Trans Med Imaging.* (2018) 37:1430–9. doi: 10.1109/TMI.2018.2823679

144. Du Q, Qiang Y, Yang W, Wang Y, Muhhammad BZ. DRGAN: a deep residual generative adversarial network for PET image reconstruction. *IET Image Processing.* (2020) 14:1690–700. doi: 10.1049/iet-ipr.2019.1107

145. Huang Q, Yang D, Wu P, Qu H, Yi J, Metaxas D, et al. reconstruction via cascaded channel-wise attention network. In: *Proceedings of the International Symposium on Biomedical Imaging (ISBI'19).* Venice (2019). doi: 10.1109/ISBI.2019.8759423

146. Zhang T, Jackson1 LH, Uus A, Clough JR, Story L, Rutherford MA, et al. Self-supervised recurrent neural network for 4D abdominal and In-utero MR imaging. In: *International Workshop on Machine Learning for Medical Image Reconstruction.* Springer (2019). doi: 10.1007/978-3-030-33843-5_2

147. Shan H, Padole A, Homayounieh F, Kruger U, Khera RD, Nitiwarangkul C, et al. Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose CT image reconstruction. *Nat Mach Intell.* (2019) 1:269–76. doi: 10.1038/s42256-019-0057-9

148. Xiang L, Wang Q, Nie D, Zhang L, Jin X, Qiao Y, et al. Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image. *Med Image Anal.* (2018) 47:31–44. doi: 10.1016/j.media.2018.03.011

149. Ge Y, Wei D, Xue Z, Wang Q, Zhou X, Zhan Y. Unpaired Mr to CT synthesis with explicit structural constrained adversarial learning. In: *Proceedings of*

the *International Symposium on Biomedical Imaging (ISBI'19)*. Venice (2019). p. 1096–99.

150. Koh PW, Liang P. Understanding black-box predictions via influence functions. In: *Proceedings of the International Conference on Machine Learning (PMLR'17)*. Sydney (2017). p. 2976–87.

151. Zhang Q, Wu Y, Zhu S-C. Interpretable convolutional neural networks. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'18)*. Salt Lake (2018). p. 8827–36. doi: 10.1109/CVPR.2018.00920

152. Pope PE, Kolouri S, Rostami M, Martin CE, Hoffmann H. Explainability methods for graph convolutional neural networks. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'19)*. Long Beach (2019). p. 10772–81. doi: 10.1109/CVPR.2019.01103

153. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'16)*. Las Vegas (2016). p. 2921–9. doi: 10.1109/CVPR.2016.319

154. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'17)*. Springer (2017). p. 618–26. doi: 10.1109/ICCV.2017.74

155. Towards trustable machine learning. *Nat Biomed Eng.* (2018) 2:709–10. doi: 10.1038/s41551-018-0315-x

156. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. (2017). *arXiv:1706.03762*.

Check for updates

AutoRadiomics: A Framework for Reproducible Radiomics Research

Piotr Woznicki*, Fabian Laqua, Thorsten Bley and Bettina Baeßler

Department of Diagnostic and Interventional Radiology, University Hospital Würzburg, Würzburg, Germany

Purpose: Machine learning based on radiomics features has seen huge success in a variety of clinical applications. However, the need for standardization and reproducibility has been increasingly recognized as a necessary step for future clinical translation. We developed a novel, intuitive open-source framework to facilitate all data analysis steps of a radiomics workflow in an easy and reproducible manner and evaluated it by reproducing classification results in eight available open-source datasets from different clinical entities.

Methods: The framework performs image preprocessing, feature extraction, feature selection, modeling, and model evaluation, and can automatically choose the optimal parameters for a given task. All analysis steps can be reproduced with a web application, which offers an interactive user interface and does not require programming skills. We evaluated our method in seven different clinical applications using eight public datasets: six datasets from the recently published WORC database, and two prostate MRI datasets—Prostate MRI and Ultrasound With Pathology and Coordinates of Tracked Biopsy (Prostate-UCLA) and PROSTATEx.

Results: In the analyzed datasets, AutoRadiomics successfully created and optimized models using radiomics features. For WORC datasets, we achieved AUCs ranging from 0.56 for lung melanoma metastases detection to 0.93 for liposarcoma detection and thereby managed to replicate the previously reported results. No significant overfitting between training and test sets was observed. For the prostate cancer detection task, results were better in the PROSTATEx dataset (AUC = 0.73 for prostate and 0.72 for lesion mask) than in the Prostate-UCLA dataset (AUC 0.61 for prostate and 0.65 for lesion mask), with external validation results varying from AUC = 0.51 to AUC = 0.77.

Conclusion: AutoRadiomics is a robust tool for radiomic studies, which can be used as a comprehensive solution, one of the analysis steps, or an exploratory tool. Its wide applicability was confirmed by the results obtained in the diverse analyzed datasets. The framework, as well as code for this analysis, are publicly available under https://github.com/pwoznicki/AutoRadiomics.

Keywords: radiomics, radiology, machine learning, reproducibility, workflow, image analysis

INTRODUCTION

Over the past decades, the search for novel, quantitative imaging biomarkers has been an emerging topic in the research landscape, with the ultimate goal of leveraging the full potential of medical imaging and enabling more personalized medical care (1, 2). Within this field, radiomics has been identified as a potential way to mathematically extract clinically meaningful quantitative imaging biomarkers (so-called features) from medical images of different modalities (3–5). Combined with machine learning (ML), radiomics classifiers have been shown to accurately predict the diagnosis (6), prognosis (7), mutational status / genetic subtypes (8–10), histopathology (8), surgery (11), or treatment response (12). Consequently, there is a huge interest in the clinical and research field to translate the diagnostic and prognostic potential of radiomics to clinical patient care.

This interest has resulted in a large number of scientific publications being issued with a similarly large variety of methods and radiomics pipelines. Besides the inherent issue of model overfitting, which comes with any ML and big data application where the number of features usually considerably exceeds the number of samples in the training set, most radiomics studies also have been proven difficult to reproduce and validate. This may be also due to the large variety of methodology and the lack of an open-science mindset within the research community, with the datasets and code rarely published alongside the results.

Fortunately, an evolving body of open-science frameworks has been accumulating in recent years, and new initiatives aiming at standardization and reproducibility of different aspects of radiomics analysis and ML have been founded. For example, the Image Biomarker Standardization Initiative (IBSI) (3) has addressed the standardization of the radiomic feature extraction process, while the Workflow for Optimal Radiomics Classification (WORC) (2) has been developed in order to automate and standardize a typical radiomics (and ML) workflow.

Performing reproducible radiomics studies usually requires programming skills, since the most prevalent tools in the research community are written in Python language (1–3). This makes it very difficult for clinicians (who will be the ones responsible for clinical translation of trained models and classifiers) to perform radiomics studies by themselves or to simply "play around" with the data.

The aim of this study was to present an intuitive, open-source framework with an interactive user interface for reproducible radiomics workflow. We evaluated its performance on eight publicly available datasets covering varying clinical applications to prove that the framework is able to reproduce previously published studies. AutoRadiomics provides tools for every step of the radiomics workflow (including image segmentation, image processing, feature extraction, classification, and evaluation) with the ability to adjust each step of the workflow as needed. We believe this framework may help to bridge the gap from programmers to clinicians and enable them to quickly experiment with their datasets in a reproducible way.

MATERIALS AND METHODS

This analysis is divided into two main parts: Section Framework describes design principles that we followed while designing AutoRadiomics, and Section Experiment provides information on experiments that were performed to evaluate its performance in publicly available tomography imaging datasets.

Framework

AutoRadiomics is an open-source Python package with an embedded web application with an interactive user interface. The framework can be accessed at https://github.com/pwoznicki/AutoRadiomics, where all the details on its development can be found. The framework is built around the standard steps of a radiomics workflow, including image processing, feature extraction, feature selection, dataset rebalancing, ML model selection, training, optimization, and evaluation. The main components of the framework are presented schematically in **Figure 1**. AutoRadiomics uses standard libraries validated in multiple radiomics studies, such as *pyradiomics* (1) for feature extraction, *scikit-learn* (13) for ML models and data splitting, and *imbalanced-learn* (14) for over-/undersampling. These reliable
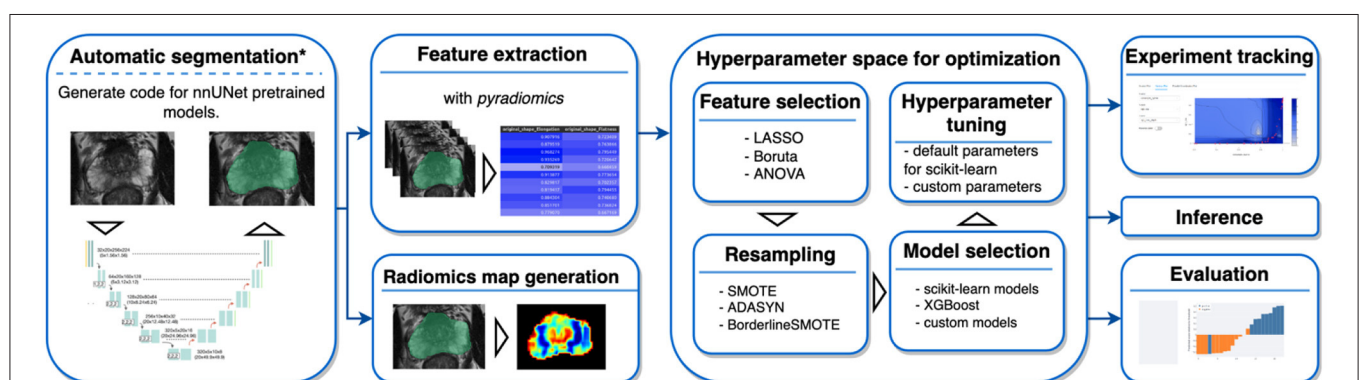
FIGURE 1 | Framework components. AutoRadiomics has a modular architecture, and its components are based on the typical steps in a radiomics analysis. *The first analysis step, automatic segmentation, is not performed inside the framework directly, but a script is generated that can be run separately.
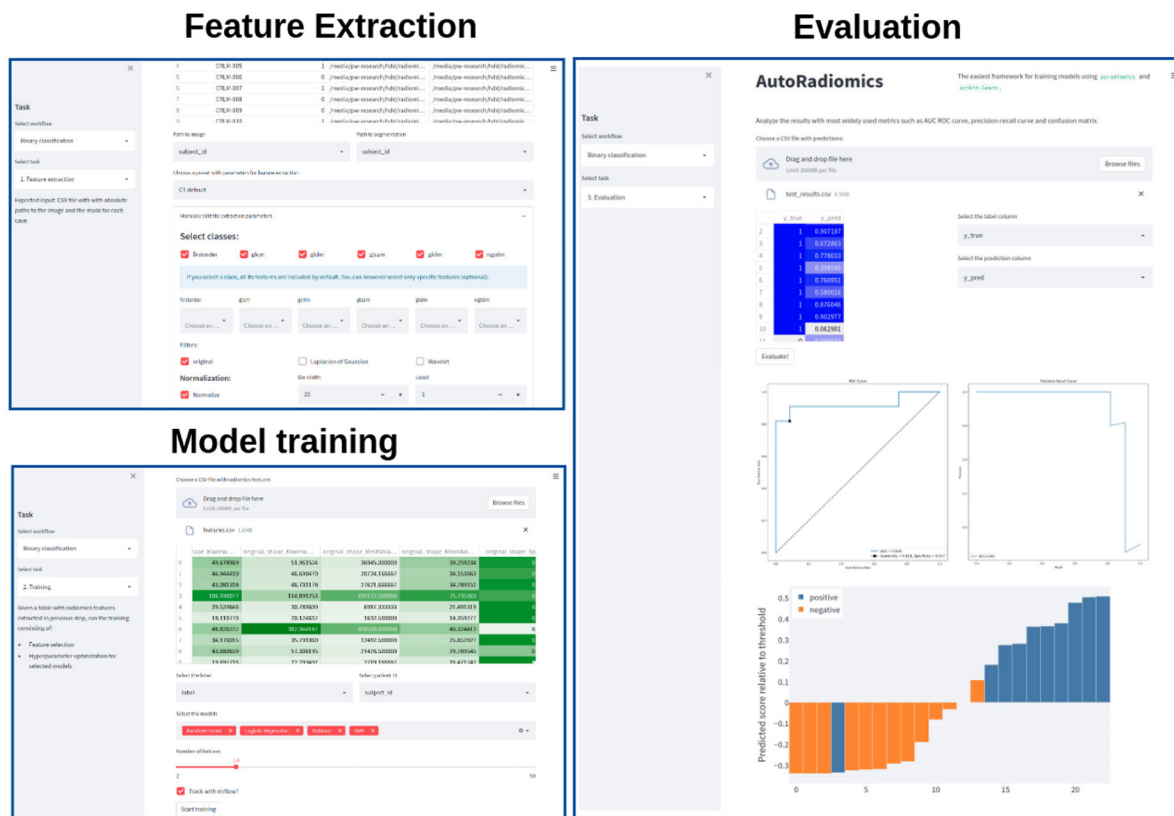
FIGURE 2 | Exemplary screenshots of the web application. The application enables users to perform all the analysis steps including feature extraction, model training, and evaluation, using standardized or custom settings.
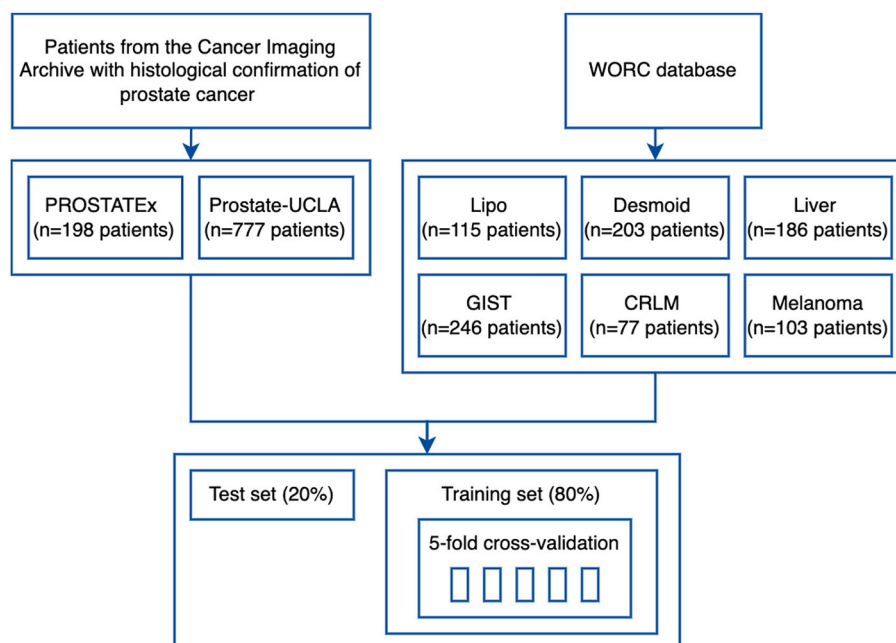
FIGURE 3 | Study flowchart.

building blocks further contribute to creating robust and reproducible workflows. The framework is available under the Apache-2.0 License.

Data Preparation and Radiomic Feature Extraction

Data splitting in AutoRadiomics is performed on provided case IDs. Depending on dataset size and application, user can choose to split the data into k folds for cross-validation (with or without a separate test set) or into training/validation/test sets. Radiomic features, including standard shape, intensity, and texture features, are extracted with *pyradiomics*, with additional parameters specified in the parameter file. A few built-in options are provided for this purpose, including extraction parameters validated in previous studies (12, 15). Additional optimizations in computing resource allocation make the extraction process more efficient.

Hyperparameter Optimization and Experiment Tracking

Hyperparameter optimization is performed using the Optuna framework (16), which dynamically constructs the search space for hyperparameters and automatically chooses optimal ones. The framework simultaneously optimizes the choice and hyperparameters of the ML classifiers as well as feature selection and oversampling methods, which greatly simplifies the training workflow. The following classifiers are included: logistic regression, support vector machines, random forest, and extreme gradient boosting (XGBoost). Experiments are tracked using an integrated MLFlow tracking dashboard, which allows the user to explore the training artifacts as well as the metrics during and after the training process.

Web Application

Recognizing the problems many non-expert users may face when being forced to use a programming interface, we developed a browser application with an interactive user interface on top of the Python package. The app can be run locally as a Docker container, satisfying the necessary privacy concerns. It adopts a straightforward, modular structure to the radiomics workflow and covers sequentially all steps of the analysis pipeline. The output of each intermediary step, training parameters, and logs are stored in the experiment's directory. That enables the user to later come back to the experiment and document the workflow. **Figure 2** presents exemplary screenshots of the app. The app also provides utilities for generating Python code that can be then executed as a separate script to perform automatic segmentation using the state-of-the-art *nnU-Net* framework (17), and for generating radiomics maps using voxel-based feature extraction.

Experiment

Data Sources

To validate the developed framework, we used eight datasets from two different sources. Firstly, we used six public datasets from the recently published WORC database (18), which includes multi-institutional annotated CT and MRI datasets with varying clinical applications. The respective classification tasks were (1) well-differentiated liposarcoma vs. lipoma, (2) desmoid-type

fibromatosis vs. extremity soft-tissue sarcoma, (3) primary solid liver tumor, malignant vs. benign, (4) gastrointestinal stromal tumor (GIST) vs. intra-abdominal gastrointestinal tumor, (5) colorectal liver metastases vs. non-metastatic tumor, and (6) lung metastases of melanoma vs. lung tumor of different etiology. The database was released together with benchmark results to facilitate reproducibility in the radiomics field and, to our knowledge, we are the first ones to replicate the previously published results (2).

Additionally, two public prostate MRI datasets, which are available on The Cancer Imaging Archive, were used: Prostate MRI and Ultrasound With Pathology and Coordinates of Tracked Biopsy (19) from the University of California, Los Angeles (UCLA) (further referred to as Prostate-UCLA) and PROSTATEx (20) with annotations from Cuocolo et al. (21). These two datasets were selected since they both had segmentations of prostate gland and lesions as well as biopsy evaluation including Gleason Score (GS) available. All lesions from the Prostate-UCLA dataset had targeted biopsy performed. For PROSTATEx, all lesions with PI-RADS ≥ 3 were biopsied.

TABLE 1 | Characteristics of training and test cohorts.

		Number of patients	
	Imaging modality	Training	Test
Lipo	T1w MRI		
well-differentiated liposarcoma		45 (49%)	11 (48%)
lipoma		46 (51%)	12 (52%)
Desmoid	T1w MRI		
desmoid-type fibromatosis		57 (35%)	15 (37%)
extremity soft-tissue sarcoma		105 (65%)	26 (63%)
Liver	T2w MRI		
malignant primary solid liver tumor		75 (51%)	19 (50%)
benign primary solid liver tumor		73 (49%)	19 (50%)
GIST	CT		
gastrointestinal stromal tumor		99 (51%)	25 (51%)
other intra-abdominal tumors		97 (49%)	24 (49%)
CLRM	CT		
colorectal liver metastases		29 (48%)	8 (50%)
other colorectal tumors		32 (52%)	8 (50%)
Melanoma	CT		
lung metastases of melanoma		38 (50%)	9 (47%)
other lung tumors		38 (50%)	10 (53%)
PROSTATEx	T2w MRI		
benign prostate lesion		80 (51%)	20 (50%)
ISUP grade 1 (GS = 6)		23 (15%)	6 (15%)
ISUP grade 2 (GS $\geq$ 7)		55 (35%)	14 (35%)
Prostate-UCLA	T2w MRI		
benign prostate lesion		142 (23%)	36 (23%)
ISUP grade 1 (GS = 6)		146 (24%)	37 (24%)
ISUP grade $\geq$ 2 (GS $\geq$ 7)		333 (54%)	83 (53%)

T1w, T1-weighted; T2w, T2-weighted; ISUP, International Society of Urological Pathology; GS, Gleason score.

TABLE 2 | Classification results.

Dataset	AUC		F1	Sensitivity	Specificity
	Five-fold CV	Test	Test	Test	Test
Lipo	0.86 ± 0.10	0.93 [0.77–1.0]	0.85 [0.67–1.0]	0.82 [0.56–1.0]	0.92 [0.73–1.0]
Desmoid	0.78 ± 0.05	0.90 [0.79–0.98]	0.77 [0.57–0.92]	0.80 [0.58–1.0]	0.84 [0.68–0.96]
Liver	0.64 ± 0.11	0.67 [0.49–0.84]	0.68 [0.48–0.82]	0.69 [0.45–0.88]	0.68 [0.47–0.88]
GIST	0.68 ± 0.03	0.69 [0.53–0.84]	0.72 [0.56–0.84]	0.72 [0.54–0.88]	0.71 [0.5–0.88]
CRLM	0.68 ± 0.14	0.75 [0.43–1.0]	0.82 [0.57–1.0]	0.88 [0.6–1.0]	0.75 [0.43–1.0]
Melanoma	0.59 ± 0.15	0.56 [0.29–0.93]	0.48 [0.15–0.76]	0.44 [0.13–0.8]	0.70 [0.42–1.0]
Prostate masks					
PROSTATEx	0.70 ± 0.03	0.73 [0.55–0.89]	0.73 [0.55–0.86]	0.75 [0.53–0.94]	0.69 [0.47–0.88]
UCLA	0.48 ± 0.09	0.61 [0.51–0.71]	0.78 [0.72–0.74]	0.75 [0.67–0.82]	0.44 [0.28–0.61]
PROSTATEx→ UCLA*		0.70 [0.62–0.79]	0.52 [0.43–0.62]	0.36 [0.27–0.45]	0.97 [0.90–1.0]
UCLA→ PROSTATEx		0.60 [0.41–0.79]	0.69 [0.52–0.83]	0.80 [0.61–0.95]	0.49 [0.27–0.73]
Lesion masks					
PROSTATEx	0.68 ± 0.10	0.72 [0.56–0.86]	0.68 [0.48–0.84]	0.55 [0.33–0.76]	0.95 [0.82–1.0]
UCLA	0.59 ± 0.03	0.65 [0.55–0.74]	0.73 [0.67–0.80]	0.64 [0.55–0.73]	0.64 [0.48–0.8]
PROSTATEx→ UCLA		0.51 [0.41–0.61]	0.31 [0.21–0.41]	0.19 [0.12–0.27]	0.92 [0.82–1.0]
UCLA→ PROSTATEx		0.77 [0.60–0.91]	0.74 [0.57–0.87]	0.70 [0.5–0.89]	0.80 [0.61–0.95]

Results are presented as mean ± std for five-fold cross validation and mean with 95% CI in brackets for test set.
*arrow denotes external validation of the model trained on PROSTATEx in the Prostate-UCLA dataset.
FIve-fold CV, five-fold cross-validation; UCLA, Prostate-UCLA dataset.

We trained radiomics models based on either the whole prostate gland or the target lesion masks in T2-weighted MR images to differentiate between benign prostate lesions and prostate cancer, as well as between clinically significant and clinically insignificant prostate cancer.

Data Processing

The study flowchart is presented in **Figure 3.** For each dataset, we split 80% of the data into training and 20% into the test set. Then, we split the training set into 5 folds to perform hyperparameter optimization using a cross-validation approach. Image and segmentation data were converted into the NIfTI format, where necessary, and no additional image preprocessing was applied. For feature extraction, we used separate extraction and image processing parameter sets for MRI and CT datasets, as recommended by the IBSI (3). Hyperparameter optimization was performed for each dataset with Optuna using 200 trials of the Tree-structured Parzen Estimator (TPE) algorithm to maximize the objective function.

Statistical Analysis

Receiver operating characteristic (ROC) curves were generated for each independent variable and the area under the curve (AUC) was calculated. The diagnostic efficacy of the model was additionally evaluated using the F1 score, sensitivity, and specificity, and was reported with 95% confidence intervals (95% CI) obtained with the bootstrap technique. The bootstrap used 1,000 resamples (with replacement) of predicted probabilities to determine the 95% CI. All analyses were performed with the AutoRadiomics framework, using Python 3.8.10.

RESULTS

All the experiments were successfully implemented using Python, but can also be reproduced using the interactive web application. **Supplementary Figure S1** shows the code extract required to run the optimization and evaluation process for a selected dataset (the implementation assumes a table with data paths is already created). The optimal configurations of models selected for each task are presented in **Supplementary Appendix S3**. The execution time of the whole pipeline, including the optimization, took around 1 h on a machine with 16 GB RAM and 8-core AMD Ryzen 5800X processor.

The details of training and test cohorts for each task are shown in **Table 1**. In total, we included 1895 patients in our analyses. In the six datasets from the WORC database, the class distribution was approximately balanced. For the two prostate datasets, the distribution of classes differed between datasets: in PROSTATEx, 50% of index lesions were classified as benign, 15% as GS 6, and 35% as GS ≥7, compared to only 23% of index lesions classified as benign, 24% as GS = 6, and 54% as GS≥7 in Prostate-UCLA.

Table 2 summarizes the classification results and **Figure 4** presents the corresponding ROC curves for all included datasets. In the following, we report the results from the test cohorts.

In the WORC database, we obtained results ranging from weak discrimination for the Melanoma dataset (AUC = 0.56 [95% CI: 0.29–0.93], F1 = 0.48 [95% CI: 0.15–0.76]) to excellent discrimination for the Lipo dataset (AUC = 0.93 [95% CI: 0.77–1.0], F1 = 0.85 [95% CI: 0.67–1.0]) and the Desmoid dataset (AUC = 0.90 [95% CI: 0.79–0.98], F1 = 0.77 [95% CI: 0.57–0.92]).
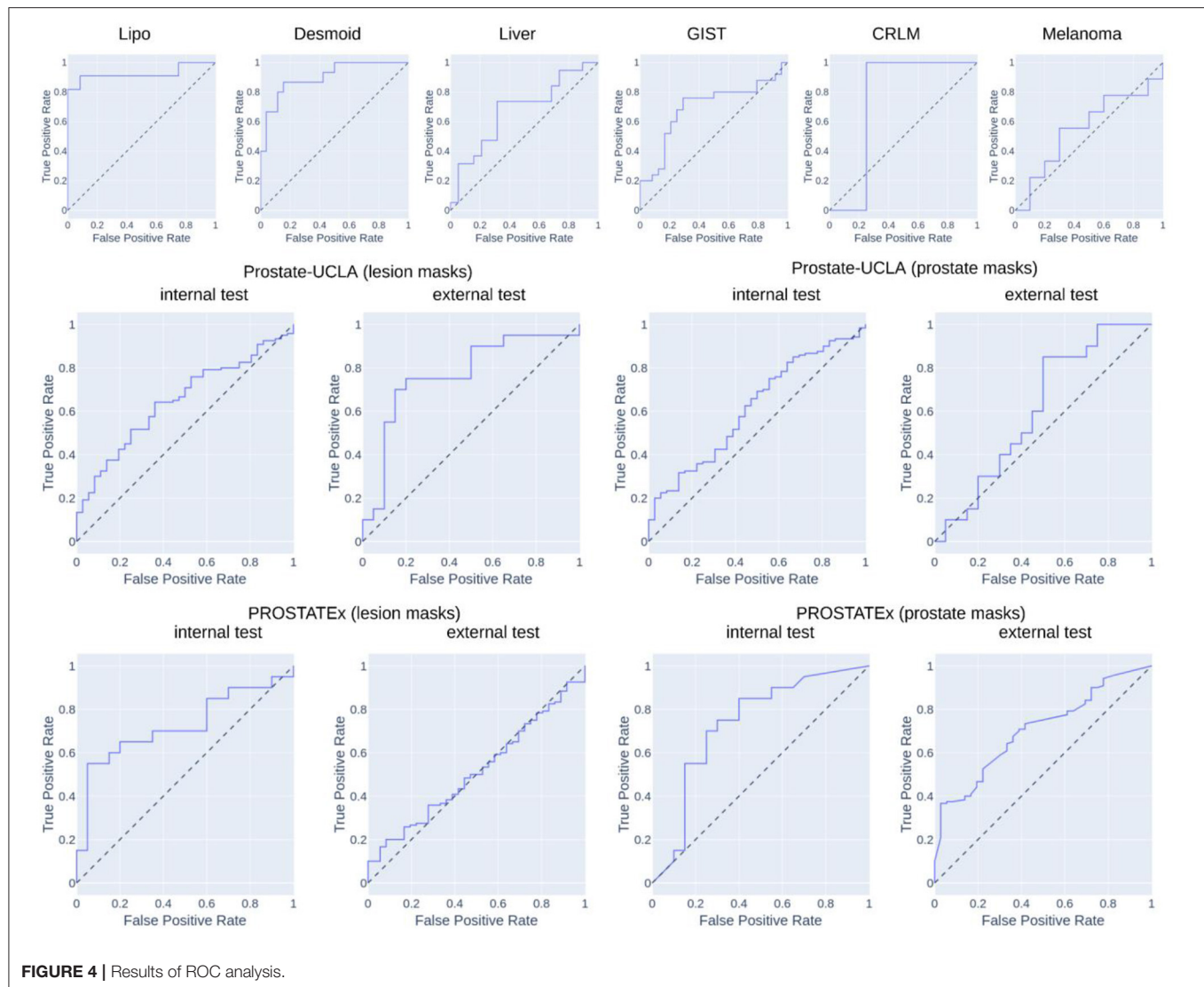
FIGURE 4 | Results of ROC analysis.

For prostate datasets, results are reported separately for classification using features from either prostate or lesion masks. The discrimination was acceptable for both prostate masks (AUC = 0.73 [95% CI: 0.55–0.89], F1 = 0.73 [0.55–0.86]) as well as lesion masks (AUC = 0.72 [95% CI: 0.56–0.86], F1 = 0.68 [95% CI: 0.48–0.84]) in the PROSTATEx dataset, and moderate for prostate masks (AUC = 0.61 [95% CI: 0.51–0.71], F1 = 0.78 [95% CI: 0.72–0.74]) and lesion masks (AUC = 0.65 [95% CI: 0.55–0.74], 0.73 [95% CI: 0.67–0.80]) in the Prostate-UCLA dataset. Both prostate datasets were additionally validated using the other dataset, and their performance varied from AUC = 0.51 [95% CI: 0.41–0.61] for the PROSTATEx model using lesion masks evaluated in Prostate-UCLA to AUC = 0.77 [95% CI: 0.60–0.91] for the Prostate-UCLA using lesion masks evaluated in PROSTATEx.

The additional evaluation of the prostate MRI datasets for differentiation between clinically significant and clinically non-significant prostate cancer is presented in the **Supplementary Table S2**. For this challenging task, the results

were worse than those for prostate cancer detection, with AUCs ranging from 0.40 [95% CI: 0.29–0.50] for the Prostate-UCLA dataset to AUC = 0.70 [95% CI: 0.33–0.97] for the PROSTATEx dataset, trained with prostate masks. The external validation results in this dataset showed AUCs in the range of 0.37 to 0.70 with high variability.

DISCUSSION

In this study, we introduced and validated a new open-source, interactive framework for reproducible radiomics research. The tool aids in selecting the optimal model for a given task, and the associated web application lowers the entry threshold for clinicians who want to contribute to the field of radiomics research and foster clinical translation.

We evaluated AutoRadiomics in six different classification tasks from the WORC database. It achieved consistently high AUCs in both cross-validation and the test set, in the direct comparison of our results vs. those reported in the original

publication on the dataset (2): 0.93 vs. 0.83 for the Lipo dataset, 0.90 vs. 0.82 for the Desmoid dataset, 0.67 vs. 0.81 for the Liver dataset, 0.69 vs. 0.77 for the GIST dataset, 0.75 vs. 0.68 for the CRLM dataset, 0.56 vs. 0.51 for the Melanoma dataset. That means, our framework achieved comparable results, higher in 4/6 tasks, and lower in 2/6 tasks. We also evaluated our framework in two public prostate MRI datasets, achieving AUCs in the range of 0.61–0.73 for internal validation, and 0.51–0.77 for external validation. Those results prove that AutoRadiomics can be successfully applied off-the-shelf and achieve competitive results with its automatic configuration. We believe the differences between ours and previously reported results may be largely explained by the relatively small sample sizes, different data splitting, and differing choice of classifiers. It has to be noted, however, that, similarly to Starmans et al. (2), we achieved best results for the Lipo dataset, and worst for the Melanoma dataset, which suggests both approaches have converged to an optimal solution.

Quantitative evaluation of disease patterns in medical images which are invisible to the human eye has shown diagnostic potential in multiple retrospective studies, but large-scale clinical validation and adoption are still missing (22). We believe that an accessible toolkit for exploratory data analysis and a standardized workflow is a key component in developing the field toward clinical translation. With this in mind, we released AutoRadiomics as an intuitive open-source framework that structures the radiomics workflow and makes it more accessible and reproducible.

Recent advances in automated ML have the potential to empower healthcare professionals with limited data science expertise (23). Inspired by those breakthroughs, new platforms for ML applied to medical imaging have recently been introduced, such as WORC (2), which focuses on automatic construction and optimization of the radiomics workflow. While this platform also provides an automated solution and is very extensive in scope, AutoRadiomics sets itself apart with its interactive web interface, state-of-the-art tooling, and additional utilities (i.e., for visualization and segmentation).

With our web application, we hope to shift the focus from metrics to interpretability, which is achieved through comprehensive visualizations and radiomics maps. We would like to point out a few scenarios, where AutoRadiomics could be especially helpful: (1) for clinicians exploring their dataset using the embedded web application to gain quick insight into their data, (2) for researchers using Python for radiomic analyses, who want to complement their current workflow or add a benchmark or reference standard, (3) for an inter-institutional collaboration as means of facilitating results sharing and workflow reproducibility.

Currently, our framework can be used only for binary classification tasks, which limits its applicability. We are planning to extend it in the future to handle multiclass classification, regression tasks, and survival data. Furthermore, some processing steps such as automatic segmentation using deep learning require GPU capability, which is why it is not integrated into our framework and only the code for performing segmentation with a nnU-Net can be generated. AutoRadiomics does not require a powerful GPU and a modern personal computer is enough to run it. One should also keep in mind that the results of any optimized model have to be considered with caution and no abstraction layer (such as our web application) may replace true expert domain knowledge.

In conclusion, we herein presented AutoRadiomics, a framework for intuitive and reproducible radiomics research. We described its key features as well as the underlying architecture, and we discuss its most promising use cases. Finally, we validated it extensively in eight public datasets to show its consistently high performance in various and diverse classification tasks. We believe that AutoRadiomics may help to improve the quality and reproducibility of future radiomics studies, and, through its accessible interface, may bring those studies closer to clinical translation.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

Ethical review and approval was not required for this study in accordance with the local legislation and institutional requirements.

AUTHOR CONTRIBUTIONS

Study conception: BB. Radiomics analysis: PW. Statistical analysis: PW and FL. Draft writing: PW and BB. Manuscript edition: FL and TB. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) (SPP 2177 to BB).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fradi.2022.919133/full#supplementary-material

REFERENCES

1. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339

2. Starmans MPA, van der Voort SR, Phil T, Timbergen MJM, Vos M, Padmos GA, et al. Reproducible radiomics through automated machine learning

validated on twelve clinical applications [Internet]. *arXiv [eess.IV].* (2021). Available online at: http://arxiv.org/abs/2108.08618

3. Zwanenburg A, Leger S, Vallières M, Löck S. Image biomarker standardisation initiative. *Radiology.* (2020) 295:328–38. doi: 10.1148/radiol.2020191145

4. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, Jong EEC de, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol.* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141

5. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

6. Baessler B, Luecke C, Lurz J, Klingel K, Roeder M, Waha S, et al. Cardiac MRI Texture Analysis of T1 and T2 maps in patients with infarctlike acute myocarditis. *Radiology.* (2018) 289:357–65. doi: 10.1148/radiol.2018180411

7. Liu Q, Li J, Liu F, Yang W, Ding J, Chen W, et al. A radiomics nomogram for the prediction of overall survival in patients with hepatocellular carcinoma after hepatectomy. *Cancer Imaging.* (2020) 20:82. doi: 10.1186/s40644-020-00360-9

8. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun.* (2020) 5:1–9. doi: 10.1038/ncomms5644

9. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol.* (2022) 19:132–46. doi: 10.1038/s41571-021-00560-7

10. Blüthgen C, Patella M, Euler A, Baessler B, Martini K, von Spiczak J, et al. Computed tomography radiomics for the prediction of thymic epithelial tumor histology, TNM stage and myasthenia gravis. *PLoS ONE.* (2021) 16:e0261401. doi: 10.1371/journal.pone.0261401

11. Mühlbauer J, Kriegmair MC, Schöning L, Egen L, Kowalewski K-F, Westhoff N, et al. Value of radiomics of perinephric fat for prediction of intraoperative complexity in renal tumor surgery. *Urol Int.* (2021) 1–12. doi: 10.1159/000520445

12. Baessler B, Nestler T, Pinto dos Santos D, Paffenholz P, Zeuch V, Pfister D, et al. Radiomics allows for detection of benign and malignant histopathology in patients with metastatic testicular germ cell tumors prior to post-chemotherapy retroperitoneal lymph node dissection. *Eur Radiol.* (2020) 30:2334–45. doi: 10.1007/s00330-019-06495-z

13. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res.* (2011) 12:2825–30. doi: 10.5555/1953048.2078195

14. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res.* (2017) 18:1–5. doi: 10.5555/3122009.3122026

15. Woznicki P, Westhoff N, Huber T, Riffel P, Froelich MF, Gresser E, et al. Multiparametric MRI for prostate cancer characterization: combined use of radiomics model with PI-RADS and clinical parameters. *Cancers.* (2020) 12:1767. doi: 10.3390/cancers12071767

16. Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. *arXiv:190710902 [cs, stat].* (2019). Available from: http://arxiv.org/abs/1907.10902 (cited April 4, 2022).

17. Baumgartner M, Jaeger PF, Isensee F, Maier-Hein KH. nnDetection: A Self-configuring Method for Medical Object Detection. *arXiv:210600817 [cs.eess].* (2021) 12905:530–9. doi: 10.1007/978-3-030-87240-3_51

18. Starmans MPA, Timbergen MJM, Vos M, Padmos GA, Grünhagen DJ, Verhoef C, et al. The WORC database: MRI and CT scans, segmentations, and clinical labels for 930 patients from six radiomics studies [Internet]. *medRxiv.* (2021). Available from: https://www.medrxiv.org/content/10.1101/2021.08.19.21262238v1 (cited March 27, 2022).

19. Sonn GA, Natarajan S, Margolis DJA, MacAiran M, Lieu P, Huang J, et al. Targeted biopsy in the detection of prostate cancer using an office based magnetic resonance ultrasound fusion device. *J Urol.* (2013) 189:86–92. doi: 10.1016/j.juro.2012.08.095

20. Litjens G, Debats O, Barentsz J, Karssemeijer N, Huisman H, TCIA Team. SPIE-AAPM PROSTATEx Challenge Data [Internet]. *The Cancer Imaging Archive.* (2017). Available from: https://wiki.cancerimagingarchive.net/x/iIFpAQ (cited March 31, 2022).

21. Cuocolo R, Stanzione A, Castaldo A, Lucia DRD, Imbriaco M. Quality control and whole-gland, zonal and lesion annotations for the PROSTATEx challenge public dataset. *Eur J Radiol.* (2021). p. 138. Available from: https://www.ejradiology.com/article/S0720-048X(21)00127-3/fulltext (cited April 4, 2022).

22. Pinto dos Santos D, Dietzel M, Baessler B. A decade of radiomics research: are images really data or just patterns in the noise? *Eur Radiol.* (2021) 31:1–4. doi: 10.1007/s00330-020-07108-w

23. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif Intell Med.* (2020) 104:101822. doi: 10.1016/j.artmed.2020.101822

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

How should studies using AI be reported? lessons from a systematic review in cardiac MRI

Ahmed Maiter[1,2†], Mahan Salehi[1†], Andrew J. Swift[1,2]
and Samer Alabed[1,2]*

[1]Department of Infection, Immunity & Cardiovascular Disease, University of Sheffield, Sheffield, United Kingdom, [2]Department of Radiology, Sheffield Teaching Hospitals, Sheffield, United Kingdom

Recent years have seen a dramatic increase in studies presenting artificial intelligence (AI) tools for cardiac imaging. Amongst these are AI tools that undertake segmentation of structures on cardiac MRI (CMR), an essential step in obtaining clinically relevant functional information. The quality of reporting of these studies carries significant implications for advancement of the field and the translation of AI tools to clinical practice. We recently undertook a systematic review to evaluate the quality of reporting of studies presenting automated approaches to segmentation in cardiac MRI (Alabed et al. 2022 Quality of reporting in AI cardiac MRI segmentation studies—a systematic review and recommendations for future studies. *Frontiers in Cardiovascular Medicine* 9:956811). 209 studies were assessed for compliance with the Checklist for AI in Medical Imaging (CLAIM), a framework for reporting. We found variable—and sometimes poor—quality of reporting and identified significant and frequently missing information in publications. Compliance with CLAIM was high for descriptions of models (100%, IQR 80%–100%), but lower than expected for descriptions of study design (71%, IQR 63–86%), datasets used in training and testing (63%, IQR 50%–67%) and model performance (60%, IQR 50%–70%). Here, we present a summary of our key findings, aimed at general readers who may not be experts in AI, and use them as a framework to discuss the factors determining quality of reporting, making recommendations for improving the reporting of research in this field. We aim to assist researchers in presenting their work and readers in their appraisal of evidence. Finally, we emphasise the need for close scrutiny of studies presenting AI tools, even in the face of the excitement surrounding AI in cardiac imaging.

KEYWORDS

artificial intelligence, machine learning, cardiac MRI, segmentation, systematic review, quality of reporting

Introduction

The development and application of artificial intelligence (AI) is an exciting frontier in radiology (1–3). AI tools promise automation of complex and time-intensive tasks, making them appealing in an era in which the demand and complexity of medical imaging are increasing. This is reflected in the recent rapid expansion in the number of studies presenting AI tools for imaging. However, there are several challenges that need to be overcome before AI can be implemented effectively in routine clinical practice (4). Transparency of model design, training and testing is critical for understanding the generalisability of tool but can be problematic where technologies are proprietary. Evaluating the performance of AI tools in relevant populations and environments is also an important step for determining their external validity. There is also growing awareness of ethical issues within the field. These include concerns about the risk of AI tools propagating human biases, including racial, that

could cause discrimination for minority population groups (5–7). These challenges are inherently linked to the manner and quality in which studies of AI tools are presented.

The ability to compare evidence underpins modern medicine and necessitates that research is presented in a transparent, consistent and reproducible manner. Poor quality of reporting can contribute to research waste, hinder advancement of the field and limit clinical applicability. It is important for all stakeholders—including researchers, radiologists using AI tools, clinicians using AI-derived information and the public—to understand what constitutes high quality reporting. Structured tools have been proposed to assist the reporting of studies using AI, including the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) (8).

AI for segmentation in CMR

The demand for cardiac imaging is growing, and with it the appetite for automation. Cardiac MRI (CMR) allows non-invasive assessment of both cardiac anatomy and function. CMR can yield quantitative metrics (such as ventricular volumes, myocardial thickness and infarct sizes) that are of diagnostic and prognostic value. However, these measurements require the accurate delineation of anatomical structures on imaging, or segmentation. Those reading CMR studies have traditionally performed manual segmentation in order to derive these metrics—a process that is laborious, time-intensive and prone to interobserver variability. The ability to automate this process using AI methods has been the focus of an increasing number of studies in recent years (9–12).

In the broadest terms, AI automates processes traditionally performed by humans. Machine learning is a major branch of AI in which a program automatically identifies relevant features in data and adapts to improve its performance at a task. Machine learning encompasses a broad range of techniques, including deep learning and neural networks. In the context of segmentation in medical imaging, this involves a program learning to identify anatomical features in an image (such as the endocardium) in order to delineate structures (such as the cardiac chambers). Although the specific approaches and model designs are myriad, they have to date shared some similarities in their development.

FIGURE 1
Violin plot indicating compliance of the 209 included studies with the CLAIM criteria, grouped into domains of study, dataset, model and performance description. Median (*solid line*) and 1st and 3rd quartile (*dashed lines*) values are indicated. From Alabed et al 2022 (13).

This typically involves three stages: training, validation and testing. During training, data is passed through the algorithm and the algorithm identifies features that enable it to undertake a task. In the validation stage, the algorithm is exposed to the unseen validation set and its performance at the task is determined. The algorithm is then adapted to optimise its performance and the training and validation steps are repeated until satisfactory performance is achieved and a final model is established. The model is then tested on new, unseen, data to yield its final performance results. This is a gross simplification of varied and complex processes, but is nonetheless important for contextualising how studies using AI are reported.

The systematic review

We recently undertook a systematic review of the quality of reporting of studies using AI methods for segmentation of structures on CMR (13). Studies presenting fully automated AI methods for the segmentation of cardiac chambers, myocardium or scar tissue on adult CMR images were eligible for inclusion. Included studies were assessed for descriptive information and compliance with CLAIM. We grouped the individual CLAIM criteria into four domains: study description, dataset description, model description and performance description. 209 studies were included, undertaken in 37 different countries and published from 2012 to 2022. The median overall compliance of studies with all CLAIM criteria was 67% [interquartile range (IQR) 59–73%]. Median compliance was highest for the model description domain (100%, IQR 80%–100%) and substantially lower for the study description (71%, IQR 63%–86%), dataset description (63%, IQR 50%–67%) and performance description (60%, IQR 50%–70%) domains (Figure 1).

The development of an AI model requires training, in which an algorithm is exposed to data (such as CMR images) to learn features (such as where different anatomical structures are located relative to each other) that enable it to undertake a task (such as to delineate the left ventricular endocardium). This process is critical and underpins the performance and validity of all AI tools. The quality, size and variation of the dataset being used for training are of particular importance when considering the clinical applicability of a model, as a model trained on data from one population or demographic group may not generalise well when applied to others. For example, a model trained solely on CMR images from younger patients may not perform as well when used on an older population. It is essential that studies describe their data sources in a clear and transparent manner so that the generalisability of their models can be understood. This includes information about cases (such as number, eligibility criteria and clinical characteristics) and the nature of the data itself (such as the type of images and how they were acquired).

We found that although most studies indicated their data sources (94%), this was a significant omission when missing. Approximately half used publicly available datasets (49%), of which most (66%) were made available through Medical Image Computing and Computer Assisted Intervention (MICCAI) challenges, emphasising their role in advancing the field. Publicly available datasets aid reproducibility and comparison between models, but as with any retrospective data source have their own selection biases. Multiple
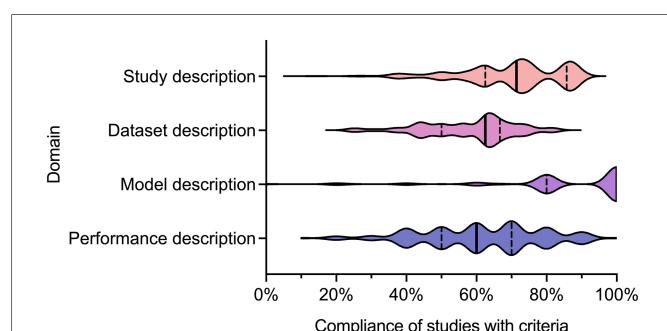
or combined datasets were used in few studies (17%) but have the potential to improve the generalisability of models by exposure to different populations. Most studies reported the number of cases used (95%), with a median of 78 and a wide range of 3 to 12,984. Insufficient case numbers and variability are likely to affect generalisability. A minority of studies failed to report the type of CMR image used for segmentation (14%), greatly limiting the interpretability of their models.

Similarly, detailed description of the structure of AI models and the training approach are important and expected in this field. Again, this should be transparent and reproducible. Understanding the model structure can help to highlight biases in performance and thus model generalisability. However, this can be challenging due to proprietary "black-box" methodologies in design. Furthermore, publications should be written in an accessible manner such that methods are not obscured. For example, studies that present a clinical message should ensure that computer science methods and concepts (such as model structure) are explained clearly for readers who may not be AI experts (and vice versa). This balance can be difficult to achieve in such a rapidly evolving and technical field. We found that compliance with the model description domain was indeed excellent. This may reflect the fact that most were published in technical (58%) and hybrid (11%) journals. Most studies provided details about the model used (95%), training approach (78%) and software used (74%). However, open source code was only provided in a minority of studies (10%). Publishing the open source code for an AI model greatly improves transparency and facilitates the comparison of different models.

Understanding how effectively AI models perform is essential for their translation into clinical practice. Performance needs to be described in a consistent manner to enable comparison between models. However, we found that descriptions of model performance were variable, with many publications failing to present key information. The way in which performance is assessed can vary and needs to be transparent. Ideally, this should involve testing a model using a distinct and external dataset (such as images from a different population acquired by a different centre). This represents an important step in ensuring that an AI model is generalisable and valid for translation into clinical use. Only a minority (22%) of the studies that we assessed tested their models on external data. It is expected that AI models can fail, and it is good practice for studies to present an analysis of failed cases to indicate how and why this occurred. This is crucial for advancement of the field and clinical implementation. A clinician using an AI model will need to understand the factors that may predispose to false results. This goes hand-in-hand with understanding measures of diagnostic accuracy (such as sensitivity and specificity), which are major determinants of clinical utility. We noted that few studies reported failure analysis of incorrectly classified cases (32%) or estimates of diagnostic accuracy (21%).

To the best of our knowledge, this study was the largest review of the AI-based cardiac imaging literature to date. There are, of course, limitations. The review had a narrow focus on AI approaches to segmentation in CMR. Only journal papers presenting fully automated techniques were included. Semi-automated techniques incorporate both manual and AI-based elements and their distinction from fully automated techniques is open to a degree of subjectivity. The exclusion of semi-automated techniques, unpublished literature and conference abstracts were important to ensure consistent and reproducible evaluation of the included studies

Preparing a publication	• Utilise a reporting framework (e.g. CLAIM) • Present the study in an accessible manner, keeping the target audience in mind • Use consistent terminology
Reporting data sources	• Ensure transparent description of data sources • Provide demographic and clinical characteristics of the cases used • Be explicit about the number of cases and imaging studies • Indicate the flow of cases and studies into different datasets (e.g. training, validation, testing) • Report details about the types of images and how they were acquired
Reporting model training	• Ensure reproducible description of the model and training process • Provide open source code for the model if possible • State what software frameworks, libraries and packages were used • Clearly state the reference standard and how it was obtained
Reporting model performance	• Ensure reproducible description of how performance was assessed • State what data was used for testing and how it differed from data used for training • Provide statistical measures of significance and uncertainty • Report failure analysis for incorrectly classified cases • State limitations of the model

FIGURE 2
Recommendations for studies based on findings of this systematic review. Adapted from Alabed et al 2022 (13) and CLAIM (8).

but did narrow the scope of the review and carried the risk of selection bias. Finally, there is an inherent risk of observer bias and interobserver variability when evaluating quality of reporting, even when using structured tools such as CLAIM; future studies may consider assessing interobserver agreement quantitatively. However, despite these limitations, our study has considered important factors for how AI studies in general are presented, and our findings are likely applicable to the broader field of AI in medical imaging.

Discussion

This systematic review identified significant and frequent gaps in the existing literature. In this paper, we have explored some of the hallmarks of high-quality AI publications in cardiac imaging. We encourage researchers and readers to bear these in mind when presenting and appraising studies using AI methods. Based on the findings in our systematic review, we make a number of recommendations for researchers to improve the quality of reporting of AI studies, which are provided in Figure 2. Study methodology should be described in sufficient detail to enable reproducibility. Information about all data sources, including clinical characteristics of all participants, should be provided in order to understand study validity and generalisability. Testing on multiple and external datasets is an important step in the translation of AI models to clinical practice. Studies in this field may have a wide readership and publications should be accessible and transparent regardless of journal type. Tools such as CLAIM may help when presenting and reviewing studies.

Author contributions

AM and SA: conceived the need for a more general discussion of the systematic review findings. AM and MS: wrote the manuscript, which was critically reviewed by AJS and SA. All authors have contributed to and read the final manuscript. All authors were also authors on the systematic review. AM and MS: contributed equally to this manuscript and should be considered as joined first authors. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. (2018) 18:500–10. doi: 10.1038/s41568-018-0016-5

2. Neri E, de Souza N, Brady A, Bayarri AA, Becker CD, Coppola F, et al. European Society of radiology (ESR). What the radiologist should know about artificial intelligence—an ESR white paper. *Insights Imaging*. (2019) 10:44. doi: 10.1186/s13244-019-0738-2

3. Reardon S. Rise of robot radiologists. *Nature*. (2019) 576:S54–8. doi: 10.1038/d41586-019-03847-z

4. Coppola F, Faggioni L, Gabelloni M, De Vietro F, Mendola V, Cattabriga A, et al. Human, all too human? An all-around appraisal of the "artificial intelligence revolution" in medical imaging. *Front Psychol*. (2021) 12:710982. doi: 10.3389/fpsyg.2021.710982

5. Char DS, Shah NH, Magnus D. Implementing machine learning in health care—addressing ethical challenges. *N Engl J Med*. (2018) 378:981–3. doi: 10.1056/NEJMp1714229

6. Wiens J, Creary M, Sjoding MW. AI Models in health care are not colour blind and we should not be either. *Lancet Digit Health*. (2022) 4:e399–400. doi: 10.1016/S2589-7500(22)00092-9

7. Gichoya JW, Banerjee I, Bhimireddy AR, Burns JL, Celi LA, Chen L-C, et al. AI Recognition of patient race in medical imaging: a modelling study. *Lancet Digit Health*. (2022) 4:e406–14. doi: 10.1016/S2589-7500(22)00063-2

8. Mongan J, Moy L, Kahn CE. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. (2020) 2:e200029. doi: 10.1148/ryai.2020200029

9. O'Regan DP. Putting machine learning into motion: applications in cardiovascular imaging. *Clin Radiol*. (2020) 75:33–7. doi: 10.1016/j.crad.2019.04.008

10. Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Front Cardiovasc Med*. (2020) 7:25. doi: 10.3389/fcvm.2020.00025

11. Alabed S, Alandejani F, Dwivedi K, Karunasaagarar K, Sharkey M, Garg P, et al. Validation of artificial intelligence cardiac MRI measurements: relationship to heart catheterization and mortality prediction. *Radiology*. (2022) 305:68–79. doi: 10.1148/radiol.212929

12. Alandejani F, Alabed S, Garg P, Goh ZM, Karunasaagarar K, Sharkey M, et al. Training and clinical testing of artificial intelligence derived right atrial cardiovascular magnetic resonance measurements. *J Cardiovasc Magn Reson Off J Soc Cardiovasc Magn Reson*. (2022) 24:25. doi: 10.1186/s12968-022-00855-3

13. Alabed S, Maiter A, Salehi M, Mahmood A, Daniel S, Jenkins S, et al. Quality of reporting in AI cardiac MRI segmentation studies—a systematic review and recommendations for future studies. *Front Cardiovasc Med*. (2022) 9:956811. doi: 10.3389/fcvm.2022.956811

Machine learning based gray-level co-occurrence matrix early warning system enables accurate detection of colorectal cancer pelvic bone metastases on MRI

Jinlian Jin*, Haiyan Zhou, Shulin Sun, Zhe Tian, Haibing Ren, Jinwu Feng and Xinping Jiang

Gezhouba Central Hospital of Sinopharm, The Third Clinical Medical College of China Three Gorges University, Yichang, Hubei, China

Objective: The mortality of colorectal cancer patients with pelvic bone metastasis is imminent, and timely diagnosis and intervention to improve the prognosis is particularly important. Therefore, this study aimed to build a bone metastasis prediction model based on Gray level Co-occurrence Matrix (GLCM) - based Score to guide clinical diagnosis and treatment.

Methods: We retrospectively included 614 patients with colorectal cancer who underwent pelvic multiparameter magnetic resonance image(MRI) from January 2015 to January 2022 in the gastrointestinal surgery department of Gezhouba Central Hospital of Sinopharm. GLCM-based Score and Machine learning algorithm, that is,artificial neural net7work model(ANNM), random forest model (RFM), decision tree model(DTM) and support vector machine model(SVMM) were used to build prediction model of bone metastasis in colorectal cancer patients. The effectiveness evaluation of each model mainly included decision curve analysis(DCA), area under the receiver operating characteristic (AUROC) curve and clinical influence curve(CIC).

Results: We captured fourteen categories of radiomics data based on GLCM for variable screening of bone metastasis prediction models. Among them, Haralick_90, IV_0, IG_90, Haralick_30, CSV, Entropy and Haralick_45 were significantly related to the risk of bone metastasis, and were listed as candidate variables of machine learning prediction models. Among them, the prediction efficiency of RFM in combination with Haralick_90, Haralick_all, IV_0, IG_90, IG_0, Haralick_30, CSV, Entropy and Haralick_45 in training set and internal verification set was [AUC: 0.926,95% CI: 0.873-0.979] and [AUC: 0.919,95% CI: 0.868-0.970] respectively. The prediction efficiency of the other four types of prediction models was between [AUC: 0.716,95% CI: 0.663-0.769] and [AUC: 0.912,95% CI: 0.859-0.965].

Conclusion: The automatic segmentation model based on diffusion-weighted imaging(DWI) using depth learning method can accurately segment the pelvic bone structure, and the subsequently established radiomics model can effectively detect bone metastases within the pelvic scope, especially the RFM algorithm, which can provide a new method for automatically evaluating the pelvic bone turnover of colorectal cancer patients.

Introduction

Worldwide, colorectal cancer is still a malignant tumor of the digestive system with a high incidence rate and mortality (1). In recent years, it is encouraging that advanced diagnostic technologies, such as computed tomography (CT) colon imaging, magnetic resonance imaging (MRI) and positron emission tomography (PET)/CT colon imaging, are beneficial to enable some early cancer patients to receive timely treatment and effectively reduce the recurrence and metastasis rate (2, 3). Nevertheless, the poor prognosis caused by colorectal cancer metastasis is still one of the important factors that can not be ignored and avoided.

Previous studies have focused on the common metastatic sites of colorectal cancer, including lymph node metastasis, liver metastasis, and so on (4–6). Vigilantly, bone metastasis is also a poor prognostic factor for colorectal cancer, with incidence rate ranging from 2% to 11% (7, 8). As one of the advanced diseases of colorectal cancer, due to the heterogeneity and complexity of bone metastasis, there are great differences in the survival and recurrence of patients (9–11). Previous studies have shown that the most common sites of bone metastasis in colorectal cancer patients are the pelvis, thoracic vertebrae and lumbar vertebrae, while there may be metastatic or implanted small lesions in the pelvis and the pelvic cavity near the sacrum. Cancer emboli can be directly transferred to the pelvis, or transferred to the sacrum through the capillaries of the sacrum, leading to vertebral bone metastasis or other sites (9–11). Therefore, it is urgently needed that the available prediction model can divide patients into different categories according to the risk score of pelvic bone metastasis, so as to select appropriate treatment methods, and can also more accurately evaluate the effectiveness of treatment measures.

Abbreviations: IQR, inter-quartile range; OA, osseous alteration; CEA, carcinoembryonic antigen; ECOG, Eastern Cooperative Oncology Group; EV, Energy value; IG_all, Inverse gap full angle; IG_0, Inverse gap 0°; IG_45, Inverse gap 45°; IG_90, Inverse gap 90°; IV_all, Inertia value full angle; IV_all_SD, Inertia value full angle SD; IV_0, Inertia value 0°; IV_45, Inertia value 45°; IV_90, Inertia value 90°; Haralick_all, Haralick full angle; Haralick_0, Haralick 0°; Haralick_30, Haralick 30°; Haralick_45, Haralick 45°; Haralick_90, Haralick 90°; CSV, Cluster shadow value; CP, Cluster prominence.

Nowadays, radiomics and advanced algorithms have been gradually applied to the medical field, of which the most widely used is clinical prediction model (12, 13). Gray level co-occurrence matrix(GLCM) has been widely used in disease diagnosis, clinical staging, treatment evaluation and prognosis evaluation (14, 15). With the help of the spatial correlation characteristics of gray level, the modified technology can describe the image texture, which can efficiently extract and model the features of a variety of medical images (16, 17). Additionally, the higher-order algorithm of machine learning, with its iterative weight distribution, can make better use of predictors to improve the diagnostic efficiency of the model.

Inspired by this, this study based on diffusion weighted imaging, on the basis of applying machine learning algorithm to automatically segment the pelvic bone structure, established radiomics model to judge whether there is bone metastasis in the pelvic bone structure of patients with colorectal cancer, in order to better serve clinical decision-making.

Materials and methods

Study population

We retrospectively included 614 patients with colorectal cancer who underwent pelvic multiparameter MRI from January 2015 to January 2022 in the gastrointestinal surgery department of Gezhouba Central Hospital of Sinopharm. Comparison against abdominopelvic CT, SPECT/CT, or histopathologic tissue sampling was used to establish the baseline ground truth for presence or absence of bone metastases at the time of enrollment. The inclusion criteria of patients are as follows: (i)Patients suspected of colorectal cancer or undergoing pelvic multi-parametric diffusion weighted imaging(mp-DWI) scan due to reexamination after colorectal cancer treatment; (ii)Patients with complete pelvic DWI images; (iii)Patients without primary pelvic bone disease (primary osteosarcoma, bone cyst, blood system disease, fracture, etc.). Exclusion criteria: (i)Patients with a history of pelvic bone structure surgery; (ii)Patients with a history of other malignancies; (iii)Patients with unsatisfied image quality, such as motion artifacts and chemical shift artifacts; (iv)Patients with

incomplete scanning scope and not including most pelvic bone structures. This retrospective study was approved by the Ethics Committee of Gezhouba Central Hospital of Sinopharm, and the research scheme was implemented according to the artificial intelligence(AI) model training specifications of the unit. All patients' personal information is encrypted to prevent leakage, and complies with the Declaration of Helsinki. All patients in this study were informed of the study protocol and approved the study by written consent. The process of incorporating patients and building prediction models was shown in Figure 1.

Acquisition of diffusion weighted imaging parameters

We used GE Discovery MR750 W3.0 T machine to perform pelvic MRI plain scan and enhanced scan. The patient was instructed to lie on his back with his head advanced, and the scanning range was from the umbilical foramen level to the pubic symphysis. The sequences included conventional transverse T1WI, (with or without) conventional transverse T2WI, (transverse, sagittal, coronal) fat compression T2WI, (with or without sagittal) transverse DWI (b value=1000s/mm, b value=2000 s/mm), transverse and sagittal fat compression LAVA enhanced sequences. Sagittal fat compression T2WI scanning parameters: TR 4 800.0 ms, TE 110.5 ms, matrix 320 × 320, layer thickness and layer spacing are 5 mm and 1 mm respectively, FOV range: 28 cm × 28 cm.

Next, we standardized the format of DWI, that is, converted the high b value DWI image in DICOM format to Nifty format, and then the radiology resident (with film reading experience of 3 years or more) used ITK-SNAP3.6.0 software(http://www.itksnap.org/pmwiki/pmwiki.php?n=Main.Publications) to manually delineate and label the DWI image along the edges of various pelvic bone structures. In addition, a radiologist (with film reading experience ≥

FIGURE 1
The flow chart of patient selection and data process.

15 years) modified and confirmed the label, and the confirmed image label was used as the gold standard of pelvic bone structure segmentation model.

Training and verification of segmentation model

We preprocessed the image and extract the texture features, and used the software GE (Shanghai) AK (Artifical Intelligence Kit, V3.2.0R version) application platform to preprocess the image: linear method is used for resampling, with X, Y, Z spacing of 1.000; Gaussian 0.50 is used for noise removal; MR bias field correction is adopted to eliminate the stray intensity change caused by the non-uniformity of magnetic field and coil; Intensity standardization adopts gray discretization, and the expected minimum value and maximum value are 0.015 and 0.255, respectively. After image preprocessing (resampling, offset field correction, intensity standardization), a total of 48 GLCM texture features are finally collected by AK software, including 8 types of parameters, namely: cluster facilitation, cluster shade, correlation, GLCM energy, GLCM entropy, Haralick correlation, inertia, inverse difference motion; 6 angles are calculated, including full angle, full angle SD, 0°, 45°, 90° and 135°.

Analysis and evaluation of pelvic bone metastasis prediction model

We preprocessed the DWI images of 614 patients, namely: size=64 × 224 × 224 (z, y, x), automatic window width and level. Patients were randomly divided into training set (70%) and verification set (30%) according to 7:3. In order to eliminate the imbalance of the classified training set data, we balance the positive/negative samples by reducing the sampling, and use Min Max to normalize the feature matrix. At the same time, we use Pierce correlation coefficient to reduce the dimension of the data, and the eigenvectors of the transformed eigenmatrix have independent features.

There are four machine learning model building methods used in this study, including artificial neural network model (ANNM), random forest model(RFM), decision tree model(DTM) and support vector machine model (SVMM) (18–21). As linear regression models are often used to build clinical prediction models, this study builds generalized linear regression model (GLRM) based on Softmax regression (22, 23), namely: $\phi k = 1 - \sum_{i-1}^{k-1} \varnothing i$ RFM, DTM, ANNM and SVM are the most commonly used algorithms in machine learning (18). In this study, a pelvic bone metastasis prediction model was built based on four supervised learning algorithms.

Before building the prediction model, we use recursive feature elimination algorithm to select features and sort them, and select the first 8 features as the best feature subset; At the same time, for the GLRM, the minimum absolute shrinkage and selection operator classifier are selected to establish a classification model for predicting pelvic bone metastasis based on DWI images (24). The effectiveness evaluation of each model mainly includes decision curve
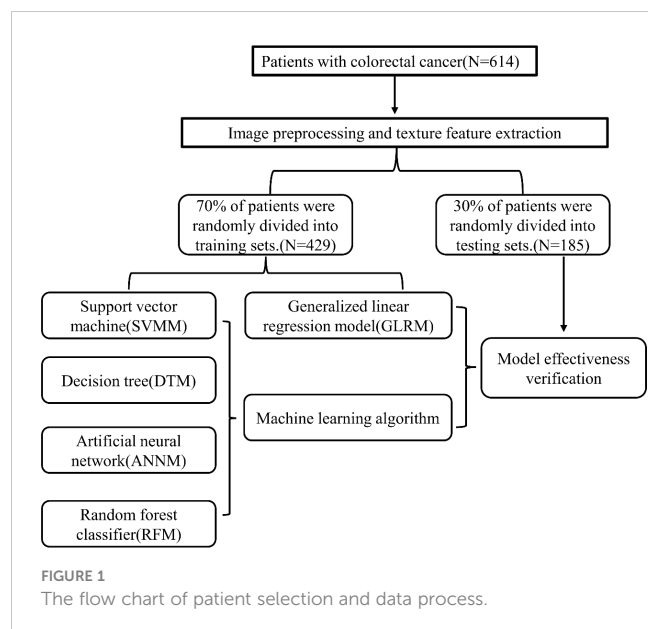
analysis(DCA) (25), area under the receiver operating characteristic (AUROC) curve and clinical influence curve(CIC) (26).

Statistical methods

The distribution of "measurement" and "counting" data in accordance with normal distribution in this study is expressed by means (interquartile interval) and percentage (%). For the independent two sample nonparametric test, the Mann Whitney rank sum test is used for the inter-group comparison that does not meet the normal distribution (27); The t-test or chi square test is used for the inter group comparison of samples from normal or nearly normal populations. In addition, the visual analysis of all charts in this study was completed with R studio software (download website: https://www.r-project.org/); Two tailed P values less than 0.05 were considered statistically significant.

Results

Patient baseline data and image segmentation characteristics

According to *Caret* software package algorithm, 614 patients included in this study were randomly divided into training set (N=429,70%) and internal verification set (N=185,30%) according to 7:3. The clinical characteristics and image data sources of patients in the data set were summarized in Table 1 and Supplementary Table 1. The average age of patients used for pelvic bone structure segmentation model training was 58 (50, 68) years. Among all the patients used to establish the pelvic bone metastasis classification histological model, 53 patients had bone metastasis [average age 55 (47, 65) years], and 561 patients had no bone metastasis [average age 58 (50, 68) years]. In addition, in the split model sample group, there was no statistically significant difference in clinical characteristics (age, pathology, osseous alteration, CEA and tumor location) between the training set and the internal validation set (P>0.05); The three types of GLCM parameters, namely correlation (full angle, 0°, 45°, 90°), inertia (full angle, full angle SD, 0°, 45°, 90°), inverse difference (full angle, 0°, 45°, 90°), cluster prominence and cluster shadow, had statistical differences (P<0.05), while energy, entropy and Haralick correlation had no statistical differences (P>0.05).

Feature variable screening based on GLCM prediction model

Then, we compared the classification features of bone metastasis lesions based on GLCM between groups, and analyzed the correlation between patients' baseline data and candidate variables of bone metastasis. The results showed that Haralick_90, Haralick_all, IV_0, IG_90, IG_0, Haralick_30, CSV, Entropy and Haralick_45 were highly positively correlated with pelvic bone metastasis in colorectal cancer patients (Figure 2A). In addition,

TABLE 1 Baseline data of patients with colorectal cancer.

Variables	Overall (N=614)
Age (median [IQR]),year	58.00 [49.25, 68.00]
sex (%)	
Male	381 (62.1)
Female	233 (37.9)
Pathology (%)	
Adenocarcinoma	293 (47.7)
Squamous cell carcinoma	231 (37.6)
Adenosquamous carcinoma	61 (9.9)
Small cell carcinoma	29 (4.7)
Tumor stage (%)	
I-II	533 (86.8)
III	39 (6.4)
IV	42 (6.8)
Differentiation (%)	
High	326 (53.1)
Moderate	189 (30.8)
Low	99 (16.1)
OA (%)	
Osteolytic	393 (64.0)
Osteogenic	188 (30.6)
Miscibility	33 (5.4)
CEA (%),ng/mL	
<100	156 (25.4)
≥100	458 (74.6)
Tumor location (%)	
Colonic segment	197 (32.1)
Rectal segment	417 (67.9)
ECOG (%)	
0-2	278 (45.3)
>2	336 (54.7)
EV (median [IQR])	0.96 [0.70, 1.22]
Entropy (median [IQR])	8.63 [8.37, 8.87]
IG_all (median [IQR])	3.06 [2.62, 3.58]
IG_0 (median [IQR])	2.20 [1.84, 2.58]
IG_45 (median [IQR])	2.96 [2.55, 3.39]
IG_90 (median [IQR])	2.30 [1.81, 2.77]
IV_all (median [IQR])	186.50 [159.00, 216.75]
IV_all_SD (median [IQR])	5229.00 [3691.25, 6978.50]

(Continued)

TABLE 1 Continued

Variables	Overall (N=614)
IV_0 (median [IQR])	159.05 [127.47, 193.60]
IV_45 (median [IQR])	159.85 [124.62, 192.50]
IV_90 (median [IQR])	131.00 [108.25, 155.00]
Haralick_all (median [IQR])	0.10 [0.09, 0.10]
Haralick_30 (median [IQR])	0.10 [0.09, 0.11]
Haralick_45 (median [IQR])	0.07 [0.07, 0.08]
Haralick_90 (median [IQR])	0.11 [0.10, 0.13]
CSV (median [IQR])	90.00 [84.00, 96.00]
CP (median [IQR])	79.00 [72.00, 85.00]

in order to build radiomics model for colorectal cancer patients to conduct classification and evaluation with and without pelvic bone metastasis, we conducted feature extraction from the labeled and segmented images and labels based on the manually labeled and automatically segmented pelvic bone structures, respectively. The extracted features were used to establish the radiomics model, and the processing steps included data equalization, data normalization, feature dimension reduction, and feature selection. As shown in Figure 2B, Haralick_90, IV_0, IG_90, Haralick_30, CSV, Entropy and Haralick_45 were the intersection candidate predictors of RFM, SVM, ANNM and DTM.

Construction of bone metastasis model based on generalized linear algorithm

According to the results of multiple logistic regression analysis, a nomograph (Figure 3A) was developed. The visual quantitative mapping tool of the prediction model was based on the scaling of each regression coefficient to 0 to 100 points in the multiple logistic regression. β The influence of the variable with the highest coefficient (absolute value) is assigned 100 points. Add the scores of all independent variables to get a total, and then convert it into the probability of predicting pelvic bone metastasis. Generally, the C index and AUC value exceeding 0.6 implied a reasonable estimate. This study showed that the C index of GLRM was 0.72, and resampling also showed that the model had an ideal robustness (Figures 3B, C).

Construction of bone metastasis model based on machine learning algorithm

As shown in Figure 4 and Supplementary Table 2, the RFM based on the "bagging" algorithm sorted the GLCM parameters, where Haralick_90, Haralick_all, IV_0, IG_90, IG_0, Haralick_30, CSV, Entropy and Haralick_45 were suitable for further model building of the RFM algorithm; The prediction efficiency of the model showed that the model still had a robust and efficient prediction efficiency (AUC: 0.926,95% CI: 0.873-0.979), even though it passed the ten fold cross validation. Consistent with the
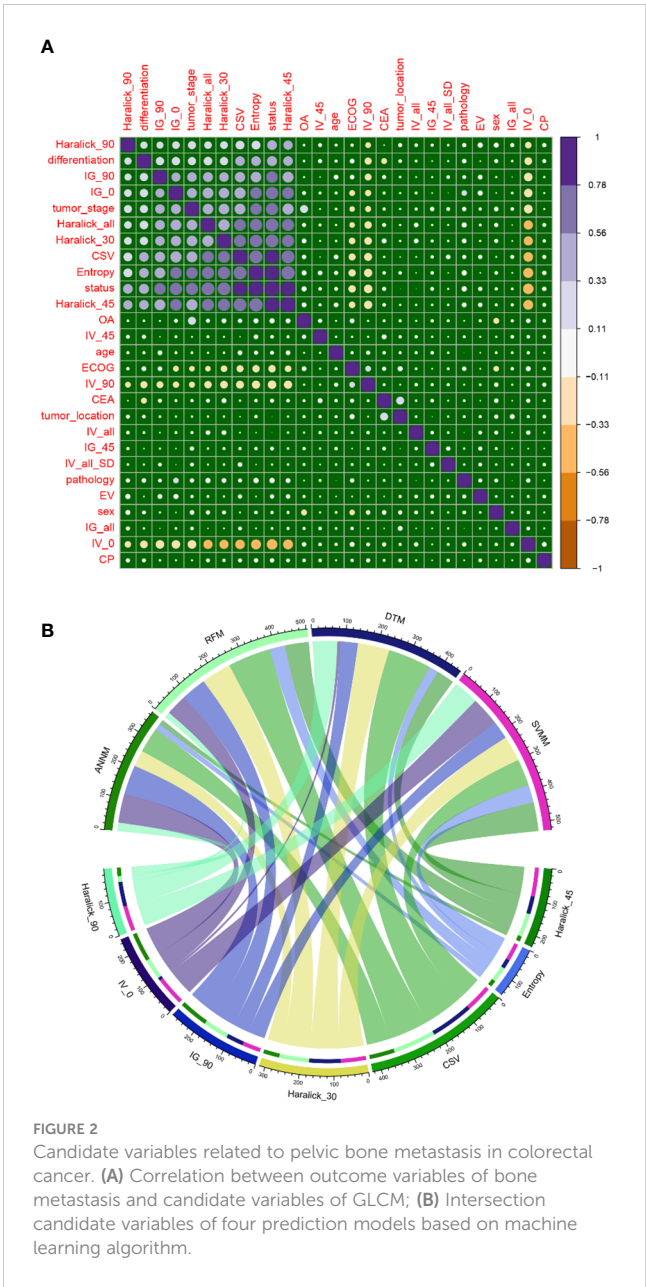
FIGURE 2
Candidate variables related to pelvic bone metastasis in colorectal cancer. **(A)** Correlation between outcome variables of bone metastasis and candidate variables of GLCM; **(B)** Intersection candidate variables of four prediction models based on machine learning algorithm.

parameter variables of the RFM model, DTM (Supplementary Figure 1) adopted CSV, Entropy and Haralick_30 as the decision factors in the "branches" of the model, and its prediction efficiency in the training set was worse than that of RFM (AUC: 0.889,95% CI: 0.836-0.942); However, ANNM included eight parameters, namely, Entropy, IG_90, IV_0, IV_45, IV_90, Haralick_al, Haralick_30, Haralick_45, Haralick_90 and CSV. The prediction efficiency obtained was AUC: 0.912, 95% CI: 0.859-0.965, which was worse than RFM, but better than DTM, SVM and GLRM.

Efficacy evaluation of five bone metastasis prediction models

DCA is a relatively new model evaluation method compared with ROC curve (28). In this study, we adopted two model
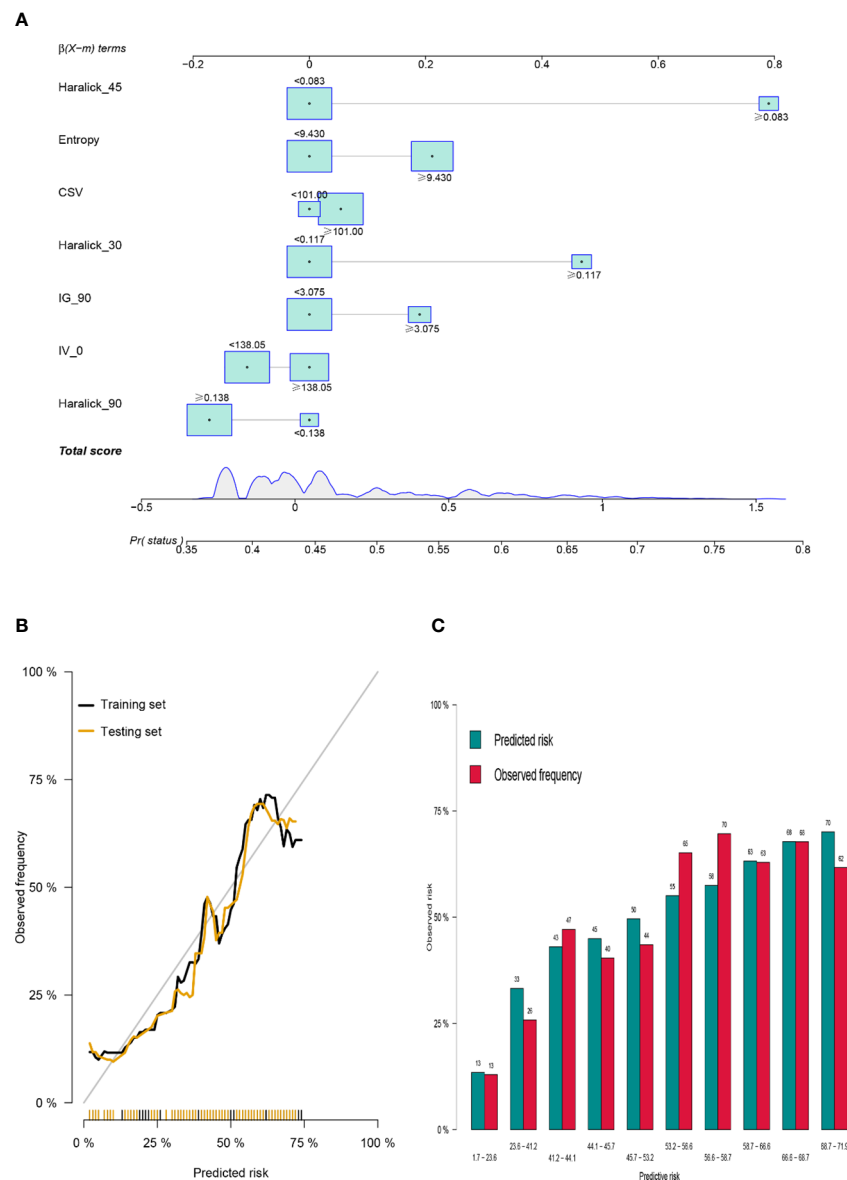
FIGURE 3
Construction of GLRM to predict bone metastasis in colorectal cancer patients. **(A)** Prediction score of bone metastasis based on Nomogram visualization; **(B)** Robustness evaluation of GLRM in training set and internal test set; **(C)** Multi-sample model prediction verification based on resampling.

effectiveness evaluation methods. We verified the linear regression model GLRM and four machine learning models (RFM, ANNM, SVM and DTM) with an internal test set. From the perspective of prediction accuracy, we can see that RFM has the largest "net benefit" in DCA(threshold probability=0.81), followed by ANNM, DTM and SVM (Figure 5 and Supplementary Table 3). GLRM was the least predictive machine learning model, with threshold probability=0.54.

At the same time, the ROC curve showed that the diagnostic efficacy of bone metastasis of RFM in training set and verification set was [AUC: 0.926,95% CI: 0.873-0.979] and [AUC: 0.919,95% CI: 0.868-0.970] respectively, while the diagnostic efficacy of bone metastasis of ANNM in training set and verification set was [AUC: 0.912,95% CI: 0.859-0.965] and [AUC: 0.894,95% CI:

0.843-0.945] respectively, which was slightly lower than that of RFM; As shown in Table 2 and Supplementary Figure 2, in general, the prediction efficiency of the prediction model for pelvic metastasis of colorectal cancer constructed by machine learning algorithm was better than that of the traditional generalized linear model. The results were consistent both in the training set and internal test set.

Prediction effectiveness evaluation of optimal prediction model

Based on the evaluation of five prediction models for pelvic metastasis of colorectal cancer, we found that RFM was the best in
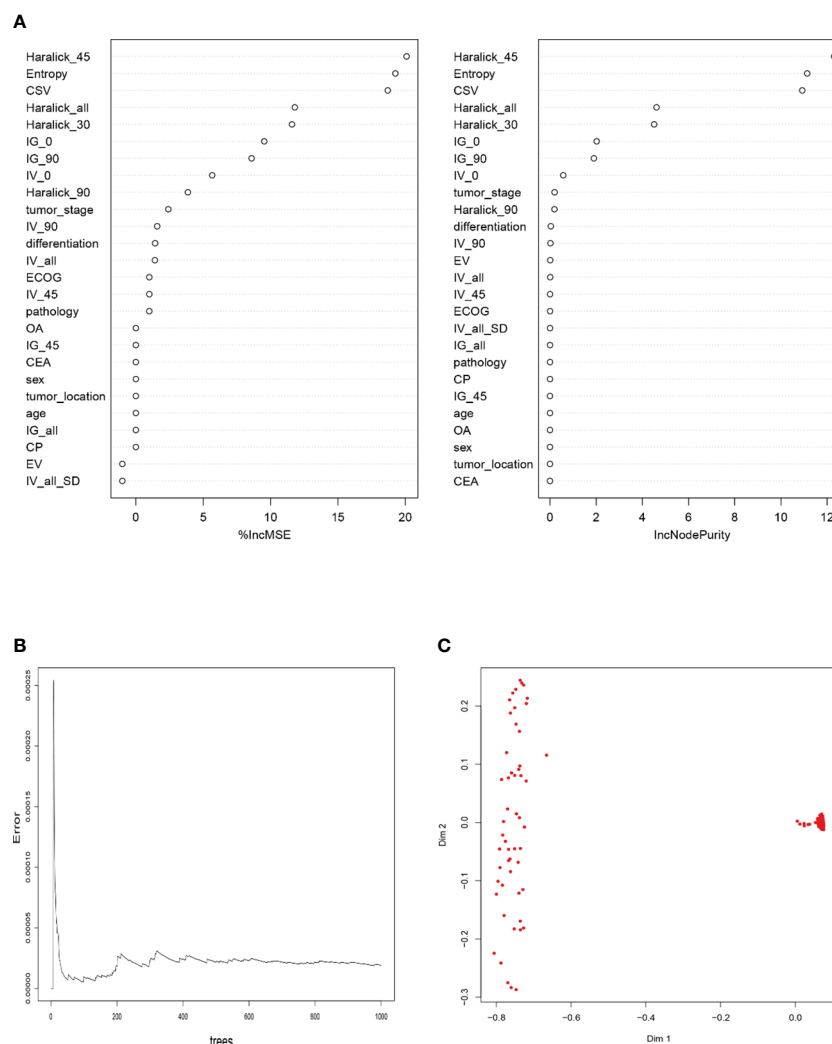
FIGURE 4

Construction of bone metastasis prediction model of colorectal cancer based on RFM. **(A)** Sorting of RFM prediction variables based on "Pruning" algorithm; **(B)** Screening optimal subset based on ten fold fold cross validation; **(C)** Recognition visualization of RFM in differentiating patients with or without pelvic bone metastasis.

terms of prediction efficiency. In order to further evaluate the differentiation efficiency of RFM, we used CIC to evaluate the "classification accuracy" of RFM in training set and internal verification set. As shown in Supplementary Figure 3, the blue curve (number high risk with outcome) indicated the number of true positives under each threshold probability, and the red curve (numberhigh risk) indicated the number of people classified as positive (high risk) by the prediction model under each threshold probability. It was credible that RFM can accurately distinguish patients with bone metastasis from those without bone metastasis, whether in training set or internal verification set, which further confirms that RFM can not only increase the interpretability of bone metastasis risk grading model for colorectal cancer patients, but also improve the grading accuracy. Therefore, RFM was suitable for stratified diagnosis and treatment.

Discussion

Invasion of colorectal cancer has always been a difficult problem in the treatment process. Because cancer cells can metastasize remotely through lymphatic vessels, blood vessels and nerves, especially vascular invasion, that is, cancer cells can metastasize remotely earlier through the portal vein and inferior vena cava (29–31). In the early stage of treatment, pay close attention to the degree of colorectal cancer invasion, and carefully check the metastatic lymph nodes (4, 32). For those with high degree of invasion or lymph node metastasis, prepare radiotherapy and chemotherapy plans in advance. Routine treatment after radical surgery can have a very important guiding value to improve the prognosis of patients. As far as we know, this is the first attempt to integrate machine learning algorithm and imaging information to build a prediction

FIGURE 5

Effectiveness evaluation of five predictive models for pelvic bone metastasis detection based on DCA. **(A)** Training set; **(B)** Internal validation set.

factors are based on liver metastasis and lung metastasis, but there is little analysis on the influencing factors of bone metastasis of colorectal cancer (5, 36). There is also a lack of reliable potential indicators that can predict bone metastasis of colorectal cancer. In view of this, this study strives to explore the main influencing factors of bone metastasis after radical resection of colorectal cancer, explore its potential bone metastasis prediction indicators, and build a bone metastasis prediction model based on advanced algorithms. As far as we know, this is the first prediction model for pelvic bone metastasis of colorectal cancer based on radiomics and machine learning. With the help of this model, we hope to better guide clinical diagnosis and treatment.

The bone metastasis of colorectal cancer is mainly osteogenic lesions, with multiple and jumping distribution, and osteogenic changes and osteolytic changes exist at the same time (37). Fortunately, mpMRI has a high sensitivity and specificity in the diagnosis of colorectal cancer bone metastasis (38). When both systemic bone phenomena and CT cannot determine the existence of bone metastasis, mpMRI is usually feasible. Generally speaking, mpMRI includes conventional sequences (TIW1 and T2W1) and functional sequences (DWI, DCE-MRI and MRS) (38, 39). Among them, DWI is more sensitive to monitoring bone metastasis of colorectal cancer than conventional sequences. DWI is an assessment of microscopic movement of water molecules in the body, and can provide quantitative (such as ADC value) and qualitative (such as signal strength) information for disease diagnosis and treatment.

Radiomics is a new image post-processing technology emerging in recent years. Through quantitative and high-throughput analysis and processing of medical images, it extracts a series of information that cannot be directly observed by the naked eye, reveals the relationship between tumor biological characteristics and images, and is used to establish descriptive and predictive models to help doctors make diagnosis (40). In this study, on the basis of segmentation of pelvic bone structure, we established radiomics model based on DWI images to detect whether colorectal cancer patients have metastatic lesions within the scope of pelvic bone structure. Encouragingly, the model is robust and accurate in the prediction of test set, which can be used to undertake early warning of colorectal cancer bone metastasis and auxiliary diagnosis before treatment.

With the in-depth development of cross field artificial intelligence machine learning, it is now possible to predict disease risks through machines, and even diagnose some diseases (41). In recent years, due to the extensive application of deep learning and

model for colorectal cancer bone metastasis. Compared with previous studies, this study extracts a series of information that cannot be directly observed by the naked eye through quantitative and high-throughput analysis and processing of medical images, which can better reveal the relationship between tumor biological characteristics and images, and can be used to establish descriptive and predictive models to help doctors make clinical decisions.

At present, studies have found that preoperative T staging, lymphatic metastasis and Duckes staging of colorectal cancer are independent factors that affect the prognosis of colorectal cancer (33–35). In addition, anesthesia, perioperative blood transfusion and treatment are also relevant factors. However, most of the above

TABLE 2 Comparison of predictive efficacy of five types of pelvic bone metastasis prediction models.

Model	Training set			Internal validation set		
	AUC Mean	AUC 95%CI	Variables[&]	AUC Mean	AUC 95%CI	Variables[&]
RFM	0.926	0.873-0.979	7	0.919	0.868-0.970	7
SVMM	0.862	0.809-0.915	11	0.841	0.790-0.892	11
DTM	0.889	0.836-0.942	5	0.839	0.788-0.890	5
ANNM	0.912	0.859-0.965	10	0.894	0.843-0.945	10
GLRM	0.716	0.663-0.769	7	0.722	0.671-0.773	7

[&]Variables included in the model.

the inclusion of rich and diverse medical images, it has become an important part of artificial intelligence machine learning diagnosis, so deep learning has also had a huge impact in medical diagnosis. Previous studies have shown that the random forest algorithm can effectively process the mixed data, missing values or outliers, and higher dimensional data in medical data, and then comprehensively classify the data through multiple decision trees, and perform correlation testing, prediction, and interpretation (42). These processing processes are not easy to appear over fitting, making the prediction accuracy more accurate. In this study, we built and validated the prediction model of colorectal cancer bone metastasis through a large sample size, among which the prediction model established through the random forest (iterative) algorithm was the best (AUC: 0.926,95% CI: 0.873-0.979). In addition, the prediction efficiency of other machine learning models (ANNM, DTM, SVMM) is also better than GLRM. The possible reason is that Logistic regression has advantages in data processing of online relationships, and machine learning is often more applicable in the face of nonlinear problems. Therefore, how to improve the extrapolation of the model needs further research.

In addition, this study inevitably has the following limitations. First of all, this study only judged whether there was bone metastasis in the pelvic cavity at the patient level, and did not discuss the bone structure of a single pelvic cavity or from the focus level. In the future, we should also detect the metastatic lesions at the bone structure level and the lesion level, so as to detect and locate the bone metastasis of colorectal cancer in the pelvic region; Second, this study did not compare the classification performance of the radiomics model with the diagnostic efficacy of radiologists. In the follow-up research, we will compare the effectiveness of artificial intelligence and human experience; Third, this study only used a single DWI sequence to classify whether there is bone metastasis. Although this sequence is essential in the process of bone metastasis detection, it still has some limitations in the detection of osteogenic changes. Therefore, we consider adding other sequences (such as ADC map, T1WI, etc.) to the model in subsequent studies to improve the prediction performance of the model for all types of metastatic lesions.

Conclusion

To sum up, this study based on depth learning segmentation DWI image pelvic bone structure of the radiomics model can better identify the pelvic range of colorectal cancer bone metastases; Among them, the predictive factors provided by RFM combined with GLCM can obtain the best predictive efficacy of bone metastasis, so it can undertake part of the work of mpMRI assisted diagnosis of colorectal cancer, so as to better guide clinical diagnosis and treatment.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Ethics statement

This retrospective study was approved by the Ethics Committee of Gezhouba Central Hospital of Sinopharm, and the research scheme was implemented according to the artificial intelligence (AI) model training specifications of the unit(No. 2020-006). In the informed consent statement of the patient involved in this article. As for oral informed consent, it generally means that if the patient's education level is low or he cannot sign in person(due to either low education level or an inability to sign), the legal guardian designated by the patient will sign on his behalf, while the patient himself has orally declared his informed consent. When we signed this Informed Consent Form, for patients with oral informed consent, it was generally signed by the patient himself and the patient's guardian at the same time, so there was no objection.

Author contributions

JJ conceived and designed the study and wrote the manuscript. HZ, SS, ZT, HR, JF and XJ collected the data, performed the data analysis, and interpreted the outcome. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1121594/full#supplementary-material

References

1. Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. *Lancet (London England).* (2019) 394(10207):1467–80. doi: 10.1016/S0140-6736(19)32319-0

2. Nasseri Y, Langenfeld SJ. Imaging for colorectal cancer. *Surg Clinics North America.* (2017) 97(3):503–13. doi: 10.1016/j.suc.2017.01.002

3. Kijima S, Sasaki T, Nagata K, Utano K, Lefor AT, Sugimoto H. Preoperative evaluation of colorectal cancer using CT colonography, MRI, and PET/CT. *World J gastroenterology.* (2014) 20(45):16964–75. doi: 10.3748/wjg.v20.i45.16964

4. Jin M, Frankel WL. Lymph node metastasis in colorectal cancer. *Surg Oncol Clinics North America.* (2018) 27(2):401–12. doi: 10.1016/j.soc.2017.11.011

5. Robinson JR, Newcomb PA, Hardikar S, Cohen SA, Phipps AI. Stage IV colorectal cancer primary site and patterns of distant metastasis. *Cancer Epidemiol* (2017) 48:92–5. doi: 10.1016/j.canep.2017.04.003

6. Bhullar DS, Barriuso J, Mullamitha S, Saunders MP, O'Dwyer ST, Aziz O. Biomarker concordance between primary colorectal cancer and its metastases. *EBioMedicine.* (2019) 40:363–74. doi: 10.1016/j.ebiom.2019.01.050

7. Stewart CL, Warner S, Ito K, Raoof M, Wu GX, Kessler J, et al. Cytoreduction for colorectal metastases: Liver, lung, peritoneum, lymph nodes, bone, brain. when does it palliate, prolong survival, and potentially cure? *Curr problems surgery.* (2018) 55 (9):330–79. doi: 10.1067/j.cpsurg.2018.08.004

8. Park HS, Chun YJ, Kim HS, Kim JH, Lee C-K, Beom S-H, et al. Clinical features and KRAS mutation in colorectal cancer with bone metastasis. *Sci Rep* (2020) 10 (1):21180. doi: 10.1038/s41598-020-78253-x

9. Qian J, Gong ZC, Zhang YN, Wu H-H, Zhao J, Wang L-T, et al. Lactic acid promotes metastatic niche formation in bone metastasis of colorectal cancer. *Cell communication signaling: CCS.* (2021) 19(1):9. doi: 10.1186/s12964-020-00667-x

10. Suresh Babu MC, Garg S, Lakshmaiah KC, Babu KG, Kumar RV, Loknatha D, et al. Colorectal cancer presenting as bone metastasis. *J Cancer Res Ther* (2017) 13 (1):80–3. doi: 10.4103/0973-1482.181177

11. Li X, Hu W, Sun H, Gou H. Survival outcome and prognostic factors for colorectal cancer with synchronous bone metastasis: A population-based study. *Clin Exp metastasis.* (2021) 38(1):89–95. doi: 10.1007/s10585-020-10069-5

12. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med image analysis.* (2016) 33:170–5. doi: 10.1016/j.media.2016.06.037

13. Van Calster B, Wynants L. Machine learning in medicine. *New Engl J Med* (2019) 380(26):2588. doi: 10.1056/NEJMc1906060

14. Iqbal N, Mumtaz R, Shafi U, Zaidi SMH. Gray Level co-occurrence matrix (GLCM) texture based crop classification using low altitude remote sensing platforms. *PeerJ Comput science.* (2021) 7:e536. doi: 10.7717/peerj-cs.536

15. Tan J, Gao Y, Liang Z, Cao W, Pomeroy MJ, Huo Y, et al. 3D-GLCM CNN: A 3-dimensional Gray-level Co-occurrence matrix-based CNN model for polyp classification *via* CT colonography. *IEEE Trans Med imaging.* (2020) 39(6):2013–24. doi: 10.1109/TMI.2019.2963177

16. Naik A, Edla DR, Dharavath R. Prediction of malignancy in lung nodules using combination of deep, fractal, and Gray-level Co-occurrence matrix features. *Big data.* (2021) 9(6):480–98. doi: 10.1089/big.2020.0190

17. Yin JD, Song LR, Lu HC, Zheng X. Prediction of different stages of rectal cancer: Texture analysis based on diffusion-weighted images and apparent diffusion coefficient maps. *World J gastroenterology.* (2020) 26(17):2082–96. doi: 10.3748/wjg.v26.i17.2082

18. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. *BMC Med Inf decision making.* (2019) 19(1):281. doi: 10.1186/s12911-019-1004-8

19. Choi RY, Coyner AS, Kalpathy-Cramer J, Chiang MF, Campbell JP. Introduction to machine learning, neural networks, and deep learning. *Trans Vision Sci technology.* (2020) 9(2):14. doi: 10.1167/tvst.9.2.14

20. Ghiasi MM, Zendehboudi S, Mohsenipour AA. Decision tree-based diagnosis of coronary artery disease: CART model. *Comput Methods programs biomedicine.* (2020) 192:105400. doi: 10.1016/j.cmpb.2020.105400

21. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genomics proteomics.* (2018) 15(1):41–51. doi: 10.21873/cgp.20063

22. Kadam VJ, Jadhav SM, Vijayakumar K. Breast cancer diagnosis using feature ensemble learning based on stacked sparse autoencoders and softmax regression. *J Med systems.* (2019) 43(8):263. doi: 10.1007/s10916-019-1397-z

23. Yang J, Bai Y, Li G, Liu M, Liu X. A novel method of diagnosing premature ventricular contraction based on sparse auto-encoder and softmax regression. *Bio-medical materials engineering.* (2015) 26 Suppl 1:S1549–1558. doi: 10.3233/BME-151454

24. Alhamzawi R, Ali HTM. The Bayesian adaptive lasso regression. *Math biosciences.* (2018) 303:75–82. doi: 10.1016/j.mbs.2018.06.004

25. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and interpreting decision curve analysis: A guide for investigators. *Eur urology.* (2018) 74(6):796–804. doi: 10.1016/j.eururo.2018.08.038

26. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost. *J Trans Med* (2020) 18(1):462. doi: 10.1186/s12967-020-02620-5

27. Fay MP, Malinovsky Y. Confidence intervals of the Mann-Whitney parameter that are compatible with the wilcoxon-Mann-Whitney test. *Stat Med* (2018) 37 (27):3991–4006. doi: 10.1002/sim.7890

28. Vickers AJ, Holland F. Decision curve analysis to evaluate the clinical benefit of prediction models. *Spine journal: Off J North Am Spine Society.* (2021) 21(10):1643–8. doi: 10.1016/j.spinee.2021.02.024

29. Piawah S, Venook AP. Targeted therapy for colorectal cancer metastases: A review of current methods of molecularly targeted therapy and the use of tumor biomarkers in the treatment of metastatic colorectal cancer. *Cancer.* (2019) 125 (23):4139–47. doi: 10.1002/cncr.32163

30. Bertocchi A, Carloni S, Ravenda PS, Bertalot G, Spadoni I, Cascio AL, et al. Gut vascular barrier impairment leads to intestinal bacteria dissemination and colorectal cancer metastasis to liver. *Cancer Cell* (2021) 39(5):708–724.e711. doi: 10.1016/j.ccell.2021.03.004

31. Tauriello DV, Calon A, Lonardo E, Batlle E. Determinants of metastatic competency in colorectal cancer. *Mol Oncol* (2017) 11(1):97–119. doi: 10.1002/1878-0261.12018

32. Ebbehøj AL, Jørgensen LN, Krarup PM, Smith HG. Histopathological risk factors for lymph node metastases in T1 colorectal cancer: meta-analysis. *Br J surgery.* (2021) 108(7):769–76. doi: 10.1093/bjs/znab168

33. Yamamoto T, Kawada K, Obama K. Inflammation-related biomarkers for the prediction of prognosis in colorectal cancer patients. *Int J Mol Sci* (2021) 22(15):112–5. doi: 10.3390/ijms22158002

34. Messersmith WA. NCCN guidelines updates: Management of metastatic colorectal cancer. *J Natl Compr Cancer Network: JNCCN.* (2019) 17(5.5):599–601. doi: 10.6004/jnccn.2019.5014

35. Wang GR, Wang ZW, Jin ZY. Application and progress of texture analysis in the therapeutic effect prediction and prognosis of neoadjuvant chemoradiotherapy for colorectal cancer. *Chin Med Sci J* (2019) 34(1):45–50. doi: 10.24920/003572

36. Allgayer H, Leupold JH, Patil N. Defining the "Metastasome": Perspectives from the genome and molecular landscape in colorectal cancer for metastasis evolution and clinical consequences. *Semin Cancer Biol* (2020) 60:1–13. doi: 10.1016/j.semcancer.2019.07.018

37. Paauwe M, Schoonderwoerd MJA, Helderman R, Harryvan TJ, Groenewoud A, Pelt GWv, et al. Endoglin expression on cancer-associated fibroblasts regulates invasion and stimulates colorectal cancer metastasis. *Clin Cancer research: an Off J Am Assoc Cancer Res* (2018) 24(24):6331–44. doi: 10.1158/1078-0432.CCR-18-0329

38. Gürses B, Böge M, Altınmakas E, Balık E. Multiparametric MRI in rectal cancer. *Diagn interventional Radiol (Ankara Turkey).* (2019) 25(3):175–82. doi: 10.5152/dir.2019.18189

39. Zhang K, Ren Y, Xu S, Lu W, Xie S, Qu J, et al. A clinical-radiomics model incorporating T2-weighted and diffusion-weighted magnetic resonance images predicts the existence of lymphovascular invasion / perineural invasion in patients with colorectal cancer. *Med physics.* (2021) 48(9):4872–82. doi: 10.1002/mp.15001

40. Wang Z, Zhao Z, Song Z, Wang Y, Zhao Z. The application of magnetic resonance imaging (MRI) for the prediction of surgical outcomes in trigeminal neuralgia. *Postgraduate Med* (2022) 134(5):480–6. doi: 10.1080/00325481.2022.2067612

41. Deo RC. Machine learning in medicine. *Circulation.* (2015) 132(20):1920–30. doi: 10.1161/CIRCULATIONAHA.115.001593

42. Savargiv M, Masoumi B, Keyvanpour MR. A new random forest algorithm based on learning automata. *Comput Intell Neurosci* (2021) 2021:5572781. doi: 10.1155/2021/5572781

uRP: An integrated research platform for one-stop analysis of medical images

Jiaojiao Wu[1†], Yuwei Xia[1†], Xuechun Wang[1†], Ying Wei[1], Aie Liu[1], Arun Innanje[2], Meng Zheng[2], Lei Chen[1], Jing Shi[1], Liye Wang[1], Yiqiang Zhan[1], Xiang Sean Zhou[1], Zhong Xue[1], Feng Shi[1]* and Dinggang Shen[1,3,4]*

[1]Department of Research and Development, Shanghai United Imaging Intelligence Co., Ltd., Shanghai, China, [2]Department of Research and Development, United Imaging Intelligence Co., Ltd., Cambridge, MA, United States, [3]School of Biomedical Engineering, ShanghaiTech University, Shanghai, China, [4]Shanghai Clinical Research and Trial Center, Shanghai, China

Introduction: Medical image analysis is of tremendous importance in serving clinical diagnosis, treatment planning, as well as prognosis assessment. However, the image analysis process usually involves multiple modality-specific software and relies on rigorous manual operations, which is time-consuming and potentially low reproducible.

Methods: We present an integrated platform - uAI Research Portal (uRP), to achieve one-stop analyses of multimodal images such as CT, MRI, and PET for clinical research applications. The proposed uRP adopts a modularized architecture to be multifunctional, extensible, and customizable.

Results and Discussion: The uRP shows 3 advantages, as it 1) spans a wealth of algorithms for image processing including semi-automatic delineation, automatic segmentation, registration, classification, quantitative analysis, and image visualization, to realize a one-stop analytic pipeline, 2) integrates a variety of functional modules, which can be directly applied, combined, or customized for specific application domains, such as brain, pneumonia, and knee joint analyses, 3) enables full-stack analysis of one disease, including diagnosis, treatment planning, and prognosis assessment, as well as full-spectrum coverage for multiple disease applications. With the continuous development and inclusion of advanced algorithms, we expect this platform to largely simplify the clinical scientific research process and promote more and better discoveries.

KEYWORDS

research platform, one-stop, medical image analysis, deep learning, semi-automatic delineation, radiomics

1. Introduction

Medical imaging is widely employed in clinical research to investigate effects on diagnosis, staging, treatment planning, and follow-up evaluations (1–4). Medical imaging contains multiple imaging sequences or modalities, such as magnetic resonance imaging (MRI), computed tomography (CT), and positron emission tomography (PET), providing complementary information (5–8). The processing and quantitative analysis of medical images ensure their clinical utility in a variety of medical applications, from general research to clinical workflows.

Most recently, machine learning- and deep learning-based intelligent imaging analyses have shown enormous advantages in providing consistent and accurate image quantifications in multiple applications, including image segmentation, registration,

classification, etc. (9–12). A series of algorithm architectures and strategies have been developed to meet different requirements. For example, U-Net (13, 14), V-Net (15), and nnU-Net (16) exhibit accurate segmentation performance; affine models (i.e., FLIRT, A-SIFT) (17) and deformable models [i.e., FNIRT, ANTS, VoxelMorph (18), Dual-PRNet (19), LDDMM (20)] assist to image registration; ResNet (21), DenseNet (22) and their variants have attracted much attention in classification tasks. Also, varied attention mechanisms and loss functions have been utilized to optimize the deep learning network and improve its robustness (23–26). The accurate analysis of medical images accelerates the development and upgrading of intelligent algorithms that can be integrated into the software to enable easy-to-use clinical research.

Numerous choices of medical image analysis tools integrating advanced algorithms are available. For example, MATLAB (27), Python (https://www.python.org/), 3D Slicer (28) (https://www. slicer.org/), and Mimics (Materialize, Leuven, Belgium) allow general image processing, while FreeSurfer (https://surfer.nmr.mgh. harvard.edu/), chest imaging platform (CIP, https:// chestimagingplatform.org/), and OpenSim (29) are proprietarily applied to the brain, lung, knee joint analyses, respectively. Meanwhile, a number of software is dedicated to a specific modality, such as resting-state fMRI data analysis toolkit (REST) (30) and statistical parametric mapping (SPM, UCL Queen Square Institute of Neurology, London, UK) designed for functional MR images; DtiStudio (31) and medical imaging interaction toolkit (MITK) (32) applied to diffusion images; SenseCare (33) provides a range of artificial intelligence (AI) toolkits for specific clinical scenarios such as lung cancer diagnosis and radiotherapy planning. Overall, the software greatly simplifies image processing and makes it easy for clinicians to understand and use.

However, users still face a series of challenges in using the software to achieve one-stop image analysis. First, complex image analysis requires introduction of multiple software to adapt to the respective modalities and organs, which makes it difficult to integrate information from different modalities organically. Second, different software relies on specific environments (e.g., Windows and Linux) and programming languages (e.g., Python, C++, and R), requiring extensive computer knowledge to be used in practice. Third, the feasibility of integrating the latest AI models into the software to iteratively optimize performance is yet to be assessed. Finally, manual contouring regions of interest (ROIs) (34) is always required in scientific research to serve as the gold standard or to extract quantitative metrics, which is time-consuming and may suffer from low reproducibility and consistency due to intra- and inter-observer variability. Therefore, it is desired to design an integrated platform for one-stop analysis of medical images, which needs to be compatible, advanced, easy to use, extensible, and reproducible. In addition, the ideal platform should offer cloud-based services (public or private) to reduce configuration requirements on the user end and allow multiple clients to work simultaneously. A dedicated data management module is also essential for organizing multiple clinical projects and massive medical data, which allows the integration of large-scale data from multiple centers to develop robust algorithms and facilitate collaborative research.

In line with the trend, we propose a multifunctional platform, called uRP (uAI research portal, https://www.uii-ai.com/en/uai/ scientific-research), to perform accurate image processing and analysis on demand. The uRP can satisfy the following requirements: (1) Integrating a variety of algorithms with respect to the image's modality (i.e., MRI, CT, PET), body part (i.e., head, chest, abdomen, pelvis), and processing task (i.e., segmentation, registration, classification), to be suitable for diverse applications; (2) Offering friendly interactive user interface (UI) to make clinicians easy to understand and independently implement complete AI-related research; (3) Possessing the extendable capability to enrich existing modules and ensure reusable and reproducible analysis across clinicians, even hospitals; (4) Achieving automatic or semi-automatic image processing (e.g., delineation) to ensure efficient and accurate analysis; (5) Supporting cloud-based computing services with high concurrency and owing a dedicated data management module.

In the below sections, we present an overview of uRP's architecture and major functional modules, including semi-automatic delineation, deep learning-based image segmentation, registration, and classification, as well as radiomics and statistics. The clinical utility of the uRP is demonstrated by three domain analyses of the brain, pneumonia, and knee joint. Representative use cases are exampled to illustrate some of the outcomes that clinicians have achieved by using the uRP. In addition, the modules of the uRP will continue to be developed and extended, and we prospect future design concepts and directions to achieve a more intelligent platform for scientific research and even clinical uses.

2. Materials and methods

In this article, we propose the uRP to realize the one-stop medical image analysis. The architecture and main modules are shown in the following parts.

2.1. Overview of uRP

From 2018, we began to build the uRP to promote one-stop advanced medical image analysis in the context of integrating AI modules. It is intended to facilitate scientific research for clinicians and is therefore designed as flexible modules that can be used directly, combined, or customized for specific application domains. Here, we start by describing its architecture and key components (**Figure 1**).

2.1.1. Architecture
The design of uRP architecture takes a modular and layered approach. Internally, the software consists of three layers: (1) The lower level is composed of hardware drivers, such as graphics processing unit (GPU) accelerated using NVIDIA CUDA, and cloud servers, such as Amazon web services (AWS), that efficiently use graphics resources of the host system; (2) At the middle level, there is application programming interface (API), primarily Python and C++, contributing a range of algorithms
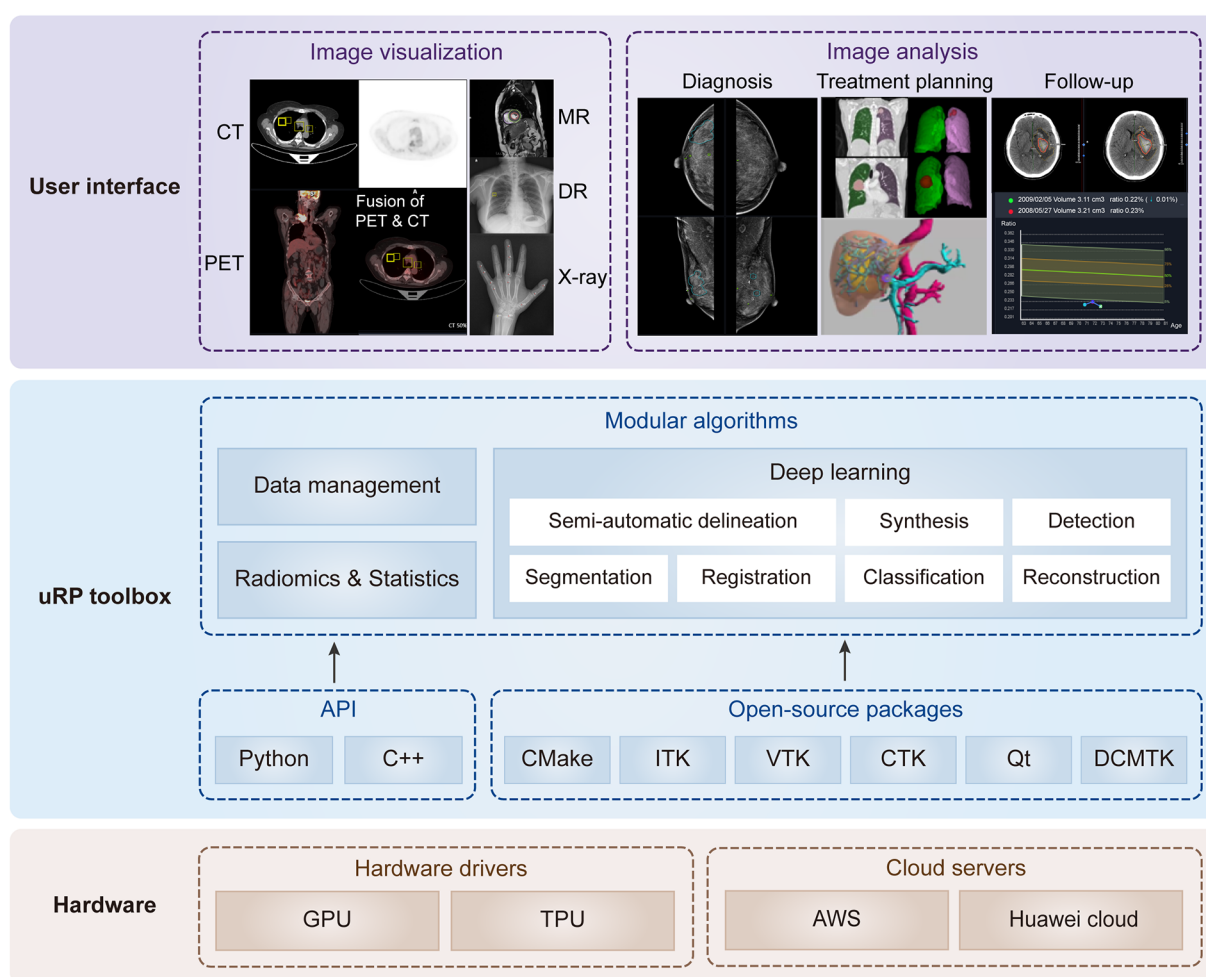
FIGURE 1
Overview of uRP with layered and modular architecture. At the lower level, it involves hardware drivers and cloud servers. At the middle level, it includes the application programming interface (API) and open-source packages to develop algorithms, ensuring versatility for data management, deep learning, as well as radiomics and statistics. At the higher level, it provides a series of user interfaces to perform different application domains covering multiple imaging modalities, full-stack analysis of one disease (i.e., diagnosis, treatment planning, and follow-up), and full-spectrum coverage for multiple diseases.

(e.g., segmentation, registration, classification) and providing higher-level functionality and abstractions. Additionally, a variety of open-source libraries are embedded in the uRP, where the DICOM toolkit (DCMTK) is used to support DICOM format data, and Qt to provide a cross-platform graphical user interface (GUI) framework; (3) The higher level presents UIs of multiple algorithms and builds blocks to the end users for domain-specific analysis (Figure 1). All computational demanding modules can run in parallel on many processors at once, using message passing between processes and/or shared-memory threads.

Since its inception, uRP has been evolving with major architecture, UIs, and functional redesigns. The uRP's version is updated every 2 months, and each release is formally tested on a variety of platform configurations to ensure its stability.

2.1.2. Interactive UI

uRP is a web-based platform supporting GPU cloud computing. It supports plug-ins to deliver task-specific functionality to the user. uRP mainly consists of three functional modules, (1) data management, supporting the upload of imaging data and non-imaging features, extraction of subject information from DICOM tags, and image search based on specific criteria; (2) image processing, including semi-automatic delineation, deep learning-based segmentation, registration, and classification; (3) radiomics and statistics, for classification and regression tasks. Images in DICOM or NIFIT formats from clinical picture archiving and communication system (PACS) or external drives are supported with various imaging modalities, including MR, CT, PET, x-ray, and digital radiography (DR) images. These technical modules can support a variety of applications on demand (Figure 1).

2.2. Semi-automatic delineation

Delineation of ROIs is essential in clinical research as the primary step for quantification and feature extraction. Manual delineation is generally considered as the gold standard, but it is

limited by its tedious, time-consuming, and error-prone characteristics, and thus difficult to achieve high-quality and efficient annotation especially with the massive amount of medical images. In view of this, various automatic delineation algorithms have been developed, including fully automatic and semi-automatic methods. It is worth noting that fully-automatic delineation is convenient but sometimes hard to reach the desired accuracy, while semi-automatic delineation allows human-computer interaction and thus can optimize the results, which is a time-saving alternative to manual delineation. Our proposed uRP platform contains a collection of tools for fully-automatic, semi-automatic, and manual delineation. As shown in Figure 2, a smart annotation tool and a ROI modification tool are integrated into the uRP to assist the delineation process.

2.2.1. Smart annotation tool

The uRP offers several smart annotation tools (SATs) for medical images, including (1) intelligent interactive segmentation and (2) annotation propagation, which enables fast extraction of the target from the complex background and 3D propagation of the annotation.

Intelligent interactive segmentation is a technique that allows the user to adjust the region of interest (ROI) by manipulating seeds. In use, the user firstly draws a rectangle to cover the ROI, where the Canny edge detection algorithm is then performed within the rectangle to generate the target boundary. After that a positive seed is generated at the centroid, and 4 negative seeds are placed at the vertices of the user-defined rectangle, serving as the control points for the boundary (Figure 2A, Supplementary Video 1). We can then update the shape and size of the generated boundary by adjusting the positive / negative seeds. Also, more positive seeds (green) can be added to enlarge the ROI by clicking on the area outside the edge, and negative seeds (red) can be added in the ROI to remove specific regions. Canny's approach is a widely used edge detection method with tweakable parameters and thus suitable to be integrated with a GUI. The algorithm can be divided into the following 5 steps: (1) Use a Gaussian filter to smooth the image and reduce the noise; (2) Calculate the gradient intensity and direction for each pixel point in the image; (3) Edge candidates are identified by applying the non-maximal or critical suppression to the gradient magnitude; (4) Apply double-threshold detection to determine real and potential edges; and (5) Finalize the edge detection by suppressing isolated weak edges (35). The implementation for the Canny edge detection algorithm could be found in the OpenCV library (36).

Another tool is annotation propagation, which applies the current annotation in one slice to its adjacent slices. Specifically, the current annotation serves as the initial mask, and the tool can automatically propagate annotations across entire image frames (Supplementary Video 2). Especially, the user can optionally keep the correct region when the ROI is propagated to other slices and divided into multiple regions. Users can save annotations for following radiomics and deep learning analyses on the uRP, or download them for future review.

2.2.2. ROI modification tool

After labeling the ROI, a variety of specific preprocessing needs to be performed on the mask to meet diverse image analyses. For example, studying tumor microenvironment requires obtaining the peritumor region, whereas studying liver fat requires extracting the ROI by intensity. The uRP offers a morphological modification tool to dilate or erode the selected ROI according to user-defined distances in the x, y, and z directions. For existing ROIs, separation can be performed by user-defined split intervals, connectivity, or minimum size (voxels). It can also merge, intersect, and complement multiple ROIs by performing linear operations on the pixels at each position of the ROI (Figure 2B). When conducting multimodal research, ROI can be duplicated across modalities. The above ROI modification tools can be applied to 2D and 3D ROI, and users can expand research directions with a variety of ROI preprocessing tools available on the platform.

2.3. Deep learning modules

With the advancement of deep learning and its wide application in medical image processing (37, 38), this technology has shown great potential in organ segmentation (39), disease diagnosis (40), etc. Some powerful online trainable deep learning modules are available on the uRP, such as the segmentation module, registration module, and classification module. The network architecture of each module is shown in Figure 3. The segmentation module can segment various organs of the whole body using a cascade coarse-to-fine framework (Figure 3A); the registration module performs unsupervised registration from a moving image to a reference image (Figure 3B), and the classification module can focus on ROI to classify the input image (Figure 3C). These deep learning modules are directly invoked through a simple parameter configuration, which can be flexibly applied to various research scenarios.

2.3.1. Segmentation

The input of the segmentation module can be single- or multimodal 2D or 3D data, where models can be trained to automatically delineate ROIs, such as organs or tumors. The uRP integrates a segmentation toolkit named VB-Net (41, 42). Briefly, we use a V-Net as the backbone for the segmentation task, which consists of a compression path, an expansion path, and skip connections (43). The compression path extracts high-level context information and the expansion path upsamples the signal to recover its original size, where skip connections allow the extracted high-level context information to be fused with fine-grained local information. To reduce model parameters and GPU memory cost, bottleneck layers are added in the down block and up block of the network (Figure 3A). Moreover, a variety of optimization strategies are embedded in the network to improve segmentation performance and extend application scenarios, as described below:
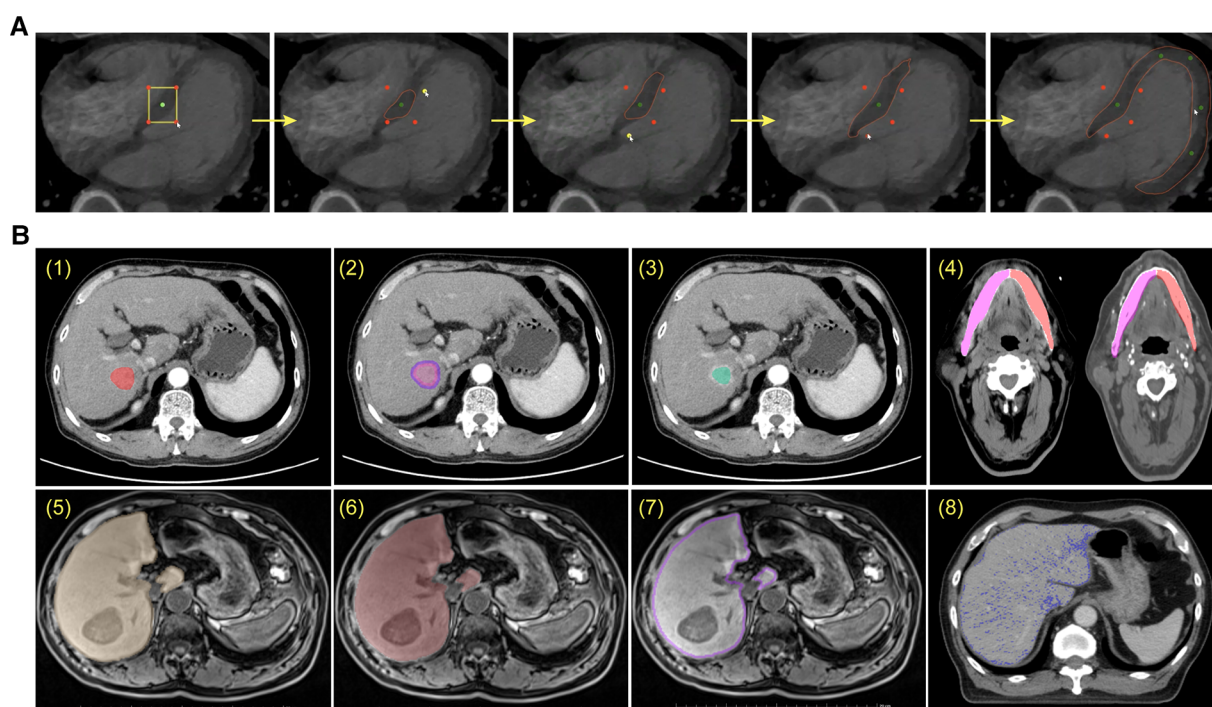
FIGURE 2
Image preprocessing by semi–automatic delineation tool. (**A**) Smart interactive segmentation. (**B**) Preprocessing methods of the original ROI (1), including dilation (2), erosion (3), duplication (4), union (5), intersection (6), complement (7), and separation by intensity (8).

(1) Adaptive input module, which adds convolutional layers to large-size images to ensure the network adapts to various input images; (2) A cascade coarse-to-fine strategy, in which the coarse-resolution model aims to localize ROIs in the original image by leveraging the global 3D context, and the fine-resolution model focuses on refining detailed boundaries of ROIs (**Figure 3A**); (3) Self-attention mechanism, which accelerates network convergence and improves segmentation accuracy; (4) Various loss functions, such as Dice loss, focal loss, and boundary loss, which can effectively constrain the segmented targets in different tasks (41).

Scientific research users can invoke the deep learning segmentation module directly through the configuration file, requiring no coding skills in the process (**Supplementary Video 3**). The configuration file provides the following functions: (1) setting paths of training data, model storage, and output results; (2) hyperparameters during the training process (e.g., GPU selection, batch size, training epochs); (3) data augmentation (e.g., flipping, rotating, scaling), (4) data preprocessing (e.g., sampling method, cropped size, padding type, normalization); (5) configurations of the loss function, optimizer, and learning rate scheduler; (6) segmentation evaluation metrics (i.e., Dice, average symmetrical surface distance, Hausdorff distance); (7) configurations of networks and strategies, e.g., cascading of different networks and selection of attention mechanisms. The configuration file should be configured in a standard format (a template is provided on the uRP) and called in the training process.

Moreover, the segmentation module integrates more than 100 high-precision organ segmentation models throughout the body.

Segmentation of organs at risk using the uRP's segmentation module has shown great advantages in terms of speed (0.7 s vs. 20 s per organ), accuracy (average Dice score 96.6% vs. 84.3%), and robustness (successful rate 98.6% vs. 83.3%) compared to conventional methods (44).

2.3.2. Registration

Image registration is to align a moving image to a reference image, which is a critical procedure in the analyses of multimodal images and longitudinal data (45). The uRP provides both traditional image registration (i.e., rigid, affine transformation) and deep learning-based nonlinear registration. Briefly, the nonlinear registration model consists of a registration network, spatial transform block, and hybrid loss calculation module. A hybrid loss is calculated to strengthen the alignment constraints of different structures, which combines image dissimilarity, deformation regularization, and segmentation dissimilarity with different weights (**Figure 3B**). The available image dissimilarity metrics include mean square difference (MSD), normalized correlation (NC), mutual information (MI), etc. Image registration can be performed between the images of the same modality or of different modalities; for example, MI is generally selected as the image dissimilarity for cross-modal registration.

Notably, the segmentation results obtained automatically from the segmentation module could also be used in the registration process. In previous studies, organ segmentations served as soft constraint in the loss function to provide auxiliary information in the training of the registration model (46, 47). Compared to
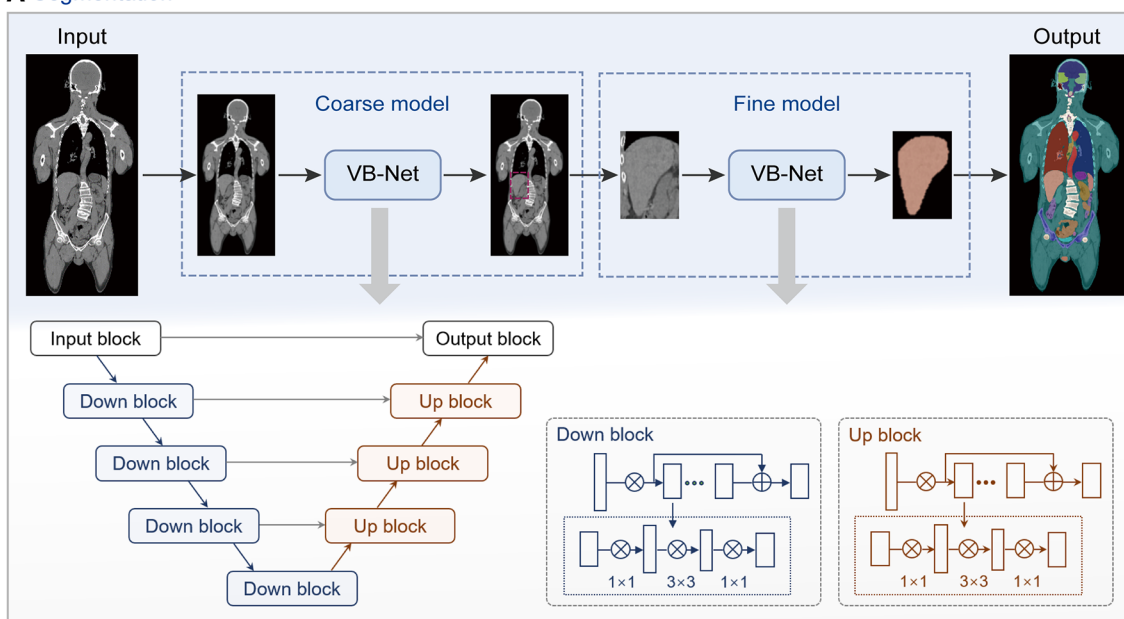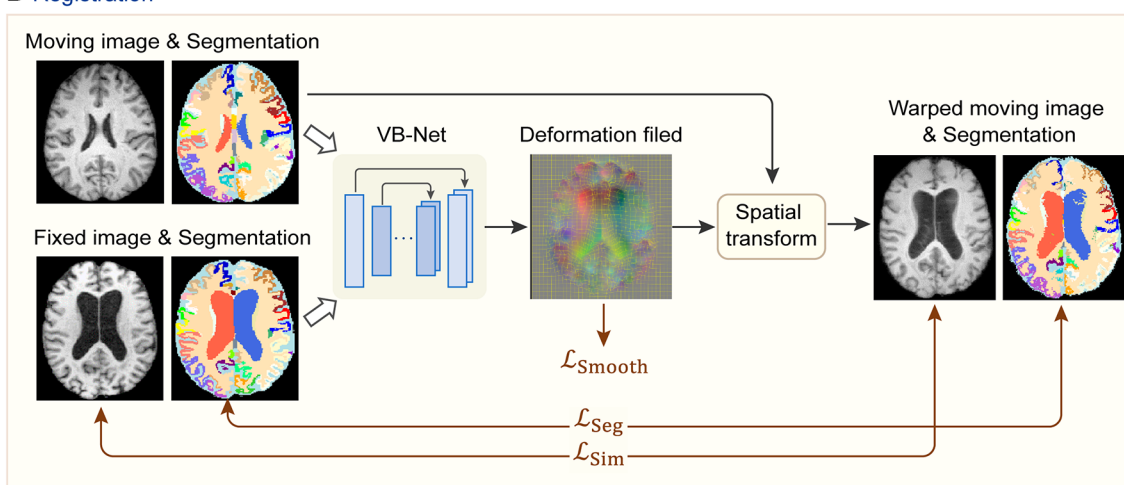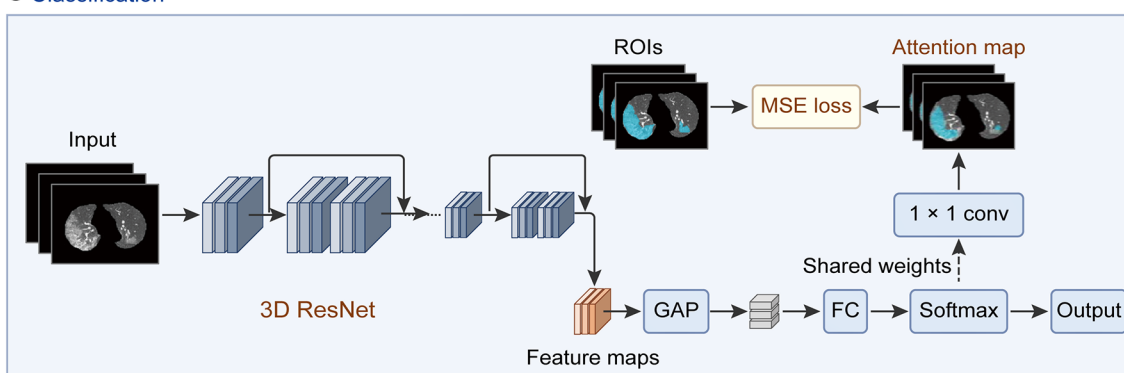
FIGURE 3

Network architectures of deep learning modules in the uRP. (**A**) Workflow of segmentation network (VB-Net) with a coarse-to-fine strategy, which first roughly locates the target area and then segments the fine boundary of the ROI. (**B**) Image registration framework with introduced region segmentations as constraints. (**C**) Classification network from images to features, with ROI attention strategy. GAP, global average pooling; FC, fully connected layer.

using only the intensity image for registration, these studies found that the large deformation can be more readily estimated with the help of the segmentation result.

2.3.3. Classification

The uRP integrates a classification module and can be used for two-class and multi-class tasks. ResNet (48) is used as the classification backbone and optimized by several strategies: (1) an online attention module as CAM (49) and Grad-CAM (50), that ensures the network to focus on ROIs and increases the model interpretability (**Figure 3C**); (2) a balanced sampling mechanism, which can alleviate the imbalanced distribution of the input data;

(3) different sampling methods for inputs; (4) various loss functions, e.g., focal loss (51), Ap loss (52), and CAM loss. A lot of classic and mature classification networks such as DenseNet (53) and EfficientNet (54) have been embedded in the uRP. Moreover, the classification module can be flexibly invoked *via* a configuration file, similar to the segmentation module.

2.4. Radiomics and statistics

Radiomics is a quantitative image analysis technique through extracting quantitative features from medical images, that aims to
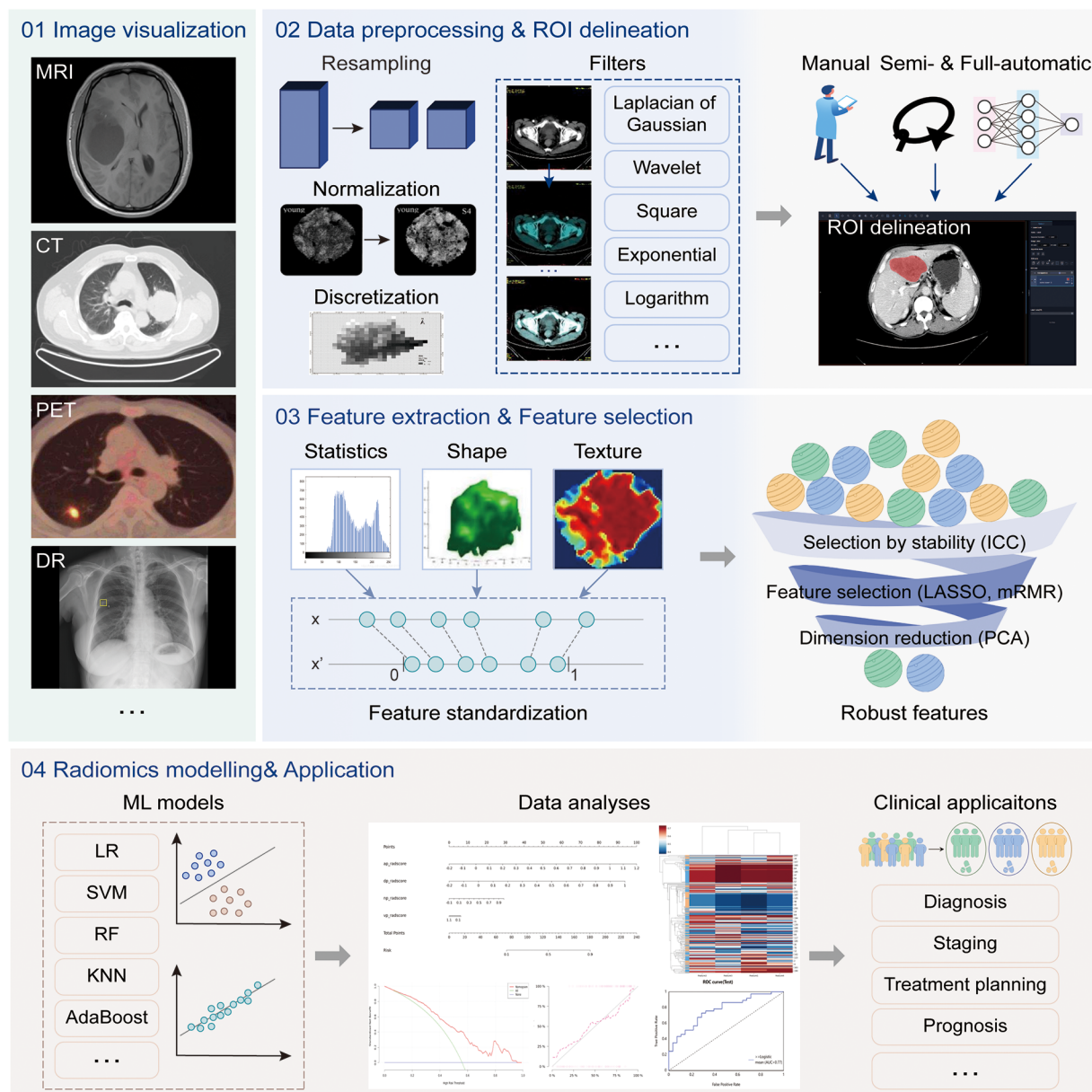
FIGURE 4
The radiomics analysis workflow. The radiomics analysis module supports four main functions, including (1) image visualization for multi-modal images, (2) data processing and region of interest (ROI) delineation, (3) feature extraction and selection, and (4) model construction and evaluation. The above analysis workflow can be used in a variety of clinical applications.

link large-scale data mining of images with clinical and biological endpoints (55). The uRP implements a one-stop analytic pipeline of radiomics, providing clinical researchers with a simple UI for image visualization, image processing, feature analysis, model construction and evaluation, and statistical analysis (**Figure 4**).

2.4.1. Data preprocessing and grouping

Previous studies have shown that radiomics features are sensitive to variations in gray level, pixel size, and slice thickness of images (56–59). However, it is difficult to standardize parameters during image acquisition for all patients in a clinical setting. The platform supports a variety of normalization algorithms to normalize image signal intensity, such as mean_std, max_min, and center_width. And it resamples the image and ROI mask to specified pixel spacing to achieve a standardized variable pixel size and slice thickness by resampling algorithms, such as the nearest neighbor, linear interpolation, and B-spline interpolation. A pre-defined bin number of 64 is used for all analyses. In image processing and feature calculation, we follow the guidelines of the imaging biomarker standardization initiative (IBSI) (60).

The uRP supports three data grouping methods, namely customized grouping, random grouping according to proportion, and cross-validation grouping. The training set is used for feature selection and model construction, and the model performance is evaluated on the testing set.

2.4.2. Feature extraction and standardization

After preprocessing the images, a total of 2,264 radiomics features can be automatically extracted from each ROI (**Supplementary Table S1**). The first-order statistics include 18 features that reflect the quantitative depiction of the distribution of voxel intensity in medical images. The shape-based features include 14 features that reflect the shape and size of a region. The textural features include 21 gray level co-occurrence matrix (GLCM) features, 16 gray level run length matrix (GLRLM) features, 16 gray level size zone matrix (GLSZM) features, 5 neighboring gray-tone difference matrix (NGTDM) features, and 14 gray levels dependent matrix (GLDM) features, which quantify regional heterogeneity differences. Additionally, the derived images are obtained by applying 24 filters (box mean, additive Gaussian noise, binomial blur, curvature flow, box-sigma, normalize, Laplacian sharpening, discrete Gaussian, mean, speckle noise, recursive Gaussian, shot noise, LoG (sigma: 0.5, 1, 1.5, 2), and wavelets (LLL, LLH, LHL, LHH, HLL, HLH, HHL, HHH)), and are used to extract first-order statistics and textural features (2,160 derived features). Most features defined in the uRP conform to feature definitions described in the IBSI (60).

To ensure the clinical utility of the model, features beyond radiomics should also be considered to improve the model's generalizability, such as demographic information and biological data. Considering that radiomics features and clinical features have different ranges, feature standardization algorithms are also provided in the uRP, such as *z*-score_scaler, min_max_scaler, quantile_transformer, yeojohnson_transformer, boxcox_transformer, L1_normalization, L2_normalization, and max_abs_scaler (61).

2.4.3. Feature selection and model construction

The feature selection module is used for feature selection or dimension reduction to improve model performance. It includes variance thresholding for removing low variance features, SelectKBest for removing high *p*-value features, as well as least absolute shrinkage and selection operator (LASSO) for sparse feature selection. It also includes factor analysis, independent component analysis (ICA), linear discriminant analysis (LDA), principal component analysis (PCA), and more than 10 algorithms for dimension reduction to get fewer new features formed by original features.

Based on the selected features, various machine learning-based models can be constructed for classification or regression tasks. Our proposed uRP integrates 13 machine learning algorithms, including adaptive boosting (AdaBoost), bagging decision tree, decision tree, Gaussian process, gradient boosting decision tree (GBDT), K-nearest neighbors (KNN), random forest (RF), logistic regression (LR), extreme gradient boosting (XGBoost), stochastic gradient descent (SGD), support vector machine (SVM), quadratic discriminant analysis (QDA), partial least squares-discriminant analysis (PLS-DA), and allows for hyperparameters adjustment. The nomogram model is also included in this module, which combines radiomics and clinical factors to facilitate the clinical utility of predictive models. It is necessary to first calculate the score of each predictive variable, obtain the total point, and then find the probability of the disease outcome corresponding to the total score.

2.4.4. Evaluation metrics

To evaluate model performance, the uRP provides two sets of quantitative metrics for classification and regression models.

(1) Classification model: The platform can automatically generate receiver operating characteristic (ROC) curves and calibration curves in the training and validation cohorts, and calculate multiple metrics, such as the area under the ROC curve (AUC, with 95% confidence interval), F1 score, precision, sensitivity, specificity, and accuracy. It also provides a visual representation of the confusion matrix and supports the comparison of multiple models. In addition, the histogram, box chart, violin chart, correlation analysis heatmap, and clustering analysis heatmap are optionally plotted to establish the relationship between features. For clinical applications, the decision curve and clinical impact curve can be plotted to assess the clinical usefulness of models by quantifying net benefits at different risk thresholds.

(2) Regression model: The platform automatically generates prediction curves and scatter plots to visualize the results, and calculates mean absolute error (MAE), mean absolute error (MSE), R-squared, and Pearson correlation coefficient for regression model evaluation.

3. Results

To illustrate the clinical utility of the uRP, we list three domain analyses (i.e., brain, pneumonia, and knee joint) and representative use cases.

3.1. Specific domain analysis

It is worth noting that uRP can be extended and applied to a variety of designated applications (**Figure 1**), e.g., brain structural analysis, pneumonia analysis, and knee joint analysis. In this section, we focus on these three specific scenarios to describe the versatility and scalability of uRP, covering segmentation, quantitative analysis of ROI, classification for disease prediction, and prognosis, where image analysis modules are integrated into sequential to form an automatic analysis pipeline.
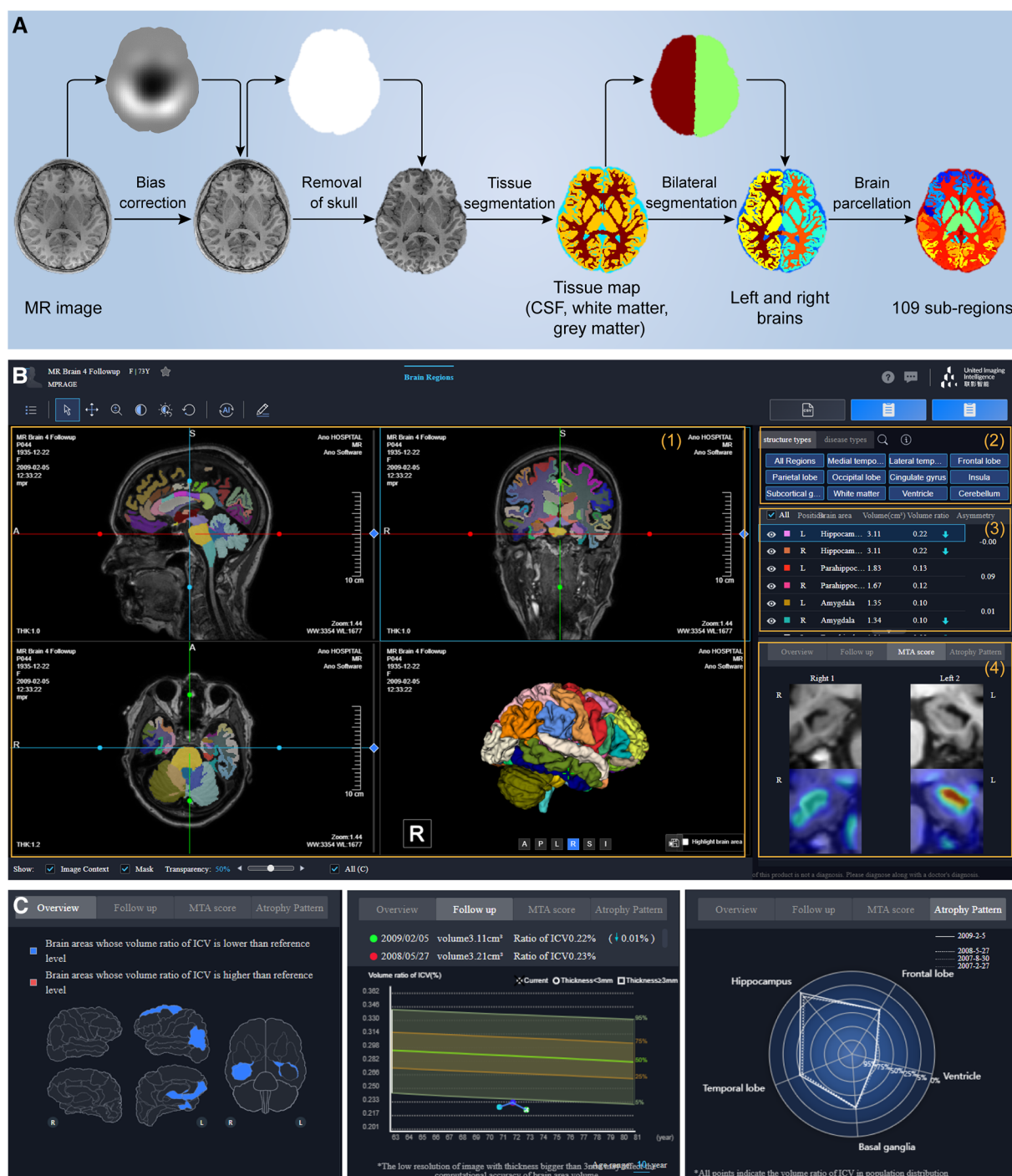
FIGURE 5
One-stop brain analysis module integrated into the uRP. (A) Segmentation pipeline of brain structure, including the bias field correction, removal of the skull, tissue segmentation of white matter, gray matter, and cerebrospinal fluid (CSF), bilateral segmentation, and parcellation of 109 sub-regions. (B) The user interface (UI) of brain analysis, including (1)−(2) image visualization of the specific region or disease, (3) quantitative volume analysis, and (4) MTA-scale score. (C) UI showing abnormal regions and follow-up analysis.

3.1.1. Brain analysis

Neuroimaging shows tremendous potential in the early diagnosis of neurodegenerative diseases (62, 63), in which structural imaging (MRI) serves as a foundation to provide brain tissue and parcellation information (Figure 5).

First, uRP can handle high-resolution MR images and hierarchically segment brain structures (Figure 5A). The workflow mainly involves: (1) the bias field correction, (2) removal of the skull, (3) tissue segmentation of white matter, gray matter, and cerebrospinal fluid (CSF), (4) bilateral segmentation, and (5) parcellation of 109 sub-regions. A total of 109 sub-regions includes 22 temporal lobe structures, 20 frontal lobe structures, 12 parietal lobe structures, 8 occipital lobe structures, 8 cingulate gyrus structures, 2 insular structures, 12 subcortical gray matter structures, cerebral white matter structures, ventricles, cerebellum, and other structures (64). To emphasize, this segmentation process depends on VB-Net, achieving efficient, precise, and end-to-end segmentation of multiple sub-regions. The model was trained on T1 images of 1,800 subjects and tested on 295 subjects with an average Dice of 0.92, where the images were acquired from the Consortium for Reliability and Reproducibility (CoRR) dataset (65) and Chinese brain molecular and functional mapping (CBMFM) project (66). Based on the segmentation results, the volume, volume ratio of each sub-region, and the asymmetry index of paired sub-region are calculated quantitatively and compared to the relevant parameters from the gender- and age-matched normal dataset (Figure 5B). Abnormal brain sub-regions are identified, where those below the 5th percentile of the normal range are considered likely to be abnormally atrophic, and those above the 95th percentile are considered abnormally enlarged (Figure 5C). On the other hand, the medial temporal lobe atrophy (MTA) score, also known as Schelten's scale, is developed with an AI model for automated assessment of the hippocampus atrophy status (67). The score ranges from 0 to 4, in which the higher the score, the more severe the hippocampal atrophy (Figure 5B). It should be noted that the segmentation and computation of the 109 sub-regions take less than 1 min.

Based on this, uRP holds enormous ability in the follow-up data analysis, i.e., (1) comparing volume changes of ROIs over time to explore pathological progression of neurodegenerative diseases; (2) constructing AI models to predict the odds of other diseases such as Parkinson's disease (PD) or mental diseases for early diagnosis and early intervention; (3) promoting the brain functional analysis with established structural ROIs, as well as fiber connectivity analysis in diffusion tensor imaging (DTI) (Figure 5C).

3.1.2. Pneumonia analysis

With the worldwide spread of coronavirus disease (COVID-19), the early diagnosis and prognostic analysis of pneumonia have become an urgent need, which inspires a large number of related researches to serve the clinic (68–71). CT has been popularly used to monitor pneumonia's progression and measure the disease severity (72–74). Based on chest CT scans, three

important issues need to be explored: (1) the location of the pneumonia infection, (2) the severity of the infection, and (3) the etiology of the disease. uRP meets these requirements by integrating segmentation, classification, and registration algorithms (Figure 6A).

The first step is to locate the infected lesions. First, the whole lung is obtained by embedded VB-Net, followed by bilateral segmentation (75). Then, the left lung is segmented into 2 lung lobes (superior and inferior lobes), while the right lung is segmented into 3 lung lobes (superior, middle, and inferior lobes). Afterward, 5 lung lobes are then finely segmented into 18 bronchopulmonary segments (Figure 6B). Infected lesions are also auto-contoured in this process, and can be visualized from the UI. Noted that the human-in-the-loop strategy is designed to iteratively update VB-Net to address the problem of limited annotated data (73). To be specific, an initial segmentation model based on a small amount of delineated data is applied to the new data, and segmentation results are manually corrected and then fed into the model, so that a more robust model will be trained through 3∼4 iterations, greatly improving the efficiency of delineation.

Following the segmentation, a diverse set of handcrafted features are calculated to quantitatively assess the severity of the pneumonia infection (75), including (1) 26 volumetric features—the volume and percentage of infections in each lobe and pulmonary segment, (2) 31 numeric features—the number of infected lobes and pulmonary segments, (3) 32 histogram features—the histogram distribution of CT intensity, (4) 7 surface features—the surface area of infections and lung boundary. A total of 96 location-specific features are displayed in the UI to reflect the severity of pneumonia infection. In addition, follow-up data can be registered with previously acquired images to extract changes in infection-specific features to monitor the progression of pneumonia and to accurately determine the severity (Figure 6C).

Importantly, uRP can also distinguish different types of pneumonia and predict possible pneumonia causes *via* uRP's built-in classification algorithms. Based on the segmented masks and extracted features, dual-sampling attention 3D ResNet is used to diagnose COVID-19 from community acquired pneumonia (CAP) (76). Moreover, masks, handcrafted features, as well as radiomics features can be used to classify the cause of pneumonia (e.g., viruses, fungi, and bacteria), and to report the corresponding probabilities.

Therefore, uRP-based pneumonia analyses involve automatic segmentation of infected lesions, extraction and visualization of quantitative metrics, and classification of different types of pneumonia, which largely accelerates scientific research on pneumonia.

3.1.3. Knee joint analysis

Knee osteoarthritis (OA), known as a degenerative joint disease, results from the wear, tear, and progressive loss of articular cartilage, which may eventually lead to disability (77). The severity staging of knee OA should be carefully taken into consideration for the treatment, which relies on a range of morphological parameters, including the volume and thickness of
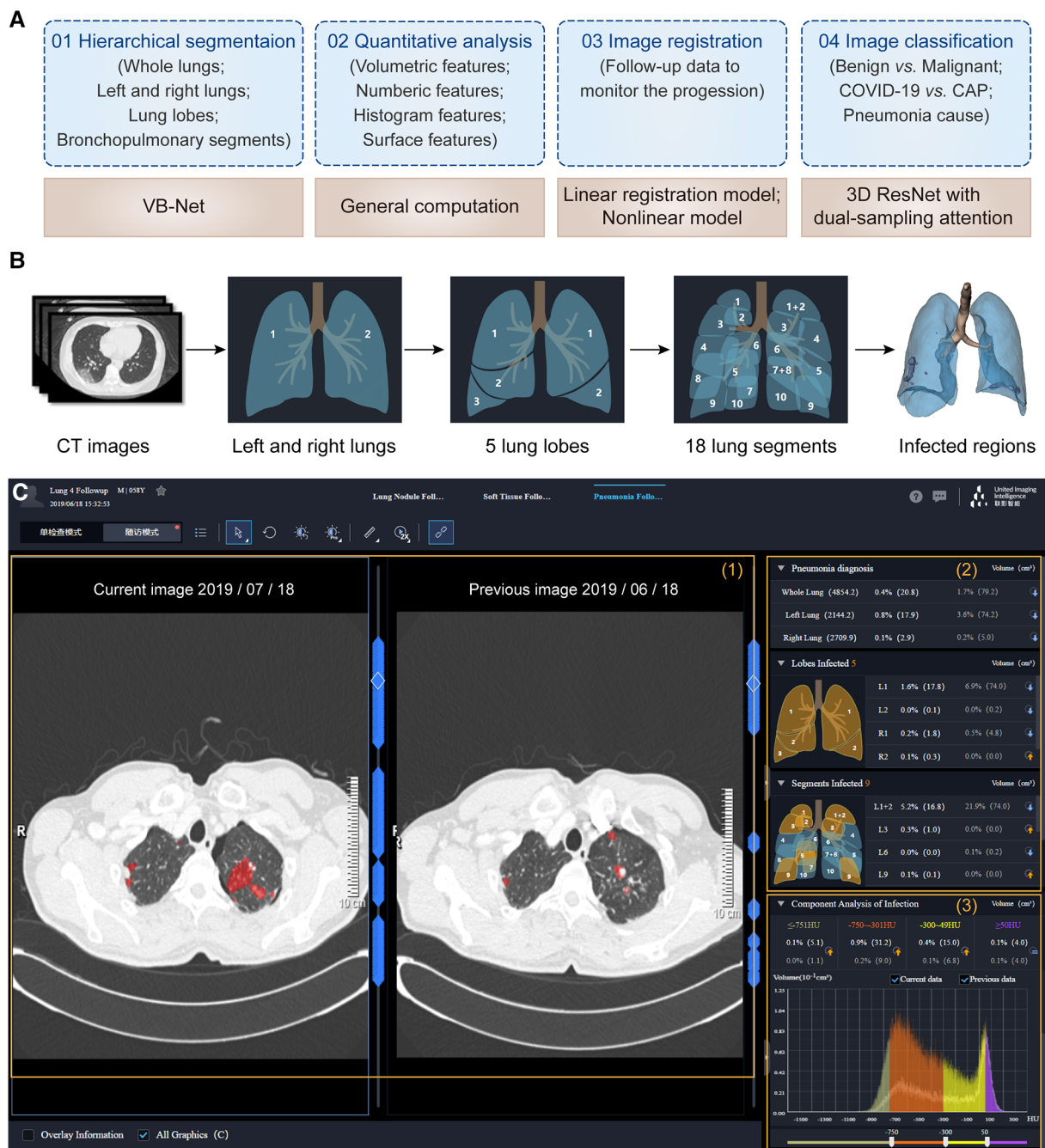
FIGURE 6
One-stop pneumonia analysis module integrated into the uRP. (A) Schematics of pneumonia analysis, including segmentation, computation, registration, and classification. (B) Hierarchical segmentation of lungs and infected lesions. (C) The user interface of pneumonia analysis on the uRP, including (1) image visualization to compare images from two-time points, (2) quantitative analysis for pneumonia diagnosis and severity assessment, and (3) histogram analysis of CT intensity distribution of images for comparison.

articular cartilage, and minimal joint space width (mJSW) (78). To benefit clinical practice, uRP implements a complete analysis pipeline that automatically segments knee tissues and calculates morphological metrics.

At first, a cascade coarse-to-fine strategy is applied to obtain fast and accurate segmentation results (Figure 3A). Through two 3D-VNet segmentation models, multiple knee joint tissues are accurately segmented, including bones (i.e., femoral bone, tibial bone, patella bone), cartilages (i.e., femoral cartilage, tibial cartilage, patella cartilage), as well as menisci, and visualized in the UI of uRP (Figure 7A).

At the same time, a series of morphological parameters are automatically calculated from corresponding segmented masks (79), including (1) volumes of cartilages, menisci, and cartilage damage, (2) the mean thicknesses of cartilages, (3) the medial and lateral mJSWs, calculated with the minimal Euclidean distance

A Knee joint segmentation
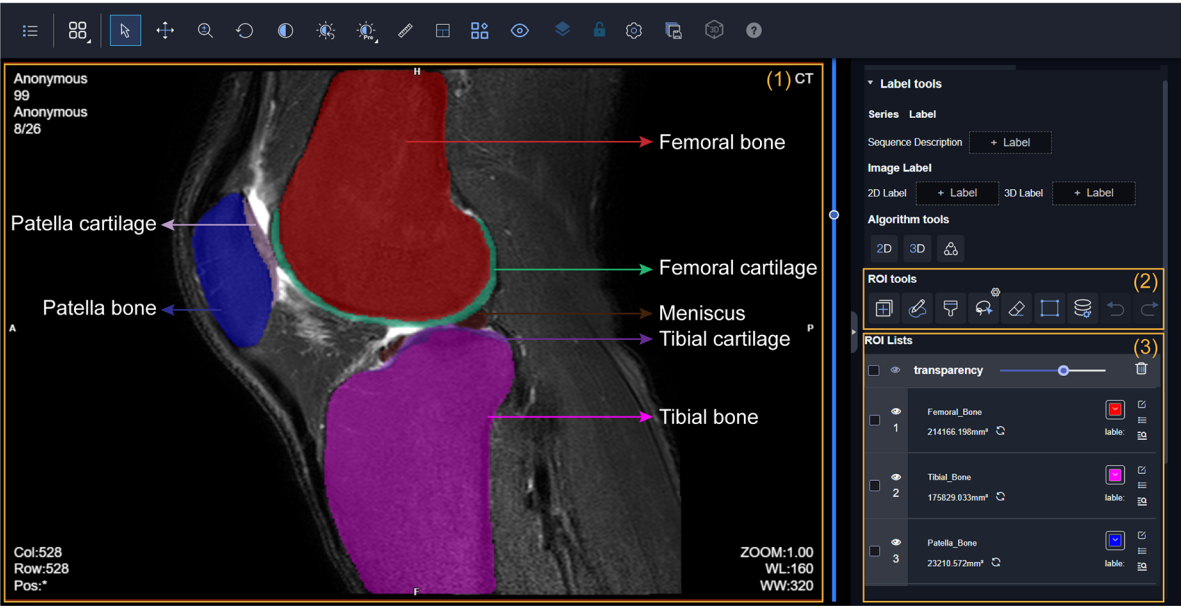
B Quantitative analysis and severity assessment
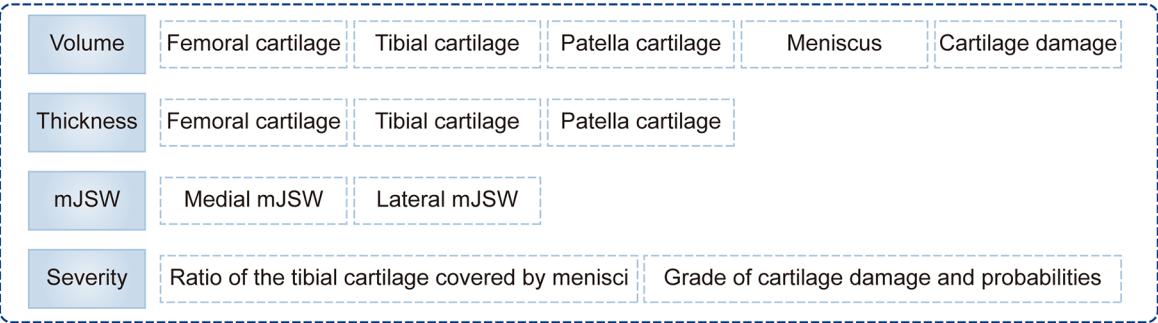
C User interface of basic features and classifaction results

FIGURE 7
Knee joint analysis module integrated into the uRP. (A) The user interface (UI) of segmentation results of the knee joint, in which multiple knee joint tissues are segmented including femoral bone and cartilage, tibial bone and cartilage, patella bone and cartilage, as well as menisci. (B) Quantitative metrics calculated for knee osteoarthritis (OA) diagnosis, i.e., volume, thicknesses, minimal joint space width (mJSW), and severity assessment. (C) UI showing quantified features and classification results for OA severity grading.

between the femoral and tibial surfaces, (4) the severity assessment— the ratio of the tibial cartilage covered by menisci, the grade of cartilage damage and corresponding probability (**Figure 7B**).

Severity grading is also performed by a classification algorithm (**Figure 7C**). All the above results can be viewed from the uRP, which is helpful for clinicians to make a quick diagnosis of knee OA.

Overall, the knee joint analysis module on uRP can handle MRI images in a fully automatic manner, in which knee joint tissues are segmented and key features (i.e., volume, thickness, mJSW) are calculated to identify the severity of knee OA, thus guiding the optimal treatment.

3.2. Use cases

To illustrate the utility of the uRP, we example several use cases and organize them into the following four parts:

3.2.1. Segmentation use

The automatic segmentation module on the uRP has been used in many medical scenarios, for example, to diagnose COVID-19 infections. Chest CT scans of 549 patients were collected from Shanghai Public Health Clinical Center and several other hospitals, and were automatically segmented and quantified the infected regions throughout the lungs (80). Researchers used the pneumonia analysis module of the uRP, yielding a Dice similarity coefficient of 91.6% ± 10.0% for the segmentation of infected areas, and a mean estimation error of the infected percentage of 0.3% for the whole lung on the validation dataset (300 patients). Besides, to predict the severity of COVID-19 patients, quantitative features of 5 lung lobes and 18 bronchopulmonary segments were calculated and used to construct a classification model based on the SVM algorithm. The best accuracy of severity prediction was 73.4% ± 1.3%, which demonstrated that uRP's pneumonia analysis module owned good performance on patient severity prediction.

3.2.2. Classification use

The image classification module built into the uRP has been experimentally explored in various applications such as disease diagnosis, risk classification, and treatment selection. Gastrointestinal stromal tumors (GISTs) are mesenchymal neoplasms with variable malignant potentials (81). In clinics, accurate preoperative risk classification is important for surgical resection and adjuvant treatment (82). Researchers from Shandong Provincial Hospital and the Affiliated Hospital of Qingdao University collected contrast-enhanced CT images and clinicopathological characteristics from 733 patients and the goal was to develop a model for predicting the GISTs risk stratification (83). A deep learning model with an attention mechanism was constructed on the uRP's classification module to divide patients into three categories (i.e., low-malignant, intermediate-malignant, and high-malignant). The obtained AUCs were 0.90, 0.80, and 0.89 on the testing sets for low-malignant, intermediate-malignant, and high-malignant GISTs, respectively. Therefore, this multi-center study demonstrated that the quantitative CT and deep learning-based approach can be an objective means of predicting the risk stratification of GISTs.

3.2.3. Registration use

Image registration is widely needed for the analyses of multimodal images and longitudinal data. For example, precise registration of dynamic contrast enhanced MR images (DCE-MEIs) with pre-contrast images can be used to obtain accurate subtraction images, which helps to better differentiate the true enhancement of residual viable tumors from coagulative necrosis. In our recent study, Qian et al. collected 3D liver DCE-MRI series from 97 patients, with each series including pre-contrast T1-weighted, post-contrast T1-weighted scan at the arterial phase, and post-contrast T1-weighted scan at the portal venous phase (84). To overcome the intensity enhancement (due to the contrast agent) and spatial distortions of the liver, the cascade registration framework integrated into uRP was used to register the post-contrast images to the pre-contrast images. The registration performance of the proposed method was compared with the traditional registration method SyN in the ANTs toolkit, and the results demonstrated that the proposed framework embedded in the registration module of uRP owned a comparable performance and significantly improved efficiency.

3.2.4. Radiomics use

The radiomics module is particularly suitable for image-based classification and regression tasks, deployed and used well in multiple centers. Radiomics analysis was conducted in the Fourth Affiliated Hospital of Harbin Medical University to discriminate acute myocardial infarction from unstable angina (85). A total of 210 patients with coronary computed tomography angiography (CCTA) images were retrospectively collected and randomly divided into the training and validation cohorts. Following the workflow of radiomics analysis in uRP, three vessel-based pericoronary adipose tissue (PCAT) radiomics features and fat attenuation index (FAI) were extracted from CCTA images, and then selected features were used to construct the classification model. Results demonstrated that the combined model achieved superior performance with AUC values of 0.97 and 0.95 for training and validation cohorts, respectively. The Affiliated Hospital of Southwest Medical University applied a radiomics model to predict the T stage, perineural invasion, and microvascular invasion of extrahepatic cholangiocarcinoma (CCA). This retrospective trial included 101 CCA patients scanned with four MR images, including T1-weighted imaging (T1WI), T2-weighted imaging (T2WI), diffusion-weighted imaging (DWI), and apparent diffusion coefficient (ADC) map. Radiomics features were extracted from four MR images, followed by dimension reduction, and selected features were used to construct three classification models corresponding to the three tasks. The AUC values of models in the testing cohort for predicting T stage, perineural invasion, and microvascular invasion were 0.962, 1.000, and 1.000, respectively (86). Similarly, researchers from Zhongshan Hospital constructed a multi-parametric radiomics nomogram for predicting the microvascular invasion (MVI) based on multiple MR sequences from 130 patients pathologically confirmed with intrahepatic CCA. The nomogram incorporating tumor size, intrahepatic duct dilatation, and the radiomics model, achieved good prediction performance with AUC values of 0.953, 0.861, and 0.819 in the training, validation ($n = 33$), and time-independent testing cohorts ($n = 24$), respectively (87).

4. Discussion

Over the years, uRP has gained broad acceptance within the medical image analysis community, which can be attributed to its breadth of functionality, extensibility, and cross-platform portability.

4.1. Extendable modules

One key strength of uRP is the modularization for customized extensibility, where plug-ins can be designed for specific purposes, and can be freely combined to accomplish complex analyses, suitable for a variety of scenarios. uRP's extensible structure means that new functionality can be integrated on top of the existing platform, rather than being created from scratch, showing significant advantages over monolithic software. The modular design benefits users in several ways:

(1) Reusability: uRP has integrated multiple algorithms that are embedded into axiomatic building blocks and are invoked for specific analysis workflows. An algorithm may participate in many tasks and perform a similar function, meaning that advanced techniques can be reused in new research areas. In addition to existing algorithms, new algorithms can also be developed and integrated into the uRP, for example, transfer learning, transformer networks, etc., to meet clinical needs. Besides, uRP can be extended with third-party software (i.e., 3D Slicer) to import historical annotation data or clinical information.

(2) Reproducibility: The design of plug-ins of uRP follows the criteria of standardization and interoperability, which can be easily shared among research groups, thus minimizing the need for duplication and facilitating reproducibility and consensus building. uRP collects detailed parameters set by the user and generates a report summarizing quantitative metrics and results so that a study can be replicated by different researchers or even different institutions. Equally important, uRP can be used for data management and fair data repositories are essential for reproducible research. Meanwhile, uRP owns reproducibility at scale, e.g., producing high-dimensional radiomics features for ROI in each image or applying the same quantitative analysis in high-throughput images.

(3) Community: uRP has an active community of more than 50 hospitals in China. The feedbacks provided by users fuel improvements of uRP, especially the development of innovative algorithms for clinical needs.

4.2. Advantages

The uRP platform has shown great potential for one-stop image analyses in multiple scientific researches. There are several other platforms for image analysis, such as 3D Slicer (28), SenseCare (33), MITK (32). The proposed uRP has several advantages. (1) uRP can provide cloud-based services allowing for high user concurrency, a data management module to facilitate collaborations, and batch processing capability for efficient analysis. (2) Some of these platforms are for clinical diagnostic and treatment planning usages, with modules such as lung cancer diagnosis, radiotherapy planning for head and neck cancers. uRP is designed for clinical scientific research, supporting image analysis and research idea validation. (3) The uRP integrates machine learning algorithms and statistical analysis methods to provide a more powerful analytical tool for clinical research. (4) The uRP also offers smart annotation tools for medical images, to remedy the time-consuming manual annotation process that serves as a prerequisite for ROI based analysis.

4.3. Outlook

Although uRP already has many applications as a one-stop medical image analysis platform, some issues are still to be addressed. Firstly, uRP currently focuses on the analysis of radiological images, while other types of images could be future supported, such as those from pathology. Secondly, most existing applications are oriented towards the adult population, while applications specific for the fetal, infant, and children are needed. We have only two applications for now, i.e., infant brain segmentation, skeletal age prediction, and would be further increased in the future. While uRP cannot meet all the scientific needs, it is fortunately a dynamic software that evolves together with the new scientific research derived from clinical problems. In the future, we will continue to develop more tools and domain-specific methods, including algorithms, statistics, as well as radiomics, to improve the efficiency, accuracy, robustness, and generalization of one-stop analysis. Represented by ChatGPT, generative AI is a hot research topic for now. It would bring many improvements to the current scientific research, such as better integrating multi-omics data to assist clinical diagnosis and prognosis assessment. We are also exploring the possibility of integrating generative AI in the platform at suitable scenarios (88). We anticipate that the uRP can be applied to diverse domains covering an increasing number of analytic pipelines for diverse pathological diseases.

5. Conclusion

In summary, uRP is a one-stop medical image analysis software for scientific research, and it not only supports versatile visualizations, but also provides advanced functionality such as automatic segmentation, registration, and classification for a variety of application domains. More specifically, it has three major merits, (1) advanced built-in algorithms (>100) applicable to multiple imaging modalities (i.e., CT, MR, PET, DR), diseases (i.e., tumor, neurodegenerative disease, pneumonia), and applications (i.e., diagnosis, treatment planning, follow-up); (2) an iterative deep learning-based training strategy for fast delineation of ROIs of large-scale datasets, thereby greatly saving

clinicians' time and obtaining novel and more robust models; (3) a modular architecture with customization and extensibility, where plugins can be designed for specific purposes. As a result, it will be necessary to investigate and develop new algorithms and strategies to expand application domains and really solve clinical problems.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding authors.

Author contributions

DS and FS contributed to conception and design of the study. YW, AL, AI, MZ, LC, JS, LW, YZ, XZ, and ZX organized the database. JW, YX, and XW wrote the sections of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Conflict of interest

JW, YX, XW, AL, YW, LC, JS, LW, YZ, XZ, ZX, FS, and DS are employees of Shanghai United Imaging Intelligence Co., Ltd, Shanghai, China. AI and MZ are employees of United Imaging Intelligence Co., Ltd., Cambridge, MA, United States. These companies have no role in designing and performing the surveillance and analyzing and interpreting the data.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fradi.2023.1153784/full#supplementary-material.

References

1. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol*. (2016) 2:1636–42. doi: 10.1001/jamaoncol.2016.2631

2. Galldiks N, Albert NL, Sommerauer M, Grosu AL, Ganswindt U, Law I, et al. PET Imaging in patients with meningioma-report of the RANO/PET group. *Neuro Oncol*. (2017) 19:1576–87. doi: 10.1093/neuonc/nox112

3. Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*. (2018) 172:1122–31. doi: 10.1016/j.cell.2018.02.010

4. Thoeny HC, Barbieri S, Froehlich JM, Turkbey B, Choyke PL. Functional and targeted lymph node imaging in prostate cancer: current status and future challenges. *Radiology*. (2017) 285:728–43. doi: 10.1148/radiol.2017161517

5. Torigian DA, Zaidi H, Kwee TC, Saboury B, Udupa JK, Cho ZH, et al. PET/MR imaging: technical aspects and potential clinical applications. *Radiology*. (2013) 267:26–44. doi: 10.1148/radiol.13121038/-/DC1

6. Fink JR, Muzi M, Peck M, Krohn KA. Multimodality brain tumor imaging: mR imaging, PET, and PET/MR imaging. *J Nucl Med*. (2015) 56:1554–61. doi: 10.2967/jnumed.113.131516

7. Fowler AM, Strigel RM. Clinical advances in PET-MRI for breast cancer. *Lancet Oncol*. (2022) 23:e32–43. doi: 10.1016/s1470-2045(21)00577-5

8. Zhang L, Wang Y, Peng Z, Weng Y, Fang Z, Xiao F, et al. The progress of multimodal imaging combination and subregion based radiomics research of cancers. *Int J Biol Sci*. (2022) 18:3458–69. doi: 10.7150/ijbs.71046

9. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. *IEEE Trans Pattern Anal Mach Intell*. (2022) 44:3523–42. doi: 10.1109/TPAMI.2021.3059968

10. Moen E, Bannon D, Kudo T, Graf W, Covert M, Van Valen D. Deep learning for cellular image analysis. *Nat Methods*. (2019) 16:1233–46. doi: 10.1038/s41592-019-0403-1

11. Jin L, Shi F, Chun Q, Chen H, Ma Y, Wu S, et al. Artificial intelligence neuropathologist for glioma classification using deep learning on hematoxylin and eosin stained slide images and molecular markers. *Neuro Oncol*. (2021) 23:44–52. doi: 10.1093/neuonc/noaa163

12. Fuyong X, Yuanpu X, Hai S, Fujun L, Lin Y. Deep learning in microscopy image analysis: a survey. *IEEE Trans Neural Netw Learn Syst*. (2018) 29:4550–68. doi: 10.1109/TNNLS.2017.2766168

13. Funke J, Tschopp F, Grisaitis W, Sheridan A, Singh C, Saalfeld S, et al. Large scale image segmentation with structured loss based deep learning for connectome reconstruction. *IEEE Trans Pattern Anal Mach Intell*. (2019) 41:1669–80. doi: 10.1109/TPAMI.2018.2835450

14. Oksuz I, Clough JR, Ruijsink B, Anton EP, Bustin A, Cruz G, et al. Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation. *IEEE Trans Med Imaging*. (2020) 39:4001–10. doi: 10.1109/TMI.2020.3008930

15. Gibson E, Giganti F, Hu Y, Bonmati E, Bandula S, Gurusamy K, et al. Automatic multi-organ segmentation on abdominal CT with dense V-networks. *IEEE Trans Med Imaging*. (2018) 37:1822–34. doi: 10.1109/tmi.2018.2806309

16. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z

17. de Vos BD, Berendsen FF, Viergever MA, Sokooti H, Staring M, Išgum I. A deep learning framework for unsupervised affine and deformable image registration. *Med Image Anal*. (2019) 52:128–43. doi: 10.1016/j.media.2018.11.010

18. Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. Voxelmorph: a learning framework for deformable medical image registration. *IEEE Trans Med Imaging*. (2019) 38:1788–800. doi: 10.1109/TMI.2019.2897538

19. Kang M, Hu X, Huang W, Scott MR, Reyes M. Dual-stream pyramid registration network. *Med Image Anal*. (2022) 78:102379. doi: 10.1016/j.media.2022.102379

20. Amor B B, Arguillere S, Shao L. ResNet-LDDMM: advancing the LDDMM framework using deep residual networks. *IEEE Trans Pattern Anal Mach Intell.* (2023) 45:3707–20. doi: 10.1109/TPAMI.2022.3174908

21. Xu J, Pan Y, Pan X, Hoi S, Yi Z, Xu Z. Regnet: self-regulated network for image classification. *IEEE Trans Neural Netw Learn Syst.* (2022) in press. doi: 10.1109/TNNLS.2022.3158966. [Epub ahead of print]

22. Wang H, Wang S, Qin Z, Zhang Y, Li R, Xia Y. Triple attention learning for classification of 14 thoracic diseases using chest radiography. *Med Image Anal.* (2021) 67:101846. doi: 10.1016/j.media.2020.101846

23. Hu JF, Huang TZ, Deng LJ, Jiang TX, Vivone G, Chanussot J. Hyperspectral image super-resolution *via* deep spatiospectral attention convolutional neural networks. *IEEE Trans Neural Netw Learn Syst.* (2021) 33:7251-65–00. doi: 10.1109/TNNLS.2021.3084682

24. Ji Z, Wang H, Han J, Pang Y. SMAN: stacked multimodal attention network for cross-modal image-text retrieval. *IEEE Trans Cybern.* (2022) 52:1086–97. doi: 10.1109/TCYB.2020.2985716

25. Tan H, Liu X, Yin B, Li X. DR-GAN: distribution regularization for text-to-image generation. *IEEE Trans Neural Netw Learn Syst.* (2022) in press. doi: 10.1109/TNNLS.2022.3165573. [Epub ahead of print]

26. Chikontwe P, Luna M, Kang M, Hong KS, Ahn JH, Park SH. Dual attention multiple instance learning with unsupervised complementary loss for COVID-19 screening. *Med Image Anal.* (2021) 72:102105. doi: 10.1016/j.media.2021.102105

27. Sobie EA. An introduction to MATLAB. *Sci Signal.* (2011) 4:tr7. doi: 10.1126/scisignal.2001984

28. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D Slicer as an image computing platform for the quantitative imaging network. *Magn Reson Imaging.* (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001

29. Chan HY, Walker PS. Opensim as a preliminary kinematic testing platform for the development of total knee arthroplasty implants. *J Biomech.* (2018) 76:53–60. doi: 10.1016/j.jbiomech.2018.05.014

30. Song XW, Dong ZY, Long XY, Li SF, Zuo XN, Zhu CZ, et al. REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One.* (2011) 6:e25031. doi: 10.1371/journal.pone.0025031

31. Jiang H, van Zijl PC, Kim J, Pearlson GD, Mori S. Dtistudio: resource program for diffusion tensor computation and fiber bundle tracking. *Comput Methods Programs Biomed.* (2006) 81:106–16. doi: 10.1016/j.cmpb.2005.08.004

32. Wolf I, Vetter M, Wegner I, Bottger T, Nolden M, Schobinger M, et al. The medical imaging interaction toolkit. *Med Image Anal.* (2005) 9:594–604. doi: 10.1016/j.media.2005.04.005

33. Duan Q, Wang G, Wang R, Fu C, Li X, Wang N, et al. SenseCare: A research platform for medical image informatics and interactive 3D visualization. *arXiv.* (2022). doi: 10.48550/arXiv.2004.07031

34. Besson FL, Henry T, Meyer C, Chevance V, Roblot V, Blanchet E, et al. Rapid contour-based segmentation for ^{18}F-FDG PET imaging of lung tumors by using ITK-SNAP: comparison to expert-based segmentation. *Radiology.* (2018) 288:277–84. doi: 10.1148/radiol.2018171756

35. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Intell.* (1986) 8:679–98. doi: 10.1109/TPAMI.1986.4767851

36. The OpenCV Library. https://opencv.org/author/opencv/

37. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature.* (2015) 521:436–44. doi: 10.1038/nature14539

38. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition. 2016 IEEE conference on computer vision and pattern recognition (CVPR)* (2016). p. 770–8. doi: 10.48550/arXiv.1512.03385

39. Ronneberger O, Fischer P, Brox T. *U-Net convolutional networks for biomedical segmentation. 2015 Medical image computing and computer-assisted intervention (MICCAI)* (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28

40. Wang X, Peng Y, Lu L, Lu Z, Bagheri M, Summers RM. *Chestx-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. 2017 IEEE conference on computer vision and pattern recognition (CVPR)* (2017). p. 3462–71. doi: 10.1109/CVPR.2017.369

41. Shi F, Hu W, Wu J, Han M, Wang J, Zhang W, et al. Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. *Nat Commun.* (2022) 13:6566. doi: 10.1038/s41467-022-34257-x

42. Han M, Yao G, Zhang W, Mu G, Zhan Y, Zhou X, et al. *Segmentation of CT thoracic organs by multi-resolution VB-nets. Proceedings of the 2019 challenge on segmentation of thoracic organs at risk in CT images (SegTHOR@ISBI 2019)* (2019).

43. Milletari F, Navab N, Ahmadi S-A. *V-Net: fully convolutional neural networks for volumetric medical image segmentation. 2016 Fourth international conference on 3D vision (3DV)* (2016). p. 565–71. doi: 10.48550/arXiv.1606.04797

44. Linte CA, Fei B, Rong C, Zhou Q, Zhou XS, Zhan YJ, et al. *Large-scale evaluation of V-net for organ segmentation in image guided radiation therapy. In Medical imaging 2019: image-guided procedures, robotic interventions, and modeling.* (2019) Vol. 10951. p. 109510O1–10O7. doi: 10.1117/12.2512318

45. Hill DLG, Batchelor PG, Holden M, Hawkes DJ. Medical image registration. *Phys Med Biol.* (2001) 46:R1–45. doi: 10.1088/0031-9155/46/3/201

46. Hu Y, Modat M, Gibson E, Li W, Ghavami N, Bonmati E, et al. Weakly-supervised convolutional neural networks for multimodal image registration. *Med Image Anal.* (2018) 49:1–13. doi: 10.1016/j.media.2018.07.002

47. Zhu W, Myronenko A, Xu Z, Li W, Roth H, Huang Y, et al. *NeurReg: neural registration and its application to image segmentation. In Proceedings of the IEEE/CVF winter conference on applications of computer vision* (2020). p. 3617–26. doi: 10.1109/WACV45572.2020.9093506

48. Hara K, Kataoka H, Satoh Y. *Can spatiotemporal 3D CNNs retrace the history of 2D CNNs and ImageNet? 2018 IEEE/CVF conference on computer vision and pattern recognition* (2018). p. 6546–55. doi: 10.1109/CVPR.2018.00685

49. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. *Learning deep features for discriminative localization. 2016 IEEE conference on computer vision and pattern recognition (CVPR)* (2016). p. 2921–9. doi: 10.48550/arXiv.1512.04150

50. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. *Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE international conference on computer vision (ICCV)* (2017). p. 618–26. doi: 10.1109/ICCV.2017.74

51. Lin TY, Goyal P, Girshick R, He K, Dollar P. Focal loss for dense object detection. *IEEE Trans Pattern Anal Mach Intell.* (2020) 42:318–27. doi: 10.1109/TPAMI.2018.2858826

52. Chen K, Lin W, Li J, See J, Wang J, Zou J. AP-loss for accurate one-stage object detection. *IEEE Trans Pattern Anal Mach Intell.* (2021) 43:3782–98. doi: 10.1109/TPAMI.2020.2991457

53. Iandola F, Moskewicz M, Karayev S, Girshick R, Darrell T, Keutzer K. DenseNet: Implementing efficient ConvNet descriptor pyramids. arXiv. (2014). doi: 10.48550/arXiv.1404.1869

54. Tan M, Le QV. EfficientNet: Rethinking models scaling for convolutional neural networks. arXiv. (2019). doi: 10.48550/arXiv.1905.11946

55. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer.* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036

56. Park SH, Lim H, Bae BK, Hahm MH, Chong GO, Jeong SY, et al. Robustness of magnetic resonance radiomic features to pixel size resampling and interpolation in patients with cervical cancer. *Cancer Imaging.* (2021) 21:19. doi: 10.1186/s40644-021-00388-5

57. Li Y, Lu L, Xiao M, Dercle L, Huang Y, Zhang Z, et al. CT Slice thickness and convolution kernel affect performance of a radiomic model for predicting EGFR status in non-small cell lung cancer: a preliminary study. *Sci Rep.* (2018) 8:17913. doi: 10.1038/s41598-018-36421-0

58. Whybra P, Parkinson C, Foley K, Staffurth J, Spezi E. Assessing radiomic feature robustness to interpolation in ^{18}F-FDG PET imaging. *Sci Rep.* (2019) 9:9649. doi: 10.1038/s41598-019-46030-0

59. Park D, Oh D, Lee M, Lee SY, Shin KM, Jun JS, et al. Importance of CT image normalization in radiomics analysis: prediction of 3-year recurrence-free survival in non-small cell lung cancer. *Eur Radiol.* (2022) 32:8716–25. doi: 10.1007/s00330-022-08869-2

60. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology.* (2020) 295:328–38. doi: 10.1148/radiol.2020191145

61. Patro SGK, sahu KK. Normalization: A preprocessing stage. arXiv. (2015). doi: 10.48550/arXiv.1503.06462

62. Hampel H, Cummings J, Blennow K, Gao P, Jack CR Jr., Vergallo A. Developing the ATX(N) classification for use across the Alzheimer disease continuum. *Nat Rev Neurol.* (2021) 17:580–9. doi: 10.1038/s41582-021-00520-w

63. Myszczynska MA, Ojamies PN, Lacoste AMB, Neil D, Saffari A, Mead R, et al. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. *Nat Rev Neurol.* (2020) 16:440–56. doi: 10.1038/s41582-020-0377-8

64. Liu S, Jie C, Zheng W, Cui J, Wang Z. Investigation of underlying association between whole brain regions and Alzheimer's Disease: a research based on an artificial intelligence model. *Front Aging Neurosci.* (2022) 14:872530. doi: 10.3389/fnagi.2022.872530

65. Zuo XN, Anderson JS, Bellec P, Birn RM, Biswal BB, Blautzik J, et al. An open science resource for establishing reliability and reproducibility in functional connectomics. *Sci Data.* (2014) 1:140049. doi: 10.1038/sdata.2014.49

66. Gu D, Shi F, Hua R, Wei Y, Li Y, Zhu J, et al. An artificial-intelligence-based age-specific template construction framework for brain structural analysis using magnetic resonance images. *Hum Brain Mapp.* (2022) 44:861–75. doi: 10.1002/hbm.26126

67. Korf ESC, Wahlund LO, Visser PJ, Scheltens P. Medial temporal lobe atrophy on MRI predicts dementia in patients with mild cognitive impairment. *Neurology.* (2004) 63:94–100. doi: 10.1212/01.wnl.0000133114.92694.93

68. Liu T, Siegel E, Shen D. Deep learning and medical image analysis for COVID-19 diagnosis and prediction. *Annu Rev Biomed Eng.* (2022) 24:179–201. doi: 10.1146/annurev-bioeng-110220-012203

69. Minaee S, Kafieh R, Sonka M, Yazdani S, Jamalipour Soufi G. Deep-COVID: predicting COVID-19 from chest x-ray images using deep transfer learning. *Med Image Anal.* (2020) 65:101794. doi: 10.1016/j.media.2020.101794

70. Wang G, Liu X, Shen J, Wang C, Li Z, Ye L, et al. A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest x-ray images. *Nat Biomed Eng.* (2021) 5:509–21. doi: 10.1038/s41551-021-00704-1

71. Wang S, Zha Y, Li W, Wu Q, Li X, Niu M, et al. A fully automatic deep learning system for COVID-19 diagnostic and prognostic analysis. *Eur Respir J.* (2020) 56:2000775. doi: 10.1183/13993003.00775-2020

72. Goncharov M, Pisov M, Shevtsov A, Shirokikh B, Kurmukov A, Blokhin I, et al. CT-based COVID-19 triage: deep multitask learning improves joint identification and severity quantification. *Med Image Anal.* (2021) 71:102054. doi: 10.1016/j.media.2021.102054

73. Wang Y, Chen Y, Wei Y, Li M, Zhang Y, Zhang N, et al. Quantitative analysis of chest CT imaging findings with the risk of ARDS in COVID-19 patients: a preliminary study. *Ann Transl Med.* (2020) 8:594. doi: 10.21037/atm-20-3554

74. Zhu X, Song B, Shi F, Chen Y, Hu R, Gan J, et al. Joint prediction and time estimation of COVID-19 developing severe symptoms using chest CT scan. *Med Image Anal.* (2021) 67:101824. doi: 10.1016/j.media.2020.101824

75. Shi F, Xia L, Shan F, Song B, Wu D, Wei Y, et al. Large-scale screening of COVID-19 from community acquired pneumonia using infection size-aware classification. *Phys Med Biol.* (2021) 66:065031. doi: 10.1088/1361-6560/abe838

76. Ouyang X, Huo J, Xia L, Shan F, Liu J, Mo Z, et al. Dual-sampling attention network for diagnosis of COVID-19 from community acquired pneumonia. *IEEE Trans Med Imaging.* (2020) 39:2595–605. doi: 10.1109/TMI.2020.2995508

77. Roos EM, Arden NK. Strategies for the prevention of knee osteoarthritis. *Nat Rev Rheumatol.* (2016) 12:92–101. doi: 10.1038/nrrheum.2015.135

78. Chaudhari AS, Kogan F, Pedoia V, Majumdar S, Gold GE, Hargreaves BA. Rapid knee MRI acquisition and analysis techniques for imaging osteoarthritis. *J Magn Reson Imaging.* (2020) 52:1321–39. doi: 10.1002/jmri.26991

79. Tang X, Guo D, Liu A, Wu D, Liu J, Xu N, et al. Fully automatic knee joint segmentation and quantitative analysis for osteoarthritis from magnetic resonance

(MR) images using a deep learning model. *Med Sci Monit.* (2022) 28:e936733. doi: 10.12659/MSM.936733

80. Shan F, Gao Y, Wang J, Shi W, Shi N, Han M, et al. Abnormal lung quantification in chest CT images of COVID-19 patients with deep learning and its application to severity prediction. *Med Phys.* (2021) 48:1633–45. doi: 10.1002/mp.14609

81. Joensuu H, Hohenberger P, Corless CL. Gastrointestinal stromal tumour. *Lancet.* (2013) 382:973–83. doi: 10.1016/s0140-6736(13)60106-3

82. Nilsson B, Bumming P, Meis-Kindblom JM, Oden A, Dortok A, Gustavsson B, et al. Gastrointestinal stromal tumors: the incidence, prevalence, clinical course, and prognostication in the preimatinib mesylate era–A population-based study in western Sweden. *Cancer.* (2005) 103:821–9. doi: 10.1002/cncr.20862

83. Kang B, Yuan X, Wang H, Qin S, Song X, Yu X, et al. Preoperative CT-based deep learning model for predicting risk stratification in patients with gastrointestinal stromal tumors. *Front Oncol.* (2021) 11:750875. doi: 10.3389/fonc.2021.750875

84. Qian L, Zhou Q, Cao X, Shen W, Suo S, Ma S, et al. A cascade-network framework for integrated registration of liver DCE-MR images. *Comput Med Imaging Graph.* (2021) 89:101887. doi: 10.1016/j.compmedimag.2021.101887

85. Si N, Shi K, Li N, Dong X, Zhu C, Guo Y, et al. Identification of patients with acute myocardial infarction based on coronary CT angiography: the value of pericoronary adipose tissue radiomics. *Eur Radiol.* (2022) 32:6868–77. doi: 10.1007/s00330-022-08812-5

86. Huang X, Shu J, Yan Y, Chen X, Yang C, Zhou T, et al. Feasibility of magnetic resonance imaging-based radiomics features for preoperative prediction of extrahepatic cholangiocarcinoma stage. *Eur J Cancer.* (2021) 155:227–35. doi: 10.1016/j.ejca.2021.06.053

87. Qian X, Lu X, Ma X, Zhang Y, Zhou C, Wang F, et al. A multi-parametric radiomics nomogram for preoperative prediction of microvascular invasion status in intrahepatic cholangiocarcinoma. *Front Oncol.* (2022) 12:838701. doi: 10.3389/fonc.2022.838701

88. Wang S, Zhao Z, Ouyang X, Wang Q, Shen D. ChatCAD: Interactive computer-aided diagnosis on medical image using large language models. arXiv. (2023). doi: 10.48550/arXiv.2302.07257

The promise and limitations of artificial intelligence in musculoskeletal imaging

Patrick Debs[1] and Laura M. Fayad[1,2,3]*

[1]The Russell H. Morgan Department of Radiology & Radiological Science, The Johns Hopkins Medical Institutions, Baltimore, MD, United States, [2]Department of Orthopaedic Surgery, Johns Hopkins University School of Medicine, Baltimore, MD, United States, [3]Department of Oncology, Johns Hopkins University School of Medicine, Baltimore, MD, United States

With the recent developments in deep learning and the rapid growth of convolutional neural networks, artificial intelligence has shown promise as a tool that can transform several aspects of the musculoskeletal imaging cycle. Its applications can involve both interpretive and non-interpretive tasks such as the ordering of imaging, scheduling, protocoling, image acquisition, report generation and communication of findings. However, artificial intelligence tools still face a number of challenges that can hinder effective implementation into clinical practice. The purpose of this review is to explore both the successes and limitations of artificial intelligence applications throughout the muscuskeletal imaging cycle and to highlight how these applications can help enhance the service radiologists deliver to their patients, resulting in increased efficiency as well as improved patient and provider satisfaction.

KEYWORDS

artificial intelligence, machine learning, neural networks, musculoskeletal imaging, image interpretation, automation

Introduction

Radiological imaging has come to play a central role in the diagnosis and management of different muscuskeletal (MSK) disorders, and both technological improvements and increased access to medical imaging have led to a rise in the utilization of common MSK imaging modalities (1, 2). As such, there is a growing need for technical innovations that can help optimize workflow and increase productivity, especially in radiology practices that are witnessing higher volumes of increasingly complex cases (3).

Artificial intelligence (AI), or the development of computer systems that can mimic human intelligence when performing human tasks, is rapidly expanding in the field of diagnostic imaging and could potentially help improve workflow efficiency (4). AI is a broad term that encompasses numerous techniques, and recent advances in the field have transformed this technology into a powerful tool with several promising applications. Nested within AI is machine learning (ML), a subfield that gives computers the ability to learn and adapt by drawing inferences from patterns in data without following explicit instructions (5). ML uses observations from data to create algorithms and subsequently makes use of these algorithms to determine future output, with the goal of designing a system that can automatically learn without any human intervention. Deep learning (DL)

Abbreviations

ACL, anterior cruciate ligament; AI, artificial intelligence; AUC, area under the curve; BMD, bone mineral density; CNN, convolutional neural network; CT, computed tomography; DL, deep learning; DXA, dual-energy x-ray absorptiometry; ML, machine learning; MRI, magnetic resonance imaging; MSK, muscuskeletal; PACS, picture archiving and communication systems.
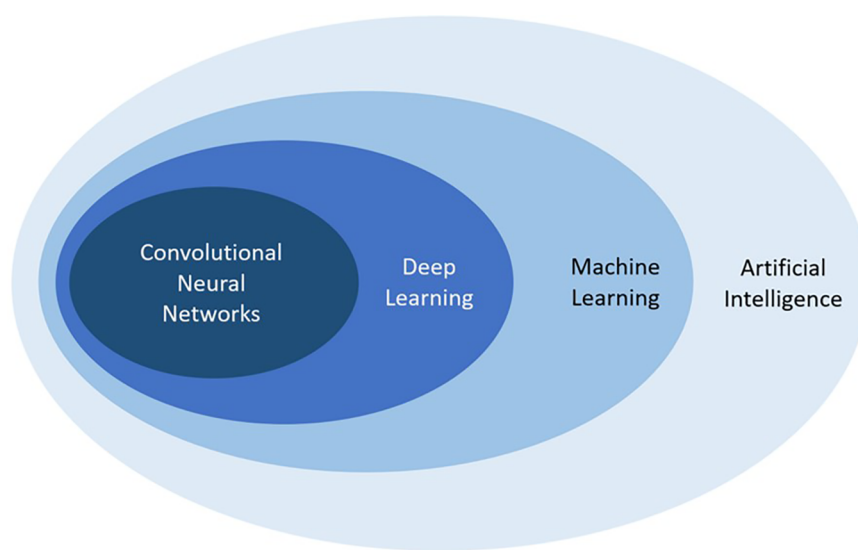
FIGURE 1
Schematic representation demonstrating the relationship between artificial intelligence, machine learning, deep learning, and convolutional neural networks, all subfields of each other.

is an even more specialized subfield within ML that uses multiple processing layers to progressively extract higher-level features from raw input presented in the form of large datasets, and the recent development of DL with convolutional neural networks (CNN) is an important technological advancement apt at solving image-based problems with reportedly outstanding performance in several key aspects of medical imaging (**Figure 1**) (6–8). CNNs are widely used in computer vision; they represent feedforward neural networks with multiple layers of non-linear transformations between inputs and outputs and can be programmed to classify an image or objects according to their features (output) by means of a training dataset with numerous images or objects (input) (4).

For AI models to be developed, large data sets with high-quality images and annotations are needed for both training of a model and validation of its performance, and, given that developers are usually not located within medical practices or hospital systems and therefore do not have access to such data, image sharing between the two becomes necessary. This multi-step process requires collaboration between clinicians and developers and, following approval from the responsible ethical committees, begins with image de-identification, storage, and resampling of resolutions (9). Images must then be appropriately labeled with ground truth definitions, and, depending on the outcome of interest, this can involve several different steps such as manual labeling of images, data extraction from medical charts and pathology reports, and detection of imaging findings from radiology reports or by radiologists' re-review of the imaging findings (9). Typically, data sets used for training are larger than data sets used for validation and testing, and, although logistically challenging in many cases, images should ideally be obtained from multiple diverse sources to increase representation of different populations and ensure generalizability of the model's performance (9).

To understand how images are used for training AI and DL models, it is important to understand the architecture of the neural networks often employed in such models. The basic building block of a deep neural network is a node, which can be considered analogous to a neuron, and one neural network is comprised of several weighted nodes arranged into layers and connected through weighted connections (10). Training data is fed to a network at an initial input layer and then propagated throughout all layers of the model: each layer performs both linear (e.g., weighted additions) and non-linear (e.g., thresholding) mathematical computations from input received from the previous layer and feeds the output to the next layer, which then performs the same computations until one final output layer is reached (10). The model then provides a prediction, which is compared to the ground truth label previously assigned. Discrepancies between the two are fed back into the network through backward propagation and gradient descent: nodal weights and connections are adjusted accordingly, and the model is refined with every data point from the training set (**Figure 2**). Once the model is sufficiently refined, a validation set is typically used to evaluate the model's generalizability and further refine predictions, after which the model is then tested using a final test set with unseen data to simulate and assess real-life performance (10). The size of the data sets needed for training, validation, and testing can vary depending on the outcome and/or the targeted population (with larger sets needed for populations with more diversity and heterogeneity) but generally follows a ratio of 80:10:10 or 70:15:15, respectively (9).

With its rapid and exponential growth, AI has the potential to significantly strengthen several steps of the MSK imaging value chain and offer applications that extend beyond imaging interpretation to assist with non-interpretive tasks such as patient scheduling, optimal protocoling, image acquisition, and data
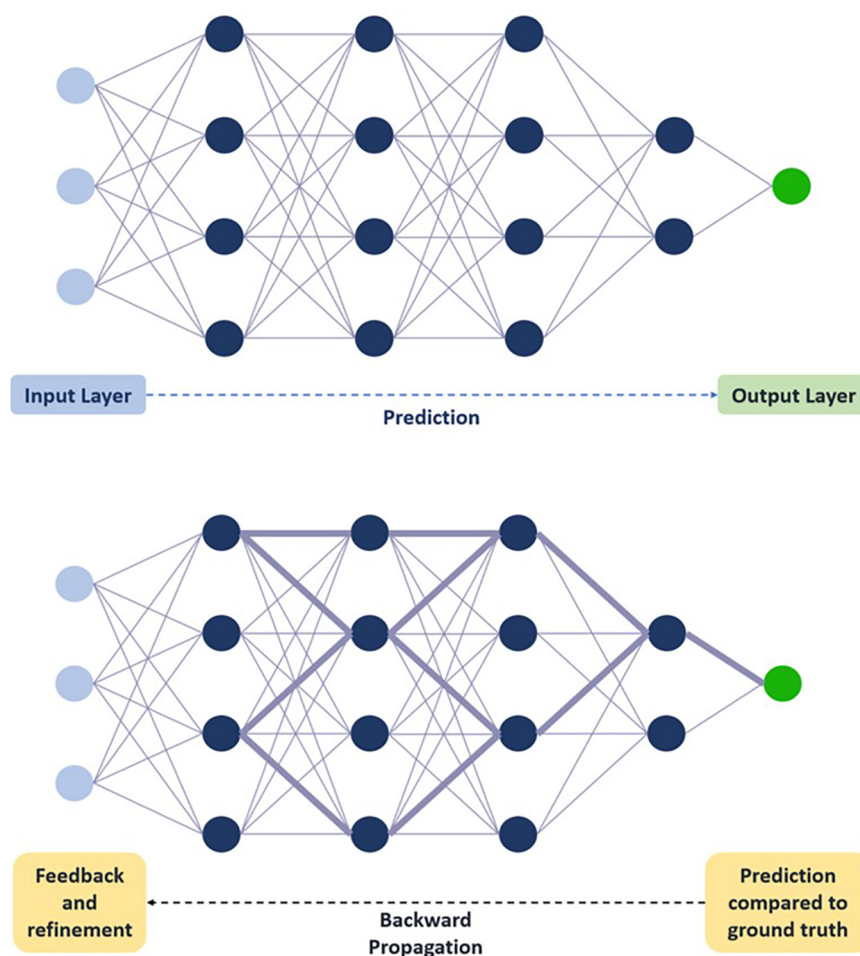
FIGURE 2
Schematic representation of deep neural network training. Training data is fed to the network at the initial input layer and propagated through subsequent layers for a prediction to be made at the final output layer. Prediction is compared to ground truth, and feedback through backward propagation leads to progressive refinement of weights. Circles represent nodes. Lines connecting circles represent weighted connections, with thickness correlating with weight magnitude. Dashed arrow represents flow of information through network.

sharing (11). AI can theoretically improve an MSK radiologist's ability to respond to the increasing workload of high-volume practices and continue delivering high-quality care by allotting more time for demanding tasks and minimizing time spent on more routine and less complex functions. However, AI is not without its pitfalls, and overutilization of this resource can pose multiple problems relating to medical errors, bias and inequality, data availability, and privacy concerns (12). The purpose of this review is to highlight the different applications AI is presently offering or can potentially offer throughout the MSK radiology imaging cycle and to discuss risks, limitations, and future directions of this important technology.

Prominent AI applications

Image appropriateness and protocoling

The first step in the MSK imaging process is to order the appropriate imaging test, the responsibility of which falls on

the referring clinician or provider confronted with a wide range of available modalities. AI, and ML in particular, could help facilitate the process: ML algorithms could be used to generate holistic clinical decision support systems that can consider various aspects from a patient's medical chart such as symptomatology, laboratory test results, physical examination findings, and previous imaging to recommend the modality best suited to address the clinical query in question (13, 14).

Protocoling comes next, and once an imaging modality is chosen, the MSK radiologist or trainee is usually responsible for ensuring that scans are performed correctly. Choosing the right protocol is crucial to reaching a proper diagnosis and optimizing patient care but can prove arduous and time-consuming for the radiologist tasked with several other responsibilities; as such, several recent studies have looked into how DL can be of assistance. Lee assessed the feasibility of using short-text classification to develop a CNN classifier capable of determining whether MRI scans should be completed following a routine or tumor protocol and, after

comparing CNN-derived protocols to those determined by MSK radiologists, reported an area under the curve (AUC) of 0.977 and an accuracy of 94.2% (15). Similarly, Trivedi et al. developed and validated a DL-based natural language classifier capable of automatically determining the need for intravenous contrast for MSK-specific MRI protocols based on the free-text clinical indication of the study and reported up to 90% agreement with human-based decisions (16). Although these studies show promising results, MSK imaging protocols are complex and diverse, given that MSK as a field encompasses localized and systemic diseases from neck to toe. More investigations could potentially explore the use of other composite classifiers such as medical history, prior imaging protocols, scanner-specific data, contrast information, and radiation exposure to help with protocoling decisions (13).

Scheduling

Given the rise in the use of medical imaging, adherence to set schedules has become more important for radiology practices, especially in the MSK setting where advanced and sometimes lengthy examinations such as MRI and CT are frequently used. No-shows or appointment cancellations can be a significant burden on practices and also represent missed opportunities for other patients to be scanned (17). There has been a growing interest in how AI can help optimize scheduling in various medical practices, and ML algorithms with predictive frameworks have been successfully used to predict missed appointments in diabetes clinics as well as urban, academic, and underserved settings (18, 19). Various ML predictive models have also been used to predict imaging no-shows effectively (20, 21), and Chong et al. demonstrated how using a pre-trained CNN with a predictive framework to predict MRI no-shows and accordingly send out proactive reminders to patients resulted in a reduced appointment no-show rate from 19.3% to 15.9% (22). ML could also help maximize patient throughput; Muelly et al. developed a feed-forward neural network that can make use of patient demographics and dynamic block lengths to estimate average MRI scan durations, resulting in decreased wait times, improved patient satisfaction, and optimized schedule fill rates (23).

Image acquisition

Magnetic resonance imaging acquisition

Given the critical need but lengthy nature of MRI scans in MSK imaging, there has always been an interest in reducing MRI acquisition times in order to decrease patient discomfort and improve scanner efficiency. Previous attempts at MRI acceleration focused on parallel imaging and compressed sensing, both of which operate by subsampling k-space and reducing the number of phase-encoding lines acquired during a scan, ultimately resulting in less data being collected (24, 25).

While efficient, these two techniques suffer from reduced image quality and increased artifacts in the reconstructed images, leading to less diagnostic imaging. ML has been proposed as a possible solution that can help mitigate the limitations of accelerated imaging by using subsampled k-space data to generate up-sampled high-resolution output images comparable to images generated from otherwise fully sampled k-space data (26). Using high quality MR images, Wang et al. trained a CNN to restore fine structural details on brain images obtained from zero-filled k-space data and were able to generate images of diagnostic quality comparable to images from a fully sampled k-space but with a fivefold increase in acquisition speed (27). Hammernik et al. were able to achieve a fourfold increase in knee MRI acquisition speed by using a DL technique that created high-quality reconstructions of under-sampled data (28), while Chaudhari et al. successfully made use of a CNN to output thin-slice knee images from thicker slices, thereby improving spatial resolution and image quality (29). Similarly, Wu et al. developed an eightfold-accelerated DL model capable of up-sampling sparsely sampled MRI data to output images with minimal artifacts and a permissible signal-to-noise ratio (30). In one study by Roh et al., DL-accelerated turbo spin echo sequences were assessed for their ability to depict acute fractures of the radius in patients wearing a splint and were shown to be effective for both increasing acquisition speed by a factor of 2 as well as improving image quality when compared to standard sequences (31). Studies are still ongoing, with AI-driven 10-fold accelerated MRI increasingly becoming within reach (32) and other exciting ML applications being explored such as the production of MR images from CT images (33) and the post-processing of a single MRI acquisition to obtain other planes and tissue weightings (34). One such advance in MSK imaging is the synthetic construction of fat-suppressed imaging from non-fat-suppressed imaging (35).

Computed tomography

Unlike MRI, CT exposes patients to ionizing radiation, and ML has shown promise as a tool that can help reduce the radiation dose of a CT scan while maintaining a high quality of images (36). The premise is similar to ML applications for MRI acquisitions, whereby the goal is to reconstruct images of diagnostic quality using lower-quality source data or reduced quantities of source data. Cross et al. demonstrated how CT images acquired at a low radiation dose and reconstructed in part using an artificial neural network were found to be similar to or improved compared to images obtained using standard radiation doses by more than 90% of the readers in the study (37). Other AI developments can also help enhance image quality by decreasing artifacts related to different factors, as demonstrated in the study by Zhang and Yu where a CNN trained to merge original- and corrected-image data was capable of suppressing metal artifacts and preserving anatomical structural integrity near metallic implants (38).

Image presentation

In radiology practices that use a Picture Archiving and Communication System (PACS), radiologists often spend a considerable amount of time manipulating image displays and toggling between sequences and viewing panes to display different imaging features in several anatomic planes. This is known as the hanging protocol and constitutes another venue that can be enhanced through the use of AI to afford radiologists more productivity and efficiency. A study by Kitamura showed how ML techniques using DenseNet-based neural network models can successfully optimize hanging protocols of lumbar spine x-rays by considering several parameters such as dynamic position and rotation correction (39). Moreover, one PACS vendor is currently using ML-based algorithms to learn a radiologist's preferences when viewing examinations, record orientations of the sequences most commonly used, suggest displays for future similar studies, and incorporate adaptations following every correction, all in an effort to improve the workflow in the reading room (40).

Image interpretation

Although AI can assist MSK radiologists with several steps of the imaging cycle, it is AI's ability to help with image interpretation, arguably a radiologist's most important responsibility, that has garnered the most attention in recent years. The next section discusses different ways AI and ML can help radiologists with MSK imaging interpretations to diagnose different conditions with greater efficiency. Table 1 provides a summary of AI tools with such applications.

Fractures

Automated fracture detection using AI can be helpful not only to radiologists but also to other clinicians (such as overnight emergency department personnel) who might not always have access to radiology services and would sometimes have to rely on their own preliminary fracture diagnosis. DL techniques have been gaining increasing attention over the past few years in their ability to detect fractures on images, as this can increase diagnostic reliability and reduce the rate of medical errors.

Some studies have shown that CNNs can outperform orthopedic surgeons when it comes to the detection of upper limb and ankle fractures on radiographs (41, 42). Additionally, multiple studies have shown promise when assessing the competence of AI in detecting both axial and appendicular skeletal fractures on radiographic and CT images (26, 43, 44, 97, 98), with one CNN model by Cheng et al. achieving an AUC of 0.98 and an accuracy of 91% for radiographic hip fracture detection (45). Rajpurkar et al. trained a 169-layer DenseNet baseline model to detect and localize fractures using a large dataset of MSK radiographs containing 40,561 manually-labeled images; when tested on a set of 207 studies, the model successfully detected finger and wrist abnormalities with an AUC

TABLE 1 AI tools in musculoskeletal imaging.

Publication	Application in musculoskeletal imaging	Algorithm type	System performance
Chung et al. (41)	Proximal humerus fracture detection	CNN	Sensitivity/specificity: 0.99/0.97 Accuracy: 96%
Olczak et al. (42)	Ankle, wrist and hand radiographic fracture detection	VGG 16-layer CNN	Accuracy: 83%
Yu et al. (43)	Hip fracture detection	CNN	Sensitivity/specificity: 97.1%/96.7%
Tomita et al. (44)	Osteoporotic vertebral fracture detection	CNN	Accuracy: 89.2%
Cheng et al. (45)	Hip fracture detection	CNN	Accuracy: 91% Sensitivity: 98%
Rajpurkar et al. (46)	Radiographic abnormality detection	169-layer DenseNet	Sensitivity/specificity: 81.5%/88.7%
Xue et al. (47)	Hip osteoarthritis detection	CNN	Sensitivity/specificity: 95.0%/90.7% Accuracy: 92.8%
Tiulpin et al. (48)	Knee osteoarthritis detection	Deep Siamese CNN	Accuracy: 66.71%
Antony et al. (49)	Knee osteoarthritis severity grading	CNN	Variable
Pedoia at el. (50)	Osteoarthritis cartilage degenerative change detection and staging	CNN	Accuracy: 80.74%, 78.02%, and 75.00% for normal, small, and complex large lesions, respectively
Liu et al. (51)	Knee cartilage lesion detection	CNN	Sensitivity/specificity: 84.1%/85.2% 80.5%/87.9%
Halabi et al. (52) Model by Cicero and Bilbily	Bone age detection	Google Inception V3 network	Mean absolute difference from ground truth: 4.265 months
Tajmir et al. (53)	Bone age detection	CNN	Accuracy: 98.6% within one year
Kim et al. (54)	Computer-assisted bone age detection	Greulich-Pyle method-based DL	69.5% correlation rate with reference bone age
Thodberg et al. (55) Martin et al. (56) Maratova et al. (57)	Bone age detection software validation	BoneXpert	Variable
Yang et al. (58)	Bone strength prediction	Scaling index method	Root mean square error: 0.869 ± 0.121 Coefficient of determination R^2: 0.68 ± 0.079
Huber et al. (59)	Bone biomechanical property detection	Scaling index method	Root mean square error: 1.021
Deniz et al. (60)	Proximal femur segmentation	CNN	Dice similarity score: 0.95 ± 0.02
Lee et al. (61)	Osteoporosis detection in panoramic radiographs	CNN	AUC values: 0.9763, 0.9991 and 0.9987

(*Continued*)

TABLE 1 Continued

Publication	Application in musculoskeletal imaging	Algorithm type	System performance
Pan et al. (62)	Osteoporosis screening using low-dose chest CT	3D U-net CNN	AUC: 0.927 for detecting osteoporosis and 0.942 for distinguishing low BMD
Jimenez-Pastor et al. (63)	Vertebrae localization and identification	Decision forests and image-based refinement	Identification rate: 79.6% for the thoracic and of 74.8% for the lumbar region
Lessmann et al. (64)	Vertebrae segmentation and identification	CNN	Dice score: 94.9 ± 2.1% for segmentation Accuracy: 93% for identification
Wimmer et al. (65)	Vertebral body and intervertebral disc localization and labeling	CNN	Detection rate: 93.6%
Jamaludin et al. (66)	Lumbar spine MRI radiological feature detection	CNN	Accuracy: 95.6% for disc detection and labeling
Lu et al. (67)	Automating lumbar vertebral segmentation, disc-level designation, and spinal stenosis grading	U-net CNN	Variable
Han et al. (68)	Spinal structure segmentation	Generative Adversarial Network	Accuracy:96.2% Dice coefficient: 87.1%, Sensitivity/ specificity: 89.1%/ 86.0%
Pan et al. (69)	Radiographic Cobb angle measurement and scoliosis detection	Mask R-CNN	Sensitivity/ specificity: 89.59%/ 70.37%
Weng et al. (70)	Radiographic sagittal vertical axis measurement	ResUNet CNN	Median absolute error: 1.183 ± 0.166 mm
Kim et al. (71)	Differentiating between tuberculous and pyogenic spondylitis	CNN	AUC: 0.802
Acar et al. (72)	Differentiating metastatic and completely responded sclerotic bone lesion in prostate cancer	Textural analysis, support vector machine, K-nearest neighbor, ensemble classifier	Variable
Lang et al. (73)	Differentiating spinal metastases origin cancer	Radiomics, CNN, CLSTM	Accuracy: 0.71 for radiomics, 0.71 for CNN, 8.81 for CLSTM
Malinauskaite et al. (74)	Differentiating soft-tissue lipoma and liposarcoma	Radiomics and ML classifier	AUC: 0.926
Zhang et al. (75)	MRI histopathological grading of soft tissue sarcomas	Radiomics, random forests, k-nearest neighbor, support vector machine	Accuracy: 0.88
He et al. (76)	Predicting recurrence of giant cell bone tumors	CNN and CNN regression models	Accuracy: 75.5% and 78.6%

(*Continued*)

TABLE 1 Continued

Publication	Application in musculoskeletal imaging	Algorithm type	System performance
Blackledge et al. (77)	Segmenting and evaluating soft tissue sarcomas after radiotherapy	Logistic regression, support vector machine, random forest, k-nearest neighbor, kernel density estimation, Naïve-Bayes, 20-node, three-layer, fully-connected neural network	Variable
Bien et al. (78)	Detecting various knee abnormalities	CNN	Variable
Liu et al. (79)	Diagnosing anterior cruciate ligament tears	CNN	Sensitivity/ specificity: 96%/ 96% AUC: 0.98
Ma et al. (80)	Diagnosing meniscal injuries of the knee	CNN	Average accuracy: 89.8%
Chang et al. (81)	Detecting complete anterior cruciate ligament tears	CNN	Accuracy: 96%
Couteaux et al. (82)	Detecting meniscal tears of the knee	Region-based CNN	AUC: 0.906
Roblot et al. (83)	Detecting meniscal tears of the knee	CNNs	AUC: 0.90
Liu et al. (84)	Segmenting knee cartilage and bone	CNNs	Variable
Norman et al. (85)	Segmenting knee cartilage and menisci	CNNs	Variable dice scores, ranging between 0.770 and 0.878 for cartilage and 0.809 and 0.753 for menisci
Balsiger et al. (86)	Peripheral nerve segmentation	CNN	Dice scores of 0.859 and 0.719
Kemnitz et al. (87)	Thigh muscle and adipose tissue segmentation	U-Net CNN	Dice score: 0.96
Yin et al. (88)	Differentiating sacral chordoma from sacral giant cell tumor	Radiomics ML classifiers	Variable
Gitto et al. (89)	Classifying deep-seated lipomas and atypical lipomatous tumors of the extremities	Radiomics-based ML	Sensitivity/ specificity: 92%/ 33%
Pfeil et al. (90)	Joint-space analysis for rheumatoid arthritis detection	Computer-aided joint space analysis	Variable
Langs et al. (91)	Erosion spotting and visualization in rheumatoid arthritis	Generative appearance model	Sensitivity/ specificity: 85%/ 84% AUC: 0.92
Liu et al. (92)	Epidural mass detection	Gaussian Mixture Model	Accuracy: 82%
Stotter et al. (93)	Radiographic measurements of the pelvis and hip	CNN	Variable
Etli et al. (94)	Sex estimation from sacrum and coccyx	Univariate discriminant	Accuracy: 67.1% for univariate

(*Continued*)

TABLE 1 Continued

Publication	Application in musculoskeletal imaging	Algorithm type	System performance
		analysis, linear discriminant function analysis, stepwise discriminant function analysis, multilayer perceptron neural networks	discriminant analysis, 82.5% for linear discriminant function analysis, 78.8% for stepwise discriminant function analysis, and 86.3% for multilayer perceptron neural networks
Yune et al. (95)	Predicting sex from hand radiographs	CNN	Agreement with phenotypic sex: 77.8%
Bowness et al. (96)	Identifying anatomical structures on ultrasound	U-Net CNN	mean highlighting scores ranging from 7.87/10 to 8.69/10

AUC, area under the curve; BMD, bone mineral density; CLSTM, convolutional long short-term memory; CNN, convolutional neural network; CT, computed tomography; DL, deep learning; ML, machine learning; MRI, magnetic resonance imaging; VGG, visual geometry group.

of 0.929 but was less competent at detecting abnormalities of the shoulder, humerus, elbow, forearm, and hand (46). Since then, their large dataset was made publicly available under the name MURA to encourage public submissions and improve fracture detection rates of the original study (99). With all the collective efforts being made to improve AI-assisted fracture detection, models are no longer just objects of research studies but have been implemented into clinical practice. Presently, Gleamer BoneView (Gleamer, Paris, France) is an FDA-approved commercially available software that can help detect fractures on radiographs and is the only AI fracture detection software to have FDA clearance for use in both adults and pediatric patients over two years of age (100).

However, despite all these promising applications, AI-assisted fracture detection still has a key limitation: each CNN model must be specifically trained on the body part being assessed using large numbers of properly labeled images, whereas humans can transfer their knowledge of one body part to another. Moreover, models can be less reliable when trying to detect less obvious fractures such as a non-displaced femoral neck fracture (98), and most models report the output in a binary fashion (fracture present or not present) without providing an in depth description of the lesion or other related findings.

Osteoarthritis

Several studies have looked into how AI can assist radiologists in evaluating images for the presence and grading of osteoarthritis. Xue et al. fine-tuned a CNN model using a set of 420 hip radiographs to detect hip osteoarthritis using a binary system and reported a performance akin to that of a radiologist with ten years of experience (47). Tiulpin et al. took advantage of the

large publicly available Osteoarthritis Initiative (OAI) and Multicenter Osteoarthritis Study (MOST) datasets to train and test a CNN model to automatically score knee osteoarthritis severity using to the Kellgren-Lawrence grading scale and reported promising results with an AUC of 0.80 (48). Interestingly, the probability distribution of KL grades was also reported to show when predicted probabilities may be comparable across two contiguous grades, rendering the model's performance more illustrative of real-life practice where arthritis severity may represent the transition between two adjacent grades instead of being neatly tiered at one single level. This was also done in a study by Antony et al. where, in an attempt to circumvent the limitation of a finite and discreet scale, knee osteoarthritis grading was redefined as a regression with continuous variables (49). Although osteoarthritis assessment has been traditionally done using radiographs, AI can also augment the quantitative and qualitative assessment of cartilage on MRI to render the evaluation of osteoarthritis more accurate, and several studies have worked on developing models capable of successfully detecting cartilage lesions and staging cartilage degenerative changes (50, 51).

Bone age

Radiographic assessment of bone age is important for pediatricians to assess the skeletal maturity and growth of a child, and efforts have been put into using AI to automate bone age assessment and avoid the use of the inflexible and error-prone traditional methods such as the Greulich-Pyle atlas and the Tanner-Whitehouse method (101, 102). The Radiological Society of North America Pediatric Bone Age Machine Learning Challenge freely provided ML developers with a dataset containing over 14,000 hand radiographs and used competitions to promote collaborative effort into designing tools competent at automating bone age assessment (52). With over 100 submissions, the winning algorithm was designed by the University of Toronto's Cicero and Bilbily who used Google's Inception V3 network for pixel information, concatenated the architecture with sex information, and added layers after concatenation for data augmentation (52). The ultimate goal would be to provide radiologists with a tool that can help them assess bone age rather than perform the task independently. Tajmir et al. revealed how radiologists assisted by AI software when assessing bone age perform better than an unaided AI model, a single radiologist working independently, and a group of expert radiologists working together (53). Moreover, Kim et al. showed how the use of AI software can reduce reading times by approximately 30%, from 1.8 to 1.38 min per study (54). Presently, BoneXpert is a commercially-available widely-used software developed by Visiana that provides automated bone assessment by delineating the distal epiphyses of several hand bones, with at least eight needed for computation (55). Using the Greulich and Pyle or Tanner-Whitehouse standards, skeletal maturity is assessed with a precision of 0.17 years, reportedly nearly three times better than human performance (14, 56, 57).

Bone fragility

Imaging is often used for the evaluation of osteoporosis, a bone disorder characterized by a decreased bone mineral density (BMD), as bone strength assessment is fundamental for clinical decision making and therapy monitoring. Several studies have coupled ML support vector machines with methods of evaluating trabecular bone microarchitecture to automate and improve quantitative bone imaging and assessment (58, 59). In one study, Yang et al. used DL algorithms to combine BMD data from dual-energy x-ray absorptiometry (DXA) with bone microarchitecture data from multi-detector CT in an attempt to predict proximal femur failure loads; analysis revealed that trabecular bone characterization and ML methods, when coupled with conventional DXA BMD data, can appreciably enhance biomechanical strength prediction (58). Huber et al. applied similar ML methods to predict proximal tibial trabecular bone strength using MRI data instead and concluded that combining ML techniques with data on bone structure can enhance MRI assessment of bone quality (59).

In the same vein, ML algorithms have been employed in an attempt to predict osteoporotic fractures from MRI data (103), with one study making use of a CNN to automate segmentation of the proximal femur and facilitate the measurement of bone quality on MRI (60). Research has also focused on developing tools that can offer opportunistic screening and assessment of bone fragility, with one study looking at a system that can evaluate bone quality on dental panoramic radiographs (61) and another describing a DL system that can measure BMD on low-dose chest CT performed for lung cancer screening (62). With all those recent developments, AI is showing promise as a tool that can help with osteoporosis diagnosis; however, further refinement of such models is still needed to better automate the objective assessment of osteoporosis, its progression, and its response to therapy (104).

Spine imaging

Given that MSK radiologists spend a considerable amount of time looking at spine imaging, efforts have been made to develop ML algorithms that can automate tasks related to spine imaging interpretation and decrease the amount of time needed to interpret individual scans (105). Multiple studies have presented AI tools that can successfully detect and label spinal vertebrae as well as intervertebral discs on MRI and CT images (63–65), obviating the need for human manual labeling and streamlining the review of images. Information from these models can be used to automate other processes, as demonstrated by Jamaludin et al. who, after presenting a model that could label vertebral bodies and intervertebral discs on MRI with a 95.6% accuracy, used a CNN to successfully provide radiologist-level assessment of several other findings such as disc narrowing, central canal stenosis, spondylolisthesis, and end plate defects (66).

In addition to that, researchers have focused on designing models that can automate segmentation of the vertebrae, with one study making use of a U-Net architecture to segment the six lumbar intervertebral disc levels (67) and another adopting an iterative instance approach whereby information on one segmented vertebra is used to iteratively detect the following one

(64). In the former study, Lu et al. also trained their model to automate spinal and foraminal stenosis grading using a large dataset obtained from 4,075 patients and reported an accuracy of 80% for grading spinal stenosis and 78% for grading neural foraminal stenosis (67). To increase concordance between automated segmentation outputs and ground truth labels, generative adversarial networks have also been used, with one resultant model concurrently segmenting the neural foramen, the vertebral bodies, and intervertebral discs (68).

Advancements in this line of research are ongoing, supported by large publicly available datasets such as SpineWeb and the MICCAI 2018 Challenge on Automatic Intervertebral Disc Localization and Segmentation dataset. The Pulse platform (NuVasive, San Diego, California, USA) is a recent FDA-approved spinal surgical automation platform that combines multiple technologies to provide intraoperative assessment during spine surgeries and can help with tasks such as neuromonitoring of nerves, improvement of screw placement, and minimizing intraoperative radiation exposure (106). Other spine imaging applications could include automating radiographic measurements of spinal alignment (69, 70) and using CNNs to distinguish tuberculous from pyogenic spondylitis (71). However, despite the promising results of all these recent developments, further research is still needed, and studies are often hindered by several limitations such as the lack of a consistent gold standard for entities where radiologists may exhibit high variability in interpretation (107).

Muscuskeletal oncology

AI can potentially have several applications in MSK oncology and may be able to help radiologists detect metastatic bone lesions, determine their origin, and assess progression and treatment response. Using CT texture analysis, Acar et al. developed an ML model with an AUC reaching up to 0.76 when differentiating metastatic bone lesions from sclerotic bone lesions with complete response in patients with prostate cancer (72). To determine tumor origin on contrast-enhanced MRI, Lang et al. used DL methods and radiomics to devise a model that successfully differentiated between spinal metastatic lesions from the lung and other origin sites with a high accuracy reaching 0.81 (73). In addition, AI can potentially help with the assessment of primary musculoskeletal tumors. For example, two studies making use of ML techniques and radiomics demonstrated how lipoma and liposarcoma could be differentiated on MRI with expert-level performance (74) and how the histopathological grades of soft tissue sarcomas can be pre-operatively and non-invasively predicted on fat-suppressed T2-weighted imaging with an accuracy reaching 0.88 (75). AI might also serve other proposed roles, such as assisting clinicians in predicting tumor recurrence as well monitoring post-treatment tumor changes on imaging (76, 77).

Cruciate ligaments and menisci

Several studies have evaluated the performance of AI models when detecting meniscal injuries and ligamentous tears of the knee. Bien et al. trained a CNN model using a set of 1,130 training and 120 validation MRI exams to recognize meniscal and anterior cruciate ligament (ACL) tears, reporting an AUC of

0.847 for meniscal tears and an AUC of 0.937 for ACL tears on the internal validation set and 0.824 on the external validation set (78). Using arthroscopy as the reference, Liu et al. trained a CNN to isolate ACL lacerations with an AUC of 0.98 and a sensitivity of 96% (51, 79), and Ma et al. trained a CNN to diagnose meniscal injury, reporting a an accuracy of 85.6% for anterior horn injury detection and 92% for posterior horn injury detection, a performance comparable to a chief physician (80).

Isolation of individual joint structures might help enhance model performance, as demonstrated by Chang et al. who, after isolating the ACL on coronal proton density 2D MRI using CNN U-Net, subsequently used a CNN classifier to evaluate the isolated ACLs for the presence of pathology and reported an AUC of 0.97 and a sensitivity of 100% (81). When testing a CNN model for meniscal segmentation on fat-suppressed MRI sequences, Pedoia et al. reported a sensitivity reaching 90%, a specificity reaching 82%, and an AUC reaching 0.89 (50). Likewise, Couteaux et al., Roblot et al., and Lassau et al. all reported similar performances, with AUC values for meniscal tear detection reaching 0.9 in all three studies (82, 83, 108).

Quantitative analysis: segmentation and radiomics

Segmentation, or the process of delineating anatomic structures, can be time-consuming but is nevertheless important for evaluating the potential degeneration of or damage to segmented structures and the resultant decline in their functionality. Semi-automated segmentation software are currently being applied in clinical cardiac and prostate MRI, but such software make use of algorithms with manually designed hand-engineered features and thus require manual adjustments to the computer-generated contours (26). As such, interest has shifted to fully automating segmentation processes using CNN, which can have a profound impact on a radiologist's functionality and efficiency in the reading room. Performance of segmentation algorithms is often assessed with a dice coefficient to assess the similarity of a segmentation to its ground truth by reporting the percentage overlap between the two regions, and a dice score of 0.95 is usually indicative of a successful algorithm (109). Recent research has heavily focused on knee segmentation, with Liu et al. designing a model that successfully segmented the different structures of the knee using a CNN combined with a 3D deformable modeling approach (84). Using both T1-rho weighted and 3D double-echo steady-state images, Norman et al. also evaluated a DL model for automated segmentation of knee cartilage and menisci but with simultaneous evaluation of cartilage relaxometry and morphology; they found the model to be adept at generating accurate segmentations and morphologic characterizations when compared to manual segmentations (85). DL techniques can have applications outside the knee as well, as demonstrated by Deniz et al. who used similar methods but shifted attention to the segmentation of the proximal femur, reporting a CNN algorithm with a dice similarity score reaching 0.95 (60). Other venues are also being explored, with AI tools showing promise in neurography segmentation (86) as well as muscle segmentation in osteoarthritis patients to help with muscular trophism evaluation (87).

Besides segmentation, AI may also have applications in radiomics, which is an emerging field in medicine that treats medical images not only as pictures intended solely for visual interpretation but also as a source of diverse quantitative characteristics extracted as mineable data that can be used for pattern identification to eventually assist with decision support, characterization, and prediction of disease processes (110). Spatial distribution of signal intensities and information on pixel interrelationships are mathematically extracted to provide and quantify textural information, which in turn can be used for quantitative imaging biomarker discovery and validation for a number of different conditions such as acute and chronic injuries, spinal abnormalities, and neoplasms (111). By uncovering imperceptible patterns in medical imaging, radiomics-bases predictive models can play different roles such as providing a detailed description of disease burden, identifying relationships between phenotypes and outcomes, and predicting diagnosis and prognosis for certain diseases, ultimately playing a key role in improving precision medicine and personalized patient management (112). ML models can identify and gather imaging characteristics such as the distribution of signal intensities and the spatial relationship of pixels that are not easily discernible with visual interpretation and that can help improve clinical care (113, 114). When testing different ML-augmented radiomics models for preoperative differentiation of sacral chordomas from sacral giant cell tumors on 3D CT; Yin et al. found contrast-enhanced CT features more optimal than non-enhanced features for helping identify the histology of the sacral tumor in question (88). In one retrospective study, Gitto et al. assessed the diagnostic performance of ML-enhanced radiomics-based MRI for the classification and differentiation of atypical lipomatous tumors of the extremities from other benign lipomas, reporting a sensitivity of 92%, a specificity of 33%, and no statistically significant difference when compared to qualitative image assessment performed by a radiologist with 7 years of experience (89). Research into the field is ongoing, and although radiomics has shown promise as a powerful and innovative tool that can help with the evaluation of different types of cancers, more research is needed to fully explore the full scope of its applications (115).

Other miscellaneous applications

Several research studies have looked into other potential applications of AI such as joint space evaluation in rheumatoid arthritis (90, 91), epidural mass detection on CT scans (92), rotator cuff pathology detection (116), femoroacetabular impingement and hip dysplasia detection (93), sex determination using CT imaging of the sacrum and coccyx (94) or hand radiographs (95), and assessment of Achilles tendon healing (117). In addition to that, AI applications can have multiple applications in MSK ultrasound (US), including but not limited to segmentation of US images (96), quantitative analysis of skeletal muscles (118), and detection of pediatric conditions such as wrist fractures and developmental dysplasia of the hip (4, 119). Research is still ongoing, and additional repetitive and time-consuming tasks might be tackled in coming years in an

attempt to automate more processes and thus accelerate the process of imaging interpretation.

Results reporting

AI can have several applications that can revolutionize the production of radiology reports and the communication of findings between physicians. Speech recognition, which has already transformed the writing of reports, could be further optimized with DL methods (120). Language processing systems can also be applied, as shown by Do et al. who presented a system capable of recognizing anatomy data from reports generated with speech recognition software to concurrently extract information on possible fractures (121) and Tan et al. who presented a system capable of scanning x-ray and MRI radiology reports to identify lumbar spine imaging findings that could be related to low back pain (122). Natural language processing (NLP) refers to the use of a computer to analyze and interpret human language. Although NLP systems are not entirely novel, recent advances in ML and neural networks have revolutionized this technology, subsequently turning it into a tool that can help with data extraction from radiology reports (123). At their core, NLP systems operate using a multistep approach, beginning with a preprocessing step whereby reports are broken down into different subsets and processing steps during which text from specific sections or differently-weighted sections is split into sentences and words (a process known as tokenization) (124). Word normalization and syntactic analysis follow, whereby spelling mistakes are fixed, medical abbreviations are fully expanded, and word roots are identified with the goal of determining grammatical structures and linking words to semantic concepts (such as symptom or disease), thus assigning meaning to the data (124). The textual features extracted are then processed by an automatic classifier using ML applications to solve the ultimate task assigned to the system (such as information extraction from reports), and ML applications have to be trained on a set of manually-annotated reports, which can be split into a training set and a validation set, both of which are needed to develop the system and assess its performance (124).

Such tools could play a number of roles, such as suggesting management recommendations to radiologists during the dictation of a report or assisting with research purposes by establishing links between different radiological findings and resultant symptomatology or prognosis. Additionally, ML applications may extend to extracting follow-up recommendations from reports, thus ensuring the adequate management of reported key findings (125).

Limitations

Although AI shows several promising applications across the entire MSK imaging cycle, this technology is still facing a number of challenges and limitations when it comes to both development of AI tools and implementation into clinical practice. Large datasets are needed to develop successful DL tools: tasks or diagnoses for which such datasets are not available might be challenging to automate, and data can be fragmented across many different systems, thus increasing the risk of errors, decreasing the comprehensiveness of datasets, and increasing the expenses of gathering complete data. Moreover, challenges in establishing reference standards, such as irregularities in contouring lesions, diagnostic uncertainties, as well as inconsistencies in human performance and labeling, can all reduce performance and hinder development. DL models being developed are usually trained to perform one single task, whereas patients seen clinically might have a number of etiologies and conditions that require complex simultaneous interpretations.

Given that large amounts of data need to be collected for the development of successful algorithms, issues pertaining to privacy and ownership of such data arise: patients may be concerned that collection of such data is a violation of privacy, especially if an AI model can predict private information about a patient without having received that information and subsequently make it available to third parties (such as life insurance companies). Large datasets can be problematic in a different way: they may be more representative of a specific subset of the population rather than the whole population and could also reflect underlying biases and inequalities in the health system. As such, algorithms trained using such datasets may propagate systemic biases and inequalities that are already present and may not be suitable for treating all patients but rather the subset with the most representation in the training dataset.

Evidently, AI models can and will make mistakes, resulting in errors and injuries to patients being treated using the model. Although medical errors are sometimes inevitable in the medical field and can occur irrespective of the use of AI, the danger of AI-related mistakes is that an underlying problem in one system might result in injuries to thousands of patients if that system becomes widespread (whereas errors from a single human provider will affect the limited number of patients being treated by that provider). Additionally, with errors arises the issue of accountability: models often do not disclose the statistical rationale behind the elaboration of their tasks, making it hard to identify the cause of the error or understand the rationale behind the final output of an algorithm and limiting implementation into medical settings. To catch errors and refine algorithms, post-implementation evaluation, maintenance, and performance monitoring of implemented AI tools is just as vital as pre-implementation development processes to the success of a model. However, such monitoring can prove to be labor-intensive, especially for smaller practices that will inevitably experience workflow disruptions due to a lack of dedicated informatics resources and an increase in the radiologists' burden (126).

Conclusion

AI, ML and DL have the potential to significantly augment several aspects of the MSK imaging chain, with applications in

the ordering of imaging, scheduling, protocoling, acquisition and presentation, image interpretation, as well as report generation and communication of findings. Although research into this technology is showing very promising results, development of tools still faces a number of challenges that impede successful implementation into clinical practice. The ultimate goal is not to design a completely independent system that replaces the need for human expertise but rather to equip radiologists and medical professionals with tools that can automate certain functions and thus alleviate some of the increasing responsibilities radiologists face, affording them more time to focus on more demanding and complex tasks. Radiologists and AI algorithms working hand in hand have the potential to increase the value provided to patients by improving imaging quality and efficiency, patient centricity, and diagnostic accuracy, all of which can greatly enhance both patient and provider satisfaction.

Author contributions

PD has made significant contributions to the conception and design of the work and drafting and revision. He provides approval for the publication of content and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. LF has made significant contributions to the conception and design of the work and drafting and critical revision for important intellectual content. She provides approval for the publication of content and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Farrell TP, Adams NC, Walsh JP, Hynes J, Eustace SK, Kavanagh E. Musculoskeletal imaging: current practice and future directions. *Semin Musculoskelet Radiol*. (2018) 22(5):564–81. doi: 10.1055/s-0038-1672193

2. Harkey P, Duszak R, Gyftopoulos S, Rosenkrantz AB. Who refers musculoskeletal extremity imaging examinations to radiologists? *Am J Roentgenol*. (2018) 210 (4):834–41. doi: 10.2214/AJR.17.18591

3. Doshi AM, Moore WH, Kim DC, Rosenkrantz AB, Fefferman NR, Ostrow DL, et al. Informatics solutions for driving an effective and efficient radiology practice. *Radiographics*. (2018) 38(6):1810–22. doi: 10.1148/rg.2018180037

4. D'Angelo T, Caudo D, Blandino A, Albrecht MH, Vogl TJ, Gruenewald LD, et al. Artificial intelligence, machine learning and deep learning in musculoskeletal imaging: current applications. *J Clin Ultrasound*. (2022) 50(9):1414–31. doi: 10.1002/jcu.23321

5. Do S, Song KD, Chung JW. Basics of deep learning: a radiologist's guide to understanding published radiology articles on deep learning. *Korean J Radiol*. (2020) 21(1):33–41. doi: 10.3348/kjr.2019.0312

6. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current applications and future impact of machine learning in radiology. *Radiology*. (2018) 288(2):318–28. doi: 10.1148/radiol.2018171820

7. Soffer S, Ben-Cohen A, Shimon O, Amitai MM, Greenspan H, Klang E. Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology*. (2019) 290(3):590–606. doi: 10.1148/radiol.2018180547

8. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ, et al. Deep learning: a primer for radiologists. *RadioGraphics*. (2017) 37(7):2113–31. doi: 10.1148/rg.2017170077

9. Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. *Radiology*. (2020) 295 (1):4–15. doi: 10.1148/radiol.2020192224

10. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. *Z für Med Phys*. (2019) 29(2):102–27. doi: 10.1016/j.zemedi.2018.11.002

11. Richardson ML, Garwood ER, Lee Y, Li MD, Lo HS, Nagaraju A, et al. Noninterpretive uses of artificial intelligence in radiology. *Acad Radiol*. (2021) 28 (9):1225–35. doi: 10.1016/j.acra.2020.01.012

12. Saw SN, Ng KH. Current challenges of implementing artificial intelligence in medical imaging. *Phys Med*. (2022) 100:12–7. doi: 10.1016/j.ejmp.2022.06.003

13. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol*. (2019) 213(3):506–13. doi: 10.2214/AJR.19.21117

14. Gorelik N, Gyftopoulos S. Applications of artificial intelligence in musculoskeletal imaging: from the request to the report. *Can Assoc Radiol J*. (2021) 72(1):45–59. doi: 10.1177/0846537120947148

15. Lee YH. Efficiency improvement in a busy radiology practice: determination of musculoskeletal magnetic resonance imaging protocol using deep-learning convolutional neural networks. *J Digit Imaging*. (2018) 31(5):604–10. doi: 10.1007/s10278-018-0066-y

16. Trivedi H, Mesterhazy J, Laguna B, Vu T, Sohn JH. Automatic determination of the need for intravenous contrast in musculoskeletal MRI examinations using IBM watson's natural language processing algorithm. *J Digit Imaging*. (2018) 31 (2):245–51. doi: 10.1007/s10278-017-0021-3

17. Mieloszyk RJ, Rosenbaum JI, Hall CS, Raghavan UN, Bhargava P. The financial burden of missed appointments: uncaptured revenue due to outpatient no-shows in radiology. *Curr Probl Diagn Radiol*. (2018) 47(5):285–6. doi: 10.1067/j.cpradiol.2018.06.001

18. Kurasawa H, Hayashi K, Fujino A, Takasugi K, Haga T, Waki K, et al. Machine-learning-based prediction of a missed scheduled clinical appointment by patients with diabetes. *J Diabetes Sci Technol*. (2016) 10(3):730–6. doi: 10.1177/1932296815614866

19. Torres O, Rothberg MB, Garb J, Ogunneye O, Onyema J, Higgins T. Risk factor model to predict a missed clinic appointment in an urban, academic, and underserved setting. *Popul Health Manag*. (2015) 18(2):131–6. doi: 10.1089/pop.2014.0047

20. Nelson A, Herron D, Rees G, Nachev P. Predicting scheduled hospital attendance with artificial intelligence. *NPJ Digit Med*. (2019) 2(1):1–7. doi: 10.1038/s41746-019-0103-3

21. Harvey HB, Liu C, Ai J, Jaworsky C, Guerrier CE, Flores E, et al. Predicting no-shows in radiology using regression modeling of data available in the electronic medical record. *J Am Coll Radiol*. (2017) 14(10):1303–9. doi: 10.1016/j.jacr.2017.05.007

22. Chong LR, Tsai KT, Lee LL, Foo SG, Chang PC. Artificial intelligence predictive analytics in the management of outpatient MRI appointment no-shows. *AJR Am J Roentgenol*. (2020) 215(5):1155–62. doi: 10.2214/AJR.19.22594

23. (ISMRM 2017) Using machine learning with dynamic exam block lengths to decrease patient wait time and optimize MRI schedule fill rate. Available at: https://archive.ismrm.org/2017/4343.html (Cited June 14, 2023).

24. Jaspan ON, Fleysher R, Lipton ML. Compressed sensing MRI: a review of the clinical literature. *Br J Radiol*. (2015) 88(1056):20150487. doi: 10.1259/bjr.20150487

25. Glockner JF, Hu HH, Stanley DW, Angelos L, King K. Parallel MR imaging: a user's guide. *Radiographics*. (2005) 25(5):1279–97. doi: 10.1148/rg.255045202

26. Chea P, Mandell JC. Current applications and future directions of deep learning in musculoskeletal radiology. *Skeletal Radiol*. (2020) 49(2):183–97. doi: 10.1007/s00256-019-03284-z

27. Wang S, Su Z, Ying L, Peng X, Zhu S, Liang F, et al. Accelerating magnetic resonance imaging via deep learning. *Proc IEEE Int Symp Biomed Imaging*. (2016) 2016:514–7. doi: 10.1109/ISBI.2016.7493320

28. Hammernik K, Klatzer T, Kobler E, Recht MP, Sodickson DK, Pock T, et al. Learning a variational network for reconstruction of accelerated MRI data. *Magn Reson Med*. (2018) 79(6):3055–71. doi: 10.1002/mrm.26977

29. Chaudhari AS, Fang Z, Kogan F, Wood J, Stevens KJ, Gibbons EK, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med*. (2018) 80(5):2139–54. doi: 10.1002/mrm.27178

30. Wu Y, Ma Y, Capaldi DP, Liu J, Zhao W, Du J, et al. Incorporating prior knowledge via volumetric deep residual network to optimize the reconstruction of sparsely sampled MRI. *Magn Reson Imaging*. (2020) 66:93–103. doi: 10.1016/j.mri.2019.03.012

31. Roh S, Park JI, Kim GY, Yoo HJ, Nickel D, Koerzdoerfer G, et al. Feasibility and clinical usefulness of deep learning-accelerated MRI for acute painful fracture patients wearing a splint: a prospective comparative study. *PLoS One*. (2023) 18(6):e0287903. doi: 10.1371/journal.pone.0287903

32. Lin DJ, Walter SS, Fritz J. Artificial intelligence–driven ultra-fast superresolution MRI : 10-fold accelerated musculoskeletal turbo spin Echo MRI within reach. *Invest Radiol*. (2023) 58(1):28–42. doi: 10.1097/RLI.0000000000000928

33. Lee JH, Han IH, Kim DH, Yu S, Lee IS, Song YS, et al. Spine computed tomography to magnetic resonance image synthesis using generative adversarial networks : a preliminary study. *J Korean Neurosurg Soc*. (2020) 63(3):386–96. doi: 10.3340/jkns.2019.0084

34. Galbusera F, Bassani T, Casaroli G, Gitto S, Zanchetta E, Costa F, et al. Generative models: an upcoming innovation in musculoskeletal radiology? A preliminary test in spine imaging. *Eur Radiol Exp*. (2018) 2:29. doi: 10.1186/s41747-018-0060-7

35. Fayad LM, Parekh VS, de Castro Luna R, Ko CC, Tank D, Fritz J, et al. A deep learning system for synthetic knee magnetic resonance imaging: is artificial intelligence-based fat-suppressed imaging feasible? *Invest Radiol*. (2021) 56(6):357–68. doi: 10.1097/RLI.0000000000000751

36. Kambadakone A. Artificial intelligence and CT image reconstruction: potential of a new era in radiation dose reduction. *J Am Coll Radiol*. (2020) 17(5):649–51. doi: 10.1016/j.jacr.2019.12.025

37. Cross NM, Kemp J, Morey J. SIIM 2017 Scientific Session Posters & Demonstrations.

38. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Trans Med Imaging*. (2018) 37(6):1370–81. doi: 10.1109/TMI.2018.2823083

39. Kitamura G. Hanging protocol optimization of lumbar spine radiographs with machine learning. *Skeletal Radiol*. (2021) 50(9):1809–19. doi: 10.1007/s00256-021-03733-8

40. Intelligent Tools for a Productive Radiologist Workflow, Universal Viewer Smart Reading Protocols. Available at: https://healthimaging.com/sponsored/1155/ge-healthcare/intelligent-tools-productive-radiologist-workflow-universal-viewer (Cited June 14, 2023).

41. Chung SW, Han SS, Lee JW, Oh KS, Kim NR, Yoon JP, et al. Automated detection and classification of the proximal humerus fracture by using deep learning algorithm. *Acta Orthop*. (2018) 89(4):468–73. doi: 10.1080/17453674.2018.1453714

42. Olczak J, Fahlberg N, Maki A, Razavian AS, Jilert A, Stark A, et al. Artificial intelligence for analyzing orthopedic trauma radiographs. *Acta Orthop*. (2017) 88(6):581–6. doi: 10.1080/17453674.2017.1344459

43. Yu JS, Yu SM, Erdal BS, Demirer M, Gupta V, Bigelow M, et al. Detection and localisation of hip fractures on anteroposterior radiographs with artificial intelligence: proof of concept. *Clin Radiol*. (2020) 75(3):237.e1–e9. doi: 10.1016/j.crad.2019.10.022

44. Tomita N, Cheung YY, Hassanpour S. Deep neural networks for automatic detection of osteoporotic vertebral fractures on CT scans. *Comput Biol Med*. (2018) 98:8–15. doi: 10.1016/j.compbiomed.2018.05.011

45. Cheng CT, Ho TY, Lee TY, Chang CC, Chou CC, Chen CC, et al. Application of a deep learning algorithm for detection and visualization of hip fractures on plain pelvic radiographs. *Eur Radiol*. (2019) 29(10):5469–77. doi: 10.1007/s00330-019-06167-y

46. Rajpurkar P, Irvin J, Bagul A, Ding D, Duan T, Mehta H, et al. MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. arXiv (2018). Available at: http://arxiv.org/abs/1712.06957 (Cited June 15, 2023).

47. Xue Y, Zhang R, Deng Y, Chen K, Jiang T. A preliminary examination of the diagnostic value of deep learning in hip osteoarthritis. *PLoS One*. (2017) 12(6):e0178992. doi: 10.1371/journal.pone.0178992

48. Tiulpin A, Thevenot J, Rahtu E, Lehenkari P, Saarakkala S. Automatic knee osteoarthritis diagnosis from plain radiographs: a deep learning-based approach. *Sci Rep*. (2018) 8(1):1727. doi: 10.1038/s41598-018-20132-7

49. Antony J, McGuinness K, Connor NEO, Moran K. Quantifying Radiographic Knee Osteoarthritis Severity using Deep Convolutional Neural Networks. arXiv (2016) Available at: http://arxiv.org/abs/1609.02469 (Cited June 15, 2023).

50. Pedoia V, Norman B, Mehany SN, Bucknor MD, Link TM, Majumdar S. 3D convolutional neural networks for detection and severity staging of meniscus and PFJ cartilage morphological degenerative changes in osteoarthritis and anterior cruciate ligament subjects. *J Magn Reson Imaging*. (2019) 49(2):400–10. doi: 10.1002/jmri.26246

51. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep learning approach for evaluating knee MR images: achieving high diagnostic performance for cartilage lesion detection. *Radiology*. (2018) 289(1):160–9. doi: 10.1148/radiol.2018172986

52. Halabi SS, Prevedello LM, Kalpathy-Cramer J, Mamonov AB, Bilbily A, Cicero M, et al. The RSNA pediatric bone age machine learning challenge. *Radiology*. (2019) 290(2):498–503. doi: 10.1148/radiol.2018180736

53. Tajmir SH, Lee H, Shailam R, Gale HI, Nguyen JC, Westra SJ, et al. Artificial intelligence-assisted interpretation of bone age radiographs improves accuracy and decreases variability. *Skeletal Radiol*. (2019) 48(2):275–83. doi: 10.1007/s00256-018-3033-2

54. Kim JR, Shim WH, Yoon HM, Hong SH, Lee JS, Cho YA, et al. Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency. *AJR Am J Roentgenol*. (2017) 209(6):1374–80. doi: 10.2214/AJR.17.18224

55. Thodberg HH, Kreiborg S, Juul A, Pedersen KD. The BoneXpert method for automated determination of skeletal maturity. *IEEE Trans Med Imaging*. (2009) 28(1):52–66. doi: 10.1109/TMI.2008.926067

56. Martin DD, Calder AD, Ranke MB, Binder G, Thodberg HH. Accuracy and self-validation of automated bone age determination. *Sci Rep*. (2022) 12(1):6388. doi: 10.1038/s41598-022-10292-y

57. Maratova K, Zemkova D, Sedlak P, Pavlikova M, Amaratunga SA, Krasnicanova H, et al. A comprehensive validation study of the latest version of BoneXpert on a large cohort of Caucasian children and adolescents. *Front Endocrinol (Lausanne)*. (2023) 14. Available at: https://www.frontiersin.org/articles/10.3389/fendo.2023.1130580 doi: 10.3389/fendo.2023.1130580

58. Yang CC, Nagarajan MB, Huber MB, Carballido-Gamio J, Bauer JS, Baum T, et al. Improving bone strength prediction in human proximal femur specimens through geometrical characterization of trabecular bone microarchitecture and support vector regression. *J Electron Imaging*. (2014) 23(1):013013. doi: 10.1117/1.JEI.23.1.013013

59. Huber MB, Lancianese SL, Nagarajan MB, Ikpot IZ, Lerner AL, Wismuller A. Prediction of biomechanical properties of trabecular bone in MR images with geometric features and support vector regression. *IEEE Trans Biomed Eng*. (2011) 58(6):1820–6. doi: 10.1109/TBME.2011.2119484

60. Deniz CM, Xiang S, Hallyburton RS, Welbeck A, Babb JS, Honig S, et al. Segmentation of the proximal femur from MR images using deep convolutional neural networks. *Sci Rep*. (2018) 8(1):16485. doi: 10.1038/s41598-018-34817-6

61. Lee JS, Adhikari S, Liu L, Jeong HG, Kim H, Yoon SJ. Osteoporosis detection in panoramic radiographs using a deep convolutional neural network-based computer-assisted diagnosis system: a preliminary study. *Dentomaxillofac Radiol*. (2019) 48(1):20170344. doi: 10.1259/dmfr.20170344

62. Pan Y, Shi D, Wang H, Chen T, Cui D, Cheng X, et al. Automatic opportunistic osteoporosis screening using low-dose chest computed tomography scans obtained for lung cancer screening. *Eur Radiol*. (2020) 30(7):4107–16. doi: 10.1007/s00330-020-06679-y

63. Jimenez-Pastor A, Alberich-Bayarri A, Fos-Guarinos B, Garcia-Castro F, Garcia-Juan D, Glocker B, et al. Automated vertebrae localization and identification by decision forests and image-based refinement on real-world CT data. *Radiol Med*. (2020) 125(1):48–56. doi: 10.1007/s11547-019-01079-9

64. Lessmann N, van Ginneken B, de Jong PA, Išgum I. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Med Image Anal*. (2019) 53:142–55. doi: 10.1016/j.media.2019.02.005

65. Wimmer M, Major D, Novikov AA, Bühler K. Fully automatic cross-modality localization and labeling of vertebral bodies and intervertebral discs in 3D spinal images. *Int J Comput Assist Radiol Surg*. (2018) 13(10):1591–603. doi: 10.1007/s11548-018-1818-3

66. Jamaludin A, Lootus M, Kadir T, Zisserman A, Urban J, Battié MC, et al. ISSLS PRIZE IN BIOENGINEERING SCIENCE 2017: automation of reading of radiological features from magnetic resonance images (MRIs) of the lumbar spine without human intervention is comparable with an expert radiologist. *Eur Spine J*. (2017) 26(5):1374–83. doi: 10.1007/s00586-017-4956-3

67. Lu JT, Pedemonte S, Bizzo B, Doyle S, Andriole KP, Michalski MH, et al. DeepSPINE: Automated Lumbar Vertebral Segmentation, Disc-level Designation, and Spinal Stenosis Grading Using Deep Learning. arXiv (2018). Available at: http://arxiv.org/abs/1807.10215 (Cited June 17, 2023).

68. Han Z, Wei B, Mercado A, Leung S, Li S. Spine-GAN: semantic segmentation of multiple spinal structures. *Med Image Anal.* (2018) 50:23–35. doi: 10.1016/j.media. 2018.08.005

69. Pan Y, Chen Q, Chen T, Wang H, Zhu X, Fang Z, et al. Evaluation of a computer-aided method for measuring the cobb angle on chest x-rays. *Eur Spine J.* (2019) 28(12):3035–43. doi: 10.1007/s00586-019-06115-w

70. Weng CH, Wang CL, Huang YJ, Yeh YC, Fu CJ, Yeh CY, et al. Artificial intelligence for automatic measurement of sagittal vertical axis using ResUNet framework. *J Clin Med.* (2019) 8(11):1826. doi: 10.3390/jcm8111826

71. Kim K, Kim S, Lee YH, Lee SH, Lee HS, Kim S. Performance of the deep convolutional neural network based magnetic resonance image scoring algorithm for differentiating between tuberculous and pyogenic spondylitis. *Sci Rep.* (2018) 8 (1):13124. doi: 10.1038/s41598-018-31486-3

72. Acar E, Leblebici A, Ellidokuz BE, Başbınar Y, Kaya GÇ. Machine learning for differentiating metastatic and completely responded sclerotic bone lesion in prostate cancer: a retrospective radiomics study. *Br J Radiol.* (2019) 92(1101):20190286. doi: 10.1259/bjr.20190286

73. Lang N, Zhang Y, Zhang E, Zhang J, Chow D, Chang P, et al. Differentiation of spinal metastases originated from lung and other cancers using radiomics and deep learning based on DCE-MRI. *Magn Reson Imaging.* (2019) 64:4–12. doi: 10.1016/j. mri.2019.02.013

74. Malinauskaite I, Hofmeister J, Burgermeister S, Neroladaki A, Hamard M, Montet X, et al. Radiomics and machine learning differentiate soft-tissue lipoma and liposarcoma better than musculoskeletal radiologists. *Sarcoma.* (2020) 2020: e7163453. doi: 10.1155/2020/7163453

75. Zhang Y, Zhu Y, Shi X, Tao J, Cui J, Dai Y, et al. Soft tissue sarcomas: preoperative predictive histopathological grading based on radiomics of MRI. *Acad Radiol.* (2019) 26(9):1262–8. doi: 10.1016/j.acra.2018.09.025

76. He Y, Guo J, Ding X, van Ooijen PMA, Zhang Y, Chen A, et al. Convolutional neural network to predict the local recurrence of giant cell tumor of bone after curettage based on pre-surgery magnetic resonance images. *Eur Radiol.* (2019) 29 (10):5441–51. doi: 10.1007/s00330-019-06082-2

77. Blackledge MD, Winfield JM, Miah A, Strauss D, Thway K, Morgan VA, et al. Supervised machine-learning enables segmentation and evaluation of heterogeneous post-treatment changes in multi-parametric MRI of soft-tissue sarcoma. *Front Oncol.* (2019) 9:941. doi: 10.3389/fonc.2019.00941

78. Bien N, Rajpurkar P, Ball RL, Irvin J, Park A, Jones E, et al. Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet. *PLoS Med.* (2018) 15(11):e1002699. doi: 10.1371/journal.pmed.1002699

79. Liu F, Guan B, Zhou Z, Samsonov A, Rosas H, Lian K, et al. Fully automated diagnosis of anterior cruciate ligament tears on knee MR images by using deep learning. *Radiology: Artificial Intelligence.* (2019) 1(3):180091. doi: 10.1148/ryai. 2019180091

80. Ma Y, Qin Y, Liang C, Li X, Li M, Wang R, et al. Visual cascaded-progressive convolutional neural network (C-PCNN) for diagnosis of Meniscus injury. *Diagnostics.* (2023) 13(12):2049. doi: 10.3390/diagnostics13122049

81. Chang PD, Wong TT, Rasiej MJ. Deep learning for detection of complete anterior cruciate ligament tear. *J Digit Imaging.* (2019) 32(6):980–6. doi: 10.1007/s10278-019-00193-4

82. Couteaux V, Si-Mohamed S, Nempont O, Lefevre T, Popoff A, Pizaine G, et al. Automatic knee meniscus tear detection and orientation classification with mask-RCNN. *Diagn Interv Imaging.* (2019) 100(4):235–42. doi: 10.1016/j.diii.2019.03.002

83. Roblot V, Giret Y, Bou Antoun M, Morillot C, Chassin X, Cotten A, et al. Artificial intelligence to diagnose meniscus tears on MRI. *Diagn Interv Imaging.* (2019) 100(4):243–9. doi: 10.1016/j.diii.2019.02.007

84. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn Reson Med.* (2018) 79 (4):2379–91. doi: 10.1002/mrm.26841

85. Norman B, Pedoia V, Majumdar S. Use of 2D U-net convolutional neural networks for automated cartilage and Meniscus segmentation of knee MR imaging data to determine relaxometry and morphometry. *Radiology.* (2018) 288(1):177–85. doi: 10.1148/radiol.2018172322

86. Balsiger F, Steindel C, Arn M, Wagner B, Grunder L, El-Koussy M, et al. Segmentation of peripheral nerves from magnetic resonance neurography: a fully-automatic, deep learning-based approach. *Front Neurol.* (2018) 9:777. doi: 10.3389/fneur.2018.00777

87. Kemnitz J, Baumgartner CF, Eckstein F, Chaudhari A, Ruhdorfer A, Wirth W, et al. Clinical evaluation of fully automated thigh muscle and adipose tissue segmentation using a U-net deep learning architecture in context of osteoarthritic knee pain. *MAGMA.* (2020) 33(4):483–93. doi: 10.1007/s10334-019-00816-5

88. Yin P, Mao N, Zhao C, Wu J, Sun C, Chen L, et al. Comparison of radiomics machine-learning classifiers and feature selection for differentiation of sacral chordoma and sacral giant cell tumour based on 3D computed tomography features. *Eur Radiol.* (2019) 29(4):1841–7. doi: 10.1007/s00330-018-5730-6

89. Gitto S, Interlenghi M, Cuocolo R, Salvatore C, Giannetta V, Badalyan J, et al. MRI radiomics-based machine learning for classification of deep-seated lipoma and atypical lipomatous tumor of the extremities. *Radiol Med.* (2023) 128(8):989–98. doi: 10.1007/s11547-023-01657-y

90. Pfeil A, Renz DM, Hansch A, Kainberger F, Lehmann G, Malich A, et al. The usefulness of computer-aided joint space analysis in the assessment of rheumatoid arthritis. *Joint Bone Spine.* (2013) 80(4):380–5. doi: 10.1016/j.jbspin.2012.10.022

91. Langs G, Peloschek P, Bischof H, Kainberger F. Model-based erosion spotting and visualization in rheumatoid arthritis. *Acad Radiol.* (2007) 14(10):1179–88. doi: 10.1016/j.acra.2007.06.013

92. Liu J, Pattanaik S, Yao J, Turkbey E, Zhang W, Zhang X, et al. Computer aided detection of epidural masses on computed tomography scans. *Comput Med Imaging Graph.* (2014) 38(7):606–12. doi: 10.1016/j.compmedimag.2014.04.007

93. Stotter C, Klestil T, Röder C, Reuter P, Chen K, Emprechtinger R, et al. Deep learning for fully automated radiographic measurements of the pelvis and hip. *Diagnostics.* (2023) 13(3):497. doi: 10.3390/diagnostics13030497

94. Etli Y, Asirdizer M, Hekimoglu Y, Keskin S, Yavuz A. Sex estimation from sacrum and coccyx with discriminant analyses and neural networks in an equally distributed population by age and sex. *Forensic Sci Int.* (2019) 303:109955. doi: 10. 1016/j.forsciint.2019.109955

95. Yune S, Lee H, Kim M, Tajmir SH, Gee MS, Do S. Beyond human perception: sexual dimorphism in hand and wrist radiographs is discernible by a deep learning model. *J Digit Imaging.* (2019) 32(4):665–71. doi: 10.1007/s10278-018-0148-x

96. Bowness J, Varsou O, Turbitt L, Burkett-St Laurent D. Identifying anatomical structures on ultrasound: assistive artificial intelligence in ultrasound-guided regional anesthesia. *Clinical Anatomy.* (2021) 34(5):802–9. doi: 10.1002/ca.23742

97. Gundry M, Knapp K, Meertens R, Meakin JR. Computer-aided detection in musculoskeletal projection radiography: a systematic review. *Radiography (Lond).* (2018) 24(2):165–74. doi: 10.1016/j.radi.2017.11.002

98. Langerhuizen DWG, Janssen SJ, Mallee WH, van den Bekerom MPJ, Ring D, Kerkhoffs GMMJ, et al. What are the applications and limitations of artificial intelligence for fracture detection and classification in orthopaedic trauma imaging? A systematic review. *Clin Orthop Relat Res.* (2019) 477(11):2482–91. doi: 10.1097/CORR.0000000000000848

99. MURA Dataset: Towards Radiologist-Level Abnormality Detection in Musculoskeletal Radiographs. Available at: https://stanfordmlgroup.github.io/competitions/mura/ (Cited June 15, 2023).

100. Diagnostic Imaging. Gleamer's BoneView Gains FDA Clearance for AI-Powered Pediatric Fracture Detection (2023). Available at: https://www.diagnosticimaging.com/view/gleamer-boneview-fda-clearance-for-ai-pediatric-fracture-detection (Cited Jun 15, 2023).

101. Greulich WW, Pyle SI. *Radiographic atlas of skeletal development of the hand and wrist.* Stanford, CA: Stanford University Press (1959). p. 288.

102. Tanner JM, Healy MJR, Cameron N, Goldstein H. *Assessment of skeletal maturity and prediction of adult height (TW3 method).* London, UK: W.B. Saunders (2001). p. 110.

103. Ferizi U, Besser H, Hysi P, Jacobs J, Rajapakse CS, Chen C, et al. Artificial intelligence applied to osteoporosis: a performance comparison of machine learning algorithms in predicting fragility fractures from MRI data. *J Magn Reson Imaging.* (2019) 49(4):1029–38. doi: 10.1002/jmri.26280

104. Ferizi U, Honig S, Chang G. Artificial intelligence, osteoporosis and fragility fractures. *Curr Opin Rheumatol.* (2019) 31(4):368–75. doi: 10.1097/BOR. 0000000000000607

105. Galbusera F, Casaroli G, Bassani T. Artificial intelligence and machine learning in spine research. *JOR Spine.* (2019) 2(1):e1044. doi: 10.1002/jsp2.1044

106. NuVasive. The Pulse platform. Available at: https://www.nuvasive.com/surgical-solutions/pulse-3/ (Cited July 18, 2023).

107. Miskin N, Gaviola GC, Huang RY, Kim CJ, Lee TC, Small KM, et al. Intra- and intersubspecialty variability in lumbar spine MRI interpretation: a multireader study comparing musculoskeletal radiologists and neuroradiologists. *Curr Probl Diagn Radiol.* (2020) 49(3):182–7. doi: 10.1067/j.cpradiol.2019.05.003

108. Lassau N, Estienne T, de Vomecourt P, Azoulay M, Cagnol J, Garcia G, et al. Five simultaneous artificial intelligence data challenges on ultrasound, CT, and MRI. *Diagn Interv Imaging.* (2019) 100(4):199–209. doi: 10.1016/j.diii.2019.02.001

109. Keles E, Irmakci I, Bagci U. Musculoskeletal MR image segmentation with artificial intelligence. *Adv Clin Radiol.* (2022) 4(1):179–88. doi: 10.1016/j.yacr.2022.04.010

110. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology.* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

111. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging—"how-to" guide and critical reflection. *Insights Imaging.* (2020) 11:91. doi: 10.1186/s13244-020-00887-2

112. Khalvati F, Zhang Y, Wong A, Haider MA. Radiomics. In: Narayan R, editors. *Encyclopedia of biomedical engineering*. Oxford: Elsevier (2019). p. 597–603. Available at: https://www.sciencedirect.com/science/article/pii/B9780128012383999641 (Cited July 18, 2023).

113. McBee MP, Awan OA, Colucci AT, Ghobadi CW, Kadom N, Kansagra AP, et al. Deep learning in radiology. *Acad Radiol*. (2018) 25(11):1472–80. doi: 10.1016/j.acra.2018.02.018

114. Fritz B, Yi PH, Kijowski R, Fritz J. Radiomics and deep learning for disease detection in musculoskeletal radiology : an overview of novel MRI- and CT-based approaches. *Invest Radiol*. (2023) 58(1):3–13. doi: 10.1097/RLI.0000000000000907

115. Klontzas ME, Triantafyllou M, Leventis D, Koltsakis E, Kalarakis G, Tzortzakakis A, et al. Radiomics analysis for multiple myeloma: a systematic review with radiomics quality scoring. *Diagnostics*. (2023) 13(12):2021. doi: 10.3390/diagnostics13122021

116. Zhan H, Teng F, Liu Z, Yi Z, He J, Chen Y, et al. Artificial intelligence aids detection of rotator cuff pathology: a systematic review. *Arthroscopy*. (2023):S0749-8063(23)00471-1. doi: 10.1016/j.arthro.2023.06.018

117. Kapiński N, Zieliński J, Borucki BA, Trzciński T, Ciszkowska-Łysoń B, Zdanowicz U, et al. Monitoring of the achilles tendon healing process: can artificial intelligence be helpful? *Acta Bioeng Biomech*. (2019) 21(1):103–11.

118. McKendrick M, Yang S, McLeod GA. The use of artificial intelligence and robotics in regional anaesthesia. *Anaesthesia*. (2021) 76(Suppl 1):171–81. doi: 10.1111/anae.15274

119. Ghasseminia S, Lim AKS, Concepcion NDP, Kirschner D, Teo YM, Dulai S, et al. Interobserver variability of hip dysplasia indices on sweep ultrasound for novices, experts, and artificial intelligence. *J Pediatr Orthop*. (2022) 42(4):e315–23. doi: 10.1097/BPO.0000000000002065

120. Hannun A, Case C, Casper J, Catanzaro B, Diamos G, Elsen E, et al. Deep Speech: Scaling up end-to-end speech recognition. arXiv (2014). Available at: http://arxiv.org/abs/1412.5567 (Cited June 18, 2023).

121. Do BH, Wu AS, Maley J, Biswal S. Automatic retrieval of bone fracture knowledge using natural language processing. *J Digit Imaging*. (2013) 26(4):709–13. doi: 10.1007/s10278-012-9531-1

122. Tan WK, Hassanpour S, Heagerty PJ, Rundell SD, Suri P, Huhdanpaa HT, et al. Comparison of natural language processing rules-based and machine-learning systems to identify lumbar spine imaging findings related to low back pain. *Acad Radiol*. (2018) 25(11):1422–32. doi: 10.1016/j.acra.2018.03.008

123. Mozayan A, Fabbri AR, Maneevese M, Tocino I, Chheang S. Practical guide to natural language processing for radiology. *RadioGraphics*. (2021) 41(5):1446–53. doi: 10.1148/rg.2021200113

124. Pons E, Braun LMM, Hunink MGM, Kors JA. Natural language processing in radiology: a systematic review. *Radiology*. (2016) 279(2):329–43. doi: 10.1148/radiol.16142770

125. Carrodeguas E, Lacson R, Swanson W, Khorasani R. Use of machine learning to identify follow-up recommendations in radiology reports. *J Am Coll Radiol*. (2019) 16(3):336–43. doi: 10.1016/j.jacr.2018.10.020

126. Daye D, Wiggins WF, Lungren MP, Alkasab T, Kottler N, Allen B, et al. Implementation of clinical artificial intelligence in radiology: who decides and how? *Radiology*. (2022) 305(3):555–63. doi: 10.1148/radiol.212151

Deep learning image segmentation approaches for malignant bone lesions: a systematic review and meta-analysis

Joseph M. Rich[1]*, Lokesh N. Bhardwaj[1], Aman Shah[2], Krish Gangal[3], Mohitha S. Rapaka[4], Assad A. Oberai[5], Brandon K. K. Fields[6], George R. Matcuk Jr[7] and Vinay A. Duddalwar[8,9]

[1]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States, [2]Department of Applied Biostatistics and Epidemiology, University of Southern California, Los Angeles, CA, United States, [3]Bridge UnderGrad Science Summer Research Program, Irvington High School, Fremont, CA, United States, [4]Department of Biology, University of Texas at Austin, Austin, TX, United States, [5]Department of Aerospace and Mechanical Engineering Department, Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, [6]Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States, [7]Department of Radiology, Cedars-Sinai Medical Center, Los Angeles, CA, United States, [8]Department of Radiology, Keck School of Medicine of the University of Southern California, Los Angeles, CA, United States, [9]Department of Radiology, USC Radiomics Laboratory, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Introduction: Image segmentation is an important process for quantifying characteristics of malignant bone lesions, but this task is challenging and laborious for radiologists. Deep learning has shown promise in automating image segmentation in radiology, including for malignant bone lesions. The purpose of this review is to investigate deep learning-based image segmentation methods for malignant bone lesions on Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Positron-Emission Tomography/CT (PET/CT).

Method: The literature search of deep learning-based image segmentation of malignant bony lesions on CT and MRI was conducted in PubMed, Embase, Web of Science, and Scopus electronic databases following the guidelines of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). A total of 41 original articles published between February 2017 and March 2023 were included in the review.

Results: The majority of papers studied MRI, followed by CT, PET/CT, and PET/MRI. There was relatively even distribution of papers studying primary vs. secondary malignancies, as well as utilizing 3-dimensional vs. 2-dimensional data. Many papers utilize custom built models as a modification or variation of U-Net. The most common metric for evaluation was the dice similarity coefficient (DSC). Most models achieved a DSC above 0.6, with medians for all imaging modalities between 0.85−0.9.

Discussion: Deep learning methods show promising ability to segment malignant osseous lesions on CT, MRI, and PET/CT. Some strategies which are commonly applied to help improve performance include data augmentation, utilization of large public datasets, preprocessing including denoising and cropping, and U-Net architecture modification. Future directions include overcoming dataset and annotation homogeneity and generalizing for clinical applicability.

KEYWORDS

bone cancer, CT, deep learning, image segmentation, MRI, PET/CT

1. Introduction

Bone is the third most common site of metastasis in the human body across all cancers, with an incidence of 18.8 cases per 100,000 each year and survival rates ranging from months to a few years (1, 2). The most common origins of bone metastases include breast, prostate, lung, and hematologic malignancies (1). Primary bone sarcomas are uncommon, with an incidence of 0.9 cases per 100,000 each year and higher survival rate (3).

Magnetic Resonance Imaging (MRI), Computed Tomography (CT), and Positron-Emission Tomography/CT (PET/CT) are commonly used to diagnose and track malignant bone lesions (Figure 1). MRI has higher sensitivity to detecting lesions in both the marrow and surrounding soft tissue structures and does not expose the patient to ionizing radiation. However, MRI requires a more expensive and laborious imaging process when compared with CT (4). CT is more sensitive to detecting changes in bone morphology and has higher spatial resolution, although it involves radiation and has poorer performance with soft-tissue and marrow imaging (5). PET/CT combines techniques of both CT (three-dimensional x-ray scanning with high spatial resolution) and PET (injection of radioactive tracer to quantify cellular metabolism), providing high sensitivity and specificity for imaging skeletal malignancies (6). These benefits make PET/CT the standard of care in bone lesion imaging, although there are still the drawbacks of higher cost and use of radiation. PET/MRI similarly offers combined benefits of both MRI and PET. Malignant bone lesions often appear as blastic (hyperdense regions indicating bone formation), lytic (hypodense regions indicating bone resorption), or a mix.

Early diagnosis of malignant bone lesions is critical for improving prognosis and treatment response. Image segmentation, in which the boundaries of a lesion are precisely delineated, allows radiologists to determine the extent of disease and accurately provide quantitative measurement for disease tracking, treatment response, and management (7). Additionally, accurate segmentation is essential for performing clinical research using radiologic images. The task of image segmentation is typically performed manually by radiologists, but this is a labor-intensive and time-consuming process, thus limiting its applicability in clinical workflows.

Machine learning has the potential to automate lesion segmentation. Some early image segmentation methods include thresholding, region-growing, edge-based segmentation, active contour models, watershed transforms, and snakes (8). All of these methods involve identifying simple features of an image such as thresholded intensity values, edges, or neighboring homogeneous regions, but are limited in analyzing more complex features (9). The progress of deep learning methods in particular, especially Convolutional Neural Networks (CNNs) (10), provides the ability to segment complex images with increasing accuracy (8, 11, 12). CNNs are deep neural networks in which convolution operations are applied as sliding filters over an image, reducing dimensionality, and identifying image features through selection of filter weights. A particularly popular CNN architecture is U-Net, which consists of an initial encoding section of convolution operations and a subsequent decoding section of transpose-convolution operations to reconstruct an image with the same dimensions as the input (13) (Figure 2). Deep learning has shown promise in image segmentation of lesions in CT and MRI scans in a wide range of contexts including lesions of the breast (14), kidney (15), and brain (16, 17).

Deep learning model performance generally improves with larger dataset sizes, with the minimal acceptable size
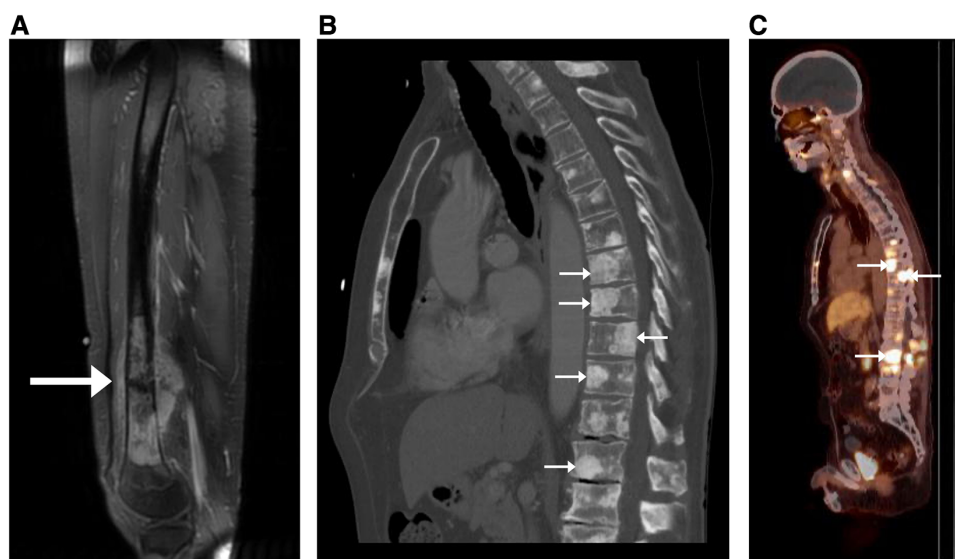
FIGURE 1
Appearance of malignant bone lesions on different imaging modalities. (A) Sagittal T1-weighted post-contrast MR image with fat suppression of the right femur in a 32-year-old female with biopsy-proven osteosarcoma of the distal femoral metadiaphysis (arrow). (B) Sagittal chest CT with bone windows showing diffuse osseous metastatic disease (arrows) in a 72-year-old male with castration-resistant prostate cancer. (C) Sagittal vertex-to-pelvis prostate-specific membrane antigen (PSMA) PET/CT fusion image showing diffuse osseous metastatic disease (arrows) in the same patient an in (B) 6 months previously. Note that in (B) and (C), not all metastatic lesions have been annotated with arrows.
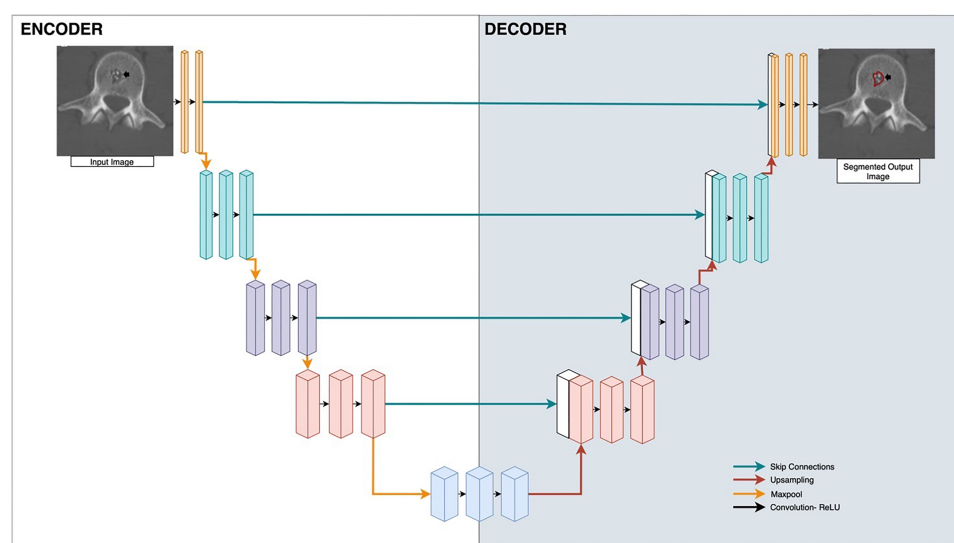
FIGURE 2
U-Net applied to bone radiology image segmentation. Input is the medical image, and output is the segmentation mask applied to the lesion. Boxes represent vectorized outputs of convolutional and pooling operations. Arrows represent mathematical operations applied to each layer. Blue arrows are skip connections, red arrows are upsampling, yellow arrows are maxpool, black arrows are Convolution-rectified linear units (ReLU).

typically being on the order of hundreds of subjects. However, this is a challenging task in the realm of medicine where the input involves patient data due to concerns regarding privacy and sharing (18). While there are some major public databases that can assist with data augmentation or transfer learning for certain clinical queries (19–22), there are many pathologies that are specific or unique enough where such datasets are not readily available. Some techniques to try to overcome this deficit include working with large pretrained models (23), data-generation techniques such as Generative Adversarial Networks (14, 24, 25), or applying domain knowledge to data preprocessing and augmentation (26, 27). There are very few public datasets or models which capture primary or metastatic skeletal lesions on CT, MRI, PET/CT, or PET/MRI.

The purpose of this systematic review and meta analysis is to describe how effective deep learning-guided image segmentation techniques are in accurately identifying and delineating malignant bone lesion on major radiologic imaging studies (CT, MRI, PET/CT, and PET/MRI), as well as to compare methods and performance across studies. We describe all algorithms and neural network architectures reported in the included studies, as well as characteristics of the datasets and additional techniques used for successful segmentation. We also note any publicly available datasets or models.

2. Materials and methods

2.1. Literature search

Our systematic literature review is in compliance with the guidelines outlined by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses 2020 (PRISMA). We performed a keyword search for papers which studied deep learning-based image segmentation of cancerous lesions of the bone on CT, MRI, PET/CT, and PET/MRI scans. Searches were performed on Pubmed, Embase, Web of Science, and Scopus. All searches were performed on May 8, 2023. The exact search criteria were as follows:

> "(CT OR CTs OR MRI OR "MR Imaging" OR "PET-CT" OR "PET/CT" OR "PET-MRI" OR "PET/MRI") AND (Segmentation) AND ("machine learning" OR "deep learning" OR "artificial intelligence" OR "neural network" OR "neural networks" OR "auto-segmentation" OR "auto segmentation") AND (bone OR skeleton OR bones OR osseous OR blastic OR lytic) AND (cancer OR cancers OR metastases OR metastasis OR neoplasm OR neoplasms OR metastatic OR tumor OR tumors OR malignant OR tumour OR tumours)"

Other inclusion criteria included a publication date range of 2010–2023, use of English language, full text availability, and only primary literature (i.e., other review articles were excluded). Exclusion criteria included segmentations performed on other imaging modalities (e.g., x-ray, bone scintigraphy, PET), other types of tissues or organs, segmentation of non-malignant features (e.g., whole bone segmentation, fracture segmentation), and non-segmentation techniques (e.g., synthetic data creation, boundary-box generation, outcome classification).

We used the Covidence platform for paper importing and screening (28). All unique papers which fit these criteria were passed through a primary screening of titles and abstracts by a single reviewer. All papers which passed the primary screen were then passed through a secondary screen involving full text review for inclusion criteria by two reviewers.

2.2. Data extraction

Categories for data extraction were chosen to describe imaging modality, model type, dataset, lesion type, and part of body in more detail. Data was extracted from each paper with the following categories (**Supplementary Table S1**):

(1) Publication date
(2) Imaging modality (CT, MRI, PET/CT, or PET/MRI)
(3) Imaging dimensionality (2-Dimensional [2D], 3-Dimensional [3D])
(4) Primary cancer type
(5) Quality of lesion (blastic, lytic, or mixed)
(6) Soft tissue component
(7) Model architecture
(8) Dataset publicity
(9) Dataset size (patients, images)
(10) Patient population (demographics)
(11) Treatment received
(12) Ground truth establishment
(13) Training-cross validation-test split
(14) Cross validation method
(15) Additional methods
(16) Metrics.

3. Results

3.1. PRISMA flowchart

The results of our literature search are shown in the PRISMA flowchart (**Figure 3**). In brief, our initial search yielded 784 papers. Covidence automatically eliminated 363 duplicates. An additional 4 duplicates were eliminated manually, leaving 421 unique manuscripts. After primary screening of titles and abstracts, 292 papers were further excluded. From the 129 papers which passed through full-text review, 41 studies were ultimately eligible for inclusion in this study (**Supplementary Table S1**) (29–69). Some of the most common reasons for exclusion included wrong tissue type, segmentation of a non-malignant feature (e.g., whole bone segmentation or fracture segmentation), wrong study design (e.g., prognosis classification, boundary box), and wrong imaging modality (e.g., bone scintigraphy, PET, x-ray).

3.2. Categorization of included studies

Of the 41 total studies, the most popular publication year was 2022 ($n = 18$ studies, 43.90% of the cohort), followed by 2023 (up until May) ($n = 9$, 21.95%). While our search criteria ranged from 2010, the oldest paper included was from 2017. The most common imaging modality studied was MRI ($n = 21$, 51.22%), followed by CT ($n = 12$, 29.27%). The most common image dimensionality method used 3D data alone ($n = 21$, 51.22%), followed by 2D alone ($n = 11$, 26.83%). Osteosarcoma was the

most common cancer primary bone malignancy ($n = 18$, 43.90%). Prostate cancer was the most secondary bone malignancy ($n = 7$, 17.07%) (**Figure 4**).

3.3. Synthesized findings of included studies

Studies were categorized primarily by dimensionality, modality, publication year, and lesion characteristics (i.e., blastic vs. lytic). All performance metrics reported by each paper, including dice similarity coefficient (DSC), F1-measure, Jaccard, accuracy, sensitivity, and specificity, were included in **Supplementary Table S1**. DSC was by far the most popular metric, recorded in 35 papers (85.3%). In order to determine statistical significance between groups, a simple two-sample t test was conducted with a power level of 95% being established prior to analysis. While there was a higher median DSC for studies which used 2D data (0.901, $n = 11$) compared to 3D data (0.856, $n = 17$), the difference was not statistically significant (**Figure 5C**, **Supplementary Table S2**). In the years 2017 through 2019, there was only a single paper published each year across the 3 years, which reported both the dimensionality method used and a DSC. Although the years 2022 and 2023 accounted for a majority of the papers within the cohort ($n = 27$, 65.85%), there was no statistically significant difference in median DSCs (**Figure 5A**). With regards to image modality, studies utilizing CT imaging generally reported higher median 2D DSCs (0.94, $n = 4$) compared to MRI (0.924, $n = 7$). In contrast, MRI generally yielded a higher 3D DSCs (0.895, $n = 10$) than studies which evaluated 3D data by CT (0.856, $n = 5$) (**Supplementary Table S2**). However, neither difference for 2D vs. 3D data was statistically significant (**Figure 5C**). Aggregating all data dimensionality, CT had a slightly higher median DSC (0.92, $n = 9$) than MRI (0.85, $n = 17$); however, there was no statistically significant difference in mean dice score between the two imaging methods ($p = 0.5469$). Papers studying lytic lesions reported higher median 2D and 3D DSCs, at 0.94 ($n = 2$) and 0.922 ($n = 5$), respectively, when compared to segmentation of blastic lesions, though this difference was similarly not statistically significant (**Figure 5D**, **Supplementary Table S2**). Papers which did not include cross-validation showed an average higher DSC (0.923, $n = 13$) than those which did (0.840, $n = 22$) ($p = 0.0038$). There was no statistically significant relationship between using data augmentation in workflow and increased DSC ($p = 0.1156$).

4. Discussion

In this systematic review and meta-analysis, we have attempted to aggregate the literature describing automated segmentation methods for primary and metastatic bone malignancies on CT and MRI. We found that most models achieved objectively good performance (DSC >0.7) on this task, with some of the most common methods including data augmentation, U-Net

FIGURE 3

PRISMA flowchart of systematic literature review.
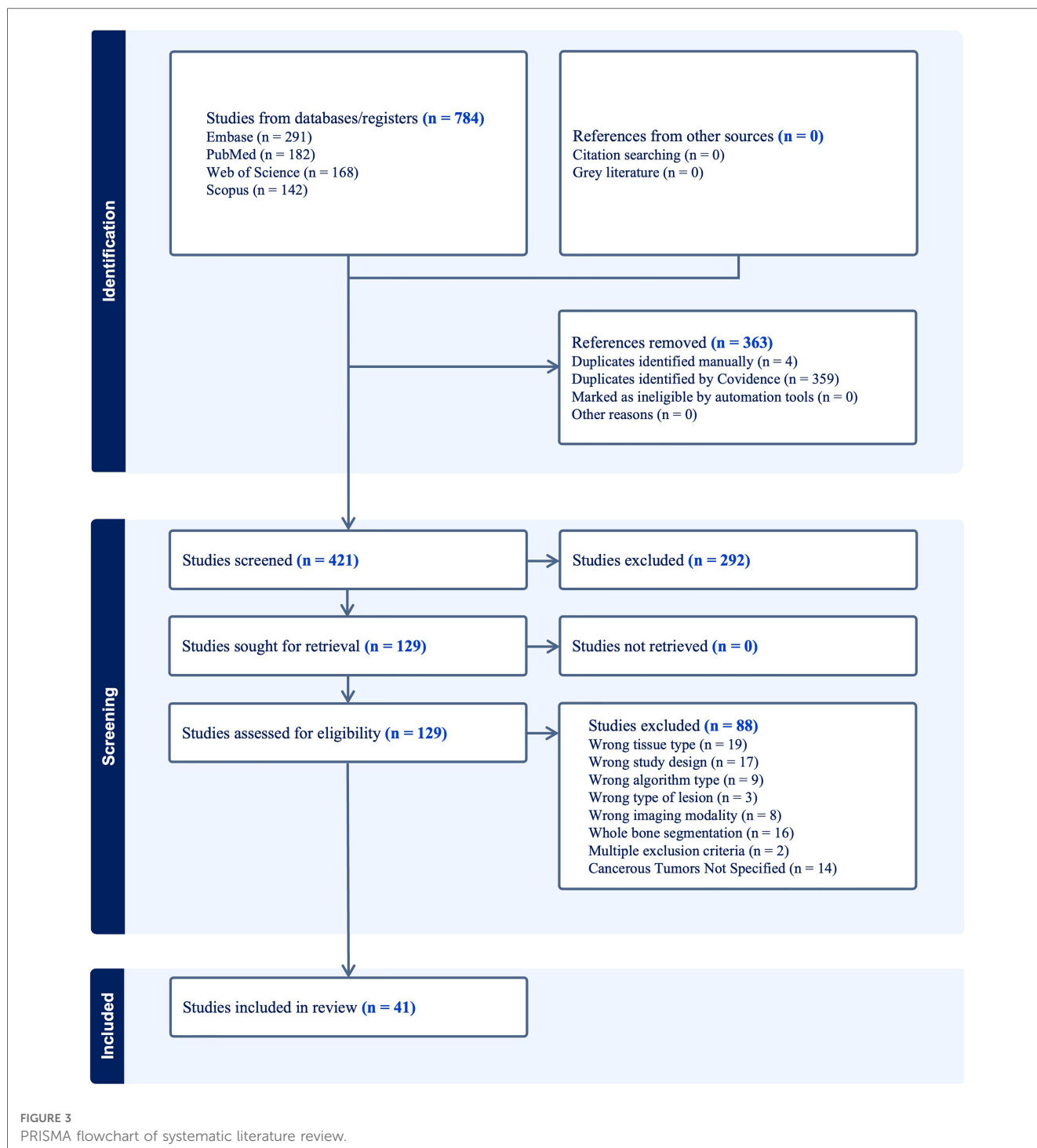
architecture modification, and preprocessing to reduce noise. We clarify the frequency of reported studies that fall into specific criteria regarding imaging approaches and lesion quality, which helps identify which problems still need to be most studied and how much precedent work exists for a specific type of problem. Overall, while small numerical differences were seen between segmentation DSCs when comparing across imaging modality, publication year, dataset dimensionality, and lesion quality (blastic vs. lytic), none of these were found to be statistically significant. The similarity in performance across these attributes indicates that these segmentation models have the capability to perform well across a range of conditions. The statistical significance in DSC improvement for papers which excluded cross validation compared to those which included it indicates the potential of an overfitting problem in these cases, highlighting the importance of test sets and external validation for generalizability. While other reviews have investigated similar segmentation performance tasks applied to various lesions or

FIGURE 4

Visualization of characteristics of included studies, showing distribution according to **(A)** publication year; **(B)** imaging modality; **(C)** image dimensionality; **(D)** type of cancer.
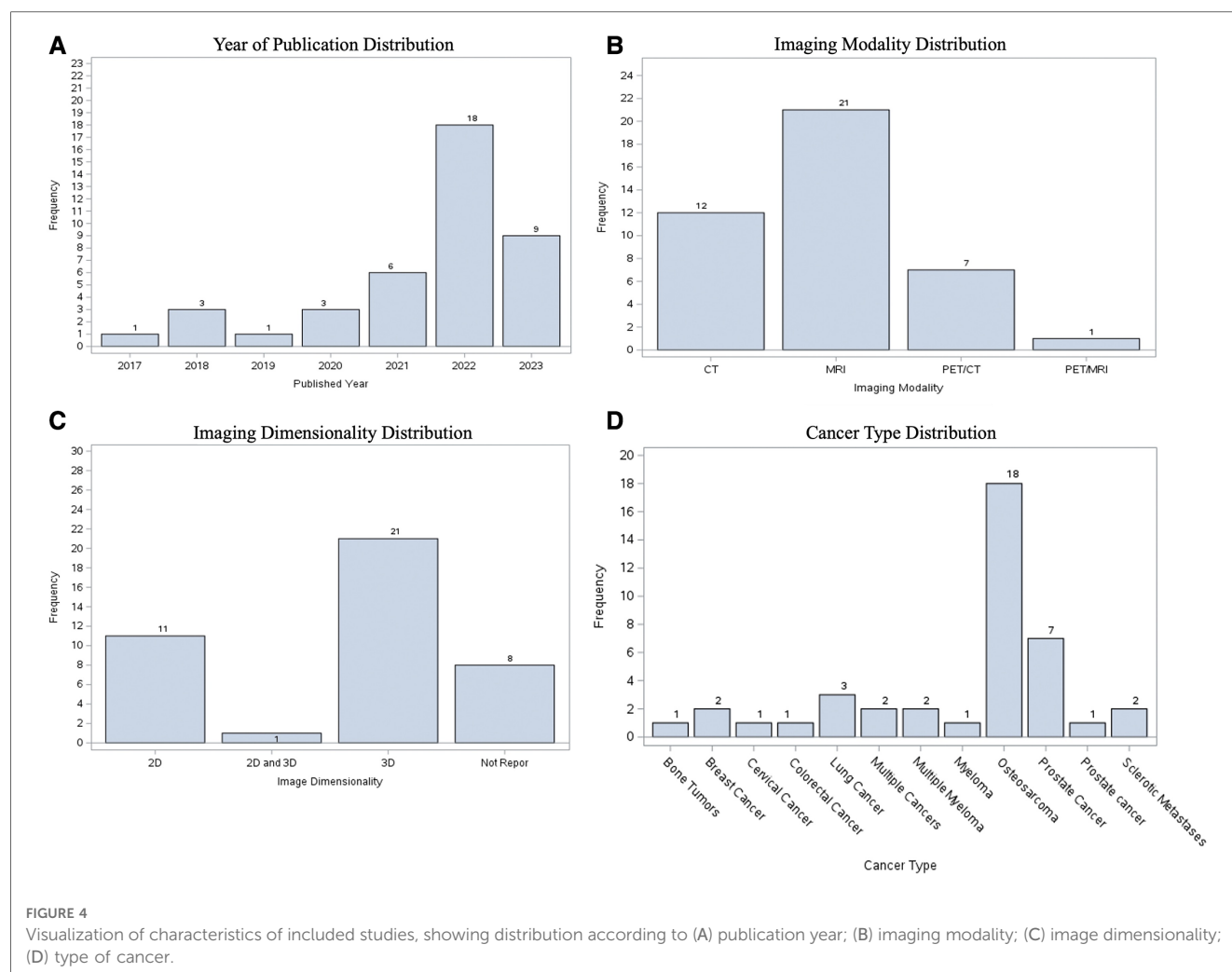
whole organs, to the best of our knowledge, ours is the first to focus on deep learning techniques applied specifically to lesions of the bone (70–77). Additionally, ours is the first which specifically evaluates differences in segmentation performance specifically as they relate to imaging modality, imaging dimensionality, and predominant lesion characteristic. Future directions include comparing further characteristics of papers (e.g., model architecture, type of cancer, dataset size, etc.) to determine which types of problems or approaches yield the best results, as well as expanding the scope of analysis to other imaging modalities or targets of segmentation to increase statistical power.

4.1. Metrics

Comparison of metrics across various studies can be difficult. Different problems or datasets may possess inherently different technical challenges even when problems appear similar, making performance comparison with metrics across studies difficult. Additionally, different metrics capture different qualities of success (Table 1). For instance, specificity is high

when there are minimal false positives (i.e., minimal areas of predicted lesions where none is present); since most lesions make up a small percentage of an image, an algorithm will achieve high specificity by predicting no lesions on an image, even though this requires no learning. Within our cohort, Zhao et al. reported an estimated DSC of 0.60, which is considerably lower than most DSCs which lie approximately within the 0.85–0.95 range (69). However, they also reported sensitivity and precision to each be 0.99, which would indicate an element of good performance. While each metric has its strengths and limitations, DSC was the most commonly reported metric by far, reported in nearly every included study. DSC's ubiquity in image segmentation is due to a few factors including its use by many others studying image segmentation techniques, its balance of precision and recall, its intuitive appeal as an approximator of percentage of overlap between ground truth and prediction, its history of being used for measuring reproducibility of manual segmentation, and its adaptability to logit transformation since its values lie between 0 and 1 (78–81). All reported metrics from each study were recorded in **Supplementary Table S1**. While a uniform dataset-agnostic success criterion cannot be established as a result of the

FIGURE 5
Performance comparison with DSC by (A) publication year; (B) imaging modality; (C) image dimensionality; (D) quality of lesion (blastic vs. lytic).
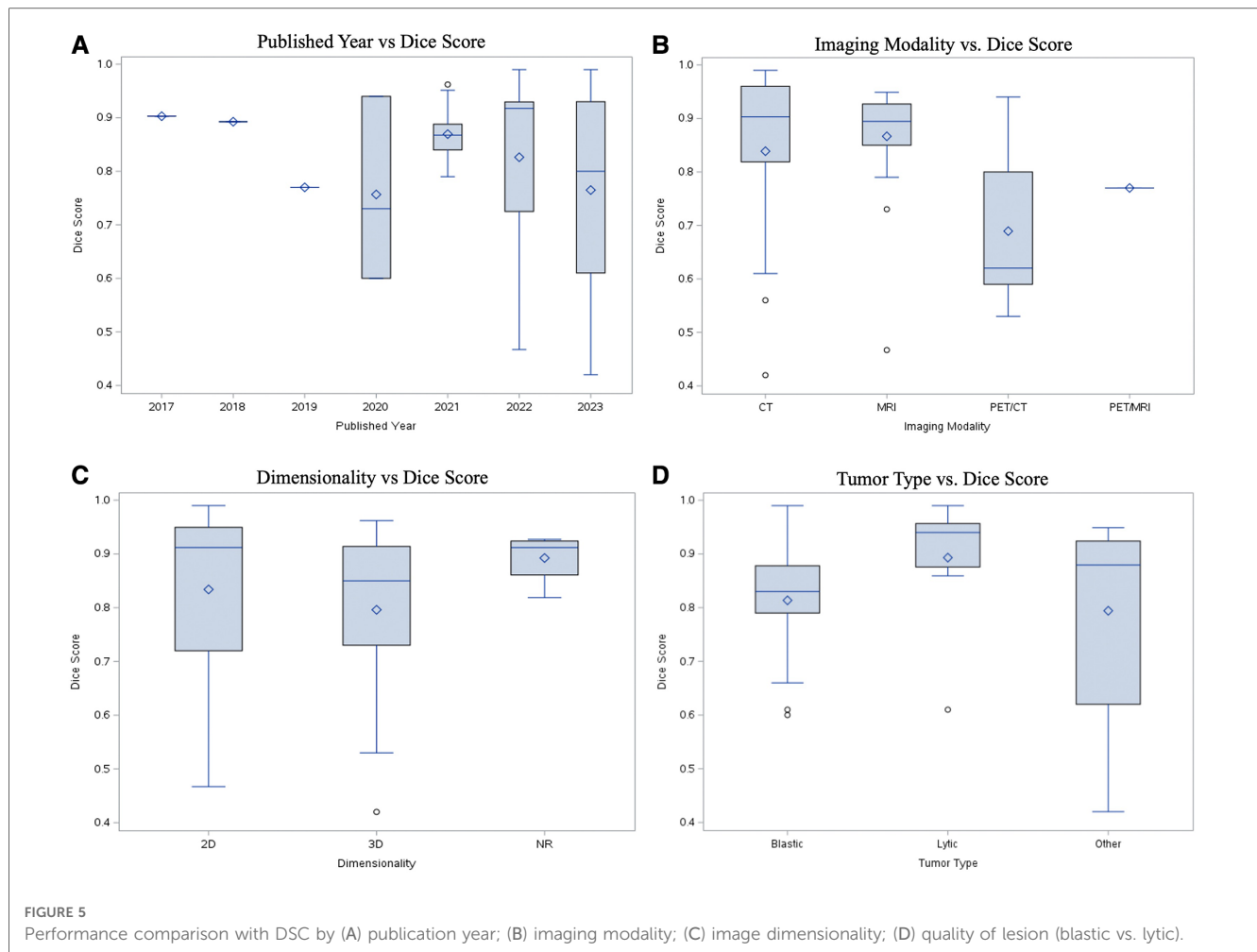
TABLE 1 Popular metrics for image segmentation.

Metric	Equation
Dice similarity coefficient	$\frac{\|P \cap L\|}{\|P\| + \|L\|} = \frac{2TP}{2TP + FP + FN}$
F1-measure	$\frac{2*\text{precision}*\text{recall}}{\text{precision} + \text{recall}}$
Jaccard index, intersection over union	$\frac{\|P \cap L\|}{\|P\| \cup \|L\|} = \frac{TP}{TP + FP + FN}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
Sensitivity, true positive rate, recall	$\frac{TP}{TP + FN}$
Specificity, true negative rate	$\frac{TN}{FP + TN}$
False positive rate	$1 - \text{specificity}$
False negative rate	$1 - \text{sensitivity}$
Precision, positive predictive value	$\frac{TP}{TP + FP}$
Negative predictive value	$\frac{TP}{TN + FP}$
Area under curve (AUC)	$\int (\text{Receiver Operating Characteristic Curve})$

P, Prediction; L, Label; TP, true positive; TN, true negative; FP, false positive; FN, false negative.

challenges described earlier, a general objective threshold for what is considered a reasonable model is to achieve a DSC around 0.7 (80), which most papers in this review surpass.

4.2. Imaging modality and dimensionality

The overwhelming majority of imaging modalities utilized throughout the paper cohort were either CT or MRI. Both CT and MRI are reasonably amenable to automated segmentation, with median DSCs between 0.85–0.95 for both modalities (**Figure 5B**, **Supplementary Table S2**). Models analyzing PET/CT and PET/MRI data demonstrate lower median DSCs than CT and MRI-trained models. PET/CT and PET/MRI combine spatial and metabolic information, providing useful context for radiologists. However, there can be noise in radioactive tracer uptake involved in PET, and errors in spatial alignment of the two scans, making data more difficult to train (82). Additionally, malignant lesions display heterogeneous metabolic activity, adding further noise to the imaging process. In order to overcome this, Hwang et al. utilized maximum-likelihood reconstruction of activity and attenuation (MLAA) algorithm as input for a CNN to improve accuracy and convergence with good results (40).

Models were able to perform well on both 2D and 3D data, with 2D data achieving slightly higher median DSCs (**Figure 5C**), although the results were notably not statistically significant. Both types of dimensionalities have pros and cons. Computer vision models were historically trained with two-

dimensional images, and 2D data is inherently generally less complex than 3D. However, given that radiologists almost always rely on 3D data for image interpretation, modern deep learning frameworks in radiology, such as nnU-Net (11), have been developed to primarily evaluate with 3D data. The third dimension adds additional spatial and contextual information that may otherwise be lost in two dimensional analysis. As a compromise, one model in our dataset utilized 2.5D data by employing two 2D encoder-decoder modules and one pseudo-3D fusion module, which extracted features from the 2D outputs (53). For clinical applications with unknown cases, considerations for determining data dimensionality for a model include spatial and contextual information, model choice, and difficulty of the segmentation problem.

4.3. Dataset size

Dataset size ranged drastically among included papers, with image count ranging from 37 (54) to 80,000 + . Generally, most papers included dataset sizes in the hundreds to low thousands of images or scans. Most studies utilized private and relatively small datasets, making generalizability of algorithms difficult. However, the one large publicly available dataset containing over 80,000 MRI scans of osteosarcoma was utilized by numerous studies (43, 45, 52, 55, 59–64). Dataset size was not a significant predictor of model performance in our cohort, as most models achieved DSCs above 0.7, and many above 0.9, at all ranges of dataset sizes.

This good performance in spite of small dataset size could be attributed in part to data augmentation techniques utilized by many papers. Some of the most popular employed techniques include random cropping, flipping, rotation, zooming, and mirroring (30–32, 35, 38, 43, 50, 52, 54, 56, 60, 67, 68). Of the 14 additional methods found within our review, 7 involved some form of data augmentation. However, as described earlier, there was no correlation between data augmentation workflow and DSC.

Transfer learning was utilized in some cases. Transfer learning is generally thought to be most effective when the transferred data is large and similar to the pathology being studied. Due to the limited nature of public radiology images, models trained on very large datasets of non-radiologic images, such as Microsoft Coco (83), may be reasonable candidates for transfer learning even for image analysis in radiology (66). Similarly, other studies utilized generative methods to create phantom images for their training sets that resembled real images (65). Data preprocessing can incorporate steps to improve model performance, such as whole-bone segmentation to allow the algorithm to have a smaller region to analyze when segmenting an osseous lesion (47).

With small datasets comes the increased risk of overfitting. There was no consensus on training-cross validation-test splits. Generally, most studies dedicated approximately 60%–80% of data to the training set, 10%–30% of data to the test set, and 0%–20% to the cross-validation set (Supplementary Table S1). Nearly half of all papers did not include a cross validation set, meaning that any hyperparameter tuning or architecture

adjustment that resulted from testing could have resulted in overfitting. The higher average DSC of papers without cross-validation (0.92) compared to those with it (0.79) supports the likelihood of overfitting in some of these cases. Only two papers utilized external validation (testing of the model on an additional dataset acquired separately from other sets used to initially train or evaluate the model), making generalizability especially difficult (47, 48). However, for both papers, the DSC on the external validation set was the same as that of the test set (at 0.79 and 0.84, respectively), demonstrating model generalizability in these cases (47, 48).

4.4. Model architecture

Most studies employed a U-Net CNN architecture for automated image segmentation. U-Net is a popular architecture type because of its ability to accurately segment small targets and fast training speed (84). Image segmentation, as opposed to classification, is especially helpful for extracting objects of interest. In particular, bone segmentation of lesions correctly identifies the spatial location of a tumor. What distinguishes U-Net from other CNNs are the encoder-decoder networks as well as the implementation of skip connectors. The encoder-decoder network ensures that the output image has the same dimensionality as the input image while skip connections ensure full recovery of details and features that may have been lost or forgotten as information passes through successive layers. This preservation of dimensionality is essential for image segmentation, where the output is a binary mask which must resemble the outlined feature on the input image (84). Another attractive feature of the U-Net is the fact that each layer of the network extracts features from a different spatial scale of the image, and by collecting results from each of these layers, the network is able to transform an input image at multiple spatial scales.

Many modifications of U-Net were created to boost model performance. For instance, dilated convolutional U-Net, which involves multiple dilated convolutions following a standard convolution, was employed in a modified U-Net with recurrent nodes in order to preserve contextual information and spatial resolution (36). Some models employed combinations of transformer models and modified U-Nets, allowing for preservation of contextual features such as edge enhancement (45, 49). Cascaded 3D U-Net likewise employ two U-Net architectures in series, with the first trained on down-sampled images and the second trained on full-resolution images, allowing for a combination of granularity and refinement of the features of choice (39).

While a majority of the papers utilized a modification of the U-Net segmentation algorithm, other alternative architectures included non-convolutional Artificial Neural Network models (41), voxel-wise classification (33), AdaBoost algorithms and Chan-Vese algorithms (37), CNN with bagging and boosting (44), and V-Net (34, 65). These alternative algorithms achieved DSCs or AUCs above 0.7, which is on par with the median

performance of the U-Net models. However, U-Net variations have been tried in a greater number of studies and demonstrated performance as high as 0.9821 in this cohort (58), indicating that U-Net may be more suitable at present day for achieving maximal performance.

4.5. Approaches to segmentation

Two approaches to delineating or segmenting regions of interest are "filling in the lesion" and "tracing precise contour". Filling in the lesion involves segmenting the entire volume of the region of interest including both the solid and necrotic components of the lesion. On the other hand, tracing precise contours involves precisely outlining the boundaries of a region of interest such that healthy tissues and other non-relevant features are excluded. While the overwhelming number of publications use lesion segmentation as the only methodology, a few studies in literature have discussed a multi step strategy "identification of lesions", viz creating bounding boxes around the lesions as a separate first step and then a subsequent strategy of precise segmentation of lesions (85, 86). Despite the different implications of these approaches, most papers did not specify which approach they followed when establishing ground truth. If establishment of ground truth was discussed at all, it was usually generally stated the number and skill level of radiologists involved in the process, but with no specific mention of methodology. Even so, Trägårdh et al. studied the importance of inter-reader heterogeneity by comparing model performance on a test set annotated by the same physician who annotated the training set as compared to separate annotators, finding substantial performance differences between sensitivities (57). Methodology of producing ground truth segmentations warrants further discussion to establish a repeatable standard in future studies. The inter-reader heterogeneity also points to the benefit of using probabilistic segmentation algorithms that would account for this variability and produce an ensemble of likely segmentations for a given input image. While these algorithms have been used for the segmentation tasks (17, 87), they have not yet been applied to bone segmentation.

One of the strengths of this review is the comprehensive analysis of all papers fitting search criteria, and the detailed data extraction to allow for comparison of methods or qualities among all papers which have studied this type of problem. Another strength is maintaining focus on clinically relevant features of model design while also keeping in mind technical details of model implementation. A limitation is the difficulty in comparing metrics across studies. Dataset quality, annotation heterogeneity, and noise can make evaluation of a good DSC specific to the specific dataset being studied. Additionally, the relatively small number of studies involved in the review made it difficult to perform any rigorous statistical analysis between subcategories.

In conclusion, deep learning shows great promise for bone lesion segmentation. Considerations include model architecture, imaging modality and dimensionality, dataset size, and establishment of ground truth. Compared to other tissues and organs, there is still much to be done to expand on the task of bone lesion segmentation. Future directions include training on larger and more diverse datasets, applying multiple methods of establishing ground truth, accounting for variability in the segmentation task, and integrating into clinical application. The success with the osteosarcoma MRI dataset from Second Xiangya Hospital of Central South University shows the importance and applicability of these large public datasets (63), and similar efforts should be undertaken from other institutions and studying other types of lesions. General image segmentation models, such as the Segment Anything Model (12), could also show promise in bone lesion segmentation, especially in conjunction with optimization processes involved in the architecture design of these studies. Deep learning-guided segmentation results have great potential to augment human performance, especially in conjunction with radiomic and pathomic data. As these models continue demonstrating success and generalizability, they will help radiologists save time and improve accuracy in delineating these lesions.

Data availability statement

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

Author contributions

JR: screened each paper, verified data extraction, and lead writing of the manuscript. LB, AS, KG, and MR: assisted in screening, data extraction, and writing portions of the discussion and boxes. AO, BF, GM, and VD: supervised and administered the study. All authors contributed to the article and approved the submitted version.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) GM and BF declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fradi.2023.1241651/full#supplementary-material

References

1. Jayarangaiah A, Kemp AK, Theetha Kariyanna P. Bone metastasis [Updated 2022 Oct 25]. In: *StatPearls* [Internet]. Treasure Island (FL): StatPearls Publishing (2023). Available at: http://www.ncbi.nlm.nih.gov/books/NBK507911/ (Cited Jun 5, 2023).

2. Ryan C, Stoltzfus KC, Horn S, Chen H, Louie AV, Lehrer EJ, et al. Epidemiology of bone metastases. *Bone*. (2022) 158:115783. doi: 10.1016/j.bone.2020.115783

3. Franchi A. Epidemiology and classification of bone tumors. *Clin Cases Miner Bone Metab*. (2012) 9(2):92–5. PMID: 23087718; PMCID: PMC3476517.

4. van Beek EJ, Hoffman EA. Functional imaging: CT and MRI. *Clin Chest Med*. (2008) 29(1):195. vii. doi: 10.1016/j.ccm.2007.12.003

5. Heindel W, Gübitz R, Vieth V, Weckesser M, Schober O, Schäfers M. The diagnostic imaging of bone metastases. *Dtsch Ärztebl Int*. (2014) 111(44):741–7. doi: 10.3238/arztebl.2014.0741

6. Kosmin M, Padhani AR, Gogbashian A, Woolf D, Ah-See ML, Ostler P, et al. Comparison of whole-body MRI, CT, and bone scintigraphy for response evaluation of cancer therapeutics in metastatic breast cancer to bone. *Radiology*. (2020) 297(3):622–9. doi: 10.1148/radiol.2020192683

7. Suetens P, Bellon E, Vandermeulen D, Smet M, Marchal G, Nuyts J, et al. Image segmentation: methods and applications in diagnostic radiology and nuclear medicine. *Eur J Radiol*. (1993) 17(1):14–21. doi: 10.1016/0720-048X(93)90023-G

8. Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image Segmentation Using Deep Learning: A Survey. *arXiv*. (2020). Available at: http://arxiv.org/abs/2001.05566 (Cited Jun 5, 2023).

9. Luo D, Zeng W, Chen J, Tang W. Deep learning for automatic image segmentation in stomatology and its clinical application. *Front Med Technol*. (2021) 3:767836. doi: 10.3389/fmedt.2021.767836

10. O'Shea K, Nash R. An Introduction to Convolutional Neural Networks. *arXiv*. (2015). Available at: http://arxiv.org/abs/1511.08458 (Cited Jun 6, 2023).

11. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. (2021) 18(2):203–11. doi: 10.1038/s41592-020-01008-z

12. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment Anything. *arXiv*. (2023). Available at: http://arxiv.org/abs/2304.02643 (Cited May 30, 2023).

13. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention—MICCAI 2015*. Cham: Springer International Publishing (2015). 234–41. p. (Lecture Notes in Computer Science).

14. Muramatsu C, Nishio M, Goto T, Oiwa M, Morita T, Yakami M, et al. Improving breast mass classification by shared data with domain transformation using a generative adversarial network. *Comput Biol Med*. (2020) 119:103698. doi: 10.1016/j.compbiomed.2020.103698

15. Anari PY, Lay N, Chaurasia A, Gopal N, Samimi S, Harmon S, et al. Automatic segmentation of clear cell renal cell tumors, kidney, and cysts in patients with von Hippel-Lindau syndrome using U-net architecture on magnetic resonance images. *ArXiv*. (2023). arXiv:2301.02538v1.

16. Rauschecker AM, Gleason TJ, Nedelec P, Duong MT, Weiss DA, Calabrese E, et al. Interinstitutional portability of a deep learning brain MRI lesion segmentation algorithm. *Radiol Artif Intell*. (2022) 4(1):e200152. doi: 10.1148/ryai.2021200152

17. Moazami S, Ray D, Pelletier D, Oberai AA. Probabilistic brain extraction in MR images via conditional generative adversarial networks. *bioRxiv*. (2022). Available at: https://www.biorxiv.org/content/10.1101/2022.03.14.484346v1 (Cited Jun 16, 2023). doi: 2022.03.14.484346

18. Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/the academy workshop. *Radiology*. (2019) 291 (3):781–91. doi: 10.1148/radiol.2019190613

19. Antonelli M, Reinke A, Bakas S, Farahani K, Kopp-Schneider A, Landman BA, et al. The medical segmentation decathlon. *Nat Commun*. (2022) 13(1):4128. doi: 10.1038/s41467-022-30695-9

20. Mei X, Liu Z, Robson PM, Marinelli B, Huang M, Doshi A, et al. Radimagenet: an open radiologic deep learning research dataset for effective transfer learning. *Radiol Artif Intell*. (2022) 4(5):e210315. doi: 10.1148/ryai.210315

21. Weinstein JN, Collisson EA, Mills GB, Shaw KM, Ozenberger BA, Ellrott K, et al. The cancer genome atlas pan-cancer analysis project. *Nat Genet*. (2013) 45 (10):1113–20. doi: 10.1038/ng.2764

22. Yan K, Wang X, Lu L, Summers RM. Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. *J Med Imaging*. (2018) 5(3):036501. doi: 10.1117/1.JMI.5.3.036501

23. Karimi D, Warfield SK, Gholipour A. Transfer learning in medical image segmentation: new insights from analysis of the dynamics of model parameters and learned representations. *Artif Intell Med*. (2021) 116:102078. doi: 10.1016/j.artmed.2021.102078

24. Jeong JJ, Tariq A, Adejumo T, Trivedi H, Gichoya JW, Banerjee I. Systematic review of generative adversarial networks (GANs) for medical image classification and segmentation. *J Digit Imaging*. (2022) 35(2):137–52. doi: 10.1007/s10278-021-00556-w

25. Skandarani Y, Jodoin PM, Lalande A. GANs for medical image synthesis: an empirical study. *J Imaging*. (2023) 9(3):69. doi: 10.3390/jimaging9030069

26. Nishio M, Fujimoto K, Matsuo H, Muramatsu C, Sakamoto R, Fujita H. Lung cancer segmentation with transfer learning: usefulness of a pretrained model constructed from an artificial dataset generated using a generative adversarial network. *Front Artif Intell*. (2021) 4:694815. doi: 10.3389/frai.2021.694815

27. Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. *J Big Data*. (2019) 6(1):60. doi: 10.1186/s40537-019-0197-0

28. Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at: www.covidence.org.

29. Baidya Kayal E, Kandasamy D, Sharma R, Bakhshi S, Mehndiratta A. Segmentation of osteosarcoma tumor using diffusion weighted MRI: a comparative study using nine segmentation algorithms. *Signal Image Video Process*. (2020) 14 (4):727–35. doi: 10.1007/s11760-019-01599-x

30. Chang CY, Buckless C, Yeh KJ, Torriani M. Automated detection and segmentation of sclerotic spinal lesions on body CTs using a deep convolutional neural network. *Skelet Radiol*. (2022) 51(2):391–9. doi: 10.1007/s00256-021-03873-x

31. Chang CY, Huber FA, Yeh KJ, Buckless C, Torriani M. Original research: utilization of a convolutional neural network for automated detection of lytic spinal lesions on body CTs. *Skelet Radiol*. (2023) 52:1377–84. doi: 10.1007/s00256-023-04283-x

32. Chen X, Ma X, Yan X, Luo F, Yang S, Wang Z, et al. Personalized auto-segmentation for magnetic resonance imaging–guided adaptive radiotherapy of prostate cancer. *Med Phys*. (2022) 49(8):4971–9. doi: 10.1002/mp.15793

33. Chmelik J, Jakubicek R, Walek P, Jan J, Ourednicek P, Lambert L, et al. Deep convolutional neural network-based segmentation and classification of difficult to define metastatic spinal lesions in 3D CT data. *Med Image Anal*. (2018) 49:76–88. doi: 10.1016/j.media.2018.07.008

34. Ding Y, Chen Z, Wang Z, Wang X, Hu D, Ma P, et al. Three-dimensional deep neural network for automatic delineation of cervical cancer in planning computed tomography images. *J Appl Clin Med Phys*. (2022) 23(4):e13566. doi: 10.1002/acm2.13566

35. Faghani S, Baffour FI, Ringler MD, Hamilton-Cave M, Rouzrokh P, Moassefi M, et al. A deep learning algorithm for detecting lytic bone lesions of multiple myeloma on CT. *Skelet Radiol*. (2023) 52(1):91–8. doi: 10.1007/s00256-022-04160-z

36. Fan X, Zhang X, Zhang Z, Jiang Y. Deep learning-based identification of spinal metastasis in lung cancer using spectral CT images. *Sci Program*. (2021) Article ID 2779390, 7. doi: 10.1155/2021/2779390

37. Fan X, Zhang X, Zhang Z, Jiang Y. Deep learning on MRI images for diagnosis of lung cancer spinal bone metastasis. *Contrast Media Mol Imaging*. (2021) 2021:5294379. doi: 10.1155/2021/5294379

38. Huang L, Xia W, Zhang B, Qiu B, Gao X. MSFCN-multiple supervised fully convolutional networks for the osteosarcoma segmentation of CT images. *Comput Meth Prog Bio*. (2017) 143:67–74. doi: 10.1016/j.cmpb.2017.02.013

39. Huo T, Xie Y, Fang Y, Wang Z, Liu P, Duan Y, et al. Deep learning-based algorithm improves radiologists' performance in lung cancer bone metastases detection on computed tomography. *Front Oncol*. (2023) 13:1125637. doi: 10.3389/fonc.2023.1125637

40. Hwang D, Kang SK, Kim KY, Seo S, Paeng JC, Lee DS, et al. Generation of PET attenuation map for whole-body time-of-flight (18)F-FDG PET/MRI using a deep

neural network trained with simultaneously reconstructed activity and attenuation maps. *J Nucl Med.* (2019) 60(8):1183–9. doi: 10.2967/jnumed.118.219493

41. Jin J, Zhou H, Sun S, Tian Z, Ren H, Feng J, et al. Machine learning based gray-level co-occurrence matrix early warning system enables accurate detection of colorectal cancer pelvic bone metastases on MRI. *Front Oncol.* (2023) 13:1121594. doi: 10.3389/fonc.2023.1121594

42. Johnsson K, Brynolfsson J, Sahlstedt H, Nickols NG, Rettig M, Probst S, et al. Analytical performance of aPROMISE: automated anatomic contextualization, detection, and quantification of [(18)F]DCFPyL (PSMA) imaging for standardized reporting. *Eur J Nucl Med Mol Imaging.* (2022) 49(3):1041–51. doi: 10.1007/s00259-021-05497-8

43. Ling Z, Yang S, Gou F, Dai Z, Wu J. Intelligent assistant diagnosis system of osteosarcoma MRI image based on transformer and convolution in developing countries. *IEEE J Biomed Health Inform.* (2022) 26(11):5563–74. doi: 10.1109/JBHI.2022.3196043

44. Lingappa E, Parvathy LR. Deep learning-based active contour technique with bagging and boosting algorithms hybrid approach for detecting bone cancer from mri scan images. *Multimed Tools Appl.* (2023). doi: 10.1007/s11042-023-14811-5

45. Liu F, Zhu J, Lv B, Yang L, Sun W, Dai Z, et al. Auxiliary segmentation method of osteosarcoma MRI image based on transformer and U-net. *Comput Intell Neurosci.* (2022) 2022:9990092. doi: 10.1155/2022/9990092

46. Liu F, Gou F, Wu J. An attention-preserving network-based method for assisted segmentation of osteosarcoma MRI images. *Mathematics.* (2022) 10:1665. doi: 10.3390/math10101665

47. Liu X, Han C, Cui Y, Xie T, Zhang X, Wang X. Detection and segmentation of pelvic bones metastases in MRI images for patients with prostate cancer based on deep learning. *Front Oncol.* (2021) 11:773299. doi: 10.3389/fonc.2021.773299

48. Liu X, Han C, Wang H, Wu J, Cui Y, Zhang X, et al. Fully automated pelvic bone segmentation in multiparameteric MRI using a 3D convolutional neural network. *Insights Imaging.* (2021) 12(1):93. doi: 10.1186/s13244-021-01044-z

49. Lv B, Liu F, Li Y, Nie J, Gou F, Wu J. Artificial intelligence-aided diagnosis solution by enhancing the edge features of medical images. *Diagnostics.* (2023) 13(6):1063. doi: 10.3390/diagnostics13061063

50. Moreau N, Rousseau C, Fourcade C, Santini G, Ferrer L, Lacombe M, et al. Deep learning approaches for bone and bone lesion segmentation on 18FDG PET/CT imaging in the context of metastatic breast cancer(). *Annu Int Conf IEEE Eng Med Biol Soc.* (2020) 2020:1532–5. doi: 10.1109/EMBC44109.2020.9175904

51. Moreau N, Rousseau C, Fourcade C, Santini G, Ferrer L, Lacombe M, et al. Influence of inputs for bone lesion segmentation in longitudinal (18)F-FDG PET/CT imaging studies. *Annu Int Conf IEEE Eng Med Biol Soc.* (2022) 2022:4736–9. doi: 10.1109/EMBC48229.2022.9871081

52. Ouyang T, Yang S, Gou F, Dai Z, Wu J. Rethinking U-net from an attention perspective with transformers for osteosarcoma MRI image segmentation. *Comput Intell Neurosci.* (2022) 2022:7973404. doi: 10.1155/2022/7973404

53. Qu Y, Li X, Yan Z, Zhao L, Zhang L, Liu C, et al. Surgical planning of pelvic tumor using multi-view CNN with relation-context representation learning. *Med Image Anal.* (2021) 69:101954. doi: 10.1016/j.media.2020.101954

54. Schott B, Weisman AJ, Perk TG, Roth AR, Liu G, Jeraj R. Comparison of automated full-body bone metastases delineation methods and their corresponding prognostic power. *Phys Med Biol.* (2023) 68(3):035011. doi: 10.1088/1361-6560/acaf22

55. Shen Y, Gou F, Dai Z. Osteosarcoma MRI image-assisted segmentation system base on guided aggregated bilateral network. *Mathematics.* (2022) 10(7):1090. doi: 10.3390/math10071090

56. Shuai L, Zou W, Hu N, Gao X, Wang J. An advanced W-shaped network with adaptive multi-scale supervision for osteosarcoma segmentation. *Bio Signal Process Control.* (2023) 80:104243. doi: 10.1016/j.bspc.2022.104243

57. Trägårdh E, Enqvist O, Ulén J, Jögi J, Bitzén U, Hedeer F, et al. Freely available, fully automated AI-based analysis of primary tumour and metastases of prostate cancer in whole-body [(18)F]-PSMA-1007 PET-CT. *Diagnostics.* (2022) 12:2101. doi: 10.3390/diagnostics12092101

58. Wang J, Shi X, Yao X, Ren J, Du X. Deep learning-based CT imaging in diagnosing myeloma and its prognosis evaluation. *J Heal Eng.* (2021) 2021:5436793. doi: 10.1155/2021/5436793

59. Wang L, Yu L, Zhu J, Tang H, Gou F, Wu J. Auxiliary segmentation method of osteosarcoma in MRI images based on denoising and local enhancement. *Healthcare.* (2022) 10:1468. doi: 10.3390/healthcare10081468

60. Wu J, Guo Y, Gou F, Dai Z. A medical assistant segmentation method for MRI images of osteosarcoma based on DecoupleSegNet. *Int J Intell Syst.* (2022) 37(11):8436–61. doi: 10.1002/int.22949

61. Wu J, Liu Z, Gou F, Zhu J, Tang H, Zhou X, et al. BA-GCA Net: boundary-aware grid contextual attention net in osteosarcoma MRI image segmentation. *Comput Intell Neurosci.* (2022) 2022:3881833. doi: 10.1155/2022/3881833

62. Wu J, Xiao P, Huang H, Gou F, Zhou Z, Dai Z. An artificial intelligence multiprocessing scheme for the diagnosis of osteosarcoma MRI images. *IEEE J Biomed Health Inf.* (2022) 26(9):4656–67. doi: 10.1109/JBHI.2022.3184930

63. Wu J, Yang S, Gou F, Zhou Z, Xie P, Xu N, et al. Intelligent segmentation medical assistance system for MRI images of osteosarcoma in developing countries. *Comput Math Methods Med.* (2022) 2022:7703583. doi: 10.1155/2022/7703583

64. Wu J, Zhou L, Gou F, Tan Y. A residual fusion network for osteosarcoma MRI image segmentation in developing countries. *Comput Intell Neurosci.* (2022) 2022:7285600. doi: 10.1155/2022/7285600

65. Xu L, Tetteh G, Lipkova J, Zhao Y, Li H, Christ P, et al. Automated whole-body bone lesion detection for multiple myeloma on (68)Ga-pentixafor PET/CT imaging using deep learning methods. *Contrast Media Mol Imaging.* (2018) 2018:2391925. doi: 10.1155/2018/2391925

66. Yildiz Potter I, Yeritsyan D, Mahar S, Wu J, Nazarian A, Vaziri A. Automated bone tumor segmentation and classification as benign or malignant using computed tomographic imaging. *J Digit Imaging.* (2023) 36:869–78. doi: 10.1007/s10278-022-00771-z

67. Zhan X, Liu J, Long H, Zhu J, Tang H, Gou F, et al. An intelligent auxiliary framework for bone malignant tumor lesion segmentation in medical image analysis. *Diagnostics.* (2023) 13(2):223. doi: 10.3390/diagnostics13020223

68. Zhang R, Huang L, Xia W, Zhang B, Qiu B, Gao X. Multiple supervised residual network for osteosarcoma segmentation in CT images. *Comput Med Imaging Graph.* (2018) 63:1–8. doi: 10.1016/j.compmedimag.2018.01.006

69. Zhao Y, Gafita A, Vollnberg B, Tetteh G, Haupt F, Afshar-Oromieh A, et al. Deep neural network for automatic characterization of lesions on (68)Ga-PSMA-11 PET/CT. *Eur J Nucl Med Mol Imaging.* (2020) 47(3):603–13. doi: 10.1007/s00259-019-04606-y

70. Paranavithana IR, Stirling D, Ros M, Field M. Systematic review of tumor segmentation strategies for bone metastases. *Cancers (Basel).* (2023) 15(6):1750. doi: 10.3390/cancers15061750

71. Faiella E, Santucci D, Calabrese A, Russo F, Vadalà G, Zobel BB, et al. Artificial intelligence in bone metastases: an MRI and CT imaging review. *Int J Environ Res Public Health.* (2022) 19(3):1880. doi: 10.3390/ijerph19031880

72. Yang C, Qin LH, Xie YE, Liao JY. Deep learning in CT image segmentation of cervical cancer: a systematic review and meta-analysis. *Radiat Oncol.* (2022) 17(1):175. doi: 10.1186/s13014-022-02148-6

73. Carvalho LE, Sobieranski AC, von Wangenheim A. 3D Segmentation algorithms for computerized tomographic imaging: a systematic literature review. *J Digit Imaging.* (2018) 31(6):799–850. doi: 10.1007/s10278-018-0101-z

74. Domingues I, Pereira G, Martins P, Duarte H, Santos J, Abreu PH. Using deep learning techniques in medical imaging: a systematic review of applications on CT and PET. *Artif Intell Rev.* (2020) 53(6):4093–160. doi: 10.1007/s10462-019-09788-3

75. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. *J Digit Imaging.* (2017) 30(4):449–59. doi: 10.1007/s10278-017-9983-4

76. Shal K, Choudhry MS. Evolution of deep learning algorithms for MRI-based brain tumor image segmentation. *Crit Rev Biomed Eng.* (2021) 49(1):77–94. doi: 10.1615/CritRevBiomedEng.2021035557

77. Gul S, Khan MS, Bibi A, Khandakar A, Ayari MA, Chowdhury MEH. Deep learning techniques for liver and liver tumor segmentation: a review. *Comput Biol Med.* (2022) 147:105620. doi: 10.1016/j.compbiomed.2022.105620

78. Fleiss J. The measurement of interrater agreement. In: *Statistical methods for rates and proportions.* 2nd ed. New York: John Wiley & Sons (1981). 212–36.

79. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging.* (2015) 15(1):29. doi: 10.1186/s12880-015-0068-x

80. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric analysis of white matter lesions in MR images: method and validation. *IEEE Trans Med Imaging.* (1994) 13(4):716–24. doi: 10.1109/42.363096

81. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, Haker SJ, et al. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol.* (2004) 11(2):178–89. doi: 10.1016/S1076-6332(03)00671-8

82. Alessio AM, Kinahan PE, Cheng PM, Vesselle H, Karp JS. PET/CT scanner instrumentation, challenges, and solutions. *Radiol Clin North Am.* (2004) 42(6):1017–32. doi: 10.1016/j.rcl.2004.08.001

83. Lin TY, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, et al. Microsoft COCO: Common Objects in Context. *arXiv.* (2015). Available at: http://arxiv.org/abs/1405.0312 (Cited Jun 11, 2023).

84. Yin XX, Sun L, Fu Y, Lu R, Zhang Y. U-Net-Based medical image segmentation. *J Healthc Eng.* (2022) 2022:e4189781. doi: 10.1155/2022/4189781

85. Carrino JA. An artificially intelligent solution for a real problem in musculoskeletal radiology: bone tumors. *Radiology.* (2021) 301(2):407–8. doi: 10.1148/radiol.2021211560

86. Von Schacky CE, Wilhelm NJ, Schäfer VS, Leonhardt Y, Gassert FG, Foreman SC, et al. Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. *Radiology.* (2021) 301(2):398–406. doi: 10.1148/radiol.2021204531

87. Zbinden L, Doorenbos L, Pissas T, Huber AT, Sznitman R, Márquez-Neila P. Stochastic Segmentation with Conditional Categorical Diffusion Models. *arXiv.* (2023). Available at: http://arxiv.org/abs/2303.08888 (Cited Jun 16, 2023).

*CORRESPONDENCE
Bino A. Varghese
✉ bino.varghese@med.usc.edu

†These authors have contributed equally to this work and share first authorship

Spatial assessments in texture analysis: what the radiologist needs to know

Bino A. Varghese[1][*][†], Brandon K. K. Fields[2][†], Darryl H. Hwang[1], Vinay A. Duddalwar[1], George R. Matcuk Jr[3] and Steven Y. Cen[1]

[1]Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States, [2]Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States, [3]Department of Radiology, Cedars-Sinai Medical Center, Los Angeles, CA, United States

To date, studies investigating radiomics-based predictive models have tended to err on the side of data-driven or exploratory analysis of many thousands of extracted features. In particular, spatial assessments of texture have proven to be especially adept at assessing for features of intratumoral heterogeneity in oncologic imaging, which likewise may correspond with tumor biology and behavior. These spatial assessments can be generally classified as spatial filters, which detect areas of rapid change within the grayscale in order to enhance edges and/or textures within an image, or neighborhood-based methods, which quantify gray-level differences of neighboring pixels/voxels within a set distance. Given the high dimensionality of radiomics datasets, data dimensionality reduction methods have been proposed in an attempt to optimize model performance in machine learning studies; however, it should be noted that these approaches should only be applied to training data in order to avoid information leakage and model overfitting. While area under the curve of the receiver operating characteristic is perhaps the most commonly reported assessment of model performance, it is prone to overestimation when output classifications are unbalanced. In such cases, confusion matrices may be additionally reported, whereby diagnostic cut points for model predicted probability may hold more clinical significance to clinical colleagues with respect to related forms of diagnostic testing.

KEYWORDS

radiomics, texture analysis, spatial assessment, machine learning, artificial intelligence

Key points

- Features of intratumoral heterogeneity are well-represented by spatial assessments of texture, which may similarly correlate with tumor biology and behavior.
- Spatial filters are used to enhance edges and/or textures of an image by identifying areas of rapid change within the grayscale.
- Neighborhood-based methods are higher-order texture approaches which quantify differences in gray-level intensities of particular regions of interest with respect to their neighbors within a set distance.

Abbreviations

GLCM, Gray-Level Co-Occurrence Matrix; GLRLM, Gray-Level Run-Length Matrix; GLSZM, Gray-Level Size-Zone Matrix; GLDM, Gray-Level Dependence Matrix; NGTDM, Neighborhood Gray-Tone Difference Matrix; FD, Fractal Dimension; PCA, Principal Component Analysis; AUC, area under the curve; ROC, receiver operating characteristic; VOI, variable of importance.

Introduction

Quantitative assessments of imaging texture characteristics have been successfully applied to answer a variety of clinically-relevant queries ranging from lesion classification to disease prognostication, often in the form of radiomics-based machine learning decision classifiers (1–13). While some approaches have previously relied on filtering of high-dimensionality data to identify the most contributory features or classes of features (14–17), recent studies have demonstrated a subset of texture metrics well-equipped to detect regions of heterogeneity in the imaging grayscale (4, 9) (**Supplementary Table S1**). These "spatial assessments" are aptly named for their ability to resolve subtle areas of voxel-to-voxel variation, or in plainer terms, what might be subjectively referred to as "coarseness" by a human interpreting radiologist (4, 7, 8, 18–23). In this review, we detail the various common approaches to spatial assessment of imaging texture, as well as their applicability and implications in future radiomics and machine learning-related studies.

Approaches to spatial assessment

Spatial filters

Spatial filters are image processing methods that enhance spatial image properties of a region of interest such as edges and/or textures (23–25). The size and shape of the filter neighborhood or convolution kernel determines the performance of the filter, and warrants standardization across multiple studies to evaluate reliability (26). Some commonly used spatial filters for texture analysis include statistical filters such as entropy filters, range filters, standard deviation filters, median filters, and average filters. However, given that use of spatial filters can lead to an increase in radiomics feature space (27), it is advised to avoid using these approaches with small sample sizes.

Directional gradients and direction invariant gradients have been used to improve edge enhancement. For example, edge filters such as Kirsch and Sobel have been reported as part of multiple radiomics panels (28, 29). Likewise, the Laplacian of Gaussian filter, which captures edges based on detecting areas of rapid change in grayscale intensity and then smooths them with a standard-deviation tunable Gaussian bandpass filter, has been reported frequently in radiomics panels to capture areas with increasingly coarse texture patterns (24, 27, 30). Kernels such as the Laws filters identify specific textures based on five fundamental vectors that emphasize features of edge, level, spot, ripple, and wave, or a combination thereof, and have been used for spatial filtering prior to feature extraction (23, 31).

In some cases, noise can be suppressed using image transforms, such as Fourier analysis (24). In this method, spatial domain information can be converted to frequency domain information and then filtered for high frequencies, low frequencies, bandpass, etc. However, while the signal to noise ratio can be improved, this technique merely suppresses the noise without improving the strength of the underlying signal (32). Wavelet transforms further build upon the Fourier technique by decomposing the original image in both spatial and frequency domains, thereby providing relatively more precise signal localization (24, 27, 33, 34). The coefficients of these decomposed sub-bands can then be weighted to enhance specific signal properties along select directions of a 3-dimensional space.

Neighborhood-based methods

Statistical characterizations of texture can also be assessed from higher-order texture methods (i.e., analysis based on both grayscale values and their spatial orientation) such as Gray-Level Co-Occurrence Matrix (GLCM), Gray-Level Run-Length Matrix (GLRLM), Gray-Level Size-Zone Matrix (GLSZM), Gray-Level Dependence Matrix (GLDM) and Neighborhood Gray-Tone Difference Matrix (NGTDM) (35, 36). In all of these methods, the metrics generated essentially quantify the differences in grayscale brightness between neighboring pixels/voxels (9, 27, 37). For example, in GLCM, texture is quantified based on how often a combination of gray-level values occur next to each other at a given distance and direction within a region of interest (23, 27, 31, 37) (**Figure 1**, top row). Some commonly reported GLCM metrics include energy, contrast, entropy, homogeneity, correlation, variance, sum average, and autocorrelation (9, 35, 36) (**Figure 2**).

In contrast to GLCM, GLRLM quantifies the pattern of gray-level intensity pixels in a fixed direction from an interference pixel (**Figure 1**, middle row). Run-length is defined as the number of adjacent pixels that have the same gray-level intensity in each direction (37). Some commonly reported GLRLM metrics include short and long run emphasis, gray-level non-uniformity, run-length non uniformity, low and high gray-level run emphasis, and their combinations (9, 35).

Similar to GLCM, in GLSZM texture is also quantified based on how often a combination of gray-level values occurs next to each other at a given distance within a region of interest (27, 37); however, in contrast to GLCM, GLSZM is direction independent (40) (**Figure 1**, bottom row). Some commonly reported GLSZM metrics include short and long zone emphasis, gray-level non-uniformity, zone-size non-uniformity, low and high gray-level zone emphasis, and their combinations (9, 35).

Likewise, GLDM quantifies the number of connected voxels within a set distance that are dependent on a center voxel (37). A neighboring voxel is considered dependent on the center voxel if the absolute difference of their respective gray-levels falls within a set value (9, 41). Some commonly reported GLDM metrics include short and long dependence emphasis, gray-level non-uniformity, dependence non-uniformity, gray-level and dependence variance, and high grey-level zone emphasis, and their combinations.

Lastly, NGTDM evaluates the difference between a particular gray-level intensity and the average gray-level intensity of its neighborhood within a given distance (23, 37, 42). Some

FIGURE 1

Top row: (A) grayscale image with four different gray-levels. (B) Digitized version of the gray-level image with unique numerical values corresponding to the gray-level or a range of gray-levels (dependent on bin size of bin width) for each theoretical pixel/voxel. (C) GLCM map of the image obtained for distance 1 and direction 0 degrees. (D) This same process is then repeated in all other directions: i.e., 45, 90, and 135 deg, respectively. To obtain direction invariant results, all results are normalized and averaged. Middle row: (A) Grayscale image with four different gray-levels. (B) Digitized version of the gray-level image with unique numerical values corresponding to the gray-level or a range of gray-levels (dependent on bin size of bin width) for each theoretical pixel/voxel. (C) GLRLM map of the image obtained for direction zero degrees. This same process is then repeated in all other directions: i.e., 45, 90, and 135 deg, respectively. To obtain direction invariant results all results are normalized and averaged. Bottom row: (A) grayscale image with four different gray-levels. (B) Digitized version of the gray-level image with unique numerical values corresponding to the gray-level or a range of gray-levels (dependent on bin size of bin width) for each theoretical pixel/voxel. (C) GLSZM map of the image. GLCM, Gray-Level Co-Occurrence Matrix; GLRLM, Gray-Level Run-Length Matrix; GLSZM, Gray-Level Size-Zone Matrix.

commonly reported NGTDM metrics include busyness, coarseness, contrast, strength, and complexity (9, 35).

Other approaches

Structural methods involve techniques of decomposing an image into basic units and then identifying the rules required to construct that given image from its basic units. For instance, Fractal Dimension (FD) is a metric that evaluates image complexity by quantifying how changes in image scale affect image detail (9, 43, 44). FD uses self-repeating structural patterns in order to quantitatively assess the homogeneity of the region of interest, and increases with greater geometric complexity (35, 43, 45, 46). This in essence functions as an objective evaluation of how consistent a shape is with itself, and thus serves as an excellent measurement of the regularity of a tumor's morphology (23, 44).

Applications in radiomics and machine learning

In oncologic imaging, radiomics analysis has shown great utility in evaluating features of intratumoral heterogeneity, which may correspondingly reflect tumor behavior (4, 5, 7–9, 11, 13, 14, 35, 47, 48). There is a growing body of literature to suggest that radiomics-based machine learning algorithms perform well with various classification tasks, including differentiating benign from malignant lesions, stratifying lesions by tumor grade, predicting risk of distant metastases, and predicting overall survival (1–13). Additional work suggests that subtle differences in the underlying texture grayscale may also correlate well with tumoral genetic and phenotypic variations, furthering the case for potential future integrations of radiomics classifiers as risk stratification schema in prospective clinical workflows (31, 35, 49, 50).

Given the sheer number of radiomics features extracted as part of standard pipeline workflows, analyses of radiomics datasets are
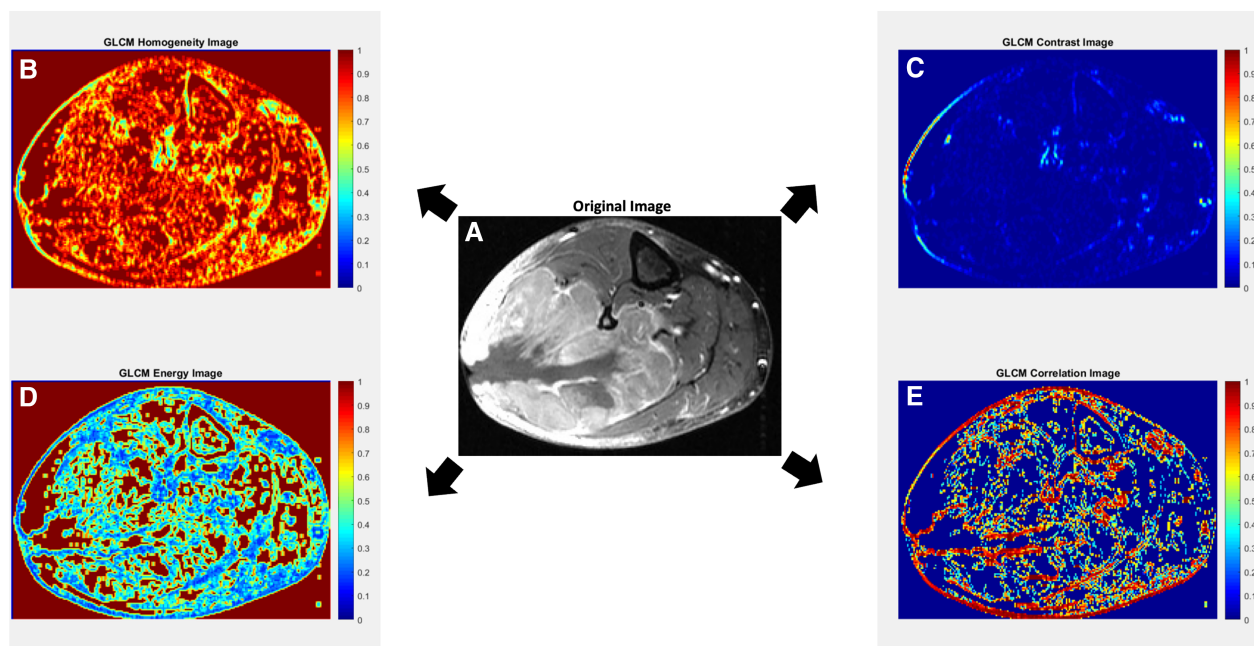
FIGURE 2

Axial contrast-enhanced T1-weighted MR image with fat suppression of a 70-year-old male with leiomyosarcoma of the posterolateral calf (A), with corresponding texture parameter maps for Gray-Level Co-Occurrence Matrix (GLCM) homogeneity (B), GLCM contrast (C), GLCM energy (D) and GLCM correlation (E). The GLCM homogeneity map (B) reflects the closeness of the distribution of elements in the GLCM map relative to the GLCM diagonal. Highly homogenous regions (i.e., regions with less variation; close to the GLCM diagonal) receive a value of 1, while highly heterogenous regions receive a value of 0. The GLCM contrast map (C) measures the intensity contrast between an index pixel and its neighborhood pixels. Regions of high contrast show high heterogeneity in values up to a maximum value of 1. A constant image receives a value of 0. In some studies, contrast may also be referred to as variance and inertia. The GLCM energy map (D) measures the sum of the squared elements in the GLCM, whereby highly homogenous regions receive values of 1 and highly heterogenous regions receive values of 0. In some studies, energy may also be referred to as angular second moment, uniformity, or uniformity of energy. The GLCM correlation map (E) reflects how correlated a given pixel is to its neighboring pixels, with highly correlated regions receiving values of 1. In general, a neighborhood of 3×3 was adopted for the GLCM approach. Original image (A) courtesy of The Cancer Genome Atlas Sarcoma Collection (TCGA-SARC) based on data generated by the TCGA research network: http://cancergenome.nih.gov/ (38, 39).

often necessarily complex and difficult to comprehend. Moreover, segmentation approaches (i.e., manual vs. semi-automated vs. fully-automated) can likewise affect the extracted radiomics parameters and—particularly in the case of manual segmentation —be a source of intra- and inter-observer variability (51, 52). Initial statistical considerations should include descriptive analyses to evaluate for skewness, kurtosis, and outlier detection, which in turn hold implications for the reproducibility of a study (53). Missing data may arise from situations where a given radiomics approach does not yield a numerical value, possibly due to image quality degradation or methodological failure. When working with sufficiently high-quality images, missing radiomics data are rarely encountered; however, missing data become much more prevalent as image quality degrades, and, in such cases, imputation methods will often be inaccurate (54, 55). Given this, we believe best practice is to simply exclude subjects with poor image quality and high numbers of missing radiomics features in order to avoid spurious associations. In cases of random missing phases in multiphase studies, we have found in our own research paradigms that imputation methods, such as the Markov Chain Monte Carlo (MCMC) method, work well given high correlation of radiomics features between contrast phases (56).

Data dimensionality reduction methods have often been described in the literature with both supervised and unsupervised machine learning constructions in an attempt to optimize classifier performance. These approaches mainly include data filtering, principal component analysis (PCA), and elimination of highly correlated features (57). However, if used, dimensionality reduction techniques must only be conducted with the training data in order to avoid information leakage, which can in turn bias the decision classifiers and lead to problems of overfitting (58). For example, PCA often suffers from poor reproducibility when applied to test data because its components are derived to maximize the variance explained in the training data (57, 59). Instead, we recommend that removal of highly correlated data (e.g., redundant features with $r > 0.8$ suggesting collinearity) should be performed as the initial approach for dimensionality reduction (60).

Reporting of machine learning performance for radiomics based models is commonly done using area under the curve (AUC) of the receiver operating characteristic (ROC). In general, while AUC can well-represent overall model prediction accuracies, it is prone to overestimating performance in cases of unbalanced classification outputs. To overcome this, a common approach is to report confusion matrices—including sensitivity

(recall), specificity (selectivity), positive predictive value (precision), and negative predictive value—corresponding to various cut points for model predicted probability. These values likewise tend to be more easily understood by clinical colleagues, whereby diagnostic cut points in some ways hold more tangible clinical significance with respect to other forms of diagnostic testing. To obtain optimal cut points, common practice includes statistical approaches such as Youden's J statistic (also referred to as Youden's index), defined as J = sensitivity + specificity—1, or simply selecting the cut points that maximize the product of sensitivity and specificity (61, 62). An arguably more sophisticated approach would be to adapt the concept of decision analysis. Decision analysis includes assessing for clinical value by also considering clinical consequences when making determinations of cut point appropriateness, such as weighing the benefits of finding a malignant tumor against the harms of unnecessary biopsies (63). Finally, reporting of machine learning performance should also highlight the variables of importance (VOIs). VOIs are defined as those metrics which are identified as having the greatest impact on classification accuracy and tend to be the most robust features for predicting the queried clinical outcomes. While different machine learning approaches have different methods for selecting VOIs, many also incorporate some form of ranking procedure based on the relative contribution of each metric or class of metrics. These rankings may in turn be useful for identifying potential correlative relationships between the investigated quantitative imaging features and phenotypic observations of disease state (64, 65).

Conclusion

Machine learning analyses of radiomics feature sets have been applied to a wide array of classification and prognostication tasks in oncologic imaging. Spatial assessments in particular have shown great potential to quantitatively evaluate features of intratumoral heterogeneity and may one day prove to be important prognostic biomarkers of phenotypic behavior in oncologic care. In this review, we discussed some of the most common approaches to spatial assessment of texture in radiologic imaging as well as

familiar reporting metrics to assess model performance in machine learning studies.

Author Contributions

BAV and BKKF conceived the idea. BAV, BKKF, and SYC drafted the manuscript with support from DHH, VAD, and GRM, who provided input and feedback throughout. All authors reviewed the manuscript and provided edits prior to submission. All authors contributed to the article and approved the submitted version.

Conflict of interest

GRM is a consultant for Canon Medical Systems, USA. VAD is a consultant for Radmetrix and Westat and serves on the advisory board for DeepTek.

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) BKKF, BAV and GRM declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

Supplementary material

References

1. Juntu J, Sijbers J, De Backer S, Rajan J, Van Dyck D. Machine learning study of several classifiers trained with texture analysis features to differentiate benign from malignant soft-tissue tumors in T1-MRI images. *J Magn Reson Imaging*. (2010) 31 (3):680–9. doi: 10.1002/jmri.22095

2. Wang H, Nie P, Wang Y, Xu W, Duan S, Chen H, et al. Radiomics nomogram for differentiating between benign and malignant soft-tissue masses of the extremities. *J Magn Reson Imaging*. (2020) 51(1):155–63. doi: 10.1002/jmri.26818

3. Wang H, Zhang J, Bao S, Liu J, Hou F, Huang Y, et al. Preoperative MRI-based radiomic machine-learning nomogram may accurately distinguish between benign and malignant soft-tissue lesions: a two-center study. *J Magn Reson Imaging*. (2020) 52(3):873–82. doi: 10.1002/jmri.27111

4. Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture analysis of imaging: what radiologists need to know. *AJR Am J Roentgenol*. (2019) 212(3):520–8. doi: 10. 2214/AJR.18.20624

5. Fields BKK, Hwang D, Cen S, Desai B, Gulati M, Hu J, et al. Quantitative magnetic resonance imaging (q-MRI) for the assessment of soft-tissue sarcoma treatment response: a narrative case review of technique development. *Clin Imaging*. (2020) 63:83–93. doi: 10.1016/j.clinimag.2020.02.016

6. Jin J, Zhou H, Sun S, Tian Z, Ren H, Feng J, et al. Machine learning based gray-level co-occurrence matrix early warning system enables accurate detection of colorectal cancer pelvic bone metastases on MRI. *Front Oncol*. (2023) 13:1121594. doi: 10.3389/fonc.2023.1121594

7. Fields BKK, Demirjian NL, Hwang DH, Varghese BA, Cen SY, Lei X, et al. Whole-tumor 3D volumetric MRI-based radiomics approach for distinguishing between benign and malignant soft tissue tumors. *Eur Radiol*. (2021) 31 (11):8522–35. doi: 10.1007/s00330-021-07914-w

8. Fields BKK, Demirjian NL, Cen SY, Varghese BA, Hwang DH, Lei X, et al. Predicting soft tissue sarcoma response to neoadjuvant chemotherapy using an

MRI-based delta-radiomics approach. *Mol Imaging Biol*. (2023) 25(4):776–87. doi: 10.1007/s11307-023-01803-y

9. Mayerhoefer ME, Materka A, Langs G, Haggstrom I, Szczypinski P, Gibbs P, et al. Introduction to radiomics. *J Nucl Med*. (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893

10. Corino VDA, Montin E, Messina A, Casali PG, Gronchi A, Marchiano A, et al. Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions. *J Magn Reson Imaging*. (2018) 47(3):829–40. doi: 10.1002/jmri.25791

11. Peeken JC, Neumann J, Asadpour R, Leonhardt Y, Moreira JR, Hippe DS, et al. Prognostic assessment in high-grade soft-tissue sarcoma patients: a comparison of semantic image analysis and radiomics. *Cancers (Basel)*. (2021) 13(8):1929. doi: 10.3390/cancers13081929

12. Vallieres M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol*. (2015) 60(14):5471–96. doi: 10.1088/0031-9155/60/14/5471

13. Urso L, Manco L, Castello A, Evangelista L, Guidi G, Castellani M, et al. PET-derived radiomics and artificial intelligence in breast cancer: a systematic review. *Int J Mol Sci*. (2022) 23(21):13409. doi: 10.3390/ijms232113409

14. Crombe A, Perier C, Kind M, De Senneville BD, Le Loarer F, Italiano A, et al. T2-based MRI delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy. *J Magn Reson Imaging*. (2019) 50(2):497–510. doi: 10.1002/jmri.26589

15. Gao Y, Kalbasi A, Hsu W, Ruan D, Fu J, Shao J, et al. Treatment effect prediction for sarcoma patients treated with preoperative radiotherapy using radiomics features from longitudinal diffusion-weighted MRIs. *Phys Med Biol*. (2020) 65(17):175006. doi: 10.1088/1361-6560/ab9e58

16. Miao L, Cao Y, Zuo L, Zhang H, Guo C, Yang Z, et al. Predicting pathological complete response of neoadjuvant radiotherapy and targeted therapy for soft tissue sarcoma by whole-tumor texture analysis of multisequence MRI imaging. *Eur Radiol*. (2023) 33(6):3984–94. doi: 10.1007/s00330-022-09362-6

17. Peeken JC, Asadpour R, Specht K, Chen EY, Klymenko O, Akinkuoroye V, et al. MRI-based delta-radiomics predicts pathologic complete response in high-grade soft-tissue sarcoma patients treated with neoadjuvant therapy. *Radiother Oncol*. (2021) 164:73–82. doi: 10.1016/j.radonc.2021.08.023

18. Cook GJ, Yip C, Siddique M, Goh V, Chicklore S, Roy A, et al. Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy? *J Nucl Med*. (2013) 54(1):19–26. doi: 10.2967/jnumed.112.107375

19. Leijenaar RT, Carvalho S, Hoebers FJ, Aerts HJ, van Elmpt WJ, Huang SH, et al. External validation of a prognostic CT-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta Oncol*. (2015) 54(9):1423–9. doi: 10.3109/0284186X.2015.1061214

20. Parmar C, Leijenaar RT, Grossmann P, Rios Velazquez E, Bussink J, Rietveld D, et al. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *Sci Rep*. (2015) 5:11044. doi: 10.1038/srep11044

21. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer. *J Nucl Med*. (2011) 52(3):369–78. doi: 10.2967/jnumed.110.082404

22. Chu Y, Li L, Goldgof DB, Qui Y, Clark RA. *Medical imaging 2003: image processing, may 15. Classification of masses on mammograms using support vector machine. Proc SPIE;5032* (2003). doi: 10.1117/12.481142

23. Parekh V, Jacobs MA. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev*. (2016) 1(2):207–26. doi: 10.1080/23808993.2016.1164013

24. Ganeshan B, Miles KA. Quantifying tumour heterogeneity with CT. *Cancer Imaging*. (2013) 13(1):140–9. doi: 10.1102/1470-7330.2013.0015

25. Depeursinge A, Andrearczyk V, Whybra P, van Griethuysen J, Müller H, Schaer R, et al. Standardised convolutional filtering for radiomics: Image Biomarker Standardisation Initiative (IBSI). (2023). version 8. Preprint at: https://arxivorg/abs/200605470 doi: 10.48550/arXiv.2006.05470

26. Coggins JM, Jain AK. A spatial filtering approach to texture analysis. *Pattern Recognit Lett*. (1985) 3(3):195–203. doi: 10.1016/0167-8655(85)90053-4

27. Shur JD, Doran SJ, Kumar S, Ap Dafydd D, Downey K, O'Connor JPB, et al. Radiomics in oncology: a practical guide. *Radiographics*. (2021) 41(6):1717–32. doi: 10.1148/rg.2021210037

28. Kirsch RA. Computer determination of the constituent structure of biological images. *Comput Biomed Res*. (1971) 4(3):315–28. doi: 10.1016/0010-4809(71)90034-6

29. Gnep K, Fargeas A, Gutierrez-Carvajal RE, Commandeur F, Mathieu R, Ospina JD, et al. Haralick textural features on T2-weighted MRI are associated with biochemical recurrence following radiotherapy for peripheral zone prostate cancer. *J Magn Reson Imaging*. (2017) 45(1):103–17. doi: 10.1002/jmri.25335

30. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology*. (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169

31. Aerts HJ, Grossmann P, Tan Y, Oxnard GR, Rizvi N, Schwartz LH, et al. Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC. *Sci Rep*. (2016) 6:33860. doi: 10.1038/srep33860

32. Varghese BA, Chen F, Hwang DH, Cen SY, Gill IS, Duddalwar VA. Differentiating solid, non-macroscopic fat containing, enhancing renal masses using fast Fourier transform analysis of multiphase CT. *Br J Radiol*. (2018) 91(1089):20170789. doi: 10.1259/bjr.20170789

33. Varghese B, Hwang D, Mohamed P, Cen S, Deng C, Chang M, Duddalwar V. *Wavelets analysis for differentiating solid, non-macroscopic fat containing, enhancing renal masses: a pilot study. Proc SPIE 10572, 13th international symposium on medical information processing and analysis, 105720 T, November 17* (2017). doi: 10.1117/12.2285948

34. Jog C, Varghese BA, Hwang DH, Cen SY, Aron M, Desai M, Duddalwar VA. *Differentiating clear cell renal cell carcinoma from oncocytoma using curvelet transform analysis of multiphase CT: preliminary study. Proc SPIE 11330, 15th international symposium on medical information processing and analysis, 1133009, January 3* (2020). doi: 10.1117/12.2540169

35. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med*. (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071

36. Lei M, Varghese B, Hwang D, Cen S, Lei X, Desai B, et al. Benchmarking Various radiomic toolkit features while applying the image biomarker standardization initiative toward clinical translation of radiomic analysis. *J Digit Imaging*. (2021) 34(5):1156–70. doi: 10.1007/s10278-021-00506-6

37. Cui Y, Yin FF. Impact of image quality on radiomics applications. *Phys Med Biol*. (2022) 67(15):15TR03. doi: 10.1088/1361-6560/ac7fd7

38. Roche C, Bonaccio E, Filippini J. The Cancer Genome Atlas Sarcoma Collection (TCGA-SARC) (Version 3) [Data set]. The Cancer Imaging Archive (2016). doi: 10.7937/K9/TCIA.2016.CX6YLSUX

39. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, et al. The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J Digit Imaging*. (2013) 26(6):1045–57. doi: 10.1007/s10278-013-9622-7

40. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Trans Biomed Eng*. (2014) 61(3):630–7. doi: 10.1109/TBME.2013.2284600

41. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Graph Image Processing*. (1982) 20(3):297. doi: 10.1016/0146-664x(82)90093-4

42. Amadasun M, King R. Textural features corresponding to textural properties. *IEEE trans syst Man Cybern*. (1989) 19(5):1264–74. doi: 10.1109/21.44046

43. Barucci A, Farnesi D, Ratto F, Pelli S, Pini R, Carpi R, Esposito M, Olmastroni M, Romei C, Taliani A, Materassi M. *Fractal-Radiomics as complexity analysis of CT and MRI cancer images. Proc IEEE 18236217, 39th institute of eletrical and electronics engineers symposium on computer applications & industrial electronics, October 10-12* (2018). doi: 10.1109/CompEng.2018.8536249

44. Yin T, Zhao T, Cen S, Lei X, Hwang D, Hajian S, Desai M, Gill I, Duddalwar V, Varghese BA. *Multi-platform fractal dimension analysis of renal masses from multiphase contrast-enhanced computed tomography as markers for tumor subtyping. Proc SPIE 12567, 18th international symposium on medical information processing and analysis, 125670I, march 6* (2023). doi: 10.1117/12.2670379

45. Kisan S, Mishra S, Rout SB. Fractal dimension in medical imaging: a review. *Int Res J Eng Technol*. (2017) 4(5):1102–6.

46. Fortin CS, Kumaresan R, Ohley WJ, Hoefer S. Fractal dimension in the analysis of medical images. *IEEE Eng Med Biol Mag*. (1992) 11(2):65–71. doi: 10.1109/51.139039

47. Demirjian NL, Varghese BA, Cen SY, Hwang DH, Aron M, Siddiqui I, et al. CT-based radiomics stratification of tumor grade and TNM stage of clear cell renal cell carcinoma. *Eur Radiol*. (2022) 32(4):2552–63. doi: 10.1007/s00330-021-08344-4

48. Aerts HJ. The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol*. (2016) 2(12):1636–42. doi: 10.1001/jamaoncol.2016.2631

49. Liu DH, Dani KA, Reddy SS, Lei X, Demirjian NL, Hwang DH, et al. Radiogenomic associations clear cell renal cell carcinoma: an exploratory study. *Oncology*. (2023) 101(6):375–88. doi: 10.1159/000530719

50. Beig N, Patel J, Prasanna P, Hill V, Gupta A, Correa R, et al. Radiogenomic analysis of hypoxia pathway is predictive of overall survival in glioblastoma. *Sci Rep*. (2018) 8(1):7. doi: 10.1038/s41598-017-18310-0

51. Haarburger C, Muller-Franzes G, Weninger L, Kuhl C, Truhn D, Merhof D. Radiomics feature reproducibility under inter-rater variability in segmentations of CT images. *Sci Rep*. (2020) 10(1):12688. doi: 10.1038/s41598-020-69534-6

52. van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical imaging-"how-to" guide and critical reflection. *Insights Imaging*. (2020) 11(1):91. doi: 10.1186/s13244-020-00887-2

53. Hair JF Jr, Hult GTM, Ringle CM, Sarstedt M. *A primer on partial least squares structural equation modeling (PLS-SEM)*. 3rd ed. Thousand Oaks, CA: SAGE Publications, Inc (2021).

54. Johansson P, Hakkinen J. Improving missing value imputation of microarray data by using spot quality weights. *BMC Bioinformatics*. (2006) 7:306. doi: 10.1186/1471-2105-7-306

55. Jager S, Allhorn A, Biessmann F. A benchmark for data imputation methods. *Front Big Data*. (2021) 4:693674. doi: 10.3389/fdata.2021.693674

56. Yap FY, Varghese BA, Cen SY, Hwang DH, Lei X, Desai B, et al. Shape and texture-based radiomics signature on CT effectively discriminates benign from malignant renal masses. *Eur Radiol*. (2021) 31(2):1011–21. doi: 10.1007/s00330-020-07158-0

57. Sotiras A, Gaonkar B, Eavani H, Honnorat N, Varol E, Dong A, et al. Machine learning as a means toward precision diagnostics and prognostics. In: Wu G, Shen D, Sabuncu MR, editors. *Machine learning and medical imaging*. London, UK: Elsevier Inc (2016). p. 299–334. doi: 10.1016/B978-0-12-804076-8.00010-4

58. Demircioglu A. Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. *Insights Imaging*. (2021) 12(1):172. doi: 10.1186/s13244-021-01115-1

59. O'Neil C, Schutt R. *Doing data science: Straight talk from the frontline*. 1st ed. Sebastopol, CA: O'Reilly Media (2014).

60. Shrestha N. Detecting multicollinearity in regression analysis. *Am J Appl Math Stat*. (2020) 8(2):39–42. doi: 10.12691/ajams-8-2-1

61. Liu X. Classification accuracy and cut point selection. *Stat Med*. (2012) 31 (23):2676–86. doi: 10.1002/sim.4509

62. Bewick V, Cheek L, Ball J. Statistics review 13: receiver operating characteristic curves. *Crit Care*. (2004) 8(6):508–12. doi: 10.1186/cc3000

63. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *Br Med J*. (2016) 352:i6. doi: 10.1136/bmj.i6

64. Jiang T, Gradus JL, Rosellini AJ. Supervised machine learning: a brief primer. *Behav Ther*. (2020) 51(5):675–87. doi: 10.1016/j.beth.2020.05.002

65. Wei P, Lu Z, Song J. Variable importance analysis: a comprehensive review. *Reliab Eng Syst Saf*. (2015) 142:399–432. doi: 10.1016/j.ress.2015.05.018

Empowering breast cancer diagnosis and radiology practice: advances in artificial intelligence for contrast-enhanced mammography

Ketki K. Kinkar[1†], Brandon K. K. Fields[2], Mary W. Yamashita[3] and Bino A. Varghese[3*†]

[1]Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, [2]Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States, [3]Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States

Artificial intelligence (AI) applications in breast imaging span a wide range of tasks including decision support, risk assessment, patient management, quality assessment, treatment response assessment and image enhancement. However, their integration into the clinical workflow has been slow due to the lack of a consensus on data quality, benchmarked robust implementation, and consensus-based guidelines to ensure standardization and generalization. Contrast-enhanced mammography (CEM) has improved sensitivity and specificity compared to current standards of breast cancer diagnostic imaging i.e., mammography (MG) and/or conventional ultrasound (US), with comparable accuracy to MRI (current diagnostic imaging benchmark), but at a much lower cost and higher throughput. This makes CEM an excellent tool for widespread breast lesion characterization for all women, including underserved and minority women. Underlining the critical need for early detection and accurate diagnosis of breast cancer, this review examines the limitations of conventional approaches and reveals how AI can help overcome them. The Methodical approaches, such as image processing, feature extraction, quantitative analysis, lesion classification, lesion segmentation, integration with clinical data, early detection, and screening support have been carefully analysed in recent studies addressing breast cancer detection and diagnosis. Recent guidelines described by Checklist for Artificial Intelligence in Medical Imaging (CLAIM) to establish a robust framework for rigorous evaluation and surveying has inspired the current review criteria.

KEYWORDS

contrast enhanced mammography, radiomics, artificial intelligence, machine learning, deep learning, quantitative analysis, breast cancer detection

1 Introduction

Breast cancer is the second most leading cause of cancer death in women globally (1), and early detection is crucial for improved prognosis (2–5). Digital Mammography (DM) is known to reduce breast cancer related deaths by 40%. However, among specific patients, heightened breast density poses a challenge in detecting early-stage small cancers, resulting in a higher rate of false positive callbacks and interval cancers (6, 7). Currently 43% of all

women, 40–85 in age, have dense breasts warranting the need for additional screening beyond DM (8). In recent years, CEM has emerged as a potential option for offering improved sensitivity and specificity compared to current standards of breast cancer diagnostic imaging i.e., mammography (MG) and/or conventional ultrasound (US) (9, 10). CEM uses iodinated contrast to visualize tumour neovascularity and dual-energy DM to create a recombined or iodine image that highlights just the enhancing lesion in the breast (11, 12). CEM has comparable sensitivity to MRI with a much higher specificity, potentially at a much lower cost and higher throughput (13–15). As a natural progression, multiple studies report of the benefits of using CEM for the screening, diagnosis of breast cancers as a cost-effective and viable alternative to the current standards, particularly in women with dense breasts and at relatively higher risk of breast cancer.

From a technical standpoint, CEM employs anode and cathode components in x-ray tubes similar to conventional DM (16). It utilizes low and high-energy x-rays to highlight contrast agent-induced differences, aiding in tissue composition and distribution assessment (17, 18). Thus, CEM employs dual-energy method to produce high-resolution, low-energy digital mammogram images. These images are recombined to create a digitally subtracted image, which can be useful to identify vascularity of a particular lesion (12). Studies have suggested, low energy mammograms obtained as part of CEM protocols is comparable to conventional mammography (19, 20) and though with the added advantage of emphasizing regions of contrast enhancement (21). CEM is currently offered on five different systems across 4 vendors (11, 22). Although a general consensus on how to perform CEM has been followed, a standardized implementation has not been established. This is a difficult task considering the differences in system characteristics across vendors.

CEM has several drawbacks (11), including the risk of mild to severe hypersensitivity reactions due to iodinated contrast administration (23). Patients should be evaluated for a history of contrast material allergy. CEM radiation dose on average requires slightly higher radiation exposure when compared to conventional mammography in phantom studies, though do tend to fall beneath the 3 mGy threshold dose limit set by Mammography Quality Standards guidelines (24, 25). Furthermore, despite enhanced sensitivity of CEM, certain breast lesions may still be undetectable due to their location within the breast; supplementary breast MRI may be required if lesions are anticipated in these areas such as region near chest wall (26). Finally, due to use subtraction techniques, certain CEM-specific artifacts may be visible on the recombined image which likewise can obscure subtle lesion detection (27–29).

In recent years, there has been significant improvement in the field of Artificial Intelligence (AI) in healthcare, leading to better and more prompt treatment for patients. AI is a useful tool to supplement the abilities and knowledge of radiologists, oncologists, and pathologists, ultimately resulting in more precise and effective identification and treatment of breast cancer. The insights offered by Mongan et al. (30) regarding the importance of systematically presenting research findings resonate deeply within the academic and scientific community. Their assertion

highlights that, beyond achieving optimal results in research, the meticulous and structured presentation of these findings is of paramount significance. These guidelines promote transparency, reproducibility, and the ability to generalize research findings. They standardize reporting, elevate research quality, and ensure clinical relevance, providing a shared foundation for researchers, reviewers, and clinicians to understand and assess deep learning studies effectively.

The goal of this review is to provide an overview of some of the basic ideas and advances in the use of for the detection of breast cancer using CEM. The limitations of conventional approaches will be addressed, as well as the ways in which these limitations can be removed using AI. Importantly, the review will include research that has looked at existing AI capabilities, as well as ideas on how these skills can be used in the clinical field.

2 Method

The literature review was conducted on the use of contrast-enhanced mammography (CEM) and artificial intelligence (AI) techniques for predicting malignancy. PubMed database was searched for articles published between 1st January 2018 and 5th October, 2023, using a query: "(contrast-enhanced mammography) AND (deep learning OR radiomics OR artificial intelligence OR quantitative analysis) AND (classification OR detection)". 53 articles that met this initial criteria were identified. Subsequently, each article was rigorously evaluated to ensure that it used CEM in conjunction with AI techniques to predict malignancy, resulting in a final selection of 14 articles. This rigorous selection process was documented in accordance with the PRISMA framework (31), which provides a transparent and structured methodology for article inclusion as shown in Figure 1. The following sections will discuss end-to-end malignancy detection pipelines using contrast-enhanced mammography. Flowchart of these methods is presented in Figure 2.

3 Image acquisition

The availability of CEM in commercial systems from vendors like GE Healthcare, Hologic, and Siemens Healthineers represents a significant advancement in breast imaging, as demonstrated in Table 1, with information sourced from Jeukens (32), Jochelson et al. (11). While optimal imaging parameters for CEM have not been extensively documented in published studies, there are a few generally accepted guidelines. Commonly, low-osmolarity iodine-based contrast in concentrations ranging from 300 mg/ml to 370 mg/ml at 1.5 ml/kg body weight (maximum 150 ml) is intravenously injected prior to image acquisition. Injection rates typically range from 2 to 3 ml per second (11). Among the reviewed studies, a total of 9 investigations made use of GE Healthcare systems, while 1 opted for Hologic systems and 3 opted to use data from both as mentioned in Table 2.

However, it is important to acknowledge that providing explicit details regarding image acquisition methods and the sources of ground truth data is essential for establishing a common

FIGURE 1
Diagram of systematic evaluation for article selection.
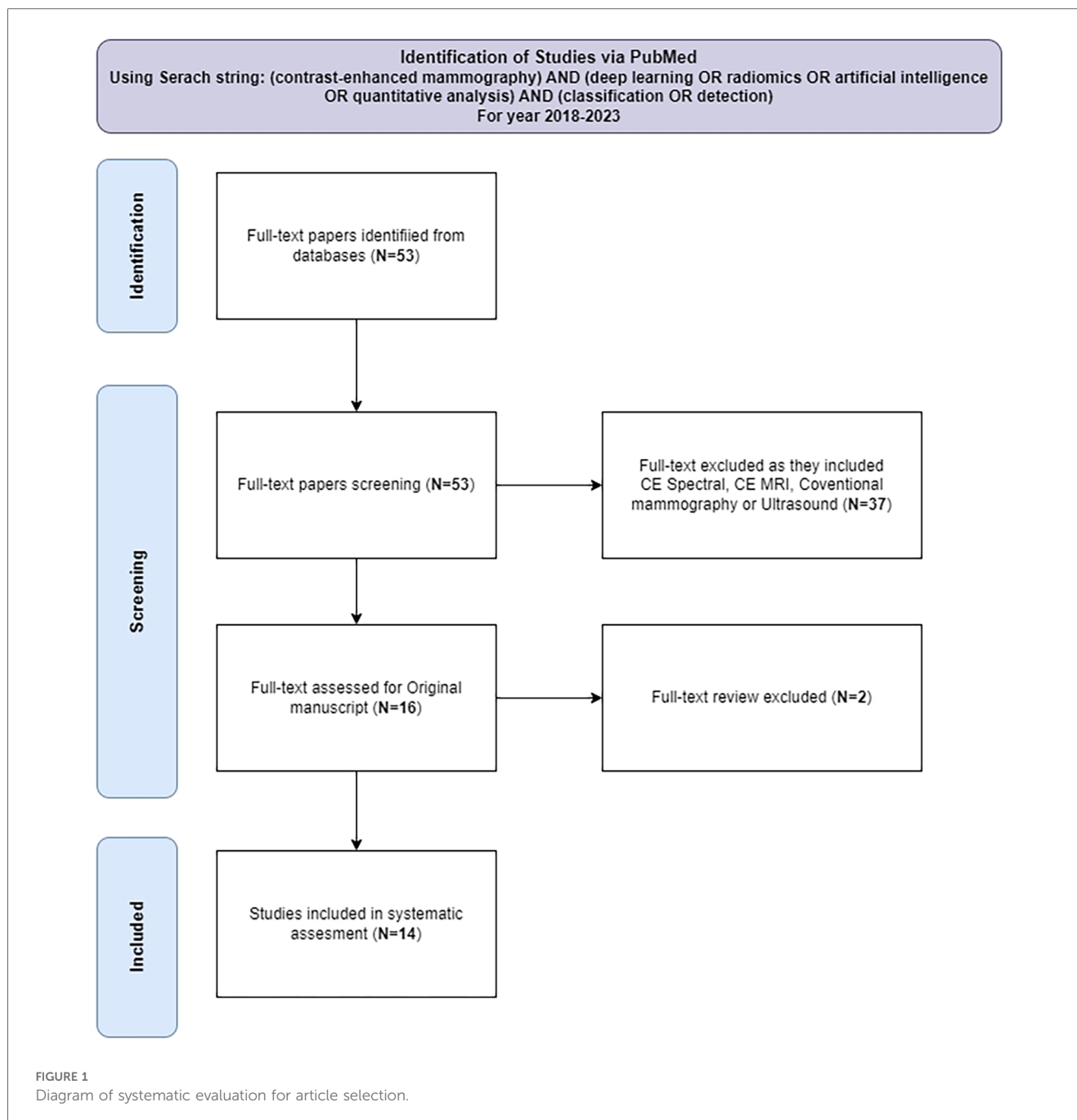
platform for comparing existing studies. While the majority of researchers have embraced transparency and rigor in their research processes, there are exceptions where such critical information remains undisclosed. This underscores the importance of robust reporting standards and transparency within the scientific community to ensure the credibility and reproducibility of research findings. For example, information regarding vendor, model, and acquisition protocol must be provided in the publications. This is crucial since across the 4 major CEM vendors, there exist different strategies for performing dual-energy mammography, using different tube voltage ranges, anode materials, filter materials/thicknesses, and image reconstruction algorithms for creating the recombined

CEM images. These differences can be a major source of inter-operator bias when using multivendor CEM within a multicenter study. Studies exploring harmonization/standardization strategies prior to using multivendor CEM data for multicenter studies are warranted.

4 Image pre-processing

Image preprocessing is crucial for models in contrast-enhanced imaging datasets, overcoming challenges like noise and artifacts. Steps like noise reduction, removal of background pixels, contrast enhancement, and data normalization improve image quality (47).
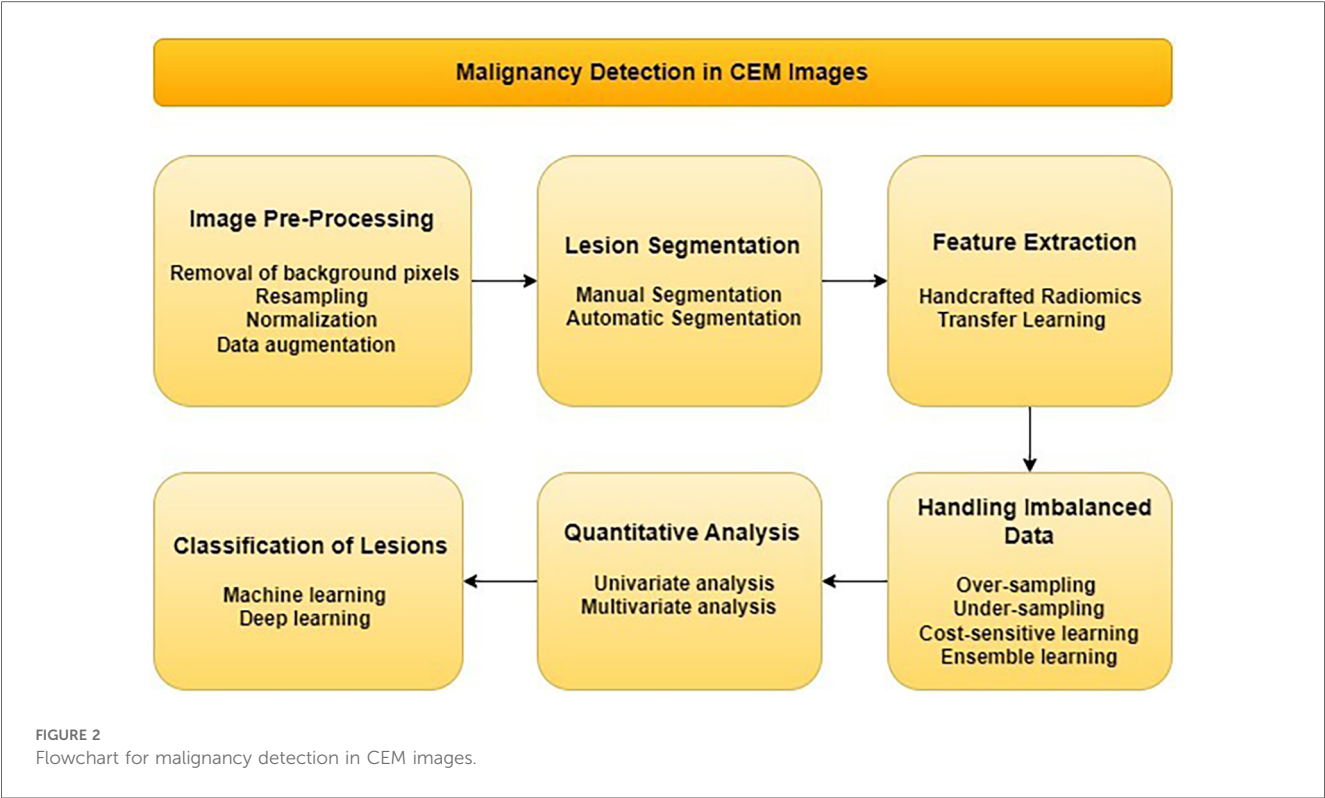
FIGURE 2
Flowchart for malignancy detection in CEM images.

TABLE 1 Vendors for CEM imaging acquisition system.

Vendors/system characteristic	Low—energy imaging: anode and filter material and thickness	High—energy imaging: anode and filter material and thickness	Mean glandular dose	Total acquisition time
Vendor 1	Mo/Mo, Mo/Rh, Rh/Rh Mo (0.03 mm), Rh (0.025 mm)	Mo/AI + Cu, Rh/AI + Cu Al (0.3 mm), Cu (0.3 mm)	1.6–2.8 mGy	3–8 s
Vendor 2	Mo/Mo, Rh/Ag Mo (0.03 mm), Ag (0.03 mm)	Mo/Cu, Rh/Cu (0.25 mm)	0.7–2.3 mGy	2–5 s
Vendor 3	W/Rh, W/Ag (0.050 mm)	W/Cu (0.3 mm)	3.0 mGy	Less than 2 s
Vendor 4	W/Rh (0.050 mm)	W/Ti (1.0 mm)	–	15–22 s

Vendor 1 offers GE Healthcare's Senographe Essential and Senobright, Vendor 2 offers GE Healthcare Pristina and Senobright HD. Vendor 3 offers Hologic Selenia Dimensions and 3Dimensions I—View, and Vendor 4 offers Siemens Healthineers Mammomat Revelation Titanium CEM system. The low-energy tube voltage range for these vendors spans 26 to 34 kV, while the high-energy range extends from 45 to 49 kV. Essential anode and filter materials include silver (Ag), aluminium (Al), copper (Cu), molybdenum (Mo), rhodium (Rh), titanium (Ti), and tungsten (W). The data within this table is sourced from Jeukens (32), Jochelson at al (11).

Techniques like data augmentation, ROI extraction, and data balancing enhance model generalization and feature detection. This preprocessing standardizes datasets, enhancing performance and accuracy. Therefore, it is critical to establish image quality standards prior to inclusion into ML/DL applications for reliable pre-processing.

4.1 Removal of background pixels

The presence of artifacts in medical images can introduce confusion or even mimic lesions, potentially leading to unnecessary medical procedures. Therefore, the removal of these artifacts plays a pivotal role in enhancing the accuracy of diagnoses. Several techniques have been developed for artifact removal, including thresholding, clustering, graph-cut algorithms,

and deep learning methods. Thresholding is particularly effective in addressing large and well-defined artifacts (47, 48). Clustering, on the other hand, groups similar pixels together to tackle artifact removal (48). Otsu's thresholding method has been applied in two notable studies (33, 43) for malignancy detection. In the case of (33), a two-step approach was employed, involving Contrast Limited Adaptive Histogram Equalization (CLAHE) before applying Otsu's thresholding. This preprocessing step, utilizing CLAHE, improved the image quality by mitigating issues related to uneven lighting conditions and varying contrast across different regions. Additionally, graph-cut algorithms provide another avenue for artifact removal, segmenting images based on pixel similarity (48). Deep learning techniques have also gained prominence, as they train neural networks to identify and subsequently remove artifacts (43). The choice of artifact removal technique hinges on the specific image characteristics

TABLE 2 Review of existing work in AI for CEM imaging.

Research work	Methodology	CEM system	Data size	Benign/ Malignant	Accuracy	Other metrics
Beuque et al. (33)	Mask RCNN with ResNet101	GE Healthcare	–	227/363 (External data set)	73%	AUC = 0.86 Sensitivity = 0.83 Specificity = 0.75
Wang et al. (34)	Least absolute shrinkage and selection operator (LASSO) logistic regression	GE Healthcare	226	101/125	88.2%	AUC = 0.96 Sensitivity = 0.90 Specificity = 0.93
Petrillo et al. (35)	Logistic Regression with LASSO	Hologic, USA and GE Healthcare	182	64/118	91.67%	Sensitivity = 0.90 Specificity = 0.92
Wang et al. (36)	Logistic regression	GE Healthcare	226	101/125	94.6%	AUC = 0.96 Sensitivity = 0.97 Specificity = 0.91
Fusco et al. (37)	Support Vector Machine	Hologic, USA and GE Healthcare	104	39/65	87%	AUC = 0.90 Sensitivity = 0.86 Specificity = 0.87
Wang et al. (38)	Least absolute shrinkage and selection operator (LASSO	GE Healthcare	223	101/122	–	AUC = 0.940
Sun et al. (39)	Least absolute shrinkage and selection operator (LASSO) regression	GE Healthcare	161	47/114	89.5%	AUC = 0.92 Sensitivity = 0.89 Specificity = 0.908
Miller et al. (40)	Penalized Linear Discriminant analysis	–	159	70/89	71.25%	AUC = 0.81 Sensitivity = 0.56 Specificity = 0.75
Gao et al. (41)	ResNet along with Convolutional Neural Network (CNN)	Hologic, USA	49	23/26	89%	AUC = 0.91 Sensitivity = 0.93 Specificity = 0.86
Jailin et al. (42)	YOLOv5 with CSPDarknet as backbone	GE HealthCare, USA	7,443	3,739/3,704 estimated	90%	AUC = 0.964 FPR = 0.128
Zheng et al. (43)	RefineNet and Xception + Pyramid pooling (PPM)	GE Healthcare	1,802	493/1,309	87.6%	Sensitivity = 0.95 Specificity = 0.70
Savaridas et al. (44)	Artificial Neural Network (ANN)	Hologic and GE Healthcare	269	14/255	91.4%	AUC = 0.97 Sensitivity = 0.95 Specificity = 0.89
Chen et al. (45)	DenseNet 121 with Convolutional Neural Network (CNN)	GE Healthcare, USA	1,903	490/1,413	87.1%	AUC = 0.912 Sensitivity = 0.947 Specificity = 0.714
Qian et al. (46)	VGG16 with Convolutional Neural Network (CNN)	GE Senographe Essential	2,496	765/1,731	85%	AUC = 0.92 Sensitivity = 0.86 Specificity = 0.85

and the desired outcome. Thresholding proves effective for larger and more distinct artifacts (49, 50) while clustering or graph-cut algorithms are better suited for smaller or grouped artifacts, offering a versatile array of tools to address artifact-related challenges in medical imaging.

4.2 Resampling

Resampling CEM images holds significance due to their high resolution, variable scan times, and diverse imaging protocols. Resampling is performed when there is a difference in the pixel resolution of an image. Image acquisition timing impacts appearance and generalization. Standardizing resolution and acquisition times enhances dataset consistency and diminishes model variance, ultimately reducing false negatives, thus improving model performance (51). Wang et al. (34) conducted a study that used data from two different centers and successfully standardized their dataset using resampling techniques. In study by Wang et al. (38) they performed resampling before feature

extraction. Resampling CEM datasets with different resolutions in multi-source data scenarios is recommended, as it is likely to improve model performance.

4.3 Normalization

Given the wide variation in study protocols, acquisition systems, and contrast injection dosages, it is clear that these factors have a significant impact on the brightness, contrast, and color balance of CEM images. Normalization is performed when there is a difference in the pixel intensity values of the image. Certain image features, such as texture and contrast, are more sensitive to fluctuations in these parameters than others. Normalization techniques offer a critical solution to mitigate these sensitivities. By normalizing CEM images, the impact of variable brightness, contrast, and color balance is minimized (52). This, in turn, enhances the reliability and precision of feature extraction processes from CEM images. For instance, in a study by Zheng et al. (43) the researchers used data from three

different sources. They used one source for training, and the other two for external testing. To ensure that the dataset was consistent, they used normalization. Qian et al. (46) enhanced CEM images by adjusting pixel values to improve contrast and highlight lesions and then performed min-max normalization. This normalization process was essential for harmonizing the diverse data sources and ensuring that the dataset was coherent and free of inconsistencies. Adopting these steps in studies is strongly recommended as they strengthen the reliability of their findings and conclusions and data integrity in multi-source studies.

4.4 Data augmentation

Being a relatively new technique, CEM studies face the challenges of being limited in size and imbalanced class distribution. These inherent characteristics pose a significant risk of overfitting, a scenario where the model becomes excessively attuned to the intricacies of the training data, hindering its ability to effectively generalize to unseen data. In response to this issue, data augmentation emerges as a valuable strategy. Data augmentation techniques, such as horizontal image mirroring, global intensity adjustments, realistic transformations of breast geometry (53), horizontal flipping, rotation, scaling, reducing size (54) and horizontal and vertical shifting have been effectively used in studies by Jailin et al. (42), Zheng et al. (43), Qian et al. (46). These techniques increase the diversity of the dataset, which improves the robustness of research findings.

4.5 Lesion segmentation

In the realm of radiomics, the extraction of features from lesion areas is a fundamental prerequisite. Achieving this necessitates the segmentation of lesions, a critical step in the process. Segmentation can be approached in two distinct ways.

4.5.1 Manual segmentation

Manual segmentation remains a widely adopted and trusted technique for precisely delineating lesions in CEM images. This method involves the meticulous outlining of lesion boundaries. Typically executed by skilled radiologists. In several studies reviewed (34, 35, 37, 38, 40, 55) manual segmentation approach was the chosen method. This approach underscored the importance of detailed and careful delineation of lesion contours, taking into account both the CC and MLO views, thus emphasizing its role in achieving precision and accuracy in radiological assessments. It is crucial to recognize that manual segmentation, despite its accuracy and reliability, demands a substantial investment of time and effort. The involvement of skilled radiologists is paramount to its success. If radiologist availability is limited, a single radiologist may need to handle segmentation. However, for high accuracy and precision demands, involving multiple radiologists to review and segment the image could be advantageous. Also, although this labor-intensive process remains indispensable for not only its inherent

precision but also its pivotal role in facilitating the development and evaluation of new automated segmentation methods.

4.5.2 Automatic segmentation

Automatic segmentation is a rapidly developing field with the potential to improve the efficiency and practicality of CEM image analysis. Automatic segmentation methods leverage the power of deep learning models to develop a comprehensive understanding of lesion features in contrast-enhanced mammography (CEM) images, enabling them to autonomously outline lesion contours. Alternatively, whole-organ analysis, the analysis of the entire breast, can be performed instead of lesion-specific segmentation. Consequently, automatic segmentation methods have the potential to reduce analysis time and effort, while also enhancing the accuracy and reproducibility of segmentation outcomes. Numerous studies have contributed to the development and evaluation of automatic segmentation methods tailored for CEM images. By merging manual segmentation with artificial intelligence, Zheng et al. (43) introduced an approach that improved lesion segmentation accuracy and efficiency. Wang et al. (56) introduced methodology that emphasizes lesion localization, providing a user-friendly and efficient alternative to conventional segmentation techniques, specifically by applying a deep learning model to detect and localize lesions in CEM images. Meanwhile, Beuque et al. (33) utilized the Mask R-CNN model (57), a region-based deep learning model that is optimized for object detection and segmentation. Jailen et al. (42) employed the YOLO v5 model, a single-stage deep learning model that is faster and more generalized than Mask R-CNN. These examples exemplify the diversity of approaches within the realm of automatic segmentation, and highlight the different trade-offs between accuracy, speed, and generalization.

However, it is crucial to acknowledge that automatic segmentation methods are still in the process of development, and several significant challenges must be addressed before they can find widespread application in clinical practice. One pressing challenge pertains to the sensitivity of these methods to the quality of the training data. In cases where training data is noisy or incomplete, the model's ability to accurately grasp lesion features may be compromised. Additionally, the computational demands of automatic segmentation methods pose a formidable hurdle, especially in clinical settings characterized by limited resources.

5 Feature extraction

Feature extraction is a critical technique for training CEM model training, enhancing the accuracy, efficiency, and interpretability of deep learning models (58). Common techniques include shape features, texture features, and kinetic features. Shape features describe the shape of the lesion, texture features describe its brightness, contrast, and homogeneity, and kinetic features describe the changes in the lesion over time. It is the foundational step that lays the groundwork for subsequent

model training. Feature extraction can be approached in two distinct ways, each bearing its own significance in the realm of medical imaging.

5.1 Handcrafted radiomics

The first method involves the extraction of handcrafted radiomics features from lesion regions, which have been meticulously annotated, segmented, or localized, as we previously discussed in the context of lesion segmentation. This approach, as observed in the reviewed studies, provides valuable insights into the characteristics of the lesion. These handcrafted features have been extracted using tools such as the PyRadiomics package and the Texture toolbox by MATLAB according to Image Biomarker Standardization Initiative (IBSI) (59), as elaborated in (33–37). Once these features are extracted, it becomes imperative to refine them to enhance data quality. This often involves normalization techniques to standardize the data and, importantly, assessing feature correlations using Spearman's coefficient. The subsequent crucial step to this feature extraction is feature selection (60) or the elimination of redundant features. The reviewed studies (34, 39, 55) have employed various methods for this purpose, such as interobserver agreement tests, Boruta's approach, Fisher criteria, maximum relevance minimum redundancy (mRMR), mutual information (MI), LASSO logistic regression (61), probability of error, pairwise correlations and average correlation (POE + ACC). Stratified 10-fold cross-validation is used in the XG Boost classifier to perform feature elimination (33). This process ensures that only the most informative and non-redundant features are retained for model training.

5.2 Transfer learning

Transfer learning is a valuable technique in deep learning pipelines for feature extraction. It utilizes pre-trained models to efficiently extract relevant features from new data, enhancing performance. This approach is particularly beneficial when working with small or noisy datasets, as it leverages knowledge learned from larger and more diverse datasets. This technique involves the use of pre-trained networks, such as Inception V3, CSP Darknet, Resnet, Xception, RetinaNet, VGG16 as observed in the reviewed studies (41–43, 45, 46, 55). Transfer learning offers computational efficiency and leverages higher-level features learned from extensive data, thus simplifying the feature extraction process from CEM images.

The choice between handcrafted radiomics and transfer learning hinges on the specific model being developed. Handcrafted radiomics requires lesion segmentation for feature extraction, while transfer learning allows for the utilization of either entire images or patches of lesions. This adaptability underscores the importance of selecting the most suitable approach based on the objectives and requirements of the model under consideration. In essence, feature extraction serves as the

linchpin in the AI pipeline for malignancy detection and segmentation, determining the quality and effectiveness of subsequent model training.

6 Handling imbalanced data

Handling data imbalance is a critical step in the AI pipeline, often underestimated but profoundly influential in obtaining accurate outputs. Failure to balance data properly can result in false positives and false negatives, as data imbalance introduces bias toward the majority class, undermining the minority class. There are several common methods to tackle this problem:

6.1 Over-sampling

This approach involves generating synthetic samples for the minority class to bolster its representation in the training dataset. Techniques like SMOTE (Synthetic Minority Over-sampling Technique) (44) and ROSE (Random Over-Sampling Examples) can be employed for this purpose. For example, in studies (33, 35) the authors utilized Adaptive Synthetic Sampling (ADASYN).

6.2 Under-sampling

In contrast, under-sampling entails removing samples from the majority class to diminish its presence in the training dataset. Various techniques, such as random under-sampling and Tomek links, can be applied to implement under-sampling effectively. As indicated, the use of under-sampling may not be advisable for CEM Images due to the issue of limited data availability. In such cases, the removal of samples from the majority class could further exacerbate the data scarcity problem, potentially leading to inadequate representation of the majority class and negatively impacting the model's performance.

6.3 Cost-sensitive learning

This method assigns different costs to the misclassification of samples from different classes. By assigning a higher cost to the minority class, this approach compels the model to give more attention to it, often resulting in improved performance on imbalanced datasets as done in study (41).

6.4 Ensemble learning

Ensemble learning entails training multiple models on different subsets of the data and then averaging their predictions. This technique helps reduce model variance and enhances performance on imbalanced datasets.

These methods illustrate the versatility required to address data imbalance effectively and emphasize the importance of choosing

the most suitable technique based on the specific dataset, as used by Gao et al. (41).

7 Quantitative analysis

Quantitative analysis of handcrafted features in CEM images encompasses the application of statistical and mathematical techniques to derive significant insights from the visual data. Following the extraction of these features from the lesion regions, it becomes imperative to subject the extracted features to rigorous measurement, quantification, and analysis before using these features for model training. Univariate and multivariate analysis represent two primary categories of quantitative methodologies extensively employed for the examination of handcrafted features within CEM images.

7.1 Univariate analysis

Univariate analysis is a fundamental statistical method focused on analyzing a single variable. It helps describe the variable's distribution, detect outliers, and identify trends, providing valuable insights into data characteristics. The non-parametric Wilcoxon-Mann-Whitney test is used for univariate analysis for handcrafted radiomics features in CEM research, as demonstrated in studies (35, 37). Its key benefits include not requiring specific data distribution assumptions, robustness against outliers, suitability for both ordinal and continuous data, and applicability to small sample sizes and non-normally distributed data. This is important because radiomics features are often non-normally distributed and can be susceptible to outliers. These attributes make it a valuable tool for comparing CEM radiomics features, ensuring robust and reliable research results. Another technique in study (37, 38) is the Intraclass Correlation Coefficient (ICC), which plays a vital role in univariate analysis for handcrafted radiomics features in CEM. The ICC assesses measurement reliability, identifies variability sources, aids in quality control, informs study design, facilitates feature reliability comparison, and determines clinical utility. By ensuring the consistency and trustworthiness of radiomics data, the ICC is essential for both research and clinical applications in CEM.

In univariate analysis, conducting Receiver Operating Characteristic (ROC) analysis and calculating the Youden index is a crucial step for determining the optimal cut-off value for each feature, also used by Wang et al. (36) to set optimal threshold for calculating accuracy and other parameters. This allows for the assessment of their discriminatory power and the identification of the point that maximizes sensitivity and specificity, which is essential for interpreting the performance of features, particularly in diagnostic or predictive modeling scenarios. Univariate analysis by Sun et al. (39) revealed that larger lesion sizes and rim or ripple artifacts were associated with more misclassifications of benign lesions and smaller lesion sizes were associated with more misclassifications of malignant lesions.

7.2 Multivariate analysis

Multivariate analysis involves the simultaneous examination of multiple variables, offering a powerful approach to uncover relationships among the features, classify data, and construct predictive models. It is a versatile tool for gaining deeper insights from complex datasets. Methods used for multivariate analysis of handcrafted radiomics features in CEM images include Principal Component Analysis (PCA) for dimensionality reduction, Linear Discriminant Analysis (LDA) (37, 62) for group discrimination, Logistic Regression (35) for binary outcome modeling, Random Forests for robust classification and regression, KNN (37) to handle outliers and non-linear relationships and Support Vector Machines (SVMs) for high-dimensional data analysis. These methods offer diverse approaches to extract insights from radiomics data, but their choice depends on research objectives and data characteristics. We recommend selecting the analytics technique that aligns with the specific criteria and research objectives. Multivariate analysis by Sun et al. (39) revealed that smaller lesion size and air trapping artifacts were associated with the misclassification of malignant lesions.

Our findings indicate that few studies have used handcrafted radiomics features, either independently or in conjunction with CEM images. Additionally, not all studies have conducted feature analysis. We strongly recommend incorporating these techniques into research endeavours. This would provide a more comprehensive understanding of the data, ultimately facilitating more effective model tuning during training.

8 Classification of lesions

After refining data from all the AI pipeline that we discussed in previous sections, the next important step in the AI pipeline for malignancy detection is to train a model to classify the data according to respective standards of ground truth. This can be done in two ways using machine learning techniques or using convolutional neural networks (CNNs).

8.1 Machine learning approach

Machine learning techniques play a vital role in malignancy detection from CEM images by distinguishing between malignant and benign lesions. In a review of 14 studies using CEM datasets as mentioned in Table 2, it was found that 7 of them used machine learning techniques for classification. Machine learning offers several advantages, including interpretability, which provides insights into how the model arrives at its outcomes. However, machine learning may not be the best choice for handling image data, such as CEM images, which are intricate and present challenges that traditional machine learning approaches may not effectively address. Machine learning is a highly suitable and effective choice for tasks where handcrafted features are used as the training data. Machine learning

techniques can effectively harness the valuable insights extracted from handcrafted features to develop robust models for making informed predictions. Here is a comprehensive overview of the key methodologies:

8.1.1 Logistic regression

Logistic regression is a binary classification technique known for its simplicity and effectiveness in distinguishing between two primary lesion categories. It plays a significant role in expediting cancer diagnosis. In study (34–36, 38, 39) it has been utilized alongside the Least Absolute Shrinkage and Selection Operator (LASSO), demonstrating good sensitivity for model outcomes. This combination of techniques provides a powerful approach for addressing classification challenges in medical research.

8.1.2 Support vector machine

Support vector machine (SVM) is versatile tool that can be used for both binary and multi-class classification tasks. It is particularly well-suited for handling the complex high-dimensional radiomics data derived from CEM images, making it an invaluable asset in the pursuit of precise malignancy detection, as used by (37).

8.1.3 Random forest

Random forest is a robust ensemble learning technique that combines multiple decision trees to improve prediction accuracy. Its innate resistance to noise and overfitting makes it dependable choices for navigating the complexities of radiomics data, emerging as steadfast allies when precision is of paramount concern as used by (36, 39).

8.1.4 Linear discriminant analysis

Linear discriminant analysis (LDA), a supervised learning algorithm, can identify optimal linear feature combinations to discriminate between different data groups. Its utility is even more significant in the realm of high-dimensional radiomics data, where it facilitates the effective categorization of lesions as used by (37, 40).

8.2 CNN approach in deep learning

Convolutional Neural Network (CNN) is a deep learning technique that uses artificial neural networks to learn from data. Neural networks are inspired by the human brain and can learn complex patterns from data. CNN is well-suited for image analysis tasks, including malignancy detection in CEM images. CNN models can learn to identify subtle features in images that may be difficult or impossible for humans to see, making them very effective at distinguishing between malignant and benign lesions. In a review of 14 studies, 7 used CNNs for model training. 6 out of 7 studies used transfer learning with a pre-trained network as the backbone for their CNN architecture. Of these, 2 studies (33, 41) used the ResNet pre-trained network. ResNet (63) pre-trained network is a popular choice for training CNNs on medical datasets due to their depth, accuracy, and

efficiency. They have been shown to be effective for a variety of medical image classification tasks and can be easily adapted to different datasets and tasks. In addition to ResNet, other pre-trained networks such as XceptionNet, CSPDarkNet, and Inception models (64) were also used in the reviewed studies.

Some studies using CNNs have not provided adequate information about key hyperparameters, such as learning rate schedule, optimization algorithm, minibatch size, dropout rates, and regularization parameters. Additionally, studies often fail to discuss why specific objective functions were chosen or how they align with the study's goals. We recommend researchers to define their criteria for selecting the best-performing model and clearly indicate when and how certain model parameters are restricted or frozen, especially in transfer learning scenarios. Adhering to these reporting standards would enhance transparency and reproducibility in CNN-based research for clinical and scientific purposes.

9 Cross validation

Cross-validation is essential for malignancy detection using CEM datasets because it prevents overfitting. CEM datasets are often small, making models more likely to overfit. Cross-validation assesses a model's ability to generalize by repeatedly testing it on different data subsets. It helps with model selection, hyperparameter tuning, and providing a robust performance estimate, ensuring reliable results in medical diagnosis. Commonly used CV methods encompass K-fold Cross-Validation, as indicated in (34, 39, 40, 43) which divides the data into subsets for rigorous evaluation. Stratified K-fold Cross-Validation is particularly beneficial for handling imbalanced datasets, ensuring that both malignant and benign cases are adequately represented. Leave-One-Out Cross-Validation, employed in (37, 41, 55) is suitable for smaller datasets but demands more computational resources due to its one-sample-at-a-time evaluation. Leave-P-Out Cross-Validation offers a middle ground for modest datasets. Repeated K-fold Cross-Validation enhances reliability by repeating the process multiple times. Nested Cross-Validation, although not cited in specific studies, plays a role in hyperparameter tuning. The choice of CV method hinges on factors like dataset size, class distribution, and research objectives, with Stratified K-fold commonly favored in CEM datasets to ensure equitable evaluation of model performance.

10 Integration with clinical data

The integration of clinical data with CEM datasets is a promising multi-modal approach for enhancing the accuracy and clinical utility of machine learning models for malignancy detection. This integration allows for a more holistic assessment of breast lesions by incorporating not only image-based features but also patient-specific clinical information. The extent to which this integration has been explored and implemented varies across studies. In addition to clinical data, some studies may also

explore the combination of CEM with other imaging modalities, such as ultrasound, MRI, etc. These multi-modal approaches seek to leverage the complementary strengths of different data sources to improve the overall performance of malignancy detection models. The specific combination of modalities can vary depending on the research objectives and data availability. In the study by Miller et al. (40), they found that incorporating demographic and clinical information into their models led to a notably improved AUC-ROC compared to using only density images, contrast images, or the combination of density and contrast images. It is observed in study by Wang et al. (36), the inclusion of clinical features to the radiomics features for model training resulted in a significant increase in both accuracy and sensitivity.

In the research article reviewed in the Table 2, we observed that all of the studies used histopathology as their reference standard for obtaining final ground truth diagnosis results, with a follow-up period of 2 weeks to 2 years, depending on the study.

11 Future scope

In the current landscape of malignancy detection research, we have explored the various strategies employed by studies to attain their results. However, there exists a compelling scope in the realm of multimodal approaches, particularly considering the persistent challenge of data scarcity in medical image datasets. The incorporation of multimodal data holds the potential to revolutionize the field by augmenting the accuracy, sensitivity, and AUC of detection models. The rationale behind exploring multimodal approaches is rooted in the inherent strengths of deep learning. This robust tool enables the extraction of intricate features from one mode of data, which can subsequently be integrated with knowledge derived from another modality. By combining different sources of medical data, researchers can overcome the limitations posed by data scarcity and achieve a more comprehensive understanding of the underlying phenomena.

Multimodal data fusion can significantly improve malignancy detection models by leveraging the unique strengths of each modality. This approach can uncover hidden patterns and correlations, leading to improved patient outcomes and clinical decision-making. The future of malignancy detection research lies in strategic utilization of multimodal data, overcoming individual limitations and paving the way for more robust and accurate detection models. The integration of multimodal approaches holds the potential to redefine malignancy detection research.

12 Conclusion

In conclusion, advances in the field of Artificial Intelligence in Contrast-Enhanced Mammography (CEM) have occurred, holding enormous potential for changing breast cancer detection and radiology practice, however, largescale validation is warranted. This review study explored the many aspects of AI in CEM, including image processing, lesion segmentation, feature

extraction, quantitative analysis, lesion classification, and integration with clinical data. The potential advantages are undeniably enormous. Timely identification and accurate diagnosis of breast abnormalities play a pivotal role in enhancing patient prognosis and minimizing unnecessary biopsy procedures. AI-powered CEM not only provides a more efficient and exact way of reaching these goals, but it also aids medical experts in to their decision-making processes. However, there is a lack of sufficient reliable labeled training data and handling variability between imaging systems, and protocols. Therefore, while AI analysis shows promise for improving CEM diagnosis, larger studies assessing its clinical value and real-world effectiveness are required. For such studies to be designed and implemented, it is critical that researchers, doctors, and technologists continue to interact and push the bounds of artificial intelligence in CEM. The synergistic partnership between AI and medical practitioners has the potential to usher in a new era of breast cancer diagnosis that prioritizes precision and efficiency. As a result, we can make great progress in lowering the burden of breast cancer and improving the lives of individuals afflicted by it.

Author contributions

KKK: Methodology, Validation, Writing – original draft, Writing – review & editing. BKKF: Validation, Writing – review & editing. MWY: Validation, Writing – review & editing. BAV: Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer MYLS declared a shared affiliation with the author BAV to the handling editor at the time of the review.

The authors BAV and BKKF declared that they were editorial board members of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Ghoncheh M, Pournamdar Z, Salehiniya H. Incidence and mortality and epidemiology of breast cancer in the world. *Asian Pac J Cancer Prev*. (2016) 17 (S3):43–6. doi: 10.7314/APJCP.2016.17.S3.43

2. DeSantis CE, Ma J, Goding Sauer A, Newman LA, Jemal A. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin*. (2017) 67 (6):439–48. doi: 10.3322/caac.21412

3. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. *Br J Cancer*. (2013) 108(11):2205–40. doi: 10.1038/bjc.2013.177

4. Pharoah PDP, Sewell B, Fitzsimmons D, Bennett HS, Pashayan N. Cost effectiveness of the NHS breast screening programme: life table model. *Br Med J*. (2013) 346:f2618. doi: 10.1136/bmj.f2618

5. Breast-cancer screening — viewpoint of the IARC working group. NEJM. Available online at: https://www.nejm.org/doi/full/10.1056/NEJMsr1504363 (Cited July 30, 2023).

6. Chetlen A, Mack J, Chan T. Breast cancer screening controversies: who, when, why, and how? *Clin Imaging*. (2016) 40(2):279–82. doi: 10.1016/j.clinimag.2015.05.017

7. Zhao H, Zou L, Geng X, Zheng S. Limitations of mammography in the diagnosis of breast diseases compared with ultrasonography: a single-center retrospective analysis of 274 cases. *Eur J Med Res*. (2015) 20(1):49. doi: 10.1186/s40001-015-0140-6

8. Monticciolo DL, Newell MS, Moy L, Niell B, Monsees B, Sickles EA. Breast cancer screening in women at higher-than-average risk: recommendations from the ACR. *J Am Coll Radiol*. (2018) 15(3, Part A):408–14. doi: 10.1016/j.jacr.2017.11.034

9. Coffey K, Jochelson MS. Contrast-enhanced mammography in breast cancer screening. *Eur J Radiol*. (2022) 156:110513. doi: 10.1016/j.ejrad.2022.110513

10. Cozzi A, Magni V, Zanardo M, Schiaffino S, Sardanelli F. Contrast-enhanced mammography: a systematic review and meta-analysis of diagnostic performance. *Radiology*. (2022) 302(3):568–81. doi: 10.1148/radiol.211412

11. Jochelson MS, Lobbes MBI. Contrast-enhanced mammography: state of the art. *Radiology*. (2021) 299(1):36–48. doi: 10.1148/radiol.2021201948

12. Dromain C, Balleyguier C. Contrast-Enhanced digital mammography. In: Bick U, Diekmann F, editors. *Digital mammography*. Berlin, Heidelberg: Springer (2010). p. 187–98. (Medical Radiology). doi: 10.1007/978-3-540-78450-0_12

13. Gelardi F, Ragaini EM, Sollini M, Bernardi D, Chiti A. Contrast-enhanced mammography versus breast magnetic resonance imaging: a systematic review and meta-analysis. *Diagnostics*. (2022) 12(8):1890. doi: 10.3390/diagnostics12081890

14. Forgia DL, Catino A, Dentamaro R, Galetta D, Losurdo L, Massafra R, et al. Role of the contrast-enhanced spectral mammography for the diagnosis of breast metastases from extramammary neoplasms. *J Buon*. (2019) 24(4):1360–6. ISSN: 1107-0625, online ISSN: 2241-6293.

15. Dominique C, Callonnec F, Berghian A, Defta D, Vera P, Modzelewski R, et al. Deep learning analysis of contrast-enhanced spectral mammography to determine histoprognostic factors of malignant breast tumours. *Eur Radiol*. (2022) 32 (7):4834–44. doi: 10.1007/s00330-022-08538-4

16. Prabhu S, Naveen DK, Bangera S, Bhat BS. Production of x-rays using x-ray tube. *J Phys Conf Ser*. (2020) 1712(1):012036. doi: 10.1088/1742-6596/1712/1/012036

17. Goo HW, Goo JM. Dual-energy CT: new horizon in medical imaging. *Korean J Radiol*. (2017) 18(4):555. doi: 10.3348/kjr.2017.18.4.555

18. Allec N, Abbaszadeh S, Scott CC, Lewin JM, Karim KS. Including the effect of motion artifacts in noise and performance analysis of dual-energy contrast-enhanced mammography. *Phys Med Biol*. (2012) 57(24):8405. doi: 10.1088/0031-9155/57/24/8405

19. Francescone MA, Jochelson MS, Dershaw DD, Sung JS, Hughes MC, Zheng J, et al. Low energy mammogram obtained in contrast-enhanced digital mammography (CEDM) is comparable to routine full-field digital mammography (FFDM). *Eur J Radiol*. (2014) 83(8):1350–5. doi: 10.1016/j.ejrad.2014.05.015

20. Lalji UC, Jeukens CRLPN, Houben I, Nelemans PJ, van Engen RE, van Wylick E, et al. Evaluation of low-energy contrast-enhanced spectral mammography images by comparing them to full-field digital mammography using EUREF image quality criteria. *Eur Radiol*. (2015) 25(10):2813–20. doi: 10.1007/s00330-015-3695-2

21. Ghaderi KF, Phillips J, Perry H, Lotfi P, Mehta TS. Contrast-enhanced mammography: current applications and future directions. *RadioGraphics*. (2019) 39(7):1907–20. doi: 10.1148/rg.2019190079

22. GE healthcare announces FDA 510(k) clearance of SenoBright contrast enhanced spectral mammography (CESM) for breast cancer diagnosis.. Available online at: https://www.gehealthcare.com/about/newsroom/press-releases/ge-healthcare-announces-fda-510k-clearance-senobright-contrast-enhanced-spectral (Cited October 15, 2023).

23. Zanardo M, Cozzi A, Trimboli RM, Labaj O, Monti CB, Schiaffino S, et al. Technique, protocols and adverse reactions for contrast-enhanced spectral

mammography (CESM): a systematic review. *Insights Imaging*. (2019) 10(1):76. doi: 10.1186/s13244-019-0756-0

24. Jeukens CRLPN, Lalji UC, Meijer E, Bakija B, Theunissen R, Wildberger JE, et al. Radiation exposure of contrast-enhanced spectral mammography compared with full-field digital mammography. *Invest Radiol*. (2014) 49(10):659. doi: 10.1097/RLI.0000000000000068

25. James JR, Pavlicek W, Hanson JA, Boltz TF, Patel BK. Breast radiation dose with CESM compared with 2D FFDM and 3D tomosynthesis mammography. *Am J Roentgenol*. (2017) 208(2):362–72. doi: 10.2214/AJR.16.16743

26. Argus A, Mahoney MC. Indications for breast MRI: case-based review. *Am J Roentgenol*. (2011) 196(3_supplement):WS1–4.

27. Sung JS, Stamler S, Brooks J, Kaplan J, Huang T, Dershaw DD, et al. Breast cancers detected at screening MR imaging and mammography in patients at high risk: method of detection reflects tumor histopathologic results. *Radiology*. (2016) 280(3):716–22. doi: 10.1148/radiol.2016151419

28. Nori J, Gill MK, Vignoli C, Bicchierai G, De Benedetto D, Di Naro F, et al. Artefacts in contrast enhanced digital mammography: how can they affect diagnostic image quality and confuse clinical diagnosis? *Insights Imaging*. (2020) 11 (1):16. doi: 10.1186/s13244-019-0811-x

29. Gluskin J, Click M, Fleischman R, Dromain C, Morris EA, Jochelson MS. Contamination artifact that mimics in-situ carcinoma on contrast-enhanced digital mammography. *Eur J Radiol*. (2017) 95:147–54. doi: 10.1016/j.ejrad.2017.08.002

30. Mongan J, Moy L, Charles E, Kahn J. Checklist for artificial intelligence in medical imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell*. (2020) 2(2):e200029. doi: 10.1148/ryai.2020200029

31. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, the PRISMA-DTA Group, et al.. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA*. (2018) 319 (4):388–96. doi: 10.1001/jama.2017.19163

32. Jeukens CRLPN. Physics of contrast-enhanced mammography. In: Lobbes M, Jochelson MS, editors. *Contrast-Enhanced Mammography*. Cham: Springer International Publishing (2019). p. 23–39. doi: 10.1007/978-3-030-11063-5_2

33. Beuque MPL, Lobbes MBI, van Wijk Y, Widaatalla Y, Primakov S, Majer M, et al. Combining deep learning and handcrafted radiomics for classification of suspicious lesions on contrast-enhanced mammograms. *Radiology*. (2023) 307(5): e221843. doi: 10.1148/radiol.221843

34. Wang S, Mao N, Duan S, Li Q, Li R, Jiang T, et al. Radiomic analysis of contrast-enhanced mammography with different image types: classification of breast lesions. *Front Oncol*. (2021) 11:1873–918. doi: 10.3389/fonc.2021.600546

35. Petrillo A, Fusco R, Di Bernardo E, Petrosino T, Barretta ML, Porto A, et al. Prediction of breast cancer histological outcome by radiomics and artificial intelligence analysis in contrast-enhanced mammography. *Cancers (Basel)*. (2022) 14(9):2132. doi: 10.3390/cancers14092132

36. Wang S, Sun Y, Mao N, Duan S, Li Q, Li R, et al. Incorporating the clinical and radiomics features of contrast-enhanced mammography to classify breast lesions: a retrospective study. *Quant Imaging Med Surg*. (2021) 11(10):4418–30. doi: 10.21037/qims-21-103

37. Fusco R, Piccirillo A, Sansone M, Granata V, Rubulotta MR, Petrosino T, et al. Radiomics and artificial intelligence analysis with textural metrics extracted by contrast-enhanced mammography in the breast lesions classification. *Diagnostics*. (2021) 11(5):815. doi: 10.3390/diagnostics11050815

38. Wang S, Sun Y, Li R, Mao N, Li Q, Jiang T, et al. Diagnostic performance of perilesional radiomics analysis of contrast-enhanced mammography for the differentiation of benign and malignant breast lesions. *Eur Radiol*. (2022) 32 (1):639–49. doi: 10.1007/s00330-021-08134-y

39. Sun Y, Wang S, Liu Z, You C, Li R, Mao N, et al. Identifying factors that may influence the classification performance of radiomics models using contrast-enhanced mammography (CEM) images. *Cancer Imaging*. (2022) 22:22. doi: 10.1186/s40644-022-00460-8

40. Miller MM, Rubaiyat AHM, Rohde GK. Predicting malignancy of breast imaging findings using quantitative analysis of contrast-enhanced mammography (CEM). *Diagnostics*. (2023) 13(6):1129. doi: 10.3390/diagnostics13061129

41. Gao F, Wu T, Li J, Zheng B, Ruan L, Shang D, et al. SD-CNN: a shallow-deep CNN for improved breast cancer diagnosis. *Comput Med Imaging Graph*. (2018) 70:53–62. doi: 10.1016/j.compmedimag.2018.09.004

42. Jailin C, Mohamed S, Iordache R, Milioni De Carvalho P, Ahmed SY, Abdel Sattar EA, et al. AI-based cancer detection model for contrast-enhanced mammography. *Bioengineering*. (2023) 10(8):974. doi: 10.3390/bioengineering10080974

43. Zheng T, Lin F, Li X, Chu T, Gao J, Zhang S, et al. Deep learning-enabled fully automated pipeline system for segmentation and classification of single-mass breast

lesions using contrast-enhanced mammography: a prospective, multicentre study. *eClinicalMedicine*. (2023) 58:101913. doi: 10.1016/j.eclinm.2023.101913

44. Savaridas SL, Agrawal U, Fagbamigbe AF, Tennant SL, McCowan C. Radiomic analysis in contrast-enhanced mammography using a multivendor data set: accuracy of models according to segmentation techniques. *Br J Radiol*. (2023) 96 (1145):20220980. doi: 10.1259/bjr.20220980

45. Chen Y, Hua Z, Lin F, Zheng T, Zhou H, Zhang S, et al. Detection and classification of breast lesions using multiple information on contrast-enhanced mammography by a multiprocess deep-learning system: a multicenter study. *Chin J Cancer Res*. (2023) 35(4):408–23. doi: 10.21147/j.issn.1000-9604.2023.04.07

46. Qian N, Jiang W, Guo Y, Zhu J, Qiu J, Yu H, et al. Breast cancer diagnosis from contrast-enhanced mammography using multi-feature fusion neural network. *Eur Radiol*. (2023) :1–11. doi: 10.1007/s00330-023-10170-9

47. Perumal S, Thambusamy V. Preprocessing by contrast enhancement techniques for medical images. *Int J Pure Appl Math*. (2018) 118:3681–8. doi: 10.3390/cancers14092132

48. Moghbel M, Ooi CY, Ismail N, Hau YW, Memari N. A review of breast boundary and pectoral muscle segmentation methods in computer-aided detection/diagnosis of breast mammography. *Artif Intell Rev*. (2020) 53(3):1873–918. doi: 10.1007/s10462-019-09721-8

49. Boss RSC, Thangavel K, Daniel DAP. Automatic mammogram image breast region extraction and removal of pectoral muscle. (2013) **4**(5).

50. Don S, Choi E, Min D. Breast mass segmentation in digital mammography using graph cuts. In: Lee G, Howard D, Ślęzak D, editors. *Convergence and Hybrid Information Technology*. Berlin, Heidelberg: Springer (2011). p. 88–96. (Communications in Computer and Information Science).

51. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and deep learning methods for radiomics. *Med Phys*. (2020) 47(5):e185–202. doi: 10.1002/mp.13678

52. Veldkamp WJH, Karssemeijer N. Normalization of local contrast in mammograms. *IEEE Trans Med Imaging*. (2000) 19(7):731–8. doi: 10.1109/42.875197

53. Caselles L, Jailin C, Muller S. *Data augmentation for breast cancer mass segmentation*. In: Su R, Zhang YD, Liu H, editors. *Proceedings of 2021 International Conference on Medical Imaging and Computer-Aided Diagnosis (MICAD 2021)*. Singapore: Springer (2022). p. 228–37. (Lecture Notes in Electrical Engineering).

54. Pérez-García F, Sparks R, Ourselin S. TorchIO: a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Comput Methods Programs Biomed*. (2021) 208:106236. doi: 10.1016/j.cmpb.2021.106236

55. Marino MA, Pinker K, Leithner D, Sung J, Avendano D, Morris EA, et al. Contrast-enhanced mammography and radiomics analysis for noninvasive breast cancer characterization: initial results. *Mol Imaging Biol*. (2020) 22(3):780–7. doi: 10.1007/s11307-019-01423-5

56. Wang K, Patel BK, Wang L, Wu T, Zheng B, Li J. A dual-mode deep transfer learning (D2TL) system for breast cancer detection using contrast enhanced digital mammograms. *IISE Trans Healthc Syst Eng*. (2019) 9(4):357–70. doi: 10.1080/24725579.2019.1628133

57. Gaiser H, Liscio E, vcarpani, viveksuryamurthy, azmi nik mohamad aizuddin bin nik. fizyr/keras-maskrcnn 0.2.2. Zenodo. (2019). Available online at: https://zenodo.org/record/3250666 (Cited August 1, 2023).

58. Zhang Y, Liu F, Zhang H, Ma H, Sun J, Zhang R, et al. Diagnostic value of radiomics analysis in contrast-enhanced spectral mammography for identifying triple-negative breast cancer. *Front Oncol*. (2021) 11:773196. doi: 10.3389/fonc.2021.773196

59. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*. (2020) 295 (2):328–38. doi: 10.1148/radiol.2020191145

60. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. (2007) 26(30):5512–28. doi: 10.1002/sim.3148

61. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B*. (1996) 58(1):267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x

62. Wang W, Mo Y, Ozolek JA, Rohde GK. Penalized fisher discriminant analysis and its application to image-based morphometry. *Pattern Recognit Lett*. (2011) 32 (15):2128–35. doi: 10.1016/j.patrec.2011.08.010

63. He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. (2015) 1026–34. Available online at: https://openaccess.thecvf.com/content_iccv_2015/html/He_Delving_Deep_into_ICCV_2015_paper.html (Cited August 15, 2023)

64. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. (2016) 2818–26. Available online at: https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Szegedy_Rethinking_the_Inception_CVPR_2016_paper.html (Cited August 22, 2023).

Artificial intelligence and machine learning applications for the imaging of bone and soft tissue tumors

Paniz Sabeghi[1], Ketki K. Kinkar[2], Gloria del Rosario Castaneda[3], Liesl S. Eibschutz[1], Brandon K. K. Fields[4], Bino A. Varghese[1], Dakshesh B. Patel[1] and Ali Gholamrezanezhad[1]*

[1]Department of Radiology, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States, [2]Viterbi School of Engineering, University of Southern California, Los Angeles, CA, United States, [3]Keck School of Medicine, University of Southern California, Los Angeles, CA, United States, [4]Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, CA, United States

Recent advancements in artificial intelligence (AI) and machine learning offer numerous opportunities in musculoskeletal radiology to potentially bolster diagnostic accuracy, workflow efficiency, and predictive modeling. AI tools have the capability to assist radiologists in many tasks ranging from image segmentation, lesion detection, and more. In bone and soft tissue tumor imaging, radiomics and deep learning show promise for malignancy stratification, grading, prognostication, and treatment planning. However, challenges such as standardization, data integration, and ethical concerns regarding patient data need to be addressed ahead of clinical translation. In the realm of musculoskeletal oncology, AI also faces obstacles in robust algorithm development due to limited disease incidence. While many initiatives aim to develop multitasking AI systems, multidisciplinary collaboration is crucial for successful AI integration into clinical practice. Robust approaches addressing challenges and embodying ethical practices are warranted to fully realize AI's potential for enhancing diagnostic accuracy and advancing patient care.

KEYWORDS
artificial intelligence, machine learning, deep learning, musculoskeletal, sarcoma

Key points

- Deep learning models have been developed for diagnosing MSK tumors and show potential to achieve diagnostic efficacy comparable to radiologists in limited classification tasks.
- AI algorithms can address issues related to variance in acquisition parameters and noise between MR scans using techniques such as edge-preserving denoising and intensity standardization.
- Multitasking AI systems that can efficiently perform multiple segmentation and analytical tasks at once hold promise for potentially useful prospective implementations in clinical practice.

Abbreviations

AI, artificial intelligence; ALTs, atypical lipomatous tumors; DL, deep learning; ML, machine learning; MSK, musculoskeletal; STS, soft-tissue sarcomas; WDLs, well-differentiated liposarcomas; CNN, convolutional neural network.

Introduction

Developments in artificial intelligence (AI) and machine learning (ML) have advanced the field of medicine and offer new and powerful digital tools to facilitate the next transformation in musculoskeletal (MSK) radiology. While it is important to acknowledge that these AI applications are still mainly in the experimental phase and need to be validated ahead of being fully integrated into standard clinical workflows, it is worth noting that they hold significant promise. In addition to streamlining radiology processes and enhancing the detection of abnormalities, AI techniques show potential for applications including predicting progression of malignancy and providing prognostic information (1–6). However, these potential advantages are not without some inherent biases and drawbacks, and radiologists must be aware of these pitfalls to allow for optimal implementation of AI tools in clinical practice (7, 8). AI may one day also enhance workflow productivity by automating repetitive processes, allowing radiologists to focus on image interpretation and clinical communication. Quality control may also come to be bolstered through enhanced automated detection of image artifacts and overall scan degradation. Finally, predictive analytics can help tailor interventions and allow for personalized modifications (9).

This review discusses key concepts and potential pitfalls of AI and ML in MSK radiology and how they can potentially be applied for diagnosis and treatment of soft tissue and bone tumors.

Artificial intelligence and machine learning in medicine

AI generally refers to computer systems that simulate or mimic human intelligence (10). Beyond imaging interpretation in radiology, AI may also have a wide range of applications ranging from augmented structured reporting and clinical support systems to radiomics-based predictive implementations (11, 12).

ML defines a field of AI in which computers learn by analyzing large amounts of aggregated data and improve algorithms by iterative exposure and performance evaluation (11–14). ML subtypes can be categorized as supervised, unsupervised, reinforced, and semi-supervised. *Supervised* learning occurs via supplied output, while *unsupervised* learning develops from pattern recognition in input data without specific feedback. *Reinforcement* learning uses punishments or rewards as decision reinforcements. *Semi-supervised* learning involves fewer explicit outputs validated against a ground truth label (11).

Deep learning (DL), a subset of ML, is a multilayered approach which leverages hierarchical arrangement of multiple algorithms. Some of the more common applications of DL in MSK radiology include detection of spinal pathology, meniscal tears, fractures, and osteoarthritis (11, 15).

Musculoskeletal radiology

MSK radiology employs a variety of imaging modalities to diagnose and assess disorders involving joints, muscles, soft tissues, and osseous structures. Imaging also plays a key role in initial assessment and treatment response characterization in bone tumors and soft-tissue sarcomas (STS) (16).

Soft tissue tumors

Imaging techniques remain a pivotal component of the diagnosis of benign and malignant soft tissue lesions. While in many cases, specific clinical and imaging features may aid in narrowing down the differential diagnoses, definitive diagnosis is often made by tissue sampling and histopathologic interpretation (17). Though malignant tumors tend to be larger, some small soft tissue masses account for a significant portion of soft tissue malignancies. These smaller masses are more likely to be missed or be under-resected at surgery (18).

Lipomas and their malignant counterparts liposarcomas are among the soft tissue masses originating from adipose tissue (19). However, while well-differentiated liposarcomas (WDLs) and atypical lipomatous tumors (ALTs) can appear similar to intramuscular lipomas on imaging, the distinction holds significant implication for prognosis and treatment (20, 21). Specifically, treatment for higher-grade liposarcomas may require wide local excision with or without neoadjuvant or adjuvant chemotherapy and/or radiotherapy (19, 20). Similarly, while benign lipomas may in many instances be amenable to clinical observation or marginal resection, ALTs/WDLs may also require wide excision and subsequent imaging surveillance (22, 23).

Even though histology remains the gold standard for diagnosis, certain imaging modalities, mainly contrast enhanced MRI, may help to narrow the differential considerations ahead of tissue sampling. Nevertheless, traditional imaging modalities do possess inherent acknowledged limitations in reliably differentiating between benign and malignant soft tissue tumors (19, 22, 24).

Bone tumors

Bone tumors can be classified into two main categories: primary tumors and secondary (metastatic) tumors. Malignant primary bone tumors arise from osseous tissues though have the potential to metastasize to other remote sites in the body (25).

Most benign bone tumors are chondrogenic in nature and are often found to be enchondromas or osteochondromas. Intermediate bone tumors such as giant cell tumors of bone may be at risk for malignant transformation. Chondrogenic and osteogenic tumors are among the most common primary bone malignancies (26).

Diagnosing bone tumors combines several approaches which consider clinical factors, histological sampling, and imaging (27).

Radiography is often the initial imaging modality employed due to its ability to localize lesions and provide rapid holistic assessment of patterns of bony destruction, margins (zones of transition), and/or presence of periosteal reaction. These destructive patterns may provide insight into lesional biological activity and aggressiveness (2, 28).

MRI is the preferred method for local evaluation and staging due to its superiority in delineating associated soft-tissue components and detecting invasion into surrounding structures (29). Fluorine-18 fluorodeoxyglucose-PET scans can evaluate tumor metabolic activity, which often correlates with aggressiveness (27). PET/CT and PET/MRI are among the most sensitive and specific modalities for evaluating skeletal metastatic disease (30).

Artificial intelligence and machine learning in musculoskeletal radiology

Various models have suggested that DL can, in relatively narrow use cases, achieve relatively similar diagnostic performance in comparison to human interpreting radiologists (31, 32). However, relative to other organ systems, the MSK system poses unique challenges for developing AI applications. The complex biomechanical interplay of the various anatomical structures makes it challenging for AI researchers to develop robust algorithms amidst the many possible scan angles and positional variations. Additionally, variability in acquisition parameters, image noise, and the field strength often necessitates complex preprocessing to improve and standardize image quality prior to AI operations (15).

Keles et al. (15) emphasize the importance of "clean data" for AI algorithms and discuss the need for preprocessing techniques. In the case of MRI, there are three main categories of challenges in medical images that need to be addressed with preprocessing, namely image nonstandardness, noise, and artifacts. The bias field artifact, also known as inhomogeneity, affects the quantitative intensity values of pixels and can in turn affect segmentation performance. In their preliminary studies, Keles et al. (15) applied generalized-scale post-processing to correct field inhomogeneities arising from the RF coil. Edge-preserving denoising was used to smooth images and thereby reduce image noise. To tackle signal intensity variations between MRI acquisitions, the authors applied an intensity standardization algorithm.

Segmentation of muscle, fat, and other regions of interest using automated techniques can be difficult due to overlapping intensity values of various tissues (15). LaLonde et al. present a DL algorithm called SegCaps, which was introduced for biomedical image segmentation tasks. The SegCaps algorithm leverages "deconvolutional capsules" in a design which purportedly requires fewer than 5% of the parameters necessary to execute the popular *U-Net* architecture (33).

Zhao et al. (34) developed three DL models utilizing ce-MRI to assist in the diagnosis of MSK neoplasms. This study's findings suggested that knowledge of the DL classifiers' predicted probability of malignancy led to increased sensitivity of imaging interpretation without significantly affecting specificity for providers of varying years of experience and training across oncology, MSK radiology, and orthopedic surgery (34).

DL may not only prove helpful in interpretation tasks but may also come to play a role in image reconstruction. Wessling et al. (35) employed a DL algorithm known as an "unrolled variational network", which leverages an iterative parallel imaging reconstruction architecture to accelerate sequence acquisition time. Their results suggested that DL-based reconstruction both improved image quality and led to reductions in acquisition times of up to 52%–59% as compared to conventional scanning parameters (35).

Artificial intelligence in imaging of soft tissue tumors

Regarding distinguishing lipomas from ALTs/WDLs in lipomatous soft tissue tumors, Leporq et al. (19) developed an MRI-based radiomics approach using fat-suppressed contrast-enhanced T1-weighted sequences and found that their classification models were able to distinguish between benign and malignant lesions. In their study, radiomics features were extracted from 2D, manually contoured tumor masks and subsequently used for machine learning. Their findings suggested that size features were most highly predictive of malignancy while intensity distribution features held the least predictive utility. However, they also found that shape features were most subject to interobserver variability.

In a recent study, Sudjai et al. (20) developed a machine-learning approach to differentiate between ALTs/WDLs and intramuscular lipomas based on radiomics features and the distance between tumor and bone on non-contrast T1-weighted MR images. The model achieved high accuracy in separating intramuscular lipomas from ALT/WDL, with an area under the curve (AUC) of 0.88. The model's performance was comparable to that of two MSK radiologists with 22 and 7 years of experience, respectively. Texture, shape, and histogram-based features were identified as most important in determining the model's predicted probability of malignancy.

Cay et al. (36) similarly found that a radiomics-based support vector machine algorithm was predictive of malignancy in lipomatous masses, with a reported sensitivity of 96.8% and a specificity of 93.72% for the machine learning approach. Regarding individual feature performance, gray-level run length matrix (GLRLM) based Run Length Non-Uniformity (RNLU) demonstrated the best performance, with an area under the curve (AUC) of 0.902 (36).

Fradet et al. (22) evaluated the relative performance of MRI radiomics with ML analysis with and without batch correction and DL models in predicting malignancy in lipomatous neoplasms. The authors performed a radiomics analysis on post-contrast fat-suppressed T1-weighted sequences with manual 3D segmentations. Best numerical results were seen with models trained on batch-corrected radiomics data (AUC of 0.80 vs. AUC

of 0.70 in the external validation cohort for gradient boosting trained on radiomics data with and without batch correction, respectively). The Random Forest and Gradient Boosting models also notably outperformed the ResNet50 DL model in external validation, the latter of which only reached an AUC of 0.64.

Wang et al. (37) developed an ML radiomics-based nomogram for detecting malignancy in unknown soft tissue masses. The nomogram combined features of tumor margin, size, and capsule, along with a calculated radiomics score, yielding AUC values of 0.96 and 0.88 in validation testing.

Similarly, Fields et al. (24) reported that predictive models developed from radiomics data using machine learning-augmented approaches demonstrated effective discriminative capabilities in correctly categorizing benign and malignant lesions on preoperative MRI scans. Models built on unfiltered radiomics datasets yielded AUC values of 0.77 for Real Adaptive Boosting and 0.72 for Random Forest, respectively. Models limited to metrics derived only from T2 fat-saturated and Short-Tau Inversion Recovery sequences yielded similar performances, with AUCs of 0.73 for Real Adaptive Boosting and 0.75 for Random Forest. These results suggest that radiomics-based models based on restricted subsets of sequences may still maintain clinical relevance, which can help limit complexity and shorten analytical processing steps in future prospective implementations of machine learning-augmented workflows.

Navarro et al. (38) similarly developed DL models to stratify between high-grade and low-grade soft tissue tumors based on pre-treatment T2-weighted fat-saturated and contrast-enhanced T1-weighted fat-saturated MRIs. Following manual segmentation, separate DL models based on a pre-trained DenseNet-161 architecture were developed for each cohort of MRI sequences, which achieved AUCs of 0.76 and 0.75 for T2-weighted fat-saturated and contrast enhanced T1-weighted fat-saturated images, respectively. The DL models notably outperformed comparator regression models based on clinical features, tumor volume, and combined tumor volume and clinical features, respectively (38).

Multi-parametric MRI is the modality of choice for evaluation of treatment response in STS. However, the highly heterogeneous nature of these changes and varying degrees of tumor cellularity often confound evaluations and can contribute to clinical uncertainty (39, 40). In a cohort study by Blackledge et al., the authors suggested that ML can aid in evaluation of tumor response to radiotherapy (41). The authors evaluated the utility of eight different machine-learning approaches in differentiating between five distinct intratumoral tissue classes. Naïve-Bayes in combination with a Markov Random Field denoising algorithm was able to quantify changes in tumor sub-regions in a limited pre- and post-treatment cohort of 8 patients. These results suggest an ability for machine-learning techniques to assess underlying changes in tumor composition even in situations when overall changes in size in response to treatment may not be overtly evident.

Despite limitations in operator dependence, ultrasound may serve as a useful adjunctive modality for evaluation of soft-tissue masses. In a study by Wang et al. (42), the authors trained a convolutional neural network (CNN) to differentiate benign and malignant soft tissue lesions on routine clinical ultrasound,

yielding an AUC of 0.91 and an accuracy of 79% on the test set. Sensitivity (90%) and specificity (74%) were not significantly different when compared to the performance of two interpreting MSK radiologists. Another CNN model was trained to distinguish between three different benign masses, namely benign nerve sheath tumors, vascular malformations, and lipomas. For the classification of lipomas, precision and recall of the model were 78% and 93%, respectively; for the classification of benign nerve sheath tumors, precision and recall were 71% and 42%, respectively; and for the classification of vascular malformations, precision and recall were 60% and 64%, respectively (42).

Artificial intelligence in imaging of bone tumors

Several studies suggest a moderate to high accuracy for AI-based predictive models in differentiating between benign, intermediate and malignant tumors (1, 2, 32). Gitto et al. (43) examined how manual segmentation variability impacted the replicability of texture analysis on CT and MRI scans of cartilaginous bone tumors. The authors conducted 2D and 3D manual segmentations on CT and MRI scans, then implemented marginal shrinkage to assess impact on feature reproducibility. Overall, contour-focused segmentation yielded higher rates of stable radiomics features for 3D (80%) as compared to 2D (75%) regions-of-interest for CT and MRI. In comparison, marginal erosion performed more poorly for 3D features ($p < 0.001$) though was not statistically significant for 2D features ($p = 0.343$) (43).

He et al. (44) utilized DL on MRI images to predict the post-curettage local recurrence of giant cell tumor of bone based on preoperative MRI examinations. The results of CNN and CNN regression models were compared against the performance of four radiologists. The authors reported 75.5% accuracy and 85.7% sensitivity for the CNN model, and 78.6% accuracy and 87.5% sensitivity for the CNN regression model. This is in comparison to 64.3% accuracy and 58.3% sensitivity for the average performance of the radiologists (44).

Challenges and future directions

Despite promising trends of applying AI and ML in MSK oncologic radiology, there exist several challenges and limitations to implementation. While there remains great interest in prospective implementations, achieving diagnostic accuracy is crucial. Practicing radiologists should strive to gain a thorough understanding of the use cases and associated challenges of AI implementation in prospective clinical workflows so as to maximize future applications in daily practice (14, 15).

Variations in implementation of AI applications is another major hurdle. Neural networks, which strive to replicate human cognition, require prolonged and frequently convoluted training and refinement stages. Furthermore, differences in implementation schemes across sites and institutions can significantly affect performance (11, 45, 46). Standardization of workflows will serve to

promote generalizability and repeatability, which will increase diagnostic confidence in prospective applications (45).

Additionally, using such large amounts of patient data raise unique considerations with respect to data use ethics (47). In order to address these concerns, federated learning offers a unique approach in allowing for local training of a centrally-maintained AI model across many participating sites, thereby obviating the need for centralized data repositories (48).

DL techniques may one day aid in supporting clinical decision-making and automating certain lower-level tasks, allowing radiologists to focus greater attention on higher level interpretive tasks. Furthemore, radiomics and ML based classifiers working alongside other -omics may synergistically work to advance many AI modalities and more holistically capture unique aspects of the patient experience (10).

Author contributions

PS: Writing – original draft, Writing – review & editing. KK: Writing – review & editing. GDRC: Writing – original draft. LE: Writing – review & editing. BKKF: Writing – review & editing. BAV: Writing – review & editing. DBP: Writing – review & editing. AG: Writing – original draft, Writing – review & editing.

Funding

Conflict of interest

BKKF received prior research grants from the RSNA R&E (2019-2020 RMS #1909; 2018-2019 RMS #1810); consulting fees from Mendaera; honorarium payments from Neurodiem (invited author) and Elsevier (book proposal reviews); RSNA and institutional support for attending meetings (RSNA RFC stipend, institutional support stipend); vice-chair of the RSNA Resident and Fellow Committee; member of the American Board of Radiology Initial Certification Advisory Committee for Diagnostic Radiology, of the RSNA Education Council, and of the Radiology: Imaging Cancer Trainee Editorial Board.

The remaining authors declare that the research was conducted in the absence of any commercial or financial Relationships that could be construed as a potential conflict of interest.

The authors BKKF, BAV, and AG declared that they were editorial board members of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Liu R, Pan D, Xu Y, Zeng H, He Z, Lin J, et al. A deep learningmachine learning fusion approach for the classification of benign, malignant, and intermediate bone tumors. *Eur Radiol*. (2022) 32(2):1371–83. doi: 10.1007/s00330-021-08195-z

2. von Schacky CE, Wilhelm NJ, Schafer VS, Leonhardt Y, Gassert FG, Foreman SC, et al. Multitask deep learning for segmentation and classification of primary bone tumors on radiographs. *Radiology*. (2021) 301(2):398–406. doi: 10.1148/radiol.2021204531

3. von Schacky CE, Wilhelm NJ, Schafer VS, Leonhardt Y, Jung M, Jungmann PM, et al. Development and evaluation of machine learning models based on x-ray radiomics for the classification and differentiation of malignant and benign bone tumors. *Eur Radiol*. (2022) 32(9):6247–57. doi: 10.1007/s00330-022-08764-w

4. Jiang J, Qu H, Zhan X, Liu D, Liang T, Chen L, et al. Identification of osteosarcoma m6A-related prognostic biomarkers using artificial intelligence: RBM15. *Sci Rep*. (2023) 13(1):5255. doi: 10.1038/s41598-023-28739-1

5. Huang Z, Hu C, Chi C, Jiang Z, Tong Y, Zhao C. An artificial intelligence model for predicting 1-year survival of bone metastases in non-small-cell lung cancer patients based on XGBoost algorithm. *Biomed Res Int*. (2020) 2020:3462363. doi: 10.1155/2020/3462363

6. Li J, Li S, Li X, Miao S, Dong C, Gao C, et al. Primary bone tumor detection and classification in full-field bone radiographs via YOLO deep learning model. *Eur Radiol*. (2023) 33(6):4237–48. doi: 10.1007/s00330-022-09289-y

7. Mun SK, Wong KH, Lo SB, Li Y, Bayarsaikhan S. Artificial intelligence for the future radiology diagnostic service. *Front Mol Biosci*. (2020) 7:614258. doi: 10.3389/fmolb.2020.614258

8. Gampala S, Vankeshwaram V, Gadula SSP. Is artificial intelligence the new friend for radiologists? A review article. *Cureus*. (2020) 12(10):e11137. doi: 10.7759/cureus.11137

9. Fritz B, Yi PH, Kijowski R, Fritz J. Radiomics and deep learning for disease detection in musculoskeletal radiology: an overview of novel MRI- and CT-based approaches. *Invest Radiol*. (2023) 58(1):3–13. doi: 10.1097/RLI.0000000000000907

10. Vogrin M, Trojner T, Kelc R. Artificial intelligence in musculoskeletal oncological radiology. *Radiol Oncol*. (2020) 55(1):1–6. doi: 10.2478/raon-2020-0068

11. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence—an ESR white paper. *Insights Imaging*. (2019) 10(1):44. doi: 10.1186/s13244-019-0738-2

12. Varghese BA, Fields BKK, Hwang DH, Duddalwar VA, Matcuk GR Jr, Cen SY. Spatial assessments in texture analysis: what the radiologist needs to know. *Front Radiol*. (2023) 3:1240544–00. doi: 10.3389/fradi.2023.1240544

13. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. (2015) 521(7553):436–44. doi: 10.1038/nature14539

14. Gyftopoulos S, Lin D, Knoll F, Doshi AM, Rodrigues TC, Recht MP. Artificial intelligence in musculoskeletal imaging: current status and future directions. *AJR Am J Roentgenol*. (2019) 213(3):506–13. doi: 10.2214/AJR.19.21117

15. Keles E, Irmakci I, Bagci U. Musculoskeletal MR image segmentation with artificial intelligence. *Adv Clin Radiol*. (2022) 4(1):179–88. doi: 10.1016/j.yacr.2022.04.010

16. Patel DB, Matcuk GR Jr. Imaging of soft tissue sarcomas. *Chin Clin Oncol*. (2018) 7(4):35. doi: 10.21037/cco.2018.07.06

17. Wu JS, Hochman MG. Soft-tissue tumors and tumorlike lesions: a systematic imaging approach. *Radiology*. (2009) 253(2):297–316. doi: 10.1148/radiol.2532081199

18. Gassert FG, Gassert FT, Specht K, Knebel C, Lenze U, Makowski MR, et al. Soft tissue masses: distribution of entities and rate of malignancy in small lesions. *BMC Cancer*. (2021) 21(1):93. doi: 10.1186/s12885-020-07769-2

19. Leporq B, Bouhamama A, Pilleul F, Lame F, Bihane C, Sdika M, et al. MRI-based radiomics to predict lipomatous soft tissue tumors malignancy: a pilot study. *Cancer Imaging*. (2020) 20(1):78. doi: 10.1186/s40644-020-00354-7

20. Sudjai N, Siriwanarangsun P, Lektrakul N, Saiviroonporn P, Maungsomboon S, Phimolsarnti R, et al. Tumor-to-bone distance and radiomic features on MRI distinguish intramuscular lipomas from well-differentiated liposarcomas. *J Orthop Surg Res*. (2023) 18(1):255. doi: 10.1186/s13018-023-03718-4

21. Malinauskaite I, Hofmeister J, Burgermeister S, Neroladaki A, Hamard M, Montet X, et al. Radiomics and machine learning differentiate soft-tissue lipoma and liposarcoma better than musculoskeletal radiologists. *Sarcoma*. (2020) 2020:7163453. doi: 10.1155/2020/7163453

22. Fradet G, Ayde R, Bottois H, El Harchaoui M, Khaled W, Drape JL, et al. Prediction of lipomatous soft tissue malignancy on MRI: comparison between machine learning applied to radiomics and deep learning. *Eur Radiol Exp*. (2022) 6 (1):41. doi: 10.1186/s41747-022-00295-9

23. Keung EZ, Rajkot N, Torres KE, Somaiah N, Hunt KK, Feig BW, et al. Evaluating the impact of surveillance follow-up intervals in patients following resection of primary well-differentiated liposarcoma of the retroperitoneum. *Ann Surg Oncol*. (2021) 28(1):570–5. doi: 10.1245/s10434-020-08582-8

24. Fields BKK, Demirjian NL, Hwang DH, Varghese BA, Cen SY, Lei X, et al. Whole-tumor 3D volumetric MRI-based radiomics approach for distinguishing between benign and malignant soft tissue tumors. *Eur Radiol*. (2021) 31 (11):8522–35. doi: 10.1007/s00330-021-07914-w

25. Tomar GB, Dave JR, Chandekar SS, Mhaske ST. Bone tumors: types and treatments. In: Rangel L, Silva PCML IV Jr, Kirubamani H, editors. *Hormone Therapy and Replacement in Cancer and Aging-Related Diseases*. Rijeka: HR: IntechOpen (2019). p. 1–22. doi: 10.5772/intechopen.86550

26. Sugiyama H, Omonishi K, Yonehara S, Ozasa K, Kajihara H, Tsuya T, et al. Characteristics of benign and malignant bone tumors registered in the Hiroshima tumor tissue registry, 1973–2012. *JB JS Open Access*. (2018) 3(2):e0064. doi: 10. 2106/JBJS.OA.17.00064

27. Salom M, Chiari C, Alessandri JMG, Willegger M, Windhager R, Sanpera I. Diagnosis and staging of malignant bone tumours in children: what is due and what is new? *J Child Orthop*. (2021) 15(4):312–21. doi: 10.1302/1863-2548.15.210107

28. Singh PK. Radiography in skeletal tumours. *J Med Sci Clin Res*. (2018) 6 (10):784–90. doi: 10.18535/jmscr/v6i10.132

29. Azad H, Ahmed A, Zafar I, Bhutta MR, Rabbani MA, Kc HR. X-ray and MRI correlation of bone tumors using histopathology as gold standard. *Cureus*. (2022) 14(7):e27262. doi: 10.7759/cureus.27262

30. O'Sullivan GJ, Carty FL, Cronin CG. Imaging of bone metastasis: an update. *World J Radiol*. (2015) 7(8):202–11. doi: 10.4329/wjr.v7.i8.202

31. Eweje FR, Bao B, Wu J, Dalal D, Liao WH, He Y, et al. Deep learning for classification of bone lesions on routine MRI. *EBioMedicine*. (2021) 68:103402. doi: 10.1016/j.ebiom.2021.103402

32. He Y, Pan I, Bao B, Halsey K, Chang M, Liu H, et al. Deep learning-based classification of primary bone tumors on radiographs: a preliminary study. *EBioMedicine*. (2020) 62:103121. doi: 10.1016/j.ebiom.2020.103121

33. LaLonde R, Xu Z, Irmakci I, Jain S, Bagci U. Capsules for biomedical image segmentation. *Med Image Anal*. (2021) 68:101889. doi: 10.1016/j.media.2020.101889

34. Zhao K, Zhang M, Xie Z, Yan X, Wu S, Liao P, et al. Deep learning assisted diagnosis of musculoskeletal tumors based on contrast-enhanced magnetic resonance imaging. *J Magn Reson Imaging*. (2022) 56(1):99–107. doi: 10.1002/jmri. 28025

35. Wessling D, Herrmann J, Afat S, Nickel D, Othman AE, Almansour H, et al. Reduction in acquisition time and improvement in image quality in T2-weighted MR imaging of musculoskeletal tumors of the extremities using a novel deep

learning-based reconstruction technique in a turbo spin Echo (TSE) sequence. *Tomography*. (2022) 8(4):1759–69. doi: 10.3390/tomography8040148

36. Cay N, Mendi BAR, Batur H, Erdogan F. Discrimination of lipoma from atypical lipomatous tumor/well-differentiated liposarcoma using magnetic resonance imaging radiomics combined with machine learning. *Jpn J Radiol*. (2022) 40(9):951–60. doi: 10. 1007/s11604-022-01278-x

37. Wang H, Zhang J, Bao S, Liu J, Hou F, Huang Y, et al. Preoperative MRI based radiomic machine-learning nomogram may accurately distinguish between benign and malignant soft-tissue lesions: a two-center study. *J Magn Reson Imaging*. (2020) 52(3):873–82. doi: 10.1002/jmri.27111

38. Navarro F, Dapper H, Asadpour R, Knebel C, Spraker MB, Schwarze V, et al. Development and external validation of deep-learning-based tumor grading models in soft- tissue sarcoma patients using MR imaging. *Cancers (Basel)*. (2021) 13(12). doi: 10.3390/cancers13122866

39. Fields BKK, Demirjian NL, Cen SY, Varghese BA, Hwang DH, Lei X, et al. Predicting soft tissue sarcoma response to neoadjuvant chemotherapy using an MRI based delta-radiomics approach. *Mol Imaging Biol*. (2023) 25(4):776–87. doi: 10. 1007/s11307-023-01803-y

40. Fields BKK, Hwang D, Cen S, Desai B, Gulati M, Hu J, et al. Quantitative magnetic resonance imaging (q-MRI) for the assessment of soft-tissue sarcoma treatment response: a narrative case review of technique development. *Clin Imaging*. (2020) 63:83–93. doi: 10.1016/j.clinimag.2020.02.016

41. Blackledge MD, Winfield JM, Miah A, Strauss D, Thway K, Morgan VA, et al. Supervised machine-learning enables segmentation and evaluation of heterogeneous post-treatment changes in multi-parametric MRI of soft-tissue sarcoma. *Front Oncol*. (2019) 9:941. doi: 10.3389/fonc.2019.00941

42. Wang B, Perronne L, Burke C, Adler RS. Artificial intelligence for classification of soft-tissue masses at US. *Radiol Artif Intell*. (2021) 3(1):e200125. doi: 10.1148/ryai. 2020200125

43. Gitto S, Cuocolo R, Emili I, Tofanelli L, Chianca V, Albano D, et al. Effects of interobserver variability on 2D and 3D CT- and MRI-based texture feature reproducibility of cartilaginous bone tumors. *J Digit Imaging*. (2021) 34(4):820–32. doi: 10.1007/s10278-021-00498-3

44. He Y, Guo J, Ding X, van Ooijen PMA, Zhang Y, Chen A, et al. Convolutional neural network to predict the local recurrence of giant cell tumor of bone after curettage based on presurgery magnetic resonance images. *Eur Radiol*. (2019) 29 (10):5441–51. doi: 10.1007/s00330-019-06082-2

45. Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture analysis of imaging: what radiologists need to know. *AJR Am J Roentgenol*. (2019) 212(3):520–8. doi: 10. 2214/AJR.18.20624

46. Strohm L, Hehakaya C, Ranschaert ER, Boon WPC, Moors EHM. Implementation of artificial intelligence (AI) applications in radiology: hindering and facilitating factors. *Eur Radiol*. (2020) 30(10):5525–32. doi: 10.1007/s00330-020-06946-y

47. Geis JR, Brady AP, Wu CC, Spencer J, Ranschaert E, Jaremko JL, et al. Ethics of artificial intelligence in radiology: summary of The Joint European and North American Multisociety Statement. *Radiology*. (2019) 293(2):436–40. doi: 10.1148/ radiol.2019191586

48. Rehman MHU, Hugo Lopez Pinaya W, Nachev P, Teo JT, Ourselin S, Cardoso MJ. Federated learning for medical imaging radiology. *Br J Radiol*. (2023) 96 (1150):20220890. doi: 10.1259/bjr.20220890

Frontiers in Radiology

Investigates new technology and applications in radiology

An exciting new journal in its field, innovating every technical aspect of radiology and radiologist's practice to improve quality, productivity and efficiency.

Discover the latest Research Topics

See more →

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact

frontiers

Frontiers in
Radiology