International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

MRI Breast Lesion Analysis Using Deep Learning Techniques

Dhaval M. Modi[1], Janhvi Chauhan[2]

[1,2]Lecturer, Electronics & Communication Department, Govt. Polytechnic, Ahmedabad

Abstract

This study aimed to evaluate the effectiveness of a deep learning model in distinguishing between benign and malignant breast lesions using magnetic resonance imaging (MRI), while also characterizing various histological subtypes of these lesions.A deep learning model was developed to simultaneously detect and characterize breast lesions. The model was trained on single 2D T1-weighted fat-suppressed post-contrast MR images selected by radiologists, acquired following the administration of a gadolinium-based contrast agent. The dataset consisted of 335 MR images from 335 patients, encompassing 17 histological subtypes categorized into four groups: mammary gland tissue, benign lesions, invasive ductal carcinoma, and other malignant lesions. Model performance was evaluated on an independent test set of 168 MR images using weighted area under the ROC curve (AUC) metrics.The model achieved a cross-validation average ROC-AUC of 0.817 across a three-shuffle, three-fold setup. On the independent test set, it achieved a weighted mean AUC of 0.8.The findings demonstrate that a supervised attention-based deep learning model can effectively analyze breast MRI for lesion detection and classification. Further validation on larger and independent datasets is recommended to confirm its clinical applicability.

Keywords: Magnetic resonance imaging (MRI); Breast lesion detection; Convolution neural networks; Transfer learning; Attention model

1. Introduction

As the volume of radiological examinations continues to rise, so does the complexity of their interpretation and the demands placed on healthcare providers [1]. Radiologists are increasingly susceptible to decision fatigue, which can lead to a greater incidence of diagnostic errors—such as missed, incorrect, or delayed diagnoses [2]. Furthermore, radiological interpretations are often subject to significant intra- and inter-observer variability [3,4].

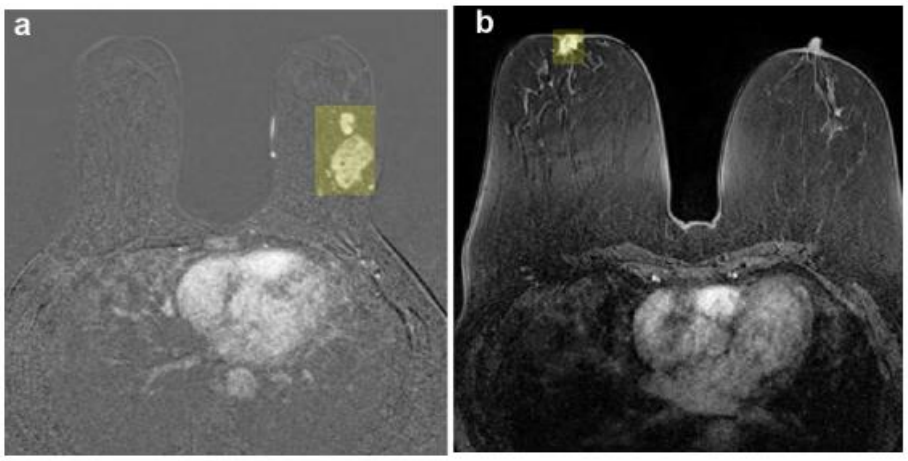

Figure 1 shows axial plane T1-weighted MR images taken following intravenous gadolinium chelate injection. (A) The MR image displaying invasive ductal cancer is overlayed with a binary mask. Heterogeneous enhancement is seen in the lesion. (B) One invasive ductal carcinoma is shown on the MR picture.

Breast magnetic resonance imaging (MRI) is now routinely used for a variety of clinical indications in the management of breast cancer. Key applications include screening high-risk individuals [5], determining the extent of disease, evaluating surgical margins, monitoring the response to neoadjuvant chemotherapy, and investigating metastatic axillary lymphadenopathy of unknown primary origin [6,7]. Breast MRI is inherently multiparametric, typically involving a combination of imaging sequences such as dynamic T1-weighted gradient-echo sequences (pre- and post-contrast with gadolinium-based agents), T2-weighted or short tau inversion recovery (STIR) sequences, and diffusion-weighted imaging (DWI) [8].

Deep learning, a subset of machine learning that employs layered artificial neural networks [9], has shown remarkable performance gains over conventional computer vision techniques [10]. In radiology, deep learning has the potential to enhance every stage of the imaging pipeline—from image reconstruction [11] and segmentation to final interpretation [12,13].

While most deep learning studies in breast imaging have focused on mammography [14,15], there is comparatively limited research on its application in breast MRI. Notably, a study comparing radiologist performance to radiomics and convolutional neural networks (CNNs) in characterizing breast lesions on MRI found that human interpretation achieved the highest AUC (0.98), outperforming CNNs (AUC = 0.88) and radiomics approaches (AUC = 0.81) [16].

Successful deployment of artificial intelligence in clinical practice requires close collaboration between radiologists and data scientists. In this study, we present an innovative deep learning tool designed to enhance the interpretation of breast MRI, potentially improving diagnostic safety and reliability. This research emerged from a collaborative effort initiated during the Journées Francophones de Radiologie data challenge held in Paris in October 2018.

The aim of the study was to evaluate the effectiveness of a deep learning model in distinguishing benign from malignant breast lesions using MRI, while also characterizing various histological subtypes of the lesions.

2. Materials and methods

2.1 Preprocessing

The dataset comprised anonymized two-dimensional T1-weighted MR breast images enhanced with gadolinium chelate, which were provided as part of the Journées Francophones de Radiologie 2018 challenge. Although the organizers had already applied a degree of standardization, the dataset remained highly heterogeneous in terms of scale (Fig. 1). To ensure uniformity across all samples, all images were resized to a fixed resolution of 240 × 345 pixels.

2.2 Automatic feature extraction

To extract features from the images, we employed a 50-layer residual neural network (ResNet-50) [17], pretrained on the ImageNet dataset. The final two layers of the network were removed to adapt it for our task. Since ResNet-50 is designed for color images, each grayscale input was replicated across three channels to mimic the red, green, and blue inputs. Given an input image of size 3 × 240 × 345, the network generated a feature map of dimensions 2048 × 8 × 11. As an initial approach, we computed the average across the spatial dimensions of this feature map, as described in Equation (1).

$$x_k = \frac{1}{8\ x\ 11} \sum_{ij=0}^{n} x_{kji} \qquad (1)$$

This method produced a 2048-dimensional feature vector for each image, which was then input into a fully connected layer with five output neurons corresponding to the classification categories: malignancy, normal tissue, other benign lesions, invasive ductal carcinoma (IDC), and other malignant lesions. However, a major limitation of this approach is its inability to distinguish between relevant and irrelevant regions, such as the thoracic area or background.

2.3 Supervised attention mechanism

One of the primary challenges in this task was the variability in both the appearance and size of breast lesions. To enhance the model's learning efficiency, we decomposed the classification process into two stages: (i) detection of abnormalities in MR images, and (ii) classification of the identified lesions.

Both steps were carried out simultaneously by two branches of a single deep learning model. For the detection phase, we generated additional localization labels consisting of bounding boxes around the lesions. These annotations were not highly detailed and were quickly drawn by a fifth-year radiology resident (P.H.) with limited experience in breast MRI (Fig. 2).

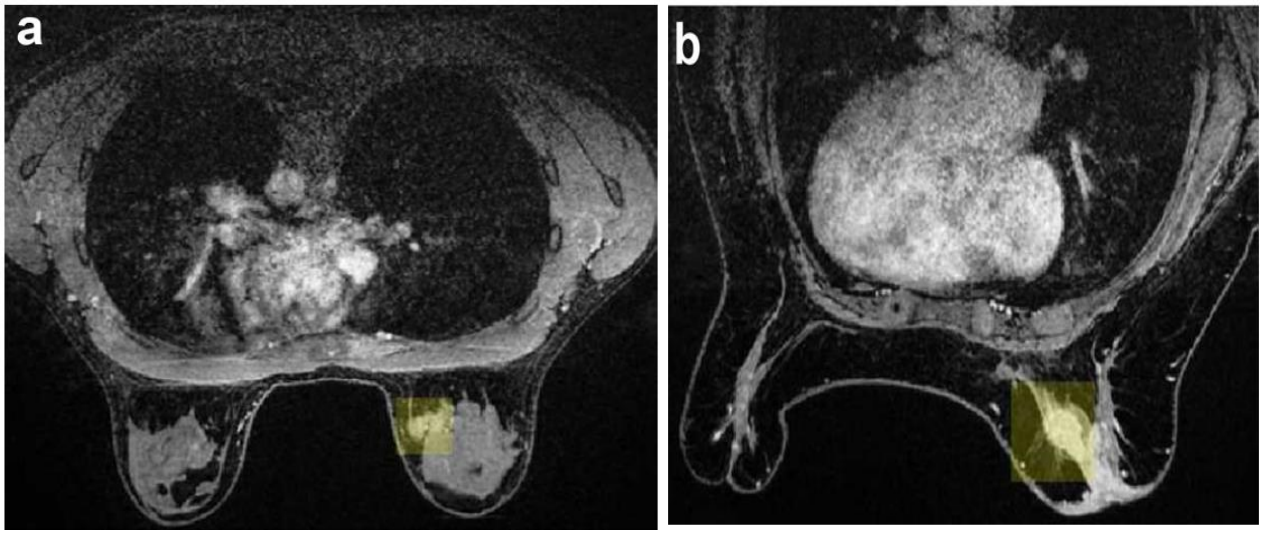

Figure 2: Axial T1-weighted MR images following intravenous gadolinium chelate injection. The MR images that display (a) a growing lesion and (b) an invasive ductal carcinoma are overlayed with annotation masks (in yellow). A radiologist created the annotations using a specialized tool that allows drawing bounding boxes.

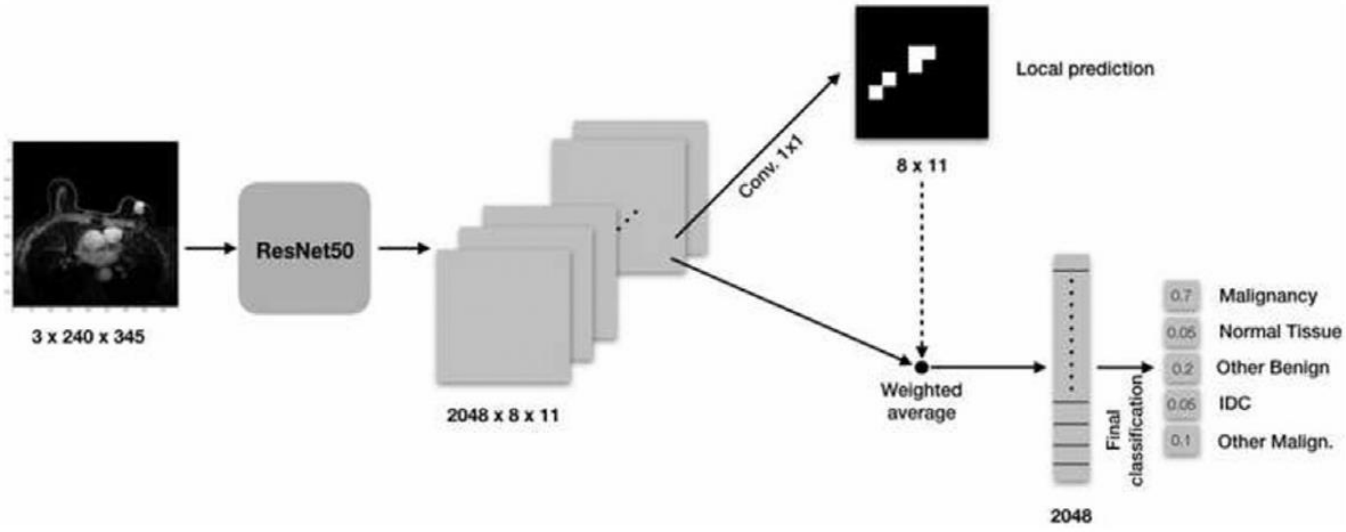

Figure 3 Model Architecture, each image (240 × 345 pixels) was processed using a ResNet-50 neural network, which produced 2048 feature maps of size 8 × 11. These maps were passed through two branches:

- **Upper branch (Attention Block):** This branch was trained to identify abnormalities within the image.
- **Lower branch:** It performed spatial averaging of feature maps over the regions identified by the attention block.

The resulting 2048-dimensional feature vector was then passed into a logistic regression layer that outputted a score between 0 and 1 for each lesion category. These scores represent the predicted probability of each lesion type being present in the image.

For every image, a binary mask of equal size (240 × 345) was created to indicate lesion presence or absence. This mask was down sampled to match the ResNet's output size (8 × 11).

Localization Module:

A 1 × 1 convolution was applied to the ResNet output (2048 × 8 × 11), producing a single-channel feature map of size 8 × 11. After applying a sigmoid function, this generated a prediction map with values between 0 and 1, approximating the binary mask from the annotations.

This localization output was then used to guide the primary classification module by calculating a weighted average over the final feature map. The weight at each spatial location (i, j) was determined by the local prediction value pijp_{ij}pij, as formalized in Eq. (2).

$$x_k = \frac{\sum_{i,j} \rho_{ij} x_{kij}}{\sum_{i,j} \rho_{ij}} \qquad (2)$$

When the localization module predicted a uniform probability distribution across the entire image, the resulting computation was equivalent to performing a simple spatial average, as in the basic model. In contrast, if the module identified a lesion with high confidence at a specific pixel, the final prediction was based solely on the feature vector extracted from that single location.

The final classification was carried out using a fully connected layer with five output neurons, each corresponding to a specific category: malignancy status, normal tissue, other benign lesions, invasive ductal carcinoma (IDC), and other malignant lesions. The complete architecture of the model is illustrated in Figure 3.

Furthermore, this attention mechanism allows interpretation of the model's predictions. To do so, we took the 8 × 11 attention map $\frac{\rho_{ij}}{\sum_{ij} \rho_{ij}}$ and resized it to the original image dimensions (i.e., 240 × 345). This map could be super imposed over the image to see the areas considered by the model to make its decision, as shown in Figure 4.

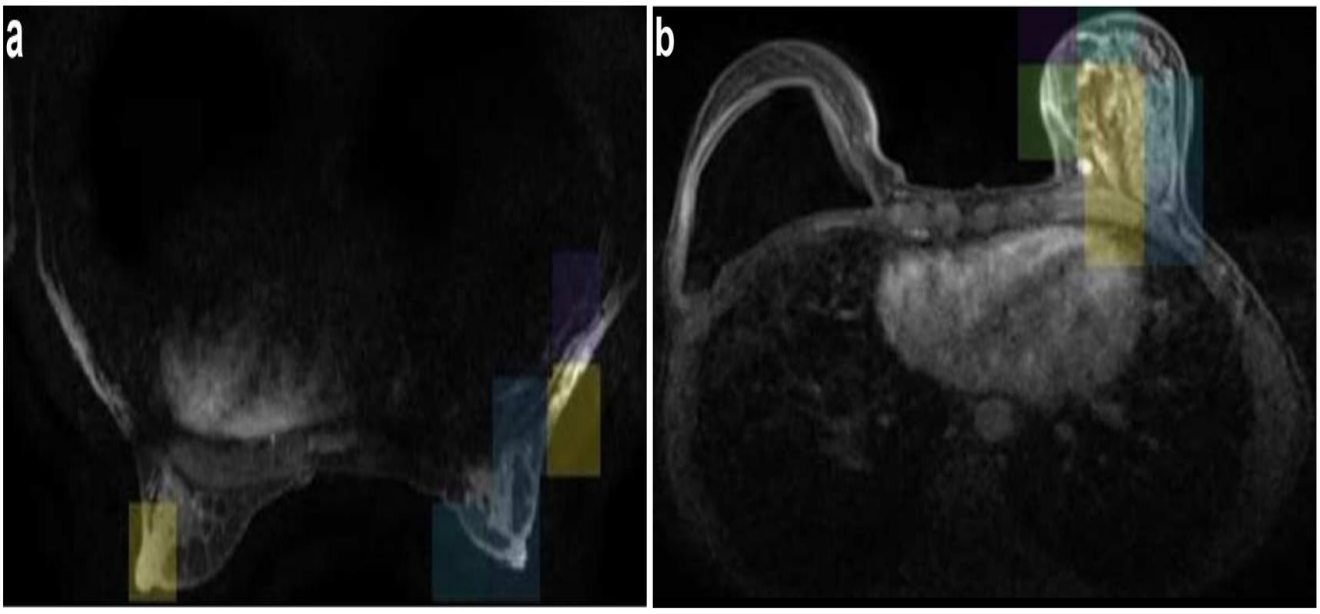

Figure 4 shows two instances of attention maps produced by the model for an invasive ductal carcinoma and glandular tissue, respectively. This demonstrates that the trained model can identify

normal tissue or lesions in a new image that it was not trained on without the need for human assistance.

2.4 Implementation

Our model was trained concurrently on the three tasks evaluated in the challenge: lesion detection, malignancy classification, and lesion subtype identification. This multitask learning approach helped reduce overfitting. However, each task progressed at a different learning rate. To address this, we saved three separate versions of the model weights, each selected based on optimal performance on a validation set.

Specifically, when the model achieved its highest AUC for lesion detection, we stored that version of the weights for use in that particular task. Training was carried out using stochastic gradient descent with Nesterov momentum.

Due to the limited size of the dataset, the model's results were highly variable. To ensure robustness, we employed three-fold cross-validation repeated across three different data splits. In total, we conducted nine training runs, each time randomly selecting 223 images from the 335-image training set for model training, and evaluating performance on the remaining 112 images using AUC scores. The final performance was assessed by averaging the results from these nine runs before applying the model to the independent test set provided by the challenge organizers.

3. Results

The number of each lesion type provided by the challenge organizers is detailed in Table 1. The number of breast lesions provided in the training set for each lesion category is presented in Table 2 and was used to determine the final score, following Eq. (3):

$$\text{Score} = 0.6 \times AUC_{\frac{benign}{malignant}} + 0.4 \times \frac{1}{4}\sum AUC_{lesion\ subgroup} \quad (3)$$

The average ROC-AUC scores obtained by our model across the repeated cross-validation process (i.e., three repetitions of three-fold cross-validation) for each lesion subgroup, along with the weighted aggregate score calculated using Equation 1 as per the challenge evaluation criteria, are presented in Table 3. The corresponding ROC curves are illustrated in Figure 5. On the independent challenge test set, the same model achieved a weighted AUC of 0.8.

Table 1 Number of breast lesions provided in the training dataset

Lesion Type	
Mammary gland	105 (30%)
Sclerosing adenosis	3 (0.9%)
Radial scar	2 (0.7%)
Fibroadenoma	24 (7.2%)
Galactophoritis	5 (1.3%)
Atypical hyperplasia	4 (1.3%)
PASH	1 (0.4%)
Papilloma	1 (0.4%)
Cyst	23 (6.9%)

Cytosteatonecrosis	13 (3.8%)
Intra-mammary lymph node	24 (7.3%)
Invasive ductal carcinoma	82 (24.6%)
Invasive lobular carcinoma	16 (4.7%)
Other proliferating lesion	8 (2.1%)
Triple negative cancer	18 (5.3%)
Intraductal carcinoma	5 (1.4%)
Mucinous carcinoma	2 (0.8%)
Total	335 (100%)
PASH indicates pseudoangiomatous stromal hyperplasia.	

Table 2 Number of breast lesions provided in the train-ing set for each lesion category.

Lesion Type	Training set	Test set
Mammary gland	104 (31.9%)	A
Other benign lesions	108 (33.9%)	A
Invasive ductal carcinoma	82 (24.8%)	A
Other malignant lesion	41 (13.6%)	A
Total	335	168

Table 3 Detailed AUC scores according to breast lesion type.

Lesion group	AUC
Malignancy	0.870 (0.024)
Mammary gland	0.721 (0.041)
Other benign lesions	0.658 (0.042)
Invasive ductal carcinoma	0.812 (0.032)
Other malignant lesions	0.786 (0.061)
Overall score (weighted sum)	0.817 (0.040)
Means are used to present data. Standard deviations are indicated by numbers in parenthesis. The mean of three shuffled three-fold cross-validations on the training set was used to calculate the AUC values. In nine experiments, our system was trained using two thirds of the training set's photos at random, and scores were calculated based on the outcomes of the final third. Standard deviations are given between brackets, and mean scores for those nine experiments are displayed.	

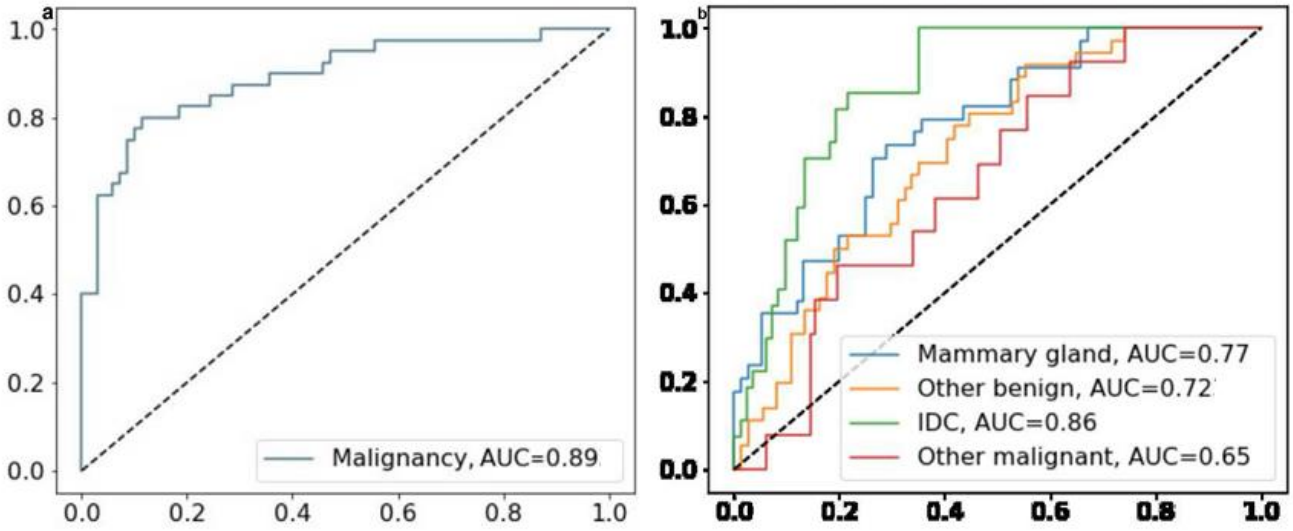

Figure: 5 ROC curves for (a) benign vs. malignant classification and (b) lesion classification using a single model are displayed

4. Discussion

Our approach achieved first place in the challenge, which is particularly encouraging given the limited dataset of only 335 training images. The incorporation of a supervised attention mechanism offered two major benefits. First, expert annotations significantly improved the interpretability of the model's outputs. The resulting heat maps allowed for better insights into the model's decision-making process, including instances of misclassification. Second, the bounding box annotations provided by a radiologist notably enhanced the model's performance.

While traditional research tools make dataset annotation a time-intensive process and are not well integrated into radiology workflows, one of the ongoing challenges in the era of deep learning in medical imaging is to develop efficient tools that produce high-performing and clinically relevant models based on annotations embedded in standard radiological practice [18]. To address this, we developed a tool that facilitated rapid image labeling—allowing the entire dataset to be annotated in under an hour—without compromising performance.

Since this challenge, convolutional neural networks (CNNs) have been applied beyond lesion characterization in breast MRI, including in tasks such as predicting breast cancer molecular subtypes [19] and assessing response to neoadjuvant therapy [20]. These developments confirm that, beyond the current enthusiasm, machine learning holds significant potential for transforming cancer management and clinical decision-making.

In clinical practice, breast lesions are assessed using the ACR BI-RADS classification on multiparametric MRI. A recent CNN-based study achieved strong performance across multiple sequences, with an AUC of 0.89 [16]. Although our study is not directly comparable—as it used a single slice from only one MRI sequence—our promising results motivate further investigation into applying our method to full three-dimensional sequences. This would allow direct comparison with the CNN approach in [16] and help determine whether our attention-based model yields improved performance on a new dataset.

In conclusion, validating deep learning models on independent datasets—especially using multiparametric and 3D MRI data rather than single 2D slices—is crucial for assessing generalizability. Further research is necessary to explore the clinical utility of such methods and to establish workflows that integrate lesion classification using BI-RADS. Access to larger datasets and multiparametric imaging is also likely to enhance model accuracy.

5. Human and animal rights

The authors confirm that the study was conducted in accordance with the World Medical Association's Declaration of Helsinki, revised in 2013, for research involving human subjects, and in compliance with the EU Directive 2010/63/EU for animal experimentation.

6. Informed consent and patient details

The authors state that this report contains no personal information that could identify any patient(s). Written informed consent was obtained from all patients and/or volunteers included in the study. The authors also confirm that all personal details of the patients and/or volunteers have been removed.

References

[1] Reiner BI, Krupinski E. The insidious problem of fatigue in medical imaging practice. J Digit Imaging 2012;25:3—6.
[2] Bruno MA, Walker EA, Abujudeh HH. Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction. Radiographics2015;35:1668—76.
[3] Muenzel D, Engels HP, Bruegel M, Kehl V, Rummeny EJ, Metz S. Intra- and inter-observer variability in measurement of target lesions: implication on response evaluation according to RECIST 1.1. Radiol Oncol 2012;46:8—18.
[4] Suzuki C, Torkzad MR, Jacobsson H, Astrom G, Sundin A, Hatschek T, et al. Interobserver and intraobserver variability in the response evaluation of cancer therapy according to RECIST and WHO-criteria. Acta Oncol 2010;49:509—14.
[5] 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 Radiol2018;15:408—14.
[6] Morris EA, Comstock CE, Lee CH. ACR BI-RADS®magneticreso nance imaging. ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System. Reston, VA. Am Coll Radiol 2013.
[7] Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS® fifth edition: a summary of changes. DiagnInterv Imaging 2017;98:179—90.
[8] Thakran S, Gupta PK, Kabra V, Saha I, Jain P, Gupta RK, et al. Characterization of breast lesion using T1-perfusion magnetic resonance imaging: qualitative vs. quantitative analysis. DiagnInterv Imaging 2018;99:633—42.
[9] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436—44.
[10] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. ImageNet large scale visual recognition challenge. Int J Comp Sci 2015;115:211—52.
[11] Zhu B, Liu JZ, Cauley SF, Rosen BR, Rosen MS. Image reconstruction by domain-transform manifold learning. Nature 2018;555:487—92.

[12] Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial intelligence in radiology. Nat Rev Cancer 2018;18: 500—10.
[13] Yasaka K, Akai H, Kunimatsu A, Kiryu S, Abe O. Deep learning with convolutional neural network in radiology. Jpn J Radiol2018;36:257—72.
[14] Becker AS, Marcon M, Ghafoor S, Wurnig MC, Frauenfelder T, Boss A. Deep learning in mammography: diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Invest Radiol2017;52:434—40.
[15] Kooi T, Litjens G, van Ginneken B, Gubern-Merida A, Sanchez CI, Mann R, et al. Large scale deep learning for computer aided detection of mammographic lesions. Med Image Anal 2017;35:303—12.
[16] Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus convolutional neural networks analysis for classification of contrast-enhancing lesions at multipara metric breast MRI. Radiology 2019;290:290—7.
[17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2016 [abs/1512.03385].
[18] Abajian AC, Levy M, Rubin DL. Informatics in radiology: improving clinical work flow through an AIM database: a sample web-based lesion tracking application. Radiographics2012;32:1543—52.
[19] Ha R, Mutasa S, Karcich J, Gupta N, Pascual Van Sant E, Nemer J, et al. Predicting breast cancer molecular sub type with MRI dataset utilizing convolutional neural network algorithm. J Digit Imaging 2019, http://dx.doi.org/10.1007/ s10278-019-00179-2.
[20] Ha R, Chin C, Karcich J, Liu MZ, Chang P, Mutasa S, et al. Prior to initiation of chemotherapy, can we predict breast tumor response? Deep learning convolutional neural networks approach using a breast MRI tumor Dataset. J Digit Imaging 2018, http://dx.doi.org/10.1007/s10278-018-0144-1