European Journal of Radiology Open 13 (2024) 100607

Contents lists available at ScienceDirect

European Journal of Radiology Open

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

Study on the classification of benign and malignant breast lesions using a multi-sequence breast MRI fusion radiomics and deep learning model

Wenjiang Wang [a], Jiaojiao Li [b], Zimeng Wang [a], Yanjun Liu [a], Fei Yang [b], Shujun Cui [b,*]

[a] Graduate Faculty, Hebei North University, Zhangjiakou, Hebei, China
[b] Department of Medical Imaging, The First Affiliated Hospital of Hebei North University, Zhangjiakou, Hebei, China

HIGHLIGHT

- The multimodal model combined the radiomics features and deep learning features from the three sequences of T1WI, T2WI, and DCE-MRI of breast MRI.
- The dataset includes various breast diseases, providing broader applicability.

ARTICLE INFO

Keywords:
Artificial Intelligence
Deep Learning
Breast Cancer
Medical Imaging

ABSTRACT

Purpose: To develop a multi-modal model combining multi-sequence breast MRI fusion radiomics and deep learning for the classification of benign and malignant breast lesions, to assist clinicians in better selecting treatment plans.

Methods: A total of 314 patients who underwent breast MRI examinations were included. They were randomly divided into training, validation, and test sets in a ratio of 7:1:2. Subsequently, features of T1-weighted images (T1WI), T2-weighted images (T2WI), and dynamic contrast-enhanced MRI (DCE-MRI) were extracted using the convolutional neural network ResNet50 for fusion, and then combined with radiomic features from the three sequences. The following models were established: T1 model, T2 model, DCE model, DCE_T1_T2 model, and DCE_T1_T2_rad model. The performance of the models was evaluated by the area under the receiver operating characteristic (ROC) curve (AUC), accuracy, sensitivity, specificity, positive predictive value, and negative predictive value. The differences between the DCE_T1_T2_rad model and the other four models were compared using the Delong test, with a P-value < 0.05 considered statistically significant.

Results: The five models established in this study performed well, with AUC values of 0.53 for the T1 model, 0.62 for the T2 model, 0.79 for the DCE model, 0.94 for the DCE_T1_T2 model, and 0.98 for the DCE_T1_T2_rad model. The DCE_T1_T2_rad model showed statistically significant differences ($P < 0.05$) compared to the other four models.

Conclusion: The use of a multi-modal model combining multi-sequence breast MRI fusion radiomics and deep learning can effectively improve the diagnostic performance of breast lesion classification.

1. Introduction

Breast cancer is the most common cancer among women worldwide and a leading cause of cancer-related deaths. Currently, breast cancer has surpassed lung cancer to become the most prevalent malignant tumor globally, and its incidence continues to rise [1]. Therefore, early screening and diagnosis of breast cancer are of utmost importance [2,3]. However, traditional methods for diagnosing breast cancer have numerous limitations [4]. Firstly, the screening stage primarily relies on imaging techniques, which require radiologists with extensive professional knowledge to make accurate diagnoses of breast cancer. Secondly, for patients with high-risk indications from imaging, core needle biopsy or fine-needle aspiration biopsy is used mainly for qualitative diagnosis of breast cancer. However, biopsy samples may not be representative of the entire tumor and involve invasive procedures [5], limiting the application of the pathological gold standard. Thus, there is an urgent

* Correspondence to: Department of Medical Imaging, The First Affiliated Hospital of Hebei North University, No. 12 Changqing Road, Qiaoxi District, Zhangjiakou, Hebei 075000, China.
E-mail address: zjkcuisj@163.com (S. Cui).

https://doi.org/10.1016/j.ejro.2024.100607
Received 12 July 2024; Received in revised form 5 October 2024; Accepted 13 October 2024

need to find a more widespread and suitable new auxiliary method.

Among imaging examination methods, magnetic resonance imaging (MRI) has a significant advantage in diagnosing breast cancer compared to mammography and breast ultrasound [6,7]. MRI utilizes magnetic fields and radio waves to generate multi-parametric images, displaying information from different soft tissues [8,9]. Therefore, it plays a crucial role in the diagnosis of breast cancer. T1-weighted MRI (T1W1) provides clearer images of the anatomical structure of breast tissue [10], while T2-weighted MRI (T2W1) offers more distinct visualization of lesions [11,12]. Additionally, dynamic contrast-enhanced MRI (DCE-MRI), which involves intravenous injection of contrast agents, enhances tissue contrast and provides excellent morphological and some functional information [13,14], showing significant advantages in the detection of early and high-risk breast cancer. Meta-analysis studies have shown that MRI has a sensitivity of 90–92 % and a specificity of 72–75 % in detecting malignant breast lesions [15,16]. However, this requires the extensive professional knowledge of radiologists.

In recent years, with the development of Artificial Intelligence (AI), early detection and diagnosis of breast cancer are undergoing significant transformations. The application of AI technology in medical image analysis, especially in breast cancer screening, has shown tremendous potential and has garnered widespread attention from researchers [17,18]. Radiomics and deep learning are two key methods in this area. Radiomics primarily includes steps such as image acquisition, region of interest segmentation, feature extraction and selection, model construction, and validation. It can capture high-dimensional features such as intensity, shape, and texture of medical images at the pixel or voxel level to build models for the diagnosis, prediction, and treatment of medical diseases. Its multimodal fusion models have shown extraordinary potential in the benign and malignant classification of breast diseases. However, radiomics still faces numerous challenges regarding feature bias and clinical relevance [19]. In contrast, deep learning utilizes advanced mathematical algorithms to extract features in an end-to-end manner. Among them, Convolutional Neural Networks (CNNs) are based on a hierarchical structure and can identify subtle differences in image intensity and shape features, providing more complex and high-dimensional abstract information [20]. Several studies have shown that deep learning is highly effective in breast cancer diagnosis [18,21], with many studies indicating that it has surpassed traditional radiomics and, in some aspects, even outperformed human-level performance, becoming an indispensable assistant in AI-assisted healthcare [22].

Currently, radiologists often need to combine information from various imaging modalities when diagnosing breast MRI, including clinical laboratory results and physical examination findings. This is because different modalities provide complementary information, and the integration of this information aids in a more comprehensive diagnosis. Previous studies have demonstrated that integrative models using multiple information sources outperform any single modality in diagnostic effectiveness [23], highlighting the significant advantages of multimodal models in disease diagnosis. In breast MRI, one study indicated that the BreastScreening-AI framework integrates multimodal imaging and AI image analysis, enhancing the efficiency of breast cancer diagnosis and reducing the workload of physicians while increasing clinician satisfaction [24]. These results suggest that the application of multiple sequences provides rich information for clinicians, and the multimodal fusion of these sequence data can effectively deliver more lesion information, aiding clinicians in making more accurate decisions. By combining AI technology with multi-sequence data, a more comprehensive analysis can be achieved. This multimodal approach not only improves the accuracy of early detection but also significantly reduces misdiagnosis and missed diagnosis rates, thereby offering patients more effective treatment options.

At present, multimodal feature fusion networks have been extensively researched and applied in breast cancer; however, the fusion methods vary significantly. Zheng et al. [25] constructed a multimodal deep learning system for predicting lymph node metastasis in breast cancer by integrating T1WI, T2WI, DCE-MRI, and clinical imaging features, achieving significant results. This study demonstrates the enormous potential of multimodal information in the clinical application of breast cancer. Similarly, Daimiel Naranjo et al. [26] proposed a multiparametric breast MRI model that fuses radiomics with traditional machine learning, successfully classifying breast lesions with an accuracy of 88.5 % and an AUC of 0.96, further highlighting the effectiveness of multimodal fusion technology. Although the aforementioned studies employed multimodal feature fusion methods, there are significant differences in specific fusion strategies and algorithm designs, affecting the comparability and generalizability of their results [27]. For instance, some studies focus on simply concatenating multimodal features, while others apply more complex feature extraction and selection mechanisms. Additionally, the weighting allocation of imaging data and radiomic features, fusion methods, and the design of deep learning network architectures can also impact the final predictive performance differently. Therefore, despite the promising prospects of multimodal feature fusion technology in the diagnosis and classification of breast diseases, optimizing feature fusion methods to enhance the model's generalization ability across different datasets and clinical scenarios remains a key challenge in the field. This not only requires methodological innovation but also demands in-depth exploration of the adaptability of different data sources, features, and algorithms in practical applications, with the aim of more accurately distinguishing between benign and malignant breast conditions and providing more reliable decision-support tools for clinical practice.

In this study, we propose a novel predictive framework that integrates radiomic features from T1WI, T2WI and DCE-MRI, along with deep learning features from these three sequences, to classify the benign and malignant lesions of breast diseases. The innovation of our research lies in employing a transfer learning approach for training after the fusion of radiomics and deep learning techniques for multi-sequence breast MRI, rather than simply merging their image features and using a classifier to output classification results. Notably, we also compared the performance of the fusion model with single-sequence models. The primary objective of this study is to establish a multimodal model for breast MRI that further enhances the accuracy of classifying the benign and malignant nature of breast diseases.

2. Materials and methods

2.1. Patient selection

This retrospective study was approved by the Ethics Committee of our hospital (Approval No.: W2024014), with a waiver of informed consent requirements. All patients who underwent breast MRI examination at the First Affiliated Hospital of Hebei North University from January 2022 to June 2024 were included. The inclusion criteria were as follows: (1) clinically diagnosed with breast disease; (2) underwent breast MRI examination within 2 weeks before surgery; (3) had complete postoperative pathological results. The exclusion criteria were: (1) underwent chemotherapy or biopsy before the breast MRI scan; (2) had multifocal breast disease; (3) DCE sequences did not show the lesion area; (4) MRI image quality did not meet the inclusion requirements. A total of 314 patients were included, with 146 benign cases and 168 malignant cases. The participant selection flowchart is shown in Fig. 1.

2.2. Imaging acquisition

All breast MRI examinations were performed on a Philips 3.0 T scanner. All patients were positioned prone, using a dedicated breast coil. The T1WI parameters were: TR = 541.7 ms, TE = 8.0 ms, slice thickness = 4 mm, and field of view (FOV) = 260×320 mm. The T2WI parameters were: TR = 3771.8 ms, TE = 90.0 ms, slice thickness = 4 mm, and FOV = 260×320 mm. The DCE-MRI parameters were: FOV =

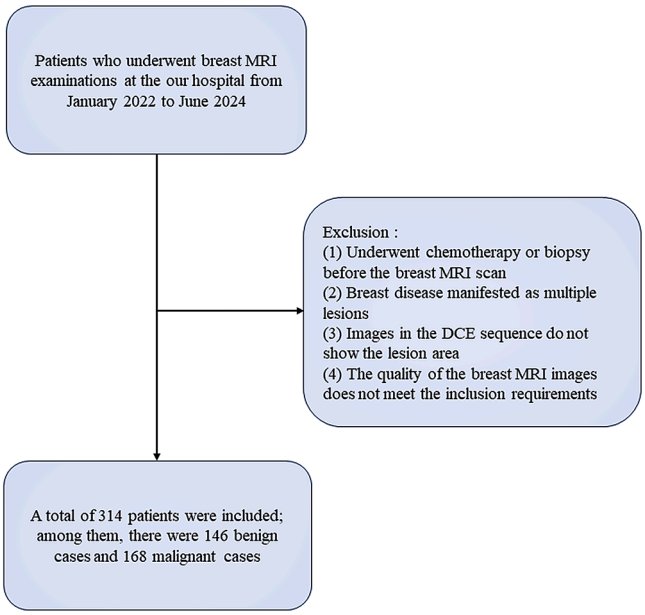

Fig. 1. Participant Selection Flowchart, illustrating the steps for participant inclusion.

340×410 mm, TR = 4.3 ms, TE = 2.1 ms, and slice thickness = 4 mm. During the DCE-MRI scan, a high-pressure injector was used to administer a contrast agent, gadopentetate dimeglumine or gadoterate meglumine, via the median cubital vein at a dose of 0.1 mmol/kg and a flow rate of 2.5 ml/s. After the contrast injection, 30 ml of saline was injected at the same flow rate. The total scan time was 10 min and 15 s. The second-phase DCE-MRI images were used in this study.

2.3. Data acquisition and ROI delineation

The T1WI, T2WI, and DCE-MRI images of eligible patients were retrieved from the PACS system and exported in DICOM format. Two radiologists with over three years of experience independently reviewed the images without knowledge of the diagnostic results and reached a consensus. Regions of interest (ROIs) of the breast lesions were delineated on the DCE-MRI images for each patient using 3D Slicer software (https://www.slicer.org). In cases of disagreement between the two radiologists, a third senior radiologist with over ten years of experience made the final decision.

2.4. Dataset partition

Each patient was labeled for breast cancer based on pathological results. The dataset was randomly divided into training, validation, and test sets in a ratio of 7:1:2, with the training set containing 215 cases, the validation set containing 38 cases, and the test set containing 61 cases. The neural network model was trained on the training set and validated on the validation set. Model training was stopped when the loss on the validation set continued to increase. Finally, the performance of the model was evaluated on the test set.

2.5. Radiomics feature extraction

The ROIs delineated on the DCE-MRI were then mapped onto the T1WI and T2WI sequences to obtain the ROIs for the T1WI and T2WI sequences. Using the PyRadiomics library (https://pyradiomics.readthedocs.io/en/latest/), the original DCE-MRI images and ROI images of each patient were input to extract radiomic features from the lesion area, resulting in 107 features (14 shape features, 18 first-order features, and 75 texture features). Radiomic features were then extracted from the T1WI and T2WI sequences of each patient. In total, 321 radiomic features were extracted for each patient from the T1WI, T2WI, and DCE-MRI sequences.

2.6. Deep learning model

The pixel values of the original breast MRI images for each patient were adjusted to a range of 0–1500, enhancing the contrast of the MRI and highlighting the lesion areas. Additionally, noise was removed from the original images using min-max normalization and mean normalization methods. The slice with the largest ROI and the slices immediately above and below it were identified for each patient. The center coordinates of the ROI were mapped to the corresponding original image slices, and the area centered on these coordinates was cropped to 224×224 pixels. These three cropped slices were then combined into a three-channel image and input into the ResNet50 [28]. Shared weights were used for deep learning feature extraction. The deep learning features extracted from the T1WI, T2WI, and DCE-MRI, along with the 321 radiomic features, were fused at the fully connected layer. The prediction was then output using the Softmax function. Additionally, online data augmentation techniques were applied to the breast MRI dataset, with a 0.5 probability of random horizontal and vertical flips and random rotations within a range of [-10°, 10°]. Based on the number of fused features used in the fully connected layer, the following models were established: T1 model, T2 model, DCE model, DCE_T1_T2, and DCE_T1_T2_rad model. The overall framework of the multimodal breast MRI model is shown in Fig. 2.

In this study, the ResNet50 model from the deep learning neural network was used for feature extraction. The ResNet50 network was loaded with pre-trained parameters from ImageNet. However, during training, all parameters except those of the final fully connected layer were frozen, and only the parameters of the last layer were trained. The cross-entropy loss function was used to construct the loss function for training the proposed network. During model training, the SGD optimizer with a weight decay factor of 5×10^{-4} and a momentum of 0.9 was used. The batch size was set to 30. The initial learning rate was 0.01. The training epoch was set to 100. The learning rate was reduced to 0.5 of the original learning rate at epochs 10, 30, 40, and 100. All experiments were conducted on an Intel(R) Core(TM) i5–13490F 2.50 GHz and NVIDIA GeForce RTX 4060ti 16 GB GPU. The software environment was Windows 10 and the Python programming language (version 3.9, https://www.python.org/). The Python packages used included gdcm, matplotlib, nibabel, numpy, opencv_python, Pillow, pydicom, scikit_learn, scipy, SimpleITK, skimage, xlrd, xlwt, and pytorch.

2.7. Statistical methods

For normally distributed measurement data, the mean ± standard deviation ($\bar{x}\pm s$) was used, while for non-normally distributed data, the median and interquartile range M (Q1, Q3) were used. The classification performance of the models was evaluated using the area under the receiver operating characteristic (ROC) curve (AUC), accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV). The DeLong test [29] was used to compare the significance levels of the AUC between the DCE_T1_T2_Rad model and the other four models. A P-value <0.05 was considered statistically significant.

3. Results

3.1. Patient characteristics

A total of 314 female patients were included in this study. Among them, 146 had benign lesions, including types such as intraductal papilloma, fibroadenoma, adenosis, benign phyllodes tumor, breast abscess, and lipoma. The 168 malignant cases included types such as ductal

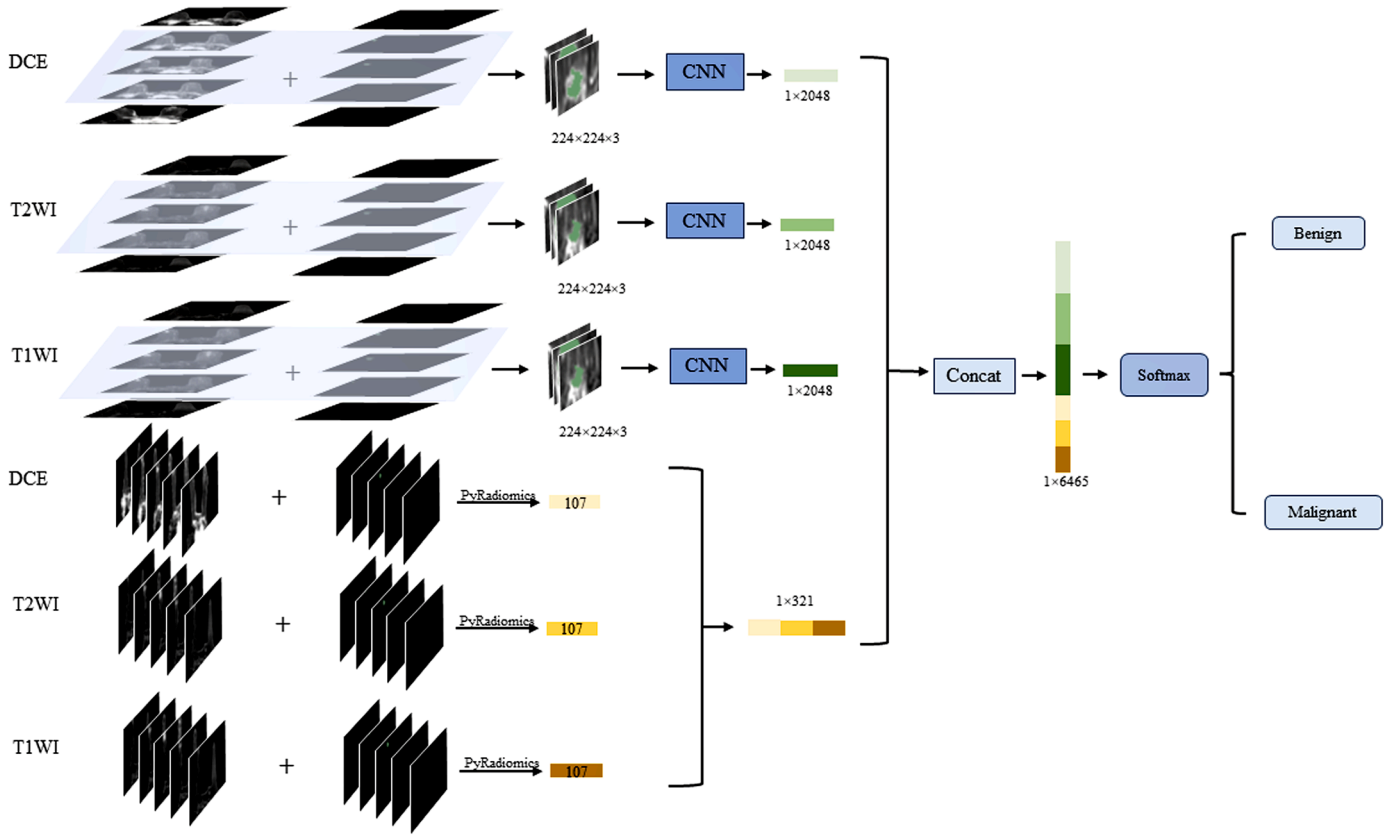

Fig. 2. Multi-modal Fusion Network Architecture; DCE: Dynamic Contrast-enhanced Breast MRI sequence; T2WI: T2-weighted Breast MRI sequence; T1WI: T1-weighted Breast MRI sequence; CNN: Convolutional Neural Network using ResNet50; PyRadiomics: Radiomics Feature Extraction; Concat: Feature Concatenation; Softmax: Activation Function; Benign: Predicted as benign; Malignant: Predicted as malignant.

carcinoma in situ, invasive ductal carcinoma, invasive lobular carcinoma, mucinous carcinoma, lymphoma, inflammatory breast cancer, and intraductal papillary carcinoma. The clinical characteristics of the dataset are detailed in Table 1.

Table 1
Clinical Features of the Data.

Classification	Age ($\bar{x}\pm s$)	Category	Quantity	Percentage
Benign (n = 146)	32.3 ±10.5	intraductal papilloma	25	45 %
		fibroadenoma	90	
		adenosis	7	
		benign phyllodes tumor	9	
		breast abscess	13	
		lipoma	2	
Malignant (n = 168)	51.6 ±12.4	ductal carcinoma in situ	25	54 %
		invasive ductal carcinoma	83	
		invasive lobular carcinoma	53	
		mucinous carcinoma	1	
		lymphoma	1	
		inflammatory breast cancer	2	
		intraductal papillary carcinoma	3	

footnote: This table describes the patient's age, classification and proportion of benign and malignant tumors, and the number of patients in each tumor category.

3.2. Train and validation loss

The loss results of the fusion model on the training and validation sets are shown in Fig. 3. When both loss curves gradually decreased and flattened, it indicated that the model was converging and stabilizing on both the training and validation sets, suggesting good fitting on the test set as well.

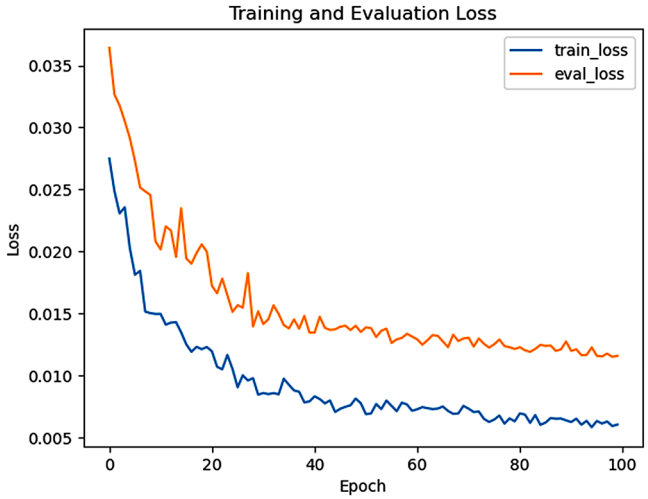

Fig. 3. Training and Validation Losses.

3.3. Performance on the test set

3.3.1. Class activation maps

Using the class activation mapping technique with generalized gradients [30], the regions of interest within the model could be visualized as heatmaps, thereby enhancing model interpretability. In the DCE_T1_T2_rad model, class activation maps were visualized for DCE-MRI images (Figure (a)), highlighting the regions of interest identified by the network model, thus enhancing the model's interpretability. Specifically, T1WI, T2WI, and DCE-MRI images were input into the network model, generating class activation maps for the DCE sequence (Figure (b)), which were then fused with the original images to obtain the combined display results (Figure (c)), as shown in Figs. 4 and 5. Fig. 4 presented an example of invasive carcinoma, while Fig. 5 presented an example of fibroadenoma. The model was observed to accurately focus on the lesion areas, enabling correct diagnosis.

3.3.2. AUC of each model

In the single-sequence models, the AUC for the T1 model was 0.53 (95 % CI: 0.39–0.68), the AUC for the T2 model was 0.62 (95 % CI: 0.48–0.76), and the AUC for the DCE model was 0.79 (95 % CI: 0.67–0.90). In the fusion models, the AUC for the DCE_T1_T2 model was 0.94 (95 % CI: 0.87–1.00), and the AUC for the DCE_T1_T2_Rad model was 0.98 (95 % CI: 0.95–1.00). The ROC curves for each model are shown in Fig. 6, and other evaluation metrics are detailed in Table 2.

3.3.3. Confusion matrix of each model

To intuitively demonstrate the superiority of the proposed fusion models, we presented the confusion matrices for the five models, as shown in Fig. 7. The confusion matrices clearly illustrated the classification performance of each model across different categories, further validating the accuracy advantage of the fusion models.

3.3.4. Comparison using Delong test

The Delong test [29] is a non-parametric statistical method used to compare the area under the ROC curve (AUC) of two or more models. Delong test was applied to assess the significance of differences in discriminative ability between the DCE_T1_T2_Rad model and the other four models. The results indicated that all pairwise comparisons yielded P-values < 0.05, suggesting significant differences among them (details in Table 3).

4. Discussion

The aim of this study was to explore the use of deep learning and radiomics to fuse features from T1WI, T2WI, and DCE-MRI in a dataset containing various breast diseases for the diagnosis of breast lesions. Overall, the proposed multi-sequence breast MRI fusion deep learning and radiomics model demonstrated superior diagnostic performance in the test set compared to single-sequence models and a fusion model with all three sequences. The model showed higher AUC and other evaluation metrics (ACC, SEN, SPE, PPV, NPV) also performed well.

In most studies, researchers often select specific datasets that may not comprehensively cover all types of breast tumors. However, in our study, the breast MRI data encompassed a wide range of types including intraductal papilloma, fibroadenoma, adenosis, benign phyllodes tumor, breast abscess, lipoma, ductal carcinoma in situ (DCIS), invasive ductal carcinoma (IDC), invasive lobular carcinoma (ILC), mucinous carcinoma, lymphoma, inflammatory breast cancer, and intraductal papillary carcinoma. This comprehensive inclusion of both common and rare types of breast cancer enhances the applicability of our study results across diverse scenarios.

In our study, we first delineated ROIs on DCE-MRI and then mapped them onto T1WI and T2WI to create ROIs for these sequences, which were used for extracting radiomic features and cropping input images for deep learning. This approach was primarily chosen due to our limited data volume, aiming to effectively reduce noise interference by focusing on ROI features. Additionally, discrepancies in lesion boundaries observed between DCE-MRI and T1WI/T2WI sequences motivated us to map the DCE-MRI findings onto T1WI and T2WI, considering that lesion areas identified on DCE-MRI, even if not visualized on T1WI and T2WI, were considered abnormal and included within the ROI. Breast MRI provides more sensitive and detailed pathophysiological information [31], with T1WI, T2WI, and DCE-MRI being crucial for breast lesion diagnosis. We utilized the ResNet50 CNN to extract deep learning features from breast MRI images for building a classification model. In single-sequence analysis, the AUC was lowest for the T1 model at 0.53, slightly higher for the T2 model at 0.62, and highest for the DCE model at 0.79, highlighting the higher diagnostic value of DCE-MRI in lesion diagnosis. Upon fusing data from all three sequences, the AUC significantly improved to 0.94, demonstrating the deep learning model's capability to learn discriminative features. Furthermore, leveraging the PyRadiomics library for radiomic feature extraction across these sequences and integrating these with deep learning features yielded a combined AUC of 0.98. This marked improvement underscores the enhanced performance of the model through the fusion of diverse imaging features, particularly integrating deep learning and radiomic features. These findings underscore the potential and significance of multimodal fusion in medical image analysis for substantially improving lesion diagnosis accuracy.

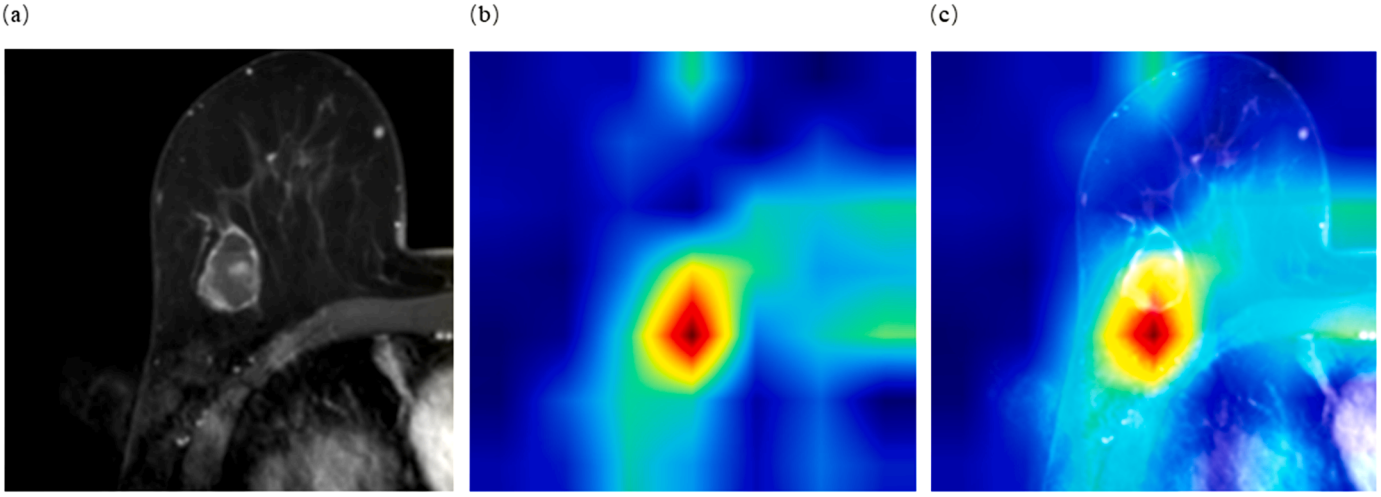

Fig. 4. Class activation map of a patient with invasive breast cancer in DCE-MRI.

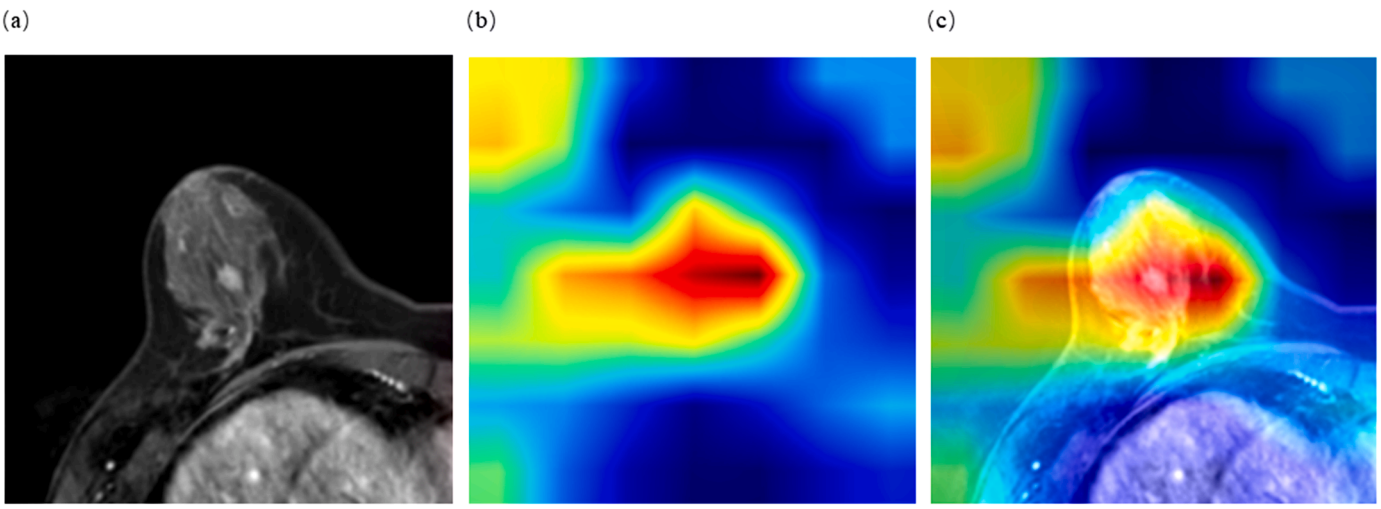

Fig. 5. Class activation map of a patient with fibroadenoma in DCE-MRI of the Breast.

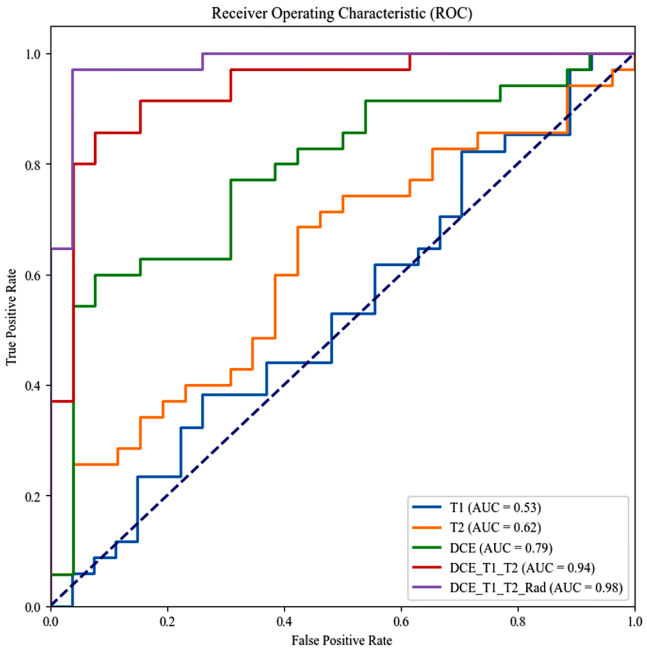

Fig. 6. ROC Curves of Various Models.

In previous studies, several deep learning methods have been proposed for the diagnosis of breast lesions, achieving promising results. Tomoyuki Fujioka et al. [32] employed multiple deep learning networks to extract DCE-MRI network features for distinguishing benign from malignant breast lesions. The best-performing InceptionResNetV2 model achieved an AUC of 0.90, comparable to the diagnostic accuracy of expert radiologists. However, this study utilized only a single DCE-MRI sequence. In our research, despite achieving an AUC of 0.79 with a single DCE-MRI sequence, integrating T1WI, T2WI, and DCE-MRI sequences through deep learning boosted the AUC to 0.94. Furthermore, by incorporating radiomic features, the AUC further increased to 0.98. This highlights that the fusion of multimodal information significantly enhances model performance. In another study, Wang et al. [33]. employed transfer learning to classify benign and malignant breast MRI DCE-MRI images, fine-tuning MobileNet within CNNs and developing MobileNetV1 and MobileNetV2 models with variations in pre-training parameter usage (V1_False, V1_True, V2_False, and V2_True). While achieving a high accuracy of 0.98 for the best-performing V1_True model, their AUC was only 0.74, indicating an imbalance in model evaluation metrics. In contrast, our study not only fused multi-sequence images but also integrated deep learning and radiomics, demonstrating balanced performance across all evaluation metrics. These findings underscore the efficacy of multimodal fusion approaches in enhancing the diagnostic accuracy of breast lesion classification models, aligning well with the standards of scientific literature.

AI has achieved revolutionary advancements in breast cancer diagnosis [34], and multimodal models have garnered significant attention due to their notable improvements in accuracy. Our fusion model demonstrates high accuracy and can be widely applied in clinical settings. Firstly, it can be used for breast cancer screening, facilitating early diagnosis and treatment, thereby improving the survival rate and quality of life for breast cancer patients [35]. Secondly, limiting pathological evaluations to only those lesions predicted to be malignant would significantly reduce the need for preoperative biopsies. On the other hand, current clinical diagnosis primarily relies on radiologists interpreting images, and in large hospitals, the high patient volume can lead to human fatigue, causing subtle features of lesions to be overlooked, resulting in misdiagnoses. Moreover, for less experienced younger doctors, our model can enhance their diagnostic accuracy regarding lesions.

Table 2
Evaluation metrics of models on the test set.

model	AUC (95 % CI)	ACC (95 % CI)	SEN (95 % CI)	SPE (95 % CI)	PPV (95 % CI)	NPV (95 % CI)
T1	0.53(0.39–0.68)	0.48(0.33–0.62)	0.44(0.29–0.59)	0.52(0.37–0.67)	0.54(0.39–0.68)	0.42(0.28–0.57)
T2	0.62(0.48–0.76)	0.56(0.41–0.70)	0.49(0.34–0.63)	0.65(0.52–0.79)	0.65(0.52–0.79)	0.49(0.34–0.63)
DCE	0.79(0.67–0.90)	0.72(0.59–0.85)	0.54(0.40–0.69)	0.96(0.91–1.00)	0.95(0.90–1.00)	0.61(0.47–0.75)
DCE_T1_T2	0.94(0.87–1.00)	0.85(0.76–0.95)	0.80(0.69–0.91)	0.92(0.86–0.99)	0.93(0.87–1.00)	0.77(0.66–0.89)
DCE_T1_T2_rad	0.98(0.95–1.00)	0.93(0.87–1.00)	0.97(0.93–1.00)	0.89(0.81–0.97)	0.92(0.85–0.99)	0.96(0.91–1.00)

The table describes the AUC, accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value (NPV), and their confidence intervals (CI) for each model.

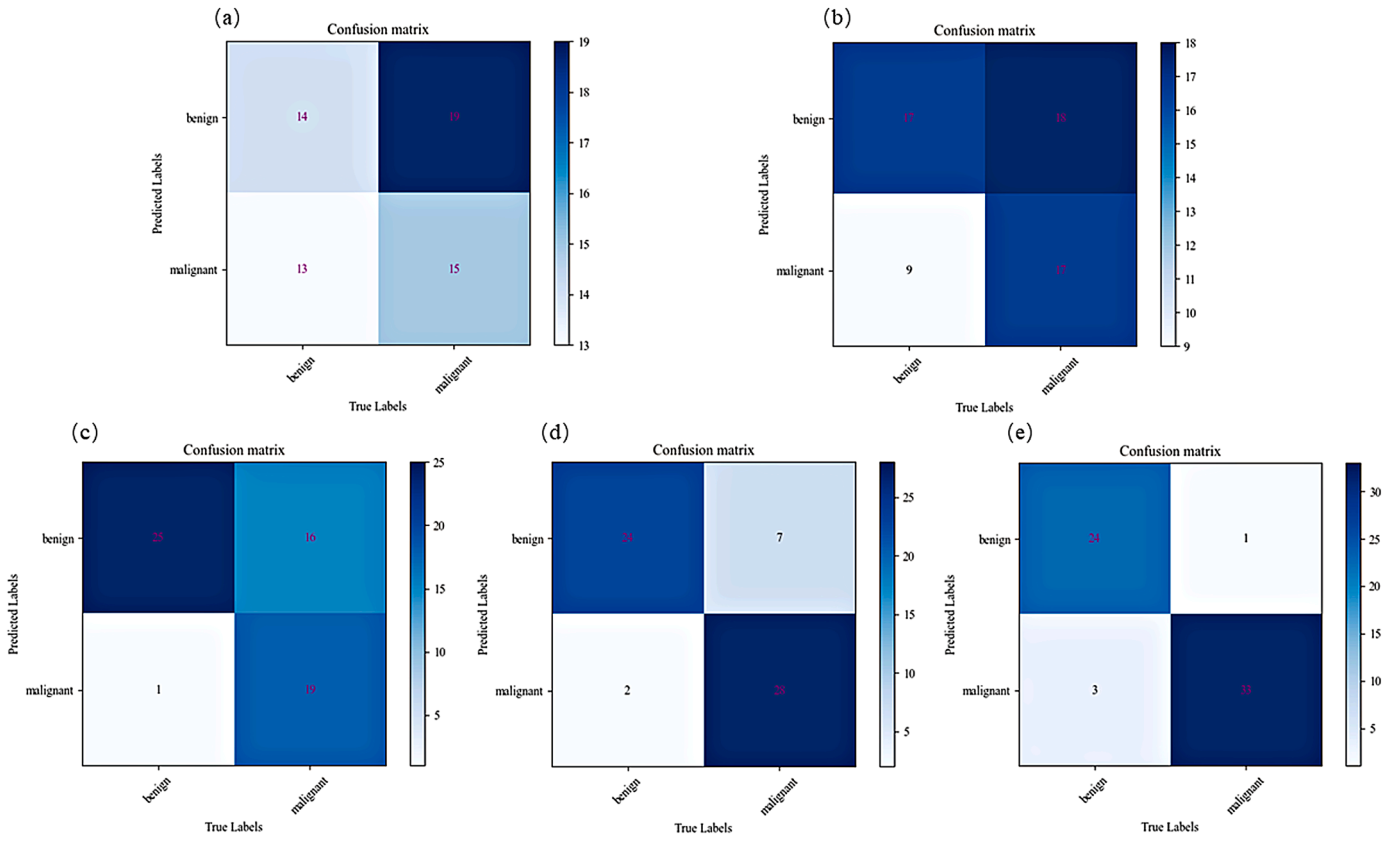

Fig. 7. Confusion Matrices of Each Model; (a) the confusion matrix for the T1 model; (b) the confusion matrix for the T2 model; (c) the confusion matrix for the DCE model; (d) the confusion matrix for the DCE_T1_T2 model; and (e) the confusion matrix for the DCE_T1_T2_rad model.

Table 3
Delong Test Comparison of DCE_T1_T2_Rad Model with Other Models.

Model	*P*-values
DCE_T1_T2_rad *VS* T1[f]	<0.01
DCE_T1_T2_rad *VS* DCE T2	<0.01
DCE_T1_T2_rad *VS* DCE	<0.01
DCE_T1_T2_rad *VS* DCE_T1_T2	<0.01

[f] *VS:* Versus

Thus, our multimodal model can better assist clinical practice [36,37].

However, our study still has some limitations. First, when extracting radiomic features, we relied on manual delineation of the ROIs, which may introduce certain subjective biases [38]. This bias not only affects the accuracy of the features but may also compromise the performance of the final model. In the future, we will explore more objective neural network approaches for ROI segmentation. Second, our data was sourced from a single center, which may lead to selection bias. Due to differences in equipment, operational procedures, and patient population characteristics among various medical institutions, a single-center dataset may not adequately represent broader clinical realities, limiting the generalizability of our findings in real-world screening [39]. Thus, the applicability of our results in different populations needs further validation. In the future, obtaining a large-scale, multicenter dataset will be crucial for training and validating deep learning models, as it will enhance the model's generalization capabilities and reliability in practical clinical applications. Third, the current application of AI in image analysis lacks unified standards and protocols. There are significant discrepancies between the existing public datasets and the T1WI, T2WI, and DCE-MRI data we used, making it difficult to directly expand our training data. Therefore, we plan to utilize neural network technology to register different image data in the future, reducing the differences between datasets and thereby improving model performance and adaptability. Fourth, this study is retrospective and still contains many uncertainties. When partitioning the dataset, we randomly divided it into training, validation, and test sets in a 7:1:2 ratio. This random division may lead to variations in results across different splits. Additionally, retrospective studies often exclude patient data that do not meet inclusion criteria, resulting in selection bias that lowers the model's simulation of real breast cancer data [40]. Therefore, further prospective validation is needed. Lastly, our research mainly focused on exploring the image features of T1WI, T2WI, and DCE-MRI, while not addressing other imaging sequences, ultrasound examinations, mammographic images, and clinical laboratory data [41]. As more datasets become available in the future, we will further investigate the integration of multimodal features to enhance the diagnostic capabilities for breast lesions.

In summary, despite some limitations, the developed multimodal fusion model achieves high accuracy in diagnosing breast lesions. For certain suspicious lesions, biopsies are often chosen to pursue aggressive treatment. If our model is used for diagnosis, predicting suspicious lesions as malignant before proceeding to pathology can reduce unnecessary biopsies. However, it is crucial to explain the diagnostic process of the AI model, as visualizing the model will enhance patient understanding. Focusing on deploying a tool that is both highly accurate and easy to interpret should be a priority.

5. Conclusions

Our study demonstrated that the multimodal fusion of breast MRI sequences using radiomics and deep learning can enhance the diagnosis of breast lesions, potentially improving the automated diagnostic performance of breast MRI. This approach aims to reduce subjective

interference and enhance diagnostic consistency.

CRediT authorship contribution statement

Yanjun Liu: Visualization, Resources, Data curation. **Fei Yang:** Supervision, Formal analysis, Data curation. **Shujun Cui:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization. **Wenjiang Wang:** Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Jiaojiao Li:** Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Zimeng Wang:** Visualization, Validation, Software, Resources, Data curation.

Ethical approval

Institutional Review Board approval was obtained (ID:W2024014).

Ethics approval and consent to participate

None.

Informed consent

Written informed consent was waived by the Institutional Review Board.

Funding

This work was supported by Hebei Natural Science Foundation (Grant No. H2023405031)

Declaration of Competing Interest

None of the authors have any personal, financial, commercial, or academic conflicts of interest.

References

[1] H. Sung, J. Ferlay, R.L. Siegel, et al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 Cancers in 185 Countries, CA A Cancer J. Clin. 71 (3) (2021) 209–249, https://doi.org/10.3322/caac.21660.
[2] Y.J. Kim, K.G. Kim, Detection and weak segmentation of masses in gray-scale breast mammogram images using deep learning, Yonsei Med. J. 63 (Suppl) (2022) S63–S73, https://doi.org/10.3349/ymj.2022.63.S63.
[3] F.D. Schwab, D.J. Huang, S.M. Schmid, A. Schötzau, U. Güth, Self-detection and clinical breast examination: comparison of the two "classical" physical examination methods for the diagnosis of breast cancer, Breast 24 (1) (2015) 90–92, https://doi.org/10.1016/j.breast.2014.11.008.
[4] S. Alimirzaie, M. Bagherzadeh, M.R. Akbari, Liquid biopsy in breast cancer: A comprehensive review, Clin. Genet. 95 (6) (2019) 643–660, https://doi.org/10.1111/cge.13514.
[5] S. Siviengphanom, Z. Gandomkar, S.J. Lewis, P.C. Brennan, Mammography-based radiomics in breast cancer: a scoping review of current knowledge and future needs, Acad. Radiol. 29 (8) (2022) 1228–1247, https://doi.org/10.1016/j.acra.2021.09.025.
[6] R.M. Mann, Breast screening: "If you really want to see it, you just make an MRI, Eur. Radiol. 33 (12) (2023) 8410–8412, https://doi.org/10.1007/s00330-023-09890-9.
[7] B.S. Alotaibi, R. Alghamdi, S. Aljaman, et al., The accuracy of breast cancer diagnostic tools, Cureus 16 (1) (2024) e51776, https://doi.org/10.7759/cureus.51776.
[8] M. Iima, M. Honda, E.E. Sigmund, A. Ohno Kishimoto, M. Kataoka, K. Togashi, Diffusion MRI of the breast: current status and future directions, J. Magn. Reson Imaging 52 (1) (2020) 70–90, https://doi.org/10.1002/jmri.26908.
[9] I. Daimiel Naranjo, R. Lo Gullo, C. Saccarelli, et al., Diagnostic value of diffusion-weighted imaging with synthetic b-values in breast tumors: comparison with dynamic contrast-enhanced and multiparametric MRI, Eur. Radiol. 31 (1) (2021) 356–367, https://doi.org/10.1007/s00330-020-07094-z.
[10] M. Wallis, Tardivon A. Helbich, T. Schreer, I, European Society of Breast Imaging. Guidelines from the European Society of Breast Imaging for diagnostic interventional breast procedures, Eur. Radio. 17 (2) (2007) 581–588, https://doi.org/10.1007/s00330-006-0408-x.
[11] C. Westra, V. Dialani, T.S. Mehta, R.L. Eisenberg, Using T2-weighted sequences to more accurately characterize breast masses seen on MRI, AJR Am. J. Roentgenol. 202 (3) (2014) W183–W190, https://doi.org/10.2214/AJR.13.11266.
[12] G. Santamaría, M. Velasco, X. Bargalló, X. Caparrós, B. Farrús, P. Luis Fernández, Radiologic and pathologic findings in breast tumors with high signal intensity on T2-weighted MR images, Radiographics 30 (2) (2010) 533–548, https://doi.org/10.1148/rg.302095044.
[13] D. Leithner, G.J. Wengert, T.H. Helbich, et al., Clinical role of breast MRI now and going forward, Clin. Radiol. 73 (8) (2018) 700–714, https://doi.org/10.1016/j.crad.2017.10.021.
[14] C.K. Kuhl, H.H. Schild, Dynamic image interpretation of MRI of the breast, J. Magn. Reson Imaging 12 (6) (2000) 965–974, https://doi.org/10.1002/1522-2586(200012)12:6<965::aid-jmri23>3.0.co;2-1.
[15] N.H.G.M. Peters, I.H.M. Borel Rinkes, N.P.A. Zuithoff, W.P.T.M. Mali, K.G. M. Moons, P.H.M. Peeters, Meta-analysis of MR imaging in the diagnosis of breast lesions, Radiology 246 (1) (2008) 116–124, https://doi.org/10.1148/radiol.2461061298.
[16] L.R. Medeiros, C.S. Duarte, D.D. Rosa, et al., Accuracy of magnetic resonance in suspicious breast lesions: a systematic quantitative review and meta-analysis, Breast Cancer Resl Treat. 126 (2) (2011) 273–285, https://doi.org/10.1007/s10549-010-1326-9.
[17] F. Gallivanone, G. Bertoli, D. Porro, Radiogenomics, breast cancer diagnosis and characterization: current status and future directions, MPs 5 (5) (2022) 78, https://doi.org/10.3390/mps5050078.
[18] Y.-M. Lei, M. Yin, M.-H. Yu, et al., Artificial intelligence in medical imaging of the breast, Front. Oncol. 11 (2021) 600557, https://doi.org/10.3389/fonc.2021.600557.
[19] N. Horvat, N. Papanikolaou, D.-M. Koh, Radiomics beyond the hype: a critical evaluation toward oncologic clinical use, Radiol.: Artif. Intell. 6 (4) (2024) e230437, https://doi.org/10.1148/ryai.230437.
[20] X. Yang, X. Fan, S. Lin, et al., Assessment of lymphovascular invasion in breast cancer using a combined mri morphological features, radiomics, and deep learning approach based on dynamic contrast-enhanced MRI, J. Magn. Reson Imaging (2023), https://doi.org/10.1002/jmri.29060.
[21] Y. Jiang, A.V. Edwards, G.M. Newstead, Artificial intelligence applied to breast MRI for improved diagnosis, Radiology 298 (1) (2021) 38–46, https://doi.org/10.1148/radiol.2020200292.
[22] F.M. Calisto, C. Santiago, N. Nunes, J.C. Nascimento, Introduction of human-centric AI assistant to aid radiologists for multimodal breast image classification, Int. J. Hum. -Comput. Stud. 150 (2021) 102607, https://doi.org/10.1016/j.ijhcs.2021.102607.
[23] C. Yoon, S. Misra, K.-J. Kim, C. Kim, B.J. Kim, Collaborative multi-modal deep learning and radiomic features for classification of strokes within 6h, Expert Syst. Appl. 228 (2023) 120473, https://doi.org/10.1016/j.eswa.2023.120473.
[24] Calisto F.M. Human-Centered Design of Personalized Intelligent Agents in Medical Imaging Diagnosis 2024. https://doi.org/10.13140/RG.2.2.28353.33126.
[25] H. Zheng, L. Jian, L. Li, W. Liu, W. Chen, Prior Clinico--Radiological Features Informed Multi--Modal MR Images Convolution Neural Network: A novel deep learning framework for prediction of lymphovascular invasion in breast cancer, Cancer Med. 13 (3) (2024) e6932, https://doi.org/10.1002/cam4.6932.
[26] I. Daimiel Naranjo, P. Gibbs, J.S. Reiner, et al., Breast lesion classification with multiparametric breast MRI using radiomics and machine learning: a comparison with radiologists' performance, Cancers (Basel) 14 (7) (2022) 1743, https://doi.org/10.3390/cancers14071743.
[27] M. Morais, F.M. Calisto, C. Santiago, C. Aleluia, J.C. NascimentoClassification of Breast Cancer in Mri with Multimodal Fusion. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). Cartagena, Colombia: IEEE; 2023;1–4. https://doi.org/10.1109/ISBI53787.2023.10230686. .
[28] He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition 2015. https://doi.org/10.48550/arXiv.1512.03385.
[29] E.R. DeLong, D.M. DeLong, D.L. Clarke-Pearson, Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach, Biometrics 44 (3) (1988) 837–845, https://doi.org/10.2307/2531595.
[3] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, IEEE Int. Conf. Comput. Vis. (ICCV) 2017 (2017) 618–626, https://doi.org/10.1109/ICCV.2017.74.
[31] R. Adam, K. Dell'Aquila, L. Hodges, T. Maldjian, T.Q. Duong, Deep learning applications to breast cancer detection by magnetic resonance imaging: a literature review, Breast Cancer Res. 25 (1) (2023) 87, https://doi.org/10.1186/s13058-023-01687-4.
[32] T. Fujioka, Y. Yashima, J. Oyama, et al., Deep-learning approach with convolutional neural network for classification of maximum intensity projections of dynamic contrast-enhanced breast magnetic resonance imaging, Magn. Reson. Imaging 75 (2021) 1–8, https://doi.org/10.1016/j.mri.2020.10.003.
[33] L. Wang, M. Zhang, G. He, D. Shen, M. Meng, Classification of breast lesions on DCE-MRI data using a fine-tuned mobilenet, Diagnostics (Basel) 13 (6) (2023) 1067, https://doi.org/10.3390/diagnostics13061067.
[34] P. Diogo, M. Morais, F.M. Calisto, C. Santiago, C. Aleluia, J.C. NascimentoWeakly-Supervised Diagnosis and Detection of Breast Cancer Using Deep Multiple Instance Learning. 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). Cartagena, Colombia: IEEE; 2023;1–4. https://doi.org/10.1109/ISBI53787.2023.10230448. .
[35] Y.-F. Zhang, C. Zhou, S. Guo, et al., Deep learning algorithm-based multimodal MRI radiomics and pathomics data improve prediction of bone metastases in primary

prostate cancer, J. Cancer Res. Clin. Oncol. 150 (2) (2024) 78, https://doi.org/10.1007/s00432-023-05574-5.
[36] K. Pierre, M. Gupta, A. Raviprasad, et al., Medical imaging and multimodal artificial intelligence models for streamlining and enhancing cancer care: opportunities and challenges, Expert Rev. Anticancer Ther. 23 (12) (2023) 1265–1279, https://doi.org/10.1080/14737140.2023.2286001.
[37] F.M. Calisto, N. Nunes, J.C. Nascimento, Modeling adoption of intelligent agents in medical imaging, Int. J. Hum. -Comput. Stud. 168 (2022) 102922, https://doi.org/10.1016/j.ijhcs.2022.102922.
[38] M. Yin, J. Lin, Y. Wang, et al., Development and validation of a multimodal model in predicting severe acute pancreatitis based on radiomics and deep learning, Int. J. Med. Inform. 184 (2024) 105341, https://doi.org/10.1016/j.ijmedinf.2024.105341.
[39] Abrantes J. External validation of a deep learning model for breast density classification 2023:1226 words. https://doi.org/10.26044/ECR2023/C-16014.
[40] P. Rauch, H. Stefanits, M. Aichholzer, et al., Deep learning-assisted radiomics facilitates multimodal prognostication for personalized treatment strategies in low-grade glioma, Sci. Rep. 13 (1) (2023) 9494, https://doi.org/10.1038/s41598-023-36298-8.
[41] Calisto F.M. Medical Imaging Multimodality Breast Cancer Diagnosis User Interface 2017. https://doi.org/10.13140_RG.2.2.15187.02084.