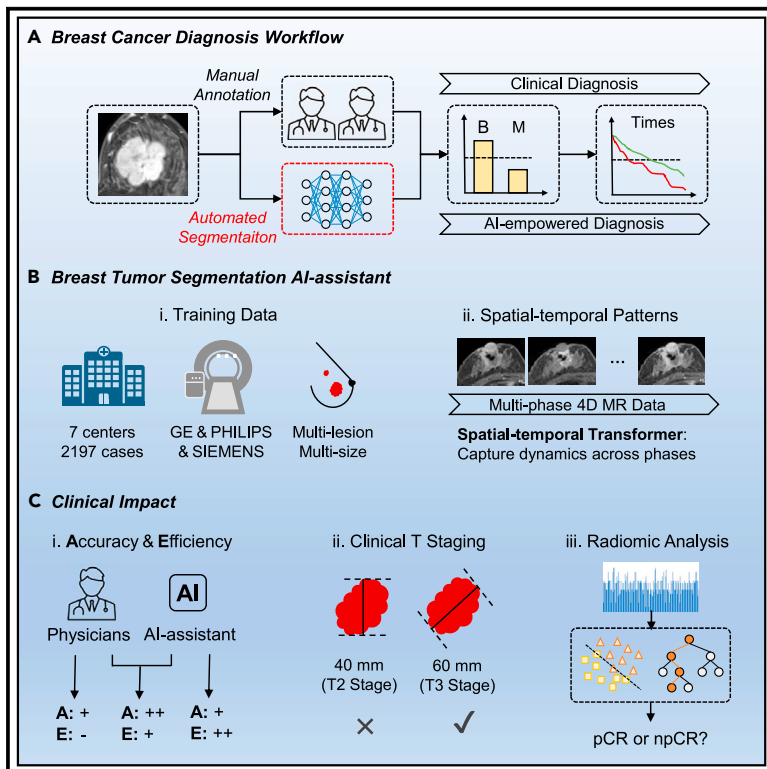# A robust and efficient AI assistant for breast tumor segmentation from DCE-MRI via a spatial-temporal framework

Graphical abstract

Authors

Jiadong Zhang, Zhiming Cui, Zhenwei Shi, ..., Zhenhui Li, Zaiyi Liu, Dinggang Shen

Correspondence

wangrongpin@126.com (R.W.),
junliu123@csu.edu.cn (J.L.),
andrewssmu@msn.com (J.Z.),
hangzhoudzx73@126.com (Z.D.),
sk12177@rjh.com.cn (K.S.),
lizhenhui@kmmu.edu.cn (Z.L.),
liuzaiyi@gdph.org.cn (Z.L.),
dgshen@shanghaitech.edu.cn (D.S.)

In brief

Clinical breast cancer diagnosis can be time-consuming and laborious. In this study, we have developed an AI assistant to automatically segment breast tumors from DCE-MRI data, trained on a very large set of data with a specially designed spatiotemporal transformer. Experimental results show that our AI assistant can produce more accurate results even than clinicians while also using significantly less time. This indicates the potential of our AI assistant in real clinical applications.

Highlights

- A robust and efficient AI assistant is developed for breast cancer segmentation

- A large set of DCE-MRI data from seven medical centers is used to train our AI assistant

- A specifically designed spatial-temporal transformer is used to capture dynamics

- Our AI assistant has potential for building an automated breast cancer diagnosis system

Patterns

Article

A robust and efficient AI assistant for breast tumor segmentation from DCE-MRI via a spatial-temporal framework

Jiadong Zhang,[1,15] Zhiming Cui,[1,15] Zhenwei Shi,[2] Yingjia Jiang,[3] Zhiliang Zhang,[4] Xiaoting Dai,[5] Zhenlu Yang,[6] Yuning Gu,[1] Lei Zhou,[7] Chu Han,[2] Xiaomei Huang,[8] Chenglu Ke,[2] Suyun Li,[2] Zeyan Xu,[2] Fei Gao,[1] Luping Zhou,[9] Rongpin Wang,[6,*] Jun Liu,[3,*] Jiayin Zhang,[5,*] Zhongxiang Ding,[10,*] Kun Sun,[11,*] Zhenhui Li,[12,*] Zaiyi Liu,[2,*] and Dinggang Shen[1,13,14,16,*]

[1]School of Biomedical Engineering, ShanghaiTech University, Shanghai 201210, China
[2]Department of Radiology, Guangdong Provincial People's Hospital, Guangdong 510080, China
[3]Department of Radiology, The Second Xiangya Hospital, Central South University, Hunan 410011, China
[4]School of Medical Imaging, Hangzhou Medical College, Zhejiang 310059, China
[5]Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, China
[6]Department of Radiology, Guizhou Provincial People's Hospital, Guizhou 550002, China
[7]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China
[8]Department of Medical Imaging, Nanfang Hospital, Southern Medical University, Guangzhou 510515, China
[9]School of Electrical and Information Engineering, The University of Sydney, Sydney, NSW 2006, Australia
[10]Department of Radiology, Key Laboratory of Clinical Cancer Pharmacology and Toxicology Research of Zhejiang Province, Hangzhou 310003, China
[11]Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China
[12]Department of Radiology, The Third Affiliated Hospital of Kunming Medical University, Kunming 650118, China
[13]Shanghai United Imaging Intelligence Co., Ltd., Shanghai 200230, China
[14]Shanghai Clinical Research and Trial Center, Shanghai 200052, China
[15]These authors contributed equally
[16]Lead contact
*Correspondence: wangrongpin@126.com (R.W.), junliu123@csu.edu.cn (J.L.), andrewssmu@msn.com (J.Z.), hangzhoudzx73@126.com (Z.D.), sk12177@rjh.com.cn (K.S.), lizhenhui@kmmu.edu.cn (Z.L.), liuzaiyi@gdph.org.cn (Z.L.), dgshen@shanghaitech.edu.cn (D.S.)
https://doi.org/10.1016/j.patter.2023.100826

THE BIGGER PICTURE Breast cancer is the most common cancer affecting women worldwide. Early detection and diagnosis are crucial for better treatment outcomes and improved survival rates. However, high incidence of breast cancer puts a significant burden on clinicians, while training experienced breast imaging clinicians is time-consuming. This raises the need for developing a robust AI assistant for assisting clinicians for breast cancer diagnosis. To this end, we collected a very large set of breast DCE-MRI data from seven medical centers for developing such a breast AI assistant, which can also be used as a foundational step for building an automated breast cancer diagnosis system to facilitate future smart medicine.

① ② ③ ④ ⑤ **Development/Pre-production:** Data science output has been rolled out/validated across multiple domains/problems

---

SUMMARY

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) allows screening, follow up, and diagnosis for breast tumor with high sensitivity. Accurate tumor segmentation from DCE-MRI can provide crucial information of tumor location and shape, which significantly influences the downstream clinical decisions. In this paper, we aim to develop an artificial intelligence (AI) assistant to automatically segment breast tumors by capturing dynamic changes in multi-phase DCE-MRI with a spatial-temporal framework. The main advantages of our AI assistant include (1) robustness, i.e., our model can handle MR data with different phase numbers and imaging intervals, as demonstrated on a large-scale dataset from seven medical centers, and (2) efficiency, i.e., our AI assistant significantly reduces the time required for manual annotation by a factor of 20, while maintaining accuracy comparable to that of physicians. More importantly, as the fundamental step to build an AI-assisted breast cancer diagnosis system, our AI assistant will promote the application of AI in more clinical diagnostic practices regarding breast cancer.

Table 1. Description of DCE-MRI from each center

COHORT	GD-hospital	YN-hospital	HZ-hospital	SH-hospital	XY-hospital	GZ-hospital	RJ-hospital
Participant information							
Patient number	638	473	273	182	171	48	405
DCE-MRI case number	642	475	274	182	171	48	405
DCE-MRI volumes	3,852	3,688	1,644	1,121	1,074	328	1,460
Examination time	2016–2020	2012–2021	2020–2022	2019–2022	2015–2020	2020–2021	2016–2018
Age (years)	50 (21, 79)	47 (24, 77)	48 (22, 85)	53 (24, 87)	51 (29, 79)	45 (27, 86)	–
Imaging protocols							
Phase interval time	1 min	1–2 min	1 min	1 min	1 min	1 min	45 s
Phase number	6	6, 8	6	6, 7, 8	6, 8	6, 8	2, 4, 5, 6, 8
Inter-slice resolution (mm)	0.44–0.98	0.33–0.70	0.75–1.17	0.66–0.93	0.50–1.10	0.63–1.43	0.70–1.00
Slice thickness (mm)	0.5–1.0	0.8–1.7	1.2	1.6	1.0–1.5	0.9–1.2	1.5
Manufacturer	Philips	Siemens, GE	Siemens	Siemens, GE	Siemens	Siemens	Siemens
Breast region characteristics							
Breast characteristic (normal/ single/ with implant)	640/2/0	472/2/1	269/4/1	178/3/1	166/3/2	47/1/0	399/6/0
Tumor size (S/M/L)	86/369/187	144/235/96	88/87/99	87/56/39	59/49/63	712/20/16	97/206/102
BIRADS category [2] (3/4/5/6)	3/185/89/112	1/23/446/5	75/105/13/81	13/68/97/4	12/36/84/39	0/3/38/7	–
Molecular subtype [2] (luminal A/ luminal B/Her2+/TN)	82/219/36/34	65/304/62/44	–	17/50/18/13	–	9/25/5/3	–

Data description includes participant information, imaging protocols, and breast region characteristic. Dash means unavailable information. Not all the patients have accurate BIRADS category or molecular subtype.

centers as the external set in our experiments. Then we randomly choose 70% of the data from the internal set to train all the models and the rest of the data (30%) as the internal testing set to evaluate the models' performance. All cases from the external set are further used to demonstrate the robustness and generalizability of our AI assistant. The detailed data partition is shown in Figure 1B.

In these experiments, three typical segmentation metrics are used to evaluate segmentation performance, including DICE similarity coefficient (DSC, 0%–100%), Sensitivity (SEN, 0%–100%), and Hausdorff distance (HD, mm). DSC measures the similarity between ground truth and the prediction mask. SEN evaluates the under-segmentation ratio. HD calculates the distance between two point sets from ground truth and predicted surfaces. Higher DSC and SEN scores indicate better segmentation performance, while a lower HD score indicates closer boundaries with better results.

Whole-breast segmentation performance

In clinical practice, physicians only focus on the breast region to localize breast tumors. According to the physician's diagnosis process, we design a whole-breast segmentation model as the first step in our breast AI assistant to facilitate the downstream tumor segmentation and diagnosis in 3-fold. First, the whole-breast mask can guide accurate patch sampling in the training stage of the tumor segmentation model. Second, it helps our AI assistant to automatically distinguish normal and abnormal breasts. Third, some over-segmentation cases in the non-breast

regions, e.g., enhanced tissues in the heart region, can be removed effectively. Therefore, whole-breast segmentation is a crucial first step to significantly improve tumor segmentation performance.

Compared with tumor segmentation, whole-breast segmentation is a simple task due to the clear intensity contrast. In this paper, we adopt the U-Net structure for whole-breast segmentation, considering its great segmentation accuracy and generalization. As expected, the overall breast segmentation performance is excellent, i.e., achieving a DSC of 92.6% on the internal testing set as well as the external sets. Representative segmentation results are given in Figure 2. For each center, we show three typical cases, including standard-sized breasts, small-sized breasts, and abnormal breast(s) (including only one breast due to surgical removal of the other breast or breast implant surgery). From Figure 2, it is evident that the quality of whole-breast segmentation is robust for different breasts from different centers. A detailed analysis about the benefits of whole-breast segmentation will be given in the ablation study section.

Breast tumor segmentation performance

Given the DCE-MRI with multiple phases, we use both the original images and the subtraction images (i.e., between the corresponding images with and without contrast agent injection) as the input to capture dynamic spatial-temporal information via a spatial-temporal transformer. Equipped with a large amount of training data and the excellent performance of transformers,[39]

A

B
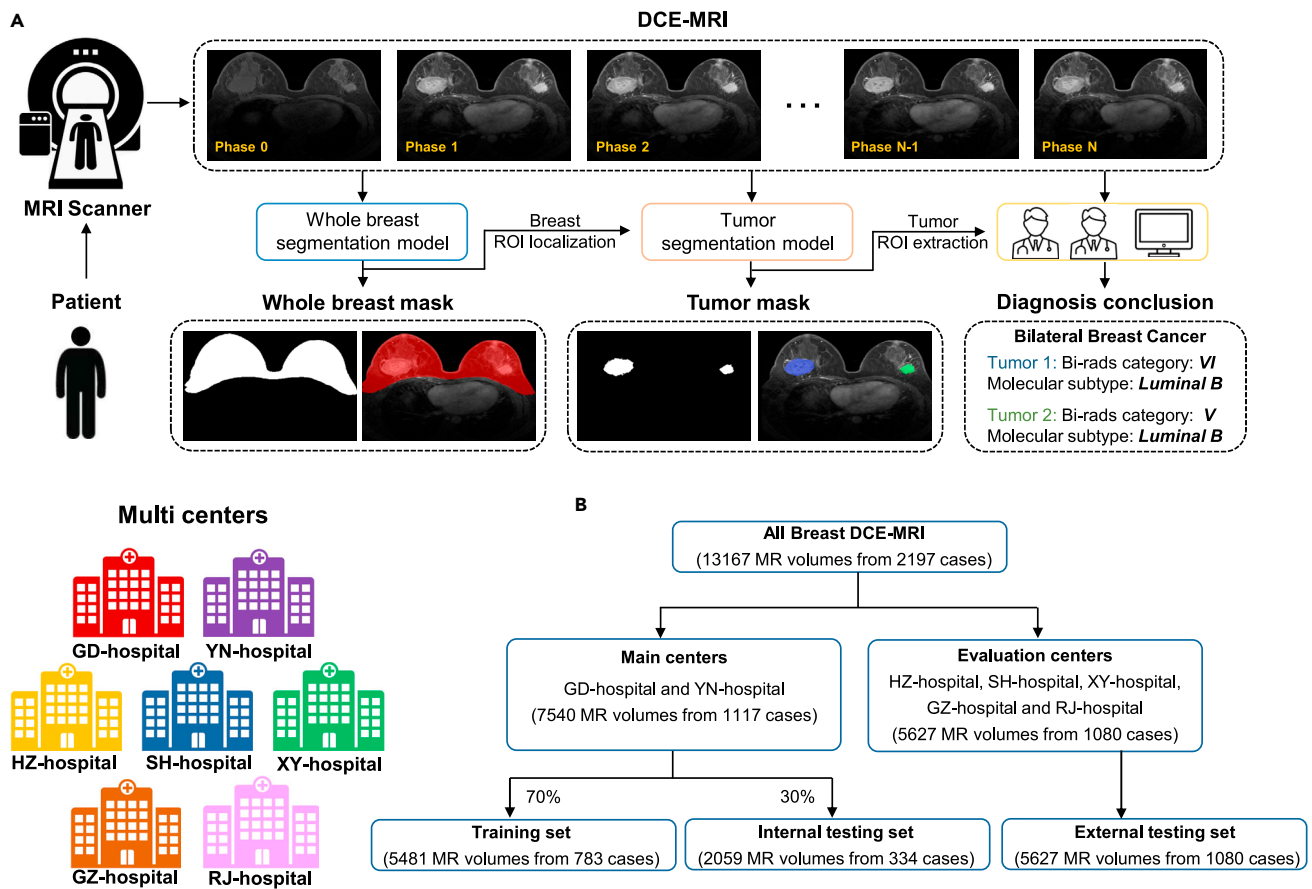
Figure 1. Overview of breast cancer diagnosis process and data information
(A) Breast cancer diagnosis process. Multi-phase DCE-MRI data are collected from MRI scanner and then used as inputs to our segmentation models for delineating the whole breast and tumor, which can be further used to guide diagnosis by physician or AI system.
(B) Data partition in experiments. Seven centers are included, and we randomly divide them into the training set, internal testing set, and external testing set.

the proposed model could be better generalized to achieve robust segmentation of images acquired by different imaging protocols and carrying varied tumor categories.

We have evaluated our tumor segmentation model on multi-center testing sets quantitatively and qualitatively. Note these multi-center data are acquired with quite different imaging settings (e.g., with different phase numbers and intervals, voxel sizes, and manufacturers). Quantitative results are provided in Figure 3, and segmentation results are shown in Figure 4, indicating good agreement between the automated segmentations by our model and the manual annotations by radiologists. Particularly, Figure 3A shows the pie chart describing data distribution among different centers (top) along the two bar charts (middle and bottom) describing the DSC and HD for each center. It can be seen that the best performance is achieved on YN-hospital (the internal testing set), with 82.7% and 3.1 mm for DSC and HD, respectively. In contrast, the segmentation results from XY-hospital (external set) are the worst with a DSC of 68.3% and HD of 12.1 mm, respectively. In addition, the segmentation results of all evaluation sets are summarized in Table 2, i.e., with a DSC of 77.7% on the internal testing set and 70.5% on the external testing set and 72.4% on all testing sets. In this table, the performance of our method is also compared with five

other methods, while ours achieves the best performance in terms of all metrics. It is noteworthy that although segmentation accuracy decreases on the external set, the DSC of 70.5% is still clinically acceptable for downstream diagnosis and treatment, as confirmed by clinicians.

We also analyze the robustness of segmentation performance with respect to different tumor sizes. As shown in Figure 3B, we use 1 mm^3 as the interval to divide tumors into 11 categories according to the sizes of the manually labeled tumors by radiologists. The segmentation performance of our AI system is robust for the medium-to-large tumors with 3–10 mm^3, as well as the large tumors above 10 mm^3, achieving an average DSC of 74.1% and HD of 4.4 mm. For the small-sized tumors, especially those extremely tiny ones (i.e., below 1 mm^3), the overall DSC decreases slightly to 63.9%, and the overall HD increases to 9.20 mm. Note that a tiny segmentation error in small-sized tumors, even with under- or over-segmentation of several voxels, can significantly affect the evaluation metrics. Overall, reasonable segmentation results can be achieved for tumors of various sizes.

We further evaluate the robustness of our AI assistant in handling DCE-MRI data with varied phase numbers. As described above, our AI assistant exploits temporal information
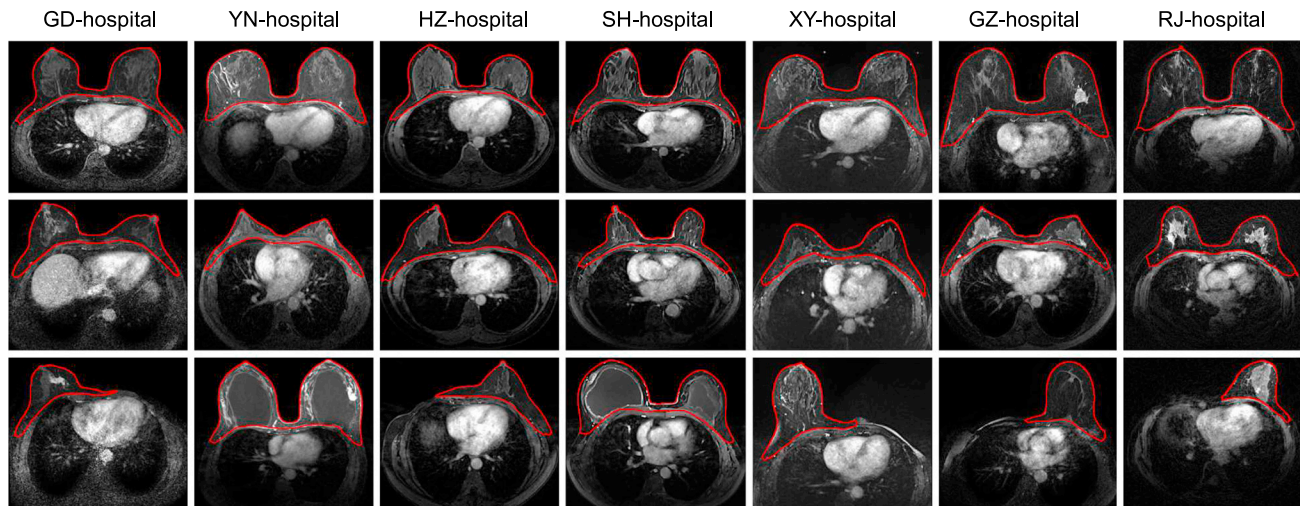
Figure 2. Visualization of whole-breast segmentation results from seven centers
Each center (column) is with three typical cases such as standard-sized breasts (top), small-sized breasts (middle), and abnormal breast(s) (bottom).

across different phases for improved segmentation performance. It is noted that not all cases have six phases available in our dataset (e.g., some cases in the RJ-hospital dataset), due to factors such as image corruption caused by the motions, although these missing-phase cases occur rarely. Fortunately, benefiting from the attention mechanism of the transformer, our framework can flexibly deal with an arbitrary number of image phases by ignoring the attention weights of missing image phases. We summarize the segmentation performance on patients with different phase numbers from RJ-hospital in Figure 3C. Consistent with our expectation, the segmentation performance gets better when more phases are involved. For example, when the phase number increases from two to six, the DSC accuracy is improved from 65.9% to 72.9%, and the HD error decreases from 6.38 to 2.61 mm. These results demonstrate the advantages of our segmentation model in that it (1) is robust to different numbers of image phases, (2) can efficiently exploit temporal relationships across different image phases to improve segmentation performance, and (3) will benefit from more image phases.

To demonstrate the superiority of our model, we also compare our model with other well-established deep-learning-based breast tumor segmentation methods on both the internal testing set and the external testing set, including U-Net,[40] MHL,[34] Tumor-sen,[36] ALMN,[41] and TSGAN.[38] Note that the U-Net, MHL, Tumor-sen, and ALMN are 3D tumor segmentation networks that only consider one or two enhancement image(s), ignoring temporal information across different phases. Although TSCAN has attempted to exploit temporal information by designing a graph attention module, the signal dynamics across different phases are still not well explored. Furthermore, it only takes 2D slices as the inputs, so the segmentation consistency across different slices cannot be guaranteed. The quantitative metrics of segmentation results are summarized in Table 2 for each method. It can be seen that our AI assistant achieves the best performance, significantly outperforming other methods, with p values (by pair t tests) lower than 0.05 for all metrics. The performance gain is mainly from our specifically designed spatial-tem-

poral framework to capture dynamic changes across different phases. Furthermore, by leveraging the whole-breast segmentation network, many artifacts (e.g., over-segmentation on non-breast regions) can be avoided, thus contributing to lower HD errors. Overall, our model is robust and generalizable on both internal and external testing sets.

Ablation study

To segment breast tumors accurately, our AI assistant system contains several key modules, including the whole-breast segmentation model, the spatial transformer, and the temporal transformer. In order to evaluate the contribution of each module, we design an ablation study on all testing sets (Table 3). When adding the whole-breast segmentation model, a significant improvement in DSC and HD metrics is achieved by removing false-positive predictions, especially in non-breast regions. Hence, it is crucial to first localize tumor regions automatically before tumor segmentation. Besides, considering that CNN kernel is a local operation and cannot model long-range relations spatially, we design a spatial transformer in bottleneck to explore long-range spatial dependency within each phase. Moreover, a temporal transformer is also proposed to capture dynamic changes across different phases. Hence, the best performance is achieved by integrating these two transformers to comprehensively exploit spatial-temporal information, leading to 5.1% and 2.7% improvement for DSC and SEN metrics, respectively.

Clinical impact

(1) Tumor segmentation compared with experienced radiologists. The proposed AI assistant for breast tumor segmentation aims to relieve physicians from tedious and cumbersome annotation processes. To demonstrate the segmentation efficiency of our AI assistant, we randomly choose 100 cases from GD-hospital and invite two radiologists (Expert-1 and Expert-2), who do *not* participate in the ground truth labeling process, to annotate breast and tumor regions from scratch and to help refine

Figure 3. Tumor segmentation performance on the testing sets
(A) Segmentation performance on seven different centers.
(B) Segmentation performance on different sized tumors.
(C) Segmentation performance with respect to different phase number. The pie charts in the first row show the data distribution, and the bar charts in the second and third rows show DSC (%) and HD (mm), respectively. The error bars indicate the 95% confidence interval.

automatic segmentations by our model, respectively. The segmentation accuracy and efficiency are summarized in Table 4. From this table, first, we can see that our proposed AI assistant can achieve comparable segmentation accuracy to experienced radiologists, but it is more than 20 times faster. This gives radiologists more time to focus on downstream diagnosis and treatment plans. On the other hand, we find significant annotation differences across different radiologists at different times, while our AI assistant can generate more robust segmentations. Meanwhile, manual refinement of automatic annotations from our AI assistant can produce more accurate results (e.g., 5% of DSC improvement from what is achieved by the radiologists or the AI assistant alone) and also significant efficiency improvement over manual annotations from scratch (i.e., seven to 10 times faster). This indicates the advantages of our AI assistant in helping radiologists with fast tumor annotation.

(2) Necessity of accurate tumor segmentation for better diagnosis. In clinical practice, physicians usually analyze tumor shapes and positions based on MR images and segmentation masks. For example, the T-stage is an important pathological index to indicate malignancy and

prognosis of breast tumors, which is derived from tumor size and position information. Specifically, T-stage has four categories, including (1) T1, i.e., the widest part of the tumor is less than 20 mm; (2) T2, i.e., the widest part of the tumor is more than 20 mm but less than 50 mm; (3) T3, i.e., the widest part of the tumor is more than 50 mm; and (4) T4, i.e., the tumor has spread to the chest wall or skin. During diagnosis, physicians often only check three cross-sectional views to select the widest axis to classify the T-stage. However, this process is inaccurate, since the widest axis does not correspond to the widest part of the actual tumor located in the 3D space. To demonstrate the clinical usefulness of our tumor segmentation mask, we compute the accuracy of T-stage classification on the 100 cases from the radiologists' classification (1) without segmentation masks, (2) based on the manually annotated segmentation masks, and (3) based on the segmentation masks produced by our AI assistant. Without segmentation masks, the classification accuracies are 81% and 78% by the two radiologists, respectively, which is not unsatisfactory for precise diagnosis. But the accuracy is significantly improved when the tumor masks are involved in T-stage classification. For example,
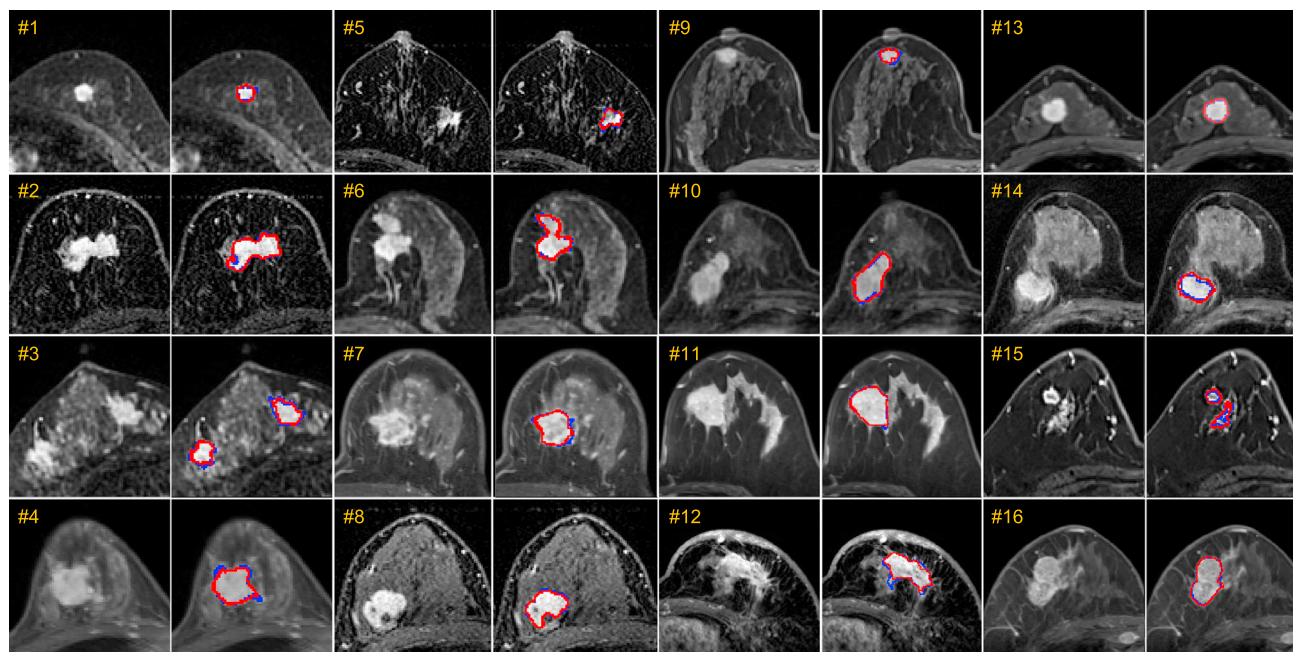
Figure 4. Typical examples of breast tumor segmentation on the testing set
Cases 1–8 are from the internal testing set, and cases 9–16 are from the external testing set. For each testing case, we show two images with and without annotations for better comparison. Blue delineation denotes ground truth annotations, while red delineation indicates automated segmentations by our model.

using manually annotated masks, the classification accuracy can be 94% and 93% from the two radiologists, with a comparable accuracy of 94% using the automatic annotations by our AI assistant, indicating the effectiveness of our AI assistant for clinical diagnosis.

(3) AI-based breast cancer diagnosis system. One important research topic in breast cancer is to predict the response after breast cancer neoadjuvant therapy using MRI-based radiomics features, which greatly relies on the tumor annotation from professional radiologists, due to lacking robust segmentation algorithms in the past. To demonstrate the clinical impact of our AI system in reducing the workload of radiologists, we extract radiomics features from manually annotated tumors by experienced radiologists and also automatically annotated tumors by our AI assistant, respectively, to train a classifier to predict neoadjuvant therapy response after standard radiotherapy process.[24] Specifically, we randomly choose 100 cases from the testing set in YN-hospital, with 34 cases of pathologic complete response (pCR) and 66 cases of non-pCR (npCR). The feature extraction model (https://pyradiomics.readthedocs.io) extracts radiomics features (e.g., shape, histogram, and texture) from DCE-MRI and segmentation masks.[42] The classifier (i.e., SVM) predicts whether the case is pCR or npCR. From the prediction results with 5-fold cross-validation, we achieve an accuracy of 66.8% with AI annotations, comparable to that of 67.1% with manual annotations. Similar results can be found using the receiver operating characteristic curve (AUC), with 0.79 (AI) versus 0.80 (human), respectively. Hence, we believe our proposed AI

assistant can be easily applied to facilitating downstream diagnosis tasks (e.g., molecular subtype prediction[43–45] and Ki-67 expression level prediction[46,47]) to promote the building of an automated all-purpose AI-based breast cancer diagnosis system.

DISCUSSION

Breast tumor segmentation allows comprehensively characterizing tumor properties (e.g., shape, size, single lesion or multiple lesions) that are crucial for accurate diagnosis and treatment plan. Currently, due to the lack of robust and automatic tumor segmentation tools, most AI-assisted breast cancer diagnosis models (e.g., for predicting breast tumor molecular subtype and response to breast cancer neoadjuvant therapy) still rely heavily on manual intervention of experienced radiologists,[25–28] which is time-consuming and tedious.

Though several works have exploited AI-assisted breast tumor segmentation on DCE-MRI data, their clinical availability and robustness are still unproven in 3-fold aspects. First, none of these works release their data, codes, or well-trained segmentation models, hindering their direct application to clinical practice. Second, the segmentation performance of the existing methods is only evaluated on their respective small internal dataset from a single center (e.g., 64 cases for TSGAN, 272 cases for MHL, and 422 cases for ALMN and Tumor-sen). Their robustness and generalizability on a multi-center dataset are unclear. For example, they may encounter over-fitting problems caused by training only on a single-site dataset. Third, most methods only pay attention to the contrast-enhanced tissues in two image phases, ignoring dynamic
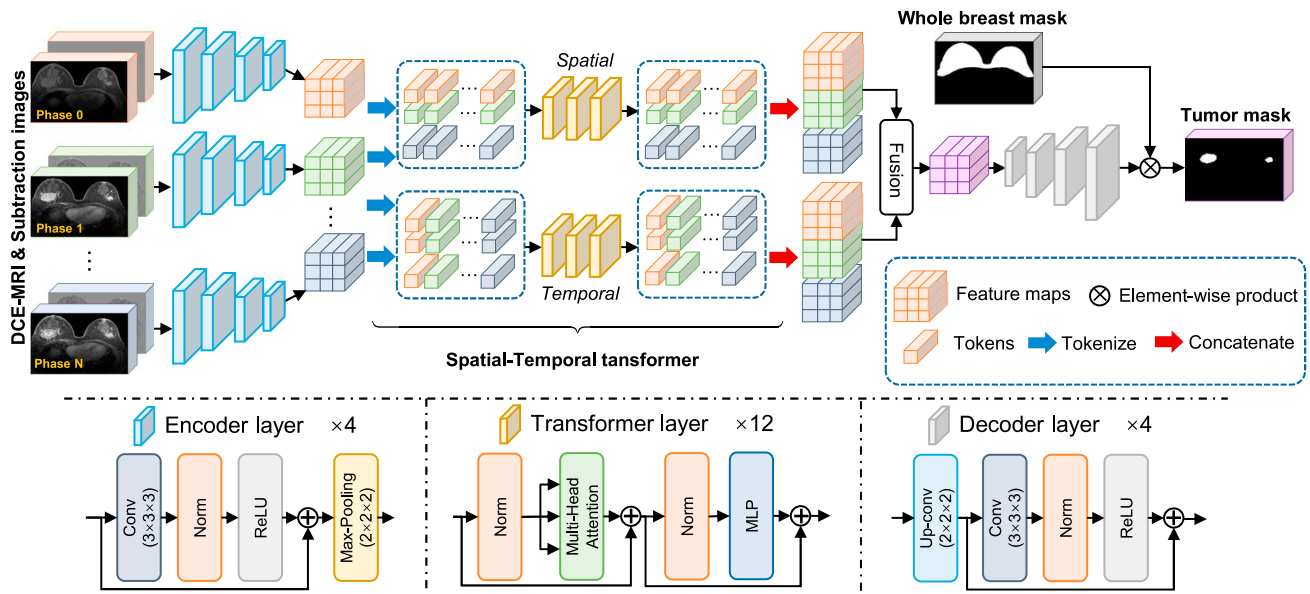
Figure 5. Overview of proposed tumor segmentation models
It tasks both original images and subtraction images from all phases as inputs. The weights from encoders in breast tumor segmentation model are shared. The detailed structures of encoder layer, transformer layer, and decoder layer are also illustrated in the figure.

changes across different phases, although these dynamic changes are crucial to detect some small tumors and distinguish them from other enhanced tissues (e.g., vessels).

Based on these observations, we propose a robust AI assistant for breast tumor segmentation by exploiting signal dynamics across different phases of DCE-MRI data. Specifically, our spatial-temporal framework can well capture both inter-phase and intra-phase long-range dynamic information. Moreover, we also design the whole-breast segmentation model to remove over-segmentations in non-breast regions to boost segmentation performance. Another essential contribution of this work is that we evaluate our AI assistant for breast tumor segmentation on a very large dataset (13,167 MR scans from 2,197 cases) collected from seven medical centers. We would like to emphasize that the data used in this study is pretty large, with various data distributions compared to most existing works. Although the authors in the paper[48] used a large-scale dataset, they collected data only from a single center without external evaluation. Hence, the advantage of multi-center evaluation in our paper makes our study more reliable. As shown in Table 2 and Figure 3, we have demonstrated impressive robustness of our method in addressing varied imaging protocols (especially different phase numbers) and tumor sizes. For all experiments conducted on this large dataset, our AI assistant achieves the best performance among all competitive segmentation methods. Most importantly, the generalizability of our AI assistant can be validated by its stable performance on the external testing dataset. In addition, we also achieve better or similar results (i.e., with DSC of 77.7%) on the internal testing set, when comparing with the reported performance of the competitive methods, such as 51.6% by random forest,[49] 63.5% by TSGAN,[38] 68.8% by SegNet,[50] 71.7% by MHL,[34] 75.88% by ST3D-net,[51] 77.0% by GOCS-DLP,[35] 78.0% by ALMN,[41] and 78.7% by Tumor-sen.[36] The code and the trained models of

our AI assistant will be released to promote breast cancer-related research or non-commercial clinical applications. Users can directly test their own data with our trained models or train the model from scratch.

Besides, our proposed tumor segmentation model is also robust to variations in DCE-MRI imaging intervals, i.e., 1 min for most centers and 45 s for RJ-hospital, by capturing dynamic signal changes across different phases with temporal information. In addition, we also demonstrate the model's robustness when some image phases are unavailable or missing (see Figure 3C for details). The specially designed temporal transformer can deal with an arbitrary number of inputs by ignoring the attention weights of missing phases, benefiting from the flexibility of the attention mechanism originated from natural language processing and widely adopted in computer vision fields.[52] It is noteworthy that motion during breast DCE-MRI imaging is often relatively small. Thus, registering all phase images brings little accuracy gain but takes much longer computational time, which is thus not adopted in our framework. This somewhat also indicates the robustness of our model to motions during the scanning process.

Finally, we would like to emphasize again that our proposed robust and efficient AI assistant is also beneficial for clinical applications. First, our accurate tumor segmentation mask can help physicians obtain more morphological information about tumors, such as tumor size, location, and malignancy, without tedious manual annotations. Second, automated segmentation is the primary step to build a smart AI diagnosis system for breast cancer. Lacking open-source breast tumor segmentation software makes existing breast tumor diagnosis research rely on manual annotations from radiologists. Our framework aims to relieve radiologists from this manual annotation step and allow their efforts to be devoted to subsequent diagnostic tasks. For example, predicting the response after breast cancer neoadjuvant therapy based on radiomics features extracted from our

Table 2. Comparison of different methods in breast tumor segmentation on internal and the external testing sets

	Internal testing set			External testing set			All testing sets		
Method	DSC	SEN	HD	DSC	SEN	HD	DSC	SEN	HD
U-Net[40]	72.4 ± 2.6	70.3 ± 2.7	6.1 ± 1.6	59.0 ± 1.7	75.1 ± 1.7	17.2 ± 1.4	62.5 ± 1.4	73.8 ± 1.1	14.3 ± 1.2
MHL[34]	66.5 ± 2.8	62.2 ± 3.0	8.4 ± 2.2	66.5 ± 1.6	67.5 ± 1.9	8.8 ± 1.3	66.1 ± 1.4	66.1 ± 1.6	8.7 ± 1.1
Tumor-sen[36]	70.9 ± 2.7	67.0 ± 2.8	4.8 ± 1.4	59.8 ± 1.7	68.9 ± 1.8	16.0 ± 1.5	62.7 ± 1.4	68.4 ± 1.6	13.1 ± 1.2
ALMN[41]	74.6 ± 2.4	72.5 ± 3.0	6.7 ± 1.2	56.4 ± 1.6	75.0 ± 1.6	19.2 ± 1.3	61.1 ± 1.4	73.7 ± 1.3	15.9 ± 1.1
TSGAN[38]	68.5 ± 1.6	69.4 ± 2.6	8.5 ± 1.5	63.3 ± 1.9	68.4 ± 1.6	19.8 ± 1.0	62.7 ± 1.9	68.7 ± 1.4	12.4 ± 1.2
Ours	77.7 ± 2.4	75.1 ± 2.5	3.7 ± 1.1	70.5 ± 1.3	76.5 ± 1.5	7.6 ± 0.9	72.4 ± 1.1	76.2 ± 1.3	6.6 ± 0.7

The evaluation metrics include DSC (%) ↑, SEN (%) ↑, and HD (mm) ↓.

tumor segmentation masks can achieve comparable performance with that using manual annotations (see more details in the clinical impact section). And we believe this robust segmentation model can be further applied to other tasks, such as predicting molecular subtypes of breast tumors.

In the section evaluating clinical impact, we compared tumor segmentation accuracy with experienced radiologists. We also evaluated two related works[34,48] with reportedly similar performance as radiologists. It is worth mentioning the difference between the segmentation evaluation methods used in our work and the two mentioned works. Specifically, two mentioned works re-used annotations from each individual radiologist as partial ground truth to train the network. While, in our work, we invited two new radiologists (who did *not* participate in the annotation process) to label images from scratch; thus our evaluation is more strict.

In conclusion, the proposed AI assistant allows robust and efficient breast tumor segmentation. We believe this model can promote more research and serve as a crucial part of the AI-assisted breast cancer diagnosis system.

Limitations of study

In this study, our AI assistant was trained on data from Asian populations, which typically have denser breasts with more fibroglandular tissue. In the future, we plan to add more diverse data from other populations to evaluate segmentation performance. Additionally, this study focused solely on segmentation from DCE-MRI, neglecting the important role of other MRI sequences (such as T2 weighted images and diffusion-weighted images) that provide tumor region characteristics from other perspectives. To further improve accuracy, a multi-parametric MRI tumor segmentation system is needed, and we are currently working on its design.

METHODOLOGY

Data pre-processing

This study was approved by the Research Ethics Committee from all centers. A detailed description of the dataset has been provided in Table 1. Considering large variations of spatial resolutions across different centers and also intensity distributions across different manufacturers, we, respectively, apply spatial normalization and intensity normalization in the data pre-processing step. Specifically, we interpolate all data to the same resolution with a voxel size of $1.0 \times 1.0 \times 1.0$ mm^3. For intensity

normalization, we clip the outliers (top 1% of the maximum values) of all images at different phases and perform min-max normalization on phase 2, which typically exhibit the strongest enhancement. The maximum and minimum values from the phase 2 images are then used to normalize the images at other phases, which makes all intensity values range from 0 to 1 while preserving the intensity change information across time sequences. Limited by the GPU memory, we extract many overlapped 3D patches to train the segmentation model. For the whole-breast segmentation model, we randomly crop patches with the size of $128 \times 128 \times 48$. For the breast tumor segmentation model, we crop patches with the size of $96 \times 96 \times 48$. Besides, since the number of background voxels significantly exceeds the foreground voxels, we require each cropped patch to contain foreground areas according to the training manual tumor annotations for efficient training of the segmentation model.

Whole-breast segmentation model

Compared with tumor segmentation, whole-breast segmentation is straightforward. We adopt the U-Net structure for breast segmentation. Patches of pre-contrast DCE-MRI data are fed into the U-Net to get the final breast mask. The detailed structure is the same as the work[40] with four downsampling layers and upsampling layers, and the skip connections of corresponding layers can exploit latent semantic information with more details. We adopt the DICE loss function[53] to supervise the model training.

Tumor segmentation model

In this paper, we propose to combine both the spatial information within each phase and the temporal information across different phases in DCE-MRI data. An overview of our network is shown in Figure 5. Briefly, we first use a convolution-based encoder **E** to get global spatial features in each phase, and we further design a spatial-temporal transformer to combine the features across different phases to capture the spatial-temporal information. The combined spatial-temporal features are then sent through a convolution-based decoder **D** to predict the tumor segmentation mask.

Specifically, we use $\{x^i\}_{i=0}^{N} \in \mathbb{R}^{(N+1) \times L \times W \times H}$ to denote multi-phase DCE-MRI data, where x^0 is the pre-contrast image and the others are the N post-contrast images. The symbols L, W, and H correspond to the length, width, and height of an input patch, respectively. We use $y \in \mathbb{R}^{1 \times L \times W \times H}$ to denote corresponding manual annotation mask. Considering the importance of subtraction images of DCE-MRI in manual tumor annotation, we also use the subtraction images at each phase as additional inputs, which can be calculated by

Table 3. Evaluation of our specially designed modules in tumor segmentation on all testing sets

| Model | Module | | | All testing sets | | |
	whole-breast segmentation model	spatial transformer	temporal transformer	DSC	SEN	HD
Baseline				63.1 ± 1.4	72.1 ± 1.2	15.9 ± 1.1
Baseline + breast mask	✔			67.3 ± 1.2	73.5 ± 1.3	7.6 ± 0.8
Baseline + breast mask + S-trans	✔	✔		69.9 ± 1.2	74.2 ± 1.3	6.9 ± 0.9
Baseline + breast mask + T-trans	✔		✔	70.6 ± 1.6	74.2 ± 1.5	7.3 ± 0.8
Baseline + breast mask + ST-trans	✔	✔	✔	72.4 ± 1.1	76.2 ± 1.3	6.6 ± 0.7

The evaluation metrics include DSC (%) ↑, SEN (%) ↑, and HD (mm) ↓. The modules include whole-breast segmentation model, spatial transformer, and temporal transformer.

$\{s^i = x^i - x^0\}_{i=0}^N \in \mathbb{R}^{(N+1) \times L \times W \times H}$. We input both original DCE-MRI and the subtraction images into the encoder **E** to get feature maps for each phase and denote them as $\{f^i = \mathbf{E}(x^i, s^i)\}_{i=0}^N \in \mathbb{R}^{(N+1) \times C \times L' \times W' \times H'}$. Note that, C, L', W', H' are the channel number, length, width, and height of feature maps, respectively. The feature maps at different phases are then tokenized to form the spatial tokens $z_{spatial} \in \mathbb{R}^{(N+1) \times K \times C}$ and the temporal tokens $z_{temporal} \in \mathbb{R}^{K \times (N+1) \times C}$, where $K = L' \times W' \times H'$. For the spatial token $z_{spatial}$, we treat the first dimension $(N+1)$ as the size of the mini-batch, the second dimension K as the number of tokens, and the last dimension C as the length of each token. In this way, we could well represent spatial information. Differently, the temporal token $z_{temporal}$ is composed of K batches, with each batch containing $N+1$ tokens, and each token having a length of C, which can comprehensively represent dynamic temporal information across different phases. The further proposed spatial-temporal transformers ($\mathbf{T_s}$ and $\mathbf{T_t}$) can exploit the latent relationship among spatial tokens and temporal tokens to improve segmentation performance via the attention mechanism. Specifically, the output tokens can be expressed as $\tilde{z}_{spatial} = \mathbf{T_s}(z_{spatial}) \in \mathbb{R}^{(N+1) \times K \times C}$ and $\tilde{z}_{temporal} = \mathbf{T_t}(z_{temporal}) \in \mathbb{R}^{K \times (N+1) \times C}$. We restore the feature maps $\tilde{f} \in \mathbb{R}^{C' \times L' \times W' \times H'}$ by combining spatial-temporal tokens together and use convolution layers to reduce phase dimension, where C' is the channel number of new feature maps. Then, we can get segmentation results ($\tilde{y} \in \mathbb{R}^{1 \times L \times W \times H}$) from the decoder **D**, which is $\tilde{y} = \mathbf{D}(\tilde{f})$. In order to remove the over-segmentation in the non-breast regions, the final tumor segmentation results are masked by the whole-breast segmentation region for effective removal of false-positive segmentations.

In our segmentation model (Figure 5), we choose the U-Net structure with residual blocks for the encoder **E** and the decoder **D**. Four downsampling and upsampling blocks are adopted. For the spatial and temporal transformers, we use the same architecture as the vision transformer,[39] and the number of transformer layers is 12. We use both DICE and binary cross-entropy loss functions to supervise the training of tumor segmentation model.

Training details

All experiments are conducted on the PyTorch platform with two NVIDIA TITAN RTX GPUs (24GB). We use the ADAM optimizer to optimize all networks. The initial learning rates of the whole-breast and tumor segmentation models are 0.002 and 0.001, respectively. And, the learning rate decays by half for every 50 epochs. A total number of 300 epochs are set for each task. We compute the training loss within 10 epochs to determine the convergence. While using the well-trained segmentation models to test the results, we use sliding windows to crop the overlapping patches, whose stride is half of the patch size. Then, we average the overlapping patches to obtain the final results.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for data should be directed to and will be fulfilled by the lead contact, Dinggang Shen (dgshen@shanghaitech.edu.cn).

Materials availability

This study did not generate new unique materials.

Ethical approval

The data were collected from seven medical centers, including Guangdong Provincial People's Hospital, Yunnan Cancer Hospital, Hangzhou First People's Hospital, Shanghai General Hospital, The Second Xiangya Hospital, Guizhou Provincial People's Hospital, and Ruijin Hospital. This study was approval by the ethical committee of each medical center. Due to the retrospective nature of this study, the informed consent was waived by the relevant institutional review board.

Data and code availability

The full datasets are protected because of privacy issues and regulation policies in hospitals. Partial data can be accessible to support the results in this study, with permission from respective medical centers. The released data

Table 4. Comparison of tumor segmentation (in terms of accuracy and efficiency) between our proposed AI assistant and two radiologists

Model	Expert-1	Expert-1[a]	Expert-2	Expert-2[a]	AI assistant
DSC (%)	81.70	86.17	81.15	86.20	82.04
SEN (%)	79.46	82.49	73.82	80.85	75.68
HD (mm)	0.56	0.52	1.69	0.78	1.25
Time (min)	14.28	2.09	15.16	1.42	0.67

[a]Indicates further manual annotation refinement based on automatic segmentations by our proposed AI assistant. It can be seen that AI + further manual annotation obtains the best result.

AUTHOR CONTRIBUTIONS

In this study, J.Z., Z.C., and D.S. designed the method and drafted the manuscript. J.Z. and Z.C. wrote the code. Z.S., Y.J., Z.Z., X.D., Z.Y., C.H., X.H., C.K., S.L., Z.X., R.W., J.L., J.Z., Z.D., K.S., Z.L., and Z.L. collected and processed the dataset. L.Z., F.G., and L.Z. provided statistical analysis and interpretation of the data. D.S. coordinated and supervised the whole work. All authors were involved in critical revisions of the manuscript and have read and approved the final version.

DECLARATION OF INTERESTS

REFERENCES

1. Venturelli, S., Leischner, C., Helling, T., Renner, O., Burkard, M., and Marongiu, L. (2022). Minerals and Cancer: overview of the possible diagnostic value. Cancers 14, 1256. https://doi.org/10.3390/cancers14051256.

2. Siegel, R.L., Miller, K.D., Fuchs, H.E., and Jemal, A. (2022). Cancer statistics, 2022. CA. CA. Cancer J. Clin. 72, 7–33. https://doi.org/10.3322/caac.21708.

3. Calnan, M. (1984). The health belief model and participation in programmes for the early detection of breast cancer: a comparative analysis. Soc. Sci. Med. 19, 823–830. https://doi.org/10.1016/0277-9536(84)90399-X.

4. Ginsburg, O., Yip, C.H., Brooks, A., Cabanes, A., Caleffi, M., Dunstan Yataco, J.A., Gyawali, B., McCormack, V., McLaughlin de Anderson, M., Mehrotra, R., et al. (2020). Breast cancer early detection: A phased approach to implementation. Cancer 126, 2379–2393. https://doi.org/10.1002/cncr.32887.

5. Nazari, S.S., and Mukherjee, P. (2018). An overview of mammographic density and its association with breast cancer. Breast Cancer 25, 259–267. https://doi.org/10.1007/s12282-018-0857-5.

6. Rodríguez-Ruiz, A., Krupinski, E., Mordang, J.J., Schilling, K., Heywang-Köbrunner, S.H., Sechopoulos, I., and Mann, R.M. (2019). Detection of breast cancer with mammography: effect of an artificial intelligence support system. Radiology 290, 305–314. https://doi.org/10.1148/radiol.2018181371.

7. Sood, R., Rositch, A.F., Shakoor, D., Ambinder, E., Pool, K.L., Pollack, E., Mollura, D.J., Mullen, L.A., and Harvey, S.C. (2019). Ultrasound for breast cancer detection globally: a systematic review and meta-analysis. J. Glob. Oncol. 5, 1–17. https://doi.org/10.1200/JGO.19.00127.

8. Ozmen, N., Dapp, R., Zapf, M., Gemmeke, H., Ruiter, N.V., and van Dongen, K.W.A. (2015). Comparing different ultrasound imaging methods for breast cancer detection. IEEE Trans. Ultrason. Ferroelectrics Freq. Control 62, 637–646. https://doi.org/10.1109/TUFFC.2014.006707.

9. Morrow, M., Waters, J., and Morris, E. (2011). MRI for breast cancer screening, diagnosis, and treatment. Lancet 378, 1804–1811. https://doi.org/10.1016/S0140-6736(11)61350-0.

10. Morris, E.A. (2002). Breast cancer imaging with MRI. Radiol. Clin. 40, 443–466. https://doi.org/10.1016/S0033-8389(01)00005-7.

11. Lehman, C.D., Gatsonis, C., Kuhl, C.K., Hendrick, R.E., Pisano, E.D., Hanna, L., Peacock, S., Smazal, S.F., Maki, D.D., Julian, T.B., et al. (2007). MRI evaluation of the contralateral breast in women with recently diagnosed breast cancer. N. Engl. J. Med. 356, 1295–1303. https://doi.org/10.1056/NEJMoa065447.

12. Yankeelov, T.E., Lepage, M., Chakravarthy, A., Broome, E.E., Niermann, K.J., Kelley, M.C., Meszoely, I., Mayer, I.A., Herman, C.R., McManus, K., et al. (2007). Integration of quantitative DCE-MRI and ADC mapping to monitor treatment response in human breast cancer: initial results. Magn. Reson. Imaging 25, 1–13. https://doi.org/10.1016/j.mri.2006.09.006.

13. Mann, R.M., Kuhl, C.K., and Moy, L. (2019). Contrast-enhanced MRI for breast cancer screening. J. Magn. Reson. Imag. 50, 377–390. https://doi.org/10.1002/jmri.26654.

14. Turnbull, L.W. (2009). Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer. NMR Biomed. 22, 28–39. https://doi.org/10.1002/nbm.1273.

15. Goetz, L.H., and Schork, N.J. (2018). Personalized medicine: motivation, challenges, and progress. Fertil. Steril. 109, 952–963. https://doi.org/10.1016/j.fertnstert.2018.05.006.

16. Guo, J., and Li, B. (2018). The application of medical artificial intelligence technology in rural areas of developing countries. Health Equity 2, 174–181. https://doi.org/10.1089/heq.2018.0037.

17. Panayides, A.S., Amini, A., Filipovic, N.D., Sharma, A., Tsaftaris, S.A., Young, A., Foran, D., Do, N., Golemati, S., Kurc, T., et al. (2020). AI in medical imaging informatics: current challenges and future directions. IEEE J. Biomed. Health Inform. 24, 1837–1857. https://doi.org/10.1109/JBHI.2020.2991043.

18. Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., Jiang, C., Sun, Y., Ma, L., Huang, J., et al. (2022). A fully automatic AI system for tooth and alveolar bone segmentation from cone-beam CT images. Nat. Commun. 13, 2096. https://doi.org/10.1038/s41467-022-29637-2.

19. Chan, H.P., Samala, R.K., and Hadjiiski, L.M. (2020). CAD and AI for breast cancer—Recent development and challenges. Br. J. Radiol. 93, 20190580. https://doi.org/10.1259/bjr.20190580.

20. McKinney, S.M., Sieniek, M., Godbole, V., Godwin, J., Antropova, N., Ashrafian, H., Back, T., Chesus, M., Corrado, G.S., Darzi, A., et al. (2020). International evaluation of an AI system for breast cancer screening. Nature 577, 89–94. https://doi.org/10.1038/s41586-019-1799-6.

21. Sheth, D., and Giger, M.L. (2020). Artificial intelligence in the interpretation of breast cancer on MRI. J. Magn. Reson. Imag. 51, 1310–1324. https://doi.org/10.1002/jmri.26878.

22. Leibig, C., Brehmer, M., Bunk, S., Byng, D., Pinker, K., and Umutlu, L. (2022). Combining the strengths of radiologists and AI for breast cancer screening: a retrospective analysis. Lancet. Digit. Health 4. e507-e519. https://doi.org/10.1016/S2589-7500(22)00070-X.

23. Carter, S.M., Rogers, W., Win, K.T., Frazer, H., Richards, B., and Houssami, N. (2020). The ethical, legal and social implications of using artificial intelligence systems in breast cancer care. Breast 49, 25–32. https://doi.org/10.1016/j.breast.2019.10.001.

24. Granzier, R.W.Y., van Nijnatten, T.J.A., Woodruff, H.C., Smidt, M.L., and Lobbes, M.B.I. (2019). Exploring breast cancer response prediction to neoadjuvant systemic therapy using MRI-based radiomics: a systematic review. Eur. J. Radiol. 121, 108736. https://doi.org/10.1016/j.ejrad.2019.108736.

25. Braman, N.M., Etesami, M., Prasanna, P., Dubchuk, C., Gilmore, H., Tiwari, P., Plecha, D., and Madabhushi, A. (2017). Intratumoral and peritumoral radiomics for the pretreatment prediction of pathological complete response to neoadjuvant chemotherapy based on breast DCE-MRI. Breast Cancer Res. 19. 57-14. https://doi.org/10.1186/s13058-017-0846-1.

26. Yoon, H.J., Kim, Y., Chung, J., and Kim, B.S. (2019). Predicting neo-adjuvant chemotherapy response and progression-free survival of locally advanced breast cancer using textural features of intratumoral heterogeneity on F-18 FDG PET/CT and diffusion-weighted MR imaging. Breast J. *25*, 373–380. https://doi.org/10.1111/tbj.13032.

27. Chamming's, F., Ueno, Y., Ferré, R., Kao, E., Jannot, A.S., Chong, J., Omeroglu, A., Mesurolle, B., Reinhold, C., and Gallix, B. (2018). Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. Radiology *286*, 412–420. https://doi.org/10.1148/radiol.2017170143.

28. Parikh, J., Selmi, M., Charles-Edwards, G., Glendenning, J., Ganeshan, B., Verma, H., Mansi, J., Harries, M., Tutt, A., and Goh, V. (2014). Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy. Radiology *272*, 100–112. https://doi.org/10.1148/radiol.14130569.

29. Agner, S.C., Xu, J., Fatakdawala, H., Ganesan, S., Madabhushi, A., Englander, S., Rosen, M., Thomas, K., Schnall, M., Feldman, M., et al. (2009). Segmentation and classification of triple negative breast cancers using DCE-MRI. In 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro (IEEE), pp. 1227–1230. https://doi.org/10.1109/ISBI.2009.5193283.

30. Zheng, Y., Baloch, S., Englander, S., Schnall, M.D., and Shen, D. (2007). Segmentation and classification of breast tumor using dynamic contrast-enhanced MR images. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007: 10th International Conference, Brisbane, Australia, October 29-November 2, 2007, Proceedings, Part II 10 (Springer Berlin Heidelberg), pp. 393–401. https://doi.org/10.1007/978-3-540-75759-7_48.

31. Zhang, L., Luo, Z., Chai, R., Arefan, D., Sumkin, J., and Wu, S. (2019). Deep-learning method for tumor segmentation in breast DCE-MRI. In Medical Imaging 2019: Imaging Informatics for Healthcare. Research, and Applications *10954*, 97–102. SPIE. https://doi.org/10.1117/12.2513090.

32. Zhang, J., Saha, A., Zhu, Z., and Mazurowski, M.A. (2018). Breast tumor segmentation in DCE-MRI using fully convolutional networks with an application in radiogenomics. Medical Imaging 2018: Computer-Aided Diagnosis *10575*, 192–196. SPIE. https://doi.org/10.1117/12.2295436.

33. Benjelloun, M., El Adoui, M., Larhmam, M.A., and Mahmoudi, S.A. (2018). Automated breast tumor segmentation in DCE-MRI using deep learning. In 2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech) (IEEE), pp. 1–6. https://doi.org/10.1109/CloudTech.2018.8713352.

34. Zhang, J., Saha, A., Zhu, Z., and Mazurowski, M.A. (2019). Hierarchical convolutional neural networks for segmentation of breast tumors in MRI with application to radiogenomics. IEEE Trans. Med. Imag. *38*, 435–447. https://doi.org/10.1109/TMI.2018.2865671.

35. Maicas, G., Carneiro, G., and Bradley, A.P. (2017). Globally optimal breast mass segmentation from DCE-MRI using deep semantic segmentation as shape prior. In 2017 IEEE 14th international symposium on biomedical imaging (IEEE), pp. 305–309. ISBI 2017. https://doi.org/10.1109/ISBI.2017.7950525.

36. Wang, S., Sun, K., Wang, L., Qu, L., Yan, F., Wang, Q., and Shen, D. (2021). Breast tumor segmentation in DCE-MRI with tumor sensitive synthesis. IEEE Transact. Neural Networks Learn. Syst. 1–12. https://doi.org/10.1109/TNNLS.2021.3129781.

37. Qiao, M., Suo, S., Cheng, F., Hua, J., Xue, D., Guo, Y., Xu, J., and Wang, Y. (2021). Three-dimensional breast tumor segmentation on DCE-MRI with a multilabel attention-guided joint-phase-learning network. Comput. Med. Imag. Graph. *90*, 101909. https://doi.org/10.1016/j.compmedimag.2021.101909.

38. Lv, T., and Pan, X. (2021). Temporal-Spatial Graph Attention Networks for DCE-MRI Breast Tumor Segmentation.

39. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S.,

et al. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.11929.

40. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (Springer International Publishing), pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.

41. Zhou, L., Wang, S., Sun, K., Zhou, T., Yan, F., and Shen, D. (2022). Three-dimensional affinity learning based multi-branch ensemble network for breast tumor segmentation in MRI. Pattern Recogn. *129*, 108723. https://doi.org/10.1016/j.patcog.2022.108723.

42. van Griethuysen, J.J.M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G.H., Fillion-Robin, J.C., Pieper, S., and Aerts, H.J.W.L. (2017). Computational radiomics system to decode the radiographic phenotype. Cancer Res. *77*. e104-e107. https://doi.org/10.1158/0008-5472.CAN-17-0339.

43. Zhang, Y., Chen, J.H., Lin, Y., Chan, S., Zhou, J., Chow, D., Chang, P., Kwong, T., Yeh, D.C., Wang, X., et al. (2021). Prediction of breast cancer molecular subtypes on DCE-MRI using convolutional neural network with transfer learning between two centers. Eur. Radiol. *31*, 2559–2567. https://doi.org/10.1007/s00330-020-07274-x.

44. Kim, J.J., Kim, J.Y., Suh, H.B., Hwangbo, L., Lee, N.K., Kim, S., Lee, J.W., Choo, K.S., Nam, K.J., Kang, T., and Park, H. (2022). Characterization of breast cancer subtypes based on quantitative assessment of intratumoral heterogeneity using dynamic contrast-enhanced and diffusion-weighted magnetic resonance imaging. Eur. Radiol. *32*, 822–833. https://doi.org/10.1007/s00330-021-08166-4.

45. Lee, J.Y., Lee, K.S., Seo, B.K., Cho, K.R., Woo, O.H., Song, S.E., Kim, E.K., Lee, H.Y., Kim, J.S., and Cha, J. (2022). Radiomic machine learning for predicting prognostic biomarkers and molecular subtypes of breast cancer using tumor heterogeneity and angiogenesis properties on MRI. Eur. Radiol. *32*, 650–660. https://doi.org/10.1007/s00330-021-08146-8.

46. Ma, W., Ji, Y., Qi, L., Guo, X., Jian, X., and Liu, P. (2018). Breast cancer Ki67 expression prediction by DCE-MRI radiomics features. Clin. Radiol. *73*. 909.e1. 909909.e5. https://doi.org/10.1016/j.crad.2018.05.027.

47. Fan, M., Yuan, W., Zhao, W., Xu, M., Wang, S., Gao, X., and Li, L. (2020). Joint prediction of breast cancer histological grade and Ki-67 expression level based on DCE-MRI and DWI radiomics. IEEE J. Biomed. Health Inform. *24*, 1632–1642. https://doi.org/10.1109/JBHI.2019.2956351.

48. Hirsch, L., Huang, Y., Luo, S., Saccarelli, C.R., Gullo, R.L., Naranjo, I.D., Bitencourt, A.G., Onishi, N., Ko, E.S., Leithner, D., et al. (2020). Deep learning achieves radiologist-level performance of tumor segmentation in breast MRI. Preprint at arXiv. https://doi.org/10.48550/arXiv.2009.09827.

49. Wu, H., Gallego-Ortiz, C., and Martel, A. (2015). Deep artificial neural network approach to automated lesion segmentation in breast. In Proceedings of the 3rd MICCAI workshop on Breast Image Analysis, pp. 73–80.

50. El Adoui, M., Mahmoudi, S.A., Larhmam, M.A., and Benjelloun, M. (2019). MRI breast tumor segmentation using different encoder and decoder CNN architectures. Computers *8*, 52. https://doi.org/10.3390/computers8030052.

51. Chen, M., Zheng, H., Lu, C., Tu, E., Yang, J., and Kasabov, N. (2018). A spatio-temporal fully convolutional network for breast lesion segmentation in DCE-MRI. In Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part VII 25 (Springer International Publishing), pp. 358–368. https://doi.org/10.1007/978-3-030-04239-4_32.

52. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. Adv. Neural Inf. Process. Syst. *30*.

53. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3 (Springer International Publishing), pp. 240–248. https://doi.org/10.1007/978-3-319-67558-9_28.

54. Zhang, J., Cui, Z., Shi, Z., Jiang, Y., Zhang, Z., Dai, X., Yang, Z., Gu, Y., Zhou, L., Han, C., et al. (2023). A robust and efficient ai assistant for breast tumor segmentation from dce-mri via a spatial-temporal framework. Zenodo. https://doi.org/10.5281/zenodo.8068383.