Generative adversarial networks-based super-resolution of
diffusion-weighted imaging: application to tumour
radiomics in breast cancer

Correspondence

Lihua Li,
Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University,
Hangzhou, China, 310018
Phone: +86-571-86878587
Fax: +86-571-87713528
Email: lilh@hdu.edu.cn

Word count: 6543 words

ABSTRACT

Diffusion-weighted imaging (DWI) is increasingly used to guide the clinical management of patients with breast tumours. However, accurate tumour characterization with DWI and the corresponding apparent diffusion coefficient (ADC) maps is challenging due to their

limited resolution. Thus, this study aimed to produce super-resolution (SR) ADC images and to assess the clinical utility of these SR images by performing a radiomic analysis for predicting the histologic grade and Ki-67 expression status of breast cancer. To this end, 322 samples of dynamic enhanced magnetic resonance imaging (DCE-MRI) and the corresponding DWI data were collected. The dataset was separated into a development dataset (n=222) for establishing a deep SR model using DCE-MR images and a validation dataset (n=100) to improve the resolution of ADC images. A super-resolution using a generative adversarial network (SRGAN) and enhanced deep SR (EDSR) method along with the bicubic interpolation method were utilized to generate SR-ADC images, from which radiomic features were extracted. The EDSR method was significantly better than the SRGAN and bicubic methods in terms of objective quality criteria. Univariate and multivariate predictive models of radiomic features were established with an area under the receiver operating characteristic (ROC) curve (AUC). Individual features from the tumour SR-ADC images showed higher performances with the EDSR and SRGAN methods than with the bicubic methods and the original images. Multivariate analysis of the collective tumour radiomics showed that the EDSR and SRGAN-based SR-ADC images performed better than the bicubic method and original images in predicting either the Ki-67 expression levels (AUCs of 0.865 and 0.828, respectively) or the tumour grade (AUCs of 0.839 and 0.819, respectively). This work demonstrates that, in addition to improving the resolution of ADC images, deep-learning-based SR methods can also improve tumour image-based diagnosis in breast cancer.

Key Words: Breast cancer, generative adversarial network, enhanced deep super-resolution, apparent diffusion coefficient, Ki-67, histological grade

List of Abbreviations:

ADC, Apparent diffusion coefficient; DCE-MRI, Dynamic contrast-enhanced magnetic resonance imaging; AUC, Area under the receiver operating characteristic curve; DWI, Diffusion-weighted imaging; SR, super-resolution; SRGAN, super-resolution using a generative adversarial network; EDSR, enhanced deep super-resolution; ResNet, residual neural network; BN, batch-normalization; ReLU, rectified linear unit; PSNR, peak-signal-to-noise ratio; SSIM, structural similarity

1. INTRODUCTION

Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) has become an indispensable part of breast cancer care, with its ability to reflect blood flow and tumour angiogenesis and, thus, has a high sensitivity for tumour diagnosis.¹ Radiomics features from DCE-MRI have been demonstrated to have utility for tumour diagnosis and treatment, such as to predict the molecular subtype,^{2,3} response to chemotherapy,⁴ histological grade⁵ and Ki-67 expression status.^{6,7} As an adjunct to DCE-MRI, diffusion-weighted imaging (DWI) is being investigated as a part of multi-parametric MRI;⁸ the derived apparent diffusion coefficient (ADC) image reflects the impedance of water molecules in the microstructural environment and tissue microstructure and cellularity and is commonly used for lesion characterization and diagnosis, especially for the prediction of receptor status and molecular subtypes.⁹⁻¹²

On the other hand, the resolution of the ADC image in routine clinical practice is quite low compared to that of the DCE-MRI, which makes it challenging to rigorously investigate ADC for tumour characterization. Moreover, due to the low spatial resolution and image quality, ADC is less frequently used in radiomic analysis.¹³ Various techniques,

such as in-plane multiband techniques, have been developed to address the issue,¹⁴ which has enhanced resolution but a reduced field of view (FOV).¹⁵ In this context, image features were extracted to be used as biomarkers for predicting response to chemotherapy^{16,17} and pathologic indicators¹⁸ in breast cancer. However, high-resolution DWI requires special technology, which generally has minimal application; hence, most DWI in clinical practice still contains substantial noise and suffers from low in-plane resolution. Therefore, there is great interest in improving DWI resolution using an established family of computer vision techniques by transforming low-resolution (LR) images into high-resolution (HR) images to facilitate more accurate tumour characterization and, hence, better clinical application.

To this end, conventional interpolation methods, such as spline and bicubic interpolation, are used to enhance the image resolution. However, these data interpolation methods inevitably lead to blurred edges and blocking artefacts. Deep-learning-based super-resolution (SR) methods have provided an alternative approach to improve image resolution based on end-to-end training from a set of LR images and their HR counterparts.¹⁹ Dong et al. used bicubic interpolation to upscale input images and trained a three-layer deep convolutional neural network to achieve state-of-the-art SR performance.^{20,21} Ledig et al. proposed SRGAN, a generative adversarial network (GAN) for image SR for natural images.²² A follow-up study of enhanced deep super-resolution (EDSR) networks further improved the SR performance by employing a better residual neural network (ResNet) architecture.²³ Recent studies have demonstrated the effectiveness of deep-learning-based SR methods in improving the image quality of neuroimaging,²⁴⁻²⁷ fluorescence microscopy²⁸ and musculoskeletal imaging.²⁹

Despite these advances, the effectiveness of image SR methods on cancer images, especially on LR ADC images, remains unknown. More importantly, the utility of the SR image must be investigated to verify its clinical implications for tumour diagnosis. The histological grade is an important clinical indicator for guiding breast cancer treatment,³⁰ while the Ki-67 is an indicator of tumour aggressiveness and proliferative activity in breast cancer.³¹ The aim of this study was to improve the spatial resolution of ADC images and to assess the clinical utility of these SR-ADC images by performing a radiomics analysis, specifically in the prediction of the commonly used biomarkers of histological grade and Ki-67 expression status in breast cancer.

2. METHODS

2.1. Framework overview

The framework of this work is shown in Figure 1. We assumed that the object structures of tumours are similar in both DCE-MRI and ADC images acquired from the same subject, and we aimed to recover such structures in the trained model. More specifically, a GAN-based deep network was trained through an end-to-end mapping between the LR image and its HR counterpart for DCE-MRI, and the trained model was in turn applied to ADC images to improve their image quality, which yielded SR-ADC images. Tumour radiomics, including morphological, statistical and texture features, were extracted on the SR-ADC images. Radiomics from the SR-ADC images were evaluated by assessing their predictive performance of tumour histological grade and Ki-67 expression status in breast cancer.

2.2. Pathologic assessment

The histological grade was identified by analysing the breast specimens independently on full-face sections by two pathologists with more than 15 years of experience who were blind to the results of MRI examination. The histological type of breast cancer was defined according to WHO classification. The tumour histological grade was assessed by the Elston–Ellis System.³² According to the St. Gallen consensus, for Ki-67, samples with more than 14% positive nuclei were determined to have high Ki-67 expression, while those with lower than 14% positive nuclei were considered to have low Ki-67 expression.³³

2.3. Imaging protocol

DCE-MR imaging was performed on all patients in the prone position with a dedicated eight-channel double-breast coil (Siemens Magnetom Verio 3.0T, Siemens Medical Systems). DCE-MR imaging was performed with a fat-suppressed T1-weighted scanning with 144 slices. The image parameters were set up as follows: repetition time (TR) = 4.51 ms, echo time (TE) = 1.61 ms, matrix = 448×448 , flip angle (FA) = 10° , field of view (FOV) = 340×340 mm, slice thickness = 1.0 mm, space between slices = 1.0 mm and in-plane resolution = 0.759×0.759 mm. After intravenous injection of contrast agent with a bolus of 0.1 mmol/kg gadobutrol at a rate of 2 mL/second, five post-contrast bilateral sagittal image acquisitions were sequentially performed with an interval of 60 seconds.

The DWI sequence was acquired before starting the pre-contrast series of MRI. The parameters were set up as follows: TR = 8400 ms, TE = 84 ms, FA = 10° , FOV = 147×359 mm, slice thickness = 4.0 mm, space between slices = 6.0 mm, pixel band width = 1263 Hz, spatial resolution = 1.64 mm, and matrix = 220×90 . DW images were obtained at $b = 50, 400$ and 800 seconds/ mm^2 , with 24 slices for each b value. ADC images were derived from DWI at the three b values, which yield 24 image slices.

2.4. Image processing

Breast cancer lesions were manually segmented by two breast radiologists in consensus with 15 years of experience. The segmentation was performed on the subtraction images between the post-contrast image series and the pre-contrast series. The breast MR images acquired from sequential MR imaging scans were registered as previously reported.³⁴ The breast area was segmented, excluding the skin and chest wall from the image.³⁵

2.5. Deep super-resolution networks

Deep-learning-based SRGAN and EDSR network structures were adopted because they have achieved great success in image SR.^{22,23} For SRGAN, the mean squared error (MSE)-based SR residual network (SRResNet) was employed for the generator, which has 16 residual blocks. In the generator, two convolutional layers were used with 3×3 kernels and 64 feature maps, followed by batch-normalization (BN) layers and rectified linear unit (ReLU) as the activation function.³⁶ The discriminator network included six convolutional layers with 4 × 4 filter kernels, two 1×1 filter kernels with leaky ReLU. The generator and discriminator networks were updated alternately. The resolution of the input image was enhanced with two sub-pixel convolution layers. The resulting 512 feature maps were followed by two dense layers and a final sigmoid activation function to obtain a probability for classification. The network structure details were previously described in Ledig et al.²² The initialization was performed using the Adam solver with 100 epochs at a learning rate of 1e-4. After that, SRGAN was trained with 500 epochs at a learning rate of 1e-4 and another 500 epochs at a lower learning rate of 1e-5.

For EDSR, the structure is similar to SRGAN; however, the BN layers are removed, as they discard the individual characteristics of each sample by normalizing the feature maps. Compared to that of the SRGAN, the generator network is increased to 32 residual blocks, and residual scaling was established with a factor of 0.1. For the resolution-increasing module, two up-sampling $\times 2$ factors were used for the $\times 4$ model. The Adam solver was used with the initial parameters of batch size = 16. The network was trained for 300000 iterations using the Adam solver with a first-order momentum of 0.9 and a second-order momentum of 0.999. The learning rate was initialized as $1e-4$ and decreased to 50% after 20000 epochs.

The goal of our training network is to generate HR ADC images based on the learned end-to-end mapping between the LR DCE-MRI and HR DCE-MRI data. For DCE-MRI, the pre-contrast, the subtraction image between the third post-contrast series and the pre-contrast images, as well as the subtraction image between the fifth post-contrast and S0, were utilized for training, termed S0, S30 and S50, respectively. DCE-MRI images with segmented breast areas that exclude skin and chest walls were used to remove redundant regions that were irrelevant to diagnosis. Therefore, the original DCE-MRI image, 448×448 in size, was reduced to 200×448 , while the LR image of DCE-MRI was reduced to 50×112 . This procedure was used to avoid large-scale convolution calculations on background pixels in the deep SR networks. The segmented breast area was used as the HR image, and the corresponding LR image was obtained after four down-sampling processes. For each image series (i.e., S0, S30 and S50), the slices that contained a tumour were used for training. Specifically, 20,616 LR DCE-MR images, with a size of 50×112 , and their corresponding HR reference images, with a size of 200×448 , are used to train the

network. The resolution of the ADC map was improved through the established SR model applied to the testing set with a batch size of 16. To obtain the HR version of the ADC image, the established deep-SR model using all the training samples was applied to LR-ADC images on the validation set, with a batch size of 16. Our implementation of the deep SR model is based on TensorFlow and was performed using an GTX 1080TI 11G GPU.³⁷

2.6. Radiomic analysis

For each lesion, image features were extracted, including statistic, morphologic and texture features using the publicly available Pyradiomics toolbox.³⁸ The histogram distribution of the pixel value in the tumour image was represented by 18 statistical features. Texture features were extracted through a grey-level co-occurrence matrix (GLCM) of 24 features, a grey-level run-length matrix (GLRLM) of 16 features, a grey-level size-zone matrix (GLSZM) of 16 features, a grey-level spatial dependence matrix (GLDM) of 14 features, and a neighbourhood grey tone difference matrix (NGTDM) of 5 features. These features were extracted on the ADC map derived from DWI. A total of 107 features were included in our study for evaluation (Table 2).

2.7. Statistical analysis

To measure the SR performance, the peak-signal-to-noise ratio (PSNR) and structural similarity (SSIM) were used; both metrics are widely used criteria for quantitative measurements of reconstructed image quality in computer vision. A deep SR model was established on the development dataset and was applied on the validation dataset to improve the ADC image resolution.

To assess individual and collective feature performance, univariate and multivariate logistic regression classifiers were performed respectively on the images of the original ADC, the SR-ADC images by methods of the bicubic interpolation, the SRGAN and the EDSR for prediction of histologic grade and Ki-67 expression status. To reduce multicollinearity in the multivariate model, the features that had high similarity (i.e., Pearson correlation coefficient greater than 0.9) with other features were removed. Thereafter, feature selection of an evolutionary algorithm (EA)-based optimization method was used to search for optimal feature subsets for the multivariate logistic regression model, similar to what was previously performed.⁴ The chromosome population, the maximum number of generations, the cross-over probability and the mutation probability were set as 500, 500, 0.1, and 0.6, respectively. A nested leave-one-out cross-validation (LOOCV) was performed to compute the AUC value of the predictive model performance using tumour radiomics on the validation dataset.³⁵ The inner loop of LOOCV was used for feature selection in combination with ten-fold cross-validation on the training data, while the outer loop produced a prediction score for the testing sample. An area under the receiver operating characteristic (ROC) curve (AUC) was used to assess the predictive model performance.

The intra-class correlation coefficient (ICC) was calculated to evaluate the radiomic feature variability among the original image and SR-ADC images based on bicubic interpolation, SRGAN and EDSR methods. P-value less than 0.05 was considered significant. A Benjamini–Hochberg procedure was used to control the false discovery rate. All statistical tests were performed using MATLAB (version R2016a, Mathworks, Inc,

Natick, MA). Machine learning methods of feature selection and validation were performed using a publicly available software platform, Weka.³⁹

3. RESULTS

3.1. Patient clinicopathological characteristics

The study was approved by the institutional review board (IRB) of the First Affiliated Hospital of Zhejiang Chinese Medical University. The requirement for informed consent was waived by the IRB due to the retrospective nature of this study. The dataset initially included 429 cases of breast cancer patients who underwent preoperative breast MRI between February 2007 and February 2011. Thereafter, 92 samples were excluded due to incomplete image series, and 15 samples were removed because they did not include malignant cancers. The final dataset included 322 patients with 322 biopsy-proven malignant tumours with invasive ductal carcinoma (size range 6-120 mm, mean 23.68 mm) (Table 1). The entire dataset was separated into a development dataset with 222 samples (68.9%) and a validation dataset with 100 (31.1%) cases. It should be noted that part of the samples had no available Ki-67 or tumour grade information. To make full use of the imaging data, we incorporated the samples that had no histological information into the development dataset, where a deep SR model was trained without the need for the histological information. Among the validation dataset (n=100) for the radiomic analysis, 44 (14%) were histologic grade 1 or 2 and 55 (17.1%) were histologic grade 3. Ki67 expression levels of more than 14% were presented in 73 samples (22.7%). The statistical test showed no significant differences regarding the Ki-67 expression level, histological

grade, menopausal status, maximum diameter and age between the development and validation datasets (Table 1).

3.2. Image super-resolution assessment

The performances of the SR methods of EDSR, SRGAN and bicubic interpolation were evaluated by the PSNR and SSIM on the development dataset (Table 3). The EDSR method achieves the best SR performance in terms of both PSNR and SSIM, whereas SRGAN achieves an intermediate performance, and the bicubic method has the lowest performance. Statistical tests of comparisons between the PSNR and SSIM values for each pair of methods (i.e., EDSR and SRGAN, SRGAN and Bicubic, and EDSR and Bicubic) showed significant differences with corrected p-values less than $1e-5$ (Figure 2).

Examples of SR methods, including the bicubic, SRGAN and EDSR methods, on the DCE-MRI and ADC images are illustrated in Figure 3. For DCE-MRI, the bicubic interpolation showed the lowest SR performance, with PSNR and SSIM values of 33.405 and 0.953, respectively. The SR DCE-MRI using the bicubic interpolation method has blurry areas (Figure 3 b), while that using the SRGAN and EDSR methods achieved a clearer appearance of the tumour (Figure 3 c, d). The SRGAN method achieved improved resolution, with PSNR and SSIM values of 34.796 and 0.960, respectively. Although the SR image using SRGAN produced sharper areas, the EDSR-based method achieved the highest performance, with PSNR and SSIM values of 35.815 and 0.974, respectively. Moreover, the SR-ADC image generated by EDSR had a clearer boundary and tumour structure compared to the other SR methods and the original LR ADC image.

3.3. Radiomic feature variability assessment

Radiomic feature variability between the original ADC image and SR-ADC images based on bicubic interpolation, SRGAN and EDSR methods was evaluated by calculating ICC values for the statistical, morphologic, and texture features (Figure 4). Among the features, all the statistical features and morphologic features of the tumour image have ICC values greater than 0.7, which suggests that these types of features are stable given LR or HR images. Additionally, the GLCM-based texture features have relatively high agreement, with a mean ICC value of 0.787 and a 95% confidence interval (CI) from 0.750 to 0.824, and 87.5% (21 out of 24) of the features had ICC values greater than 0.7. In contrast, the GLSZM-based texture features showed relatively low agreement between original images and SR-ADC images with these methods, with a mean ICC value of 0.716 and a 95% CI from 0.653 to 0.778, and 9 out of 14 (64.3%) features had ICC values greater than 0.7.

3.4. Individual feature performance for the EDSR-based SR method

We further examined the effectiveness of SR-ADC images for breast cancer diagnosis. Table 4 shows the individual feature performance (in terms of AUC values) of the original images and SR-ADC images using EDSR, SGRAN, and bicubic interpolation methods for prediction of tumour grade and Ki-67 expression level, with the averaged increment of AUC values larger than 0.1. The distribution of the individual AUC values of the 107 features based on the original ADC images and resolution enhanced images using the bicubic, SRGAN and EDSR methods is illustrated in Figure 5. In the prediction of Ki-67 status, individual features from the tumour SR-ADC images using the SRGAN method showed significantly higher AUC values than those using either the original images or the bicubic or EDSR-generated SR-ADC images, with p-values of 1.805e-5, 4.108e-4 and 0.015, respectively. Moreover, the AUC values using the EDSR-generated SR-ADC

images were significantly higher than those using the original images, with a p-value of 0.049. In the prediction of tumour grade, the AUC values of radiomic features from the EDSR-generated SR-ADC images were significantly higher than those from the original images, with a p-value of 0.026.

We also show examples of individual features with enhanced prediction performance using the EDSR method. The image feature values of the grey level variance, inverse difference, high grey level zone emphasis and short run high grey level emphasis are shown in Figure 6. The feature value of the grey level variance is increased in the EDSR-based image, while the feature value increment is greater in the high-grade group than in the low-grade group (Figure 6 a, c, d). On the other hand, the inverse difference feature value of the SR images in the low-grade group is similar to that of the original images, while higher feature values were observed in the high-grade group (Figure 6 b). This makes the differences in the inverse difference features of the EDSR-based images between the low-grade and high-grade groups more significant, with a p-value of 8.022×10^{-5} .

3.5. Multivariate model based on the original ADC and SR-ADC images for the prediction of Ki-67 and histological grade

The collective performance of the imaging features based on the original image, the SR-ADC images using the bicubic method, the SRGAN and EDSR methods are shown in Figure 7. The SR-ADC image generated by EDSR showed the highest AUC values of 0.865 compared to that based on the SRGAN, bicubic and original images, with AUCs of 0.828, 0.821 and 0.795, respectively. Similarly, for the Ki-67 prediction, the images from the EDSR-based method achieved the highest prediction performance (AUC = 0.839), and that based on the original images had the lowest performance, with an AUC of 0.781. We

observed an ascending trend of these AUC values for the original images and SR images using the bicubic, SRGAN and EDSR methods, which is consistent with that of the SR performance. In other words, the greater resolution improvement (in terms of the PSNR and SSIM) may be correlated with the higher predictive performance for the tumour grade and the Ki-67 expression level classification.

4. DISCUSSION

We have applied and compared multiple SR methods, including the bicubic, SRGAN and EDSR methods, to improve the image quality of ADC images and to assess the clinical utility of the SR-ADC images using radiomic analyses. Our results showed the superior performance of deep-learning-based methods over the bicubic interpolation method in terms of visual quality and the objective quality criteria of PSNR and SSIM. More importantly, the predictive model achieved better performance based on images with higher SR performance than based on the original images and the bicubic interpolation method, with its lower SR performance. Our results suggest a promising method of deep SR for enhancing image quality toward improving the diagnosis of breast cancer.

Previous studies have applied deep learning SR methods in applications for musculoskeletal MRI.²⁹ A recent study used deep SR and outperformed a tricubic interpolation method to improve the image quality.⁴⁰ A previous study proposed a method for the prediction of HR 7T-like MR images from LR 3T MRIs, which showed better spatial resolution compared to 3T MR images and the highest accuracy for segmentation.⁴¹ However, previous studies have not examined SR images in terms of clinical utility through radiomic analysis. In this study, we systematically applied SR methods to MR images and

assessed their clinical applications in breast cancer diagnosis. Moreover, we trained a deep SR model using LR as well as HR DCE-MR images and applied the model to the ADC images to improve the image quality. This was performed based on the observation that ADC images and their corresponding DCE-MRI are derived from the same sources and, hence, show similar tumour shape. Specifically, we used HR DCE-MRI data to guide the SR of the ADC images. The visual inspection showed an improved image quality of the SR images.

In our study, the image resolution was improved by using interpolation and deep learning methods. We observed the superiority of the deep SR method over the bicubic interpolation method; simultaneously, we found that the prediction performance of tumour radiomics on SR-ADC is better than that for the original images and those generated by bicubic interpolation. The results suggested that improved diagnosis accuracy was associated with higher image resolution. Moreover, 107 radiomic features were calculated using publicly available software to ensure the repeatability of this study.³⁸ The statistic features and morphologic features are less affected by the image resolution in terms of the ICC value (greater than 0.7) of all features, whereas the texture features, such as GLSZM features, showed relatively low agreement, with low ICC values.

Our study has several limitations. First, the SR model is built based on the 2D analysis of the DCE-MRI slices. A further study based on the 3D SR model should be performed to examine the effectiveness of the method. Second, in this study, we examined radiomic features in tumour ROI on the ADC image and the enhanced tumour ADC. It is interesting to determine whether the imaging features in the parenchymal region that surrounds the tumour could be improved for the prediction tasks. Third, in this study, we trained deep SR

1
2
3 methods using DCE-MRI series of S0, S30 and S50, based on which the ADC map was
4
5 tested for an HR image. This is an indirect method of SR because the DWI in routine
6
7 clinical usage does not have an HR version. Specifically, we used HR DCE-MR images to
8
9 guide the SR model for improving the image quality of ADC.
10
11

12
13 **5. CONCLUSIONS**
14
15

16 In an effort to illustrate the utility of a deep-learning SR neural network on both image
17
18 resolution improvement and image-based diagnosis, we have quantitatively shown that
19
20 deep SR methods enhance image quality to a better extend compared with conventional
21
22 methods. In particular, radiomic features derived from tumour SR-ADC images have better
23
24 predictive performance than those derived from conventional methods. Based on these
25
26 observations, deep SR may be a promising technique for improving the effectiveness of
27
28 DWI-based diagnoses of breast cancer.
29
30

31
32 **Funding information**
33
34

35 This work was supported by the National Natural Science Foundation of China
36
37 (61731008, 61871428), the Natural Science Foundation of Zhejiang Province of China
38
39 (LJ19H180001), and by the King Abdullah University of Science and Technology
40
41 (KAUST) Office of Sponsored Research (OSR) under award no. URF/1/3450-01.
42
43
44
45

46
47 **REFERENCES**
48
49

50 1. Pinker K, Helbich TH, Morris EA. The potential of multiparametric MRI of the breast. Br
51 J Radiol 2017; 90:20160715.
52
53
54
55
56
57
58
59
60

2. Fan M, Li H, Wang S, Zheng B, Zhang J, Li L. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. *PLoS One* 2017; 12:e0171683.
3. Sutton EJ, Dashevsky BZ, Oh JH, et al. Breast cancer molecular subtype classifier that incorporates MRI features. *J Magn Reson Imaging* 2016; 44:122-129.
4. Fan M, Wu G, Cheng H, Zhang J, Shao G, Li L. Radiomic analysis of DCE-MRI for prediction of response to neoadjuvant chemotherapy in breast cancer patients. *Eur J Radiol* 2017; 94:140-147.
5. Fan M, Liu Z, Xie S, et al. Integration of dynamic contrast-enhanced magnetic resonance imaging and T2-weighted imaging radiomic features by a canonical correlation analysis-based feature fusion method to predict histological grade in ductal breast carcinoma. *Phys Med Biol* 2019; 64:215001.
6. Fan M, Yuan W, Zhao W, et al. Joint prediction of breast cancer histological grade and Ki-67 expression level based on DCE-MRI and DWI radiomics. *IEEE J Biomed Health Inform* 2019.
7. Fan M, Cheng H, Zhang P, et al. DCE-MRI texture analysis with tumor subregion partitioning for predicting Ki-67 status of estrogen receptor-positive breast cancers. *J Magn Reson Imaging* 2018; 48:237-247.
8. Partridge S, Rahbar H, Murthy R, et al. Improved diagnostic accuracy of breast MRI through combined apparent diffusion coefficients and dynamic contrast-enhanced kinetics. 2011; 65:1759-1767.
9. Tan SL, Rahmat K, Rozalli FI, et al. Differentiation between benign and malignant breast lesions using quantitative diffusion-weighted sequence on 3 T MRI. *Clin Radiol* 2014; 69:63-71.
10. Fan M, He T, Zhang P, Zhang J, Li L. Heterogeneity of Diffusion-Weighted Imaging in Tumours and the Surrounding Stroma for Prediction of Ki-67 Proliferation Status in Breast Cancer. *Sci Rep* 2017; 7:2875.

11. Fan M, He T, Zhang P, et al. Diffusion-weighted imaging features of breast tumours and the surrounding stroma reflect intrinsic heterogeneous characteristics of molecular subtypes in breast cancer. *NMR Biomed* 2018 31(2).

12. Leithner D, Bernard-Davila B, Martinez DF, et al. Radiomic Signatures Derived from Diffusion-Weighted Imaging for the Assessment of Breast Cancer Receptor Status and Molecular Subtypes. *Mol Imaging Biol* 2019.

13. Vidic I, Egnell L, Jerome NP, et al. Support vector machine for breast cancer classification using diffusion-weighted MRI histogram features: Preliminary study. *J Magn Reson Imaging* 2017.

14. Sinha S, Sinha U. Functional magnetic resonance of human breast tumors: diffusion and perfusion imaging. *Ann N Y Acad Sci* 2002; 980:95-115.

15. Taviani V, Alley MT, Banerjee S, et al. High-resolution diffusion-weighted imaging of the breast with multiband 2D radiofrequency pulses and a generalized parallel imaging reconstruction. *Magn Reson Med* 2017; 77:209-220.

16. McLaughlin RL, Newitt DC, Wilmes LJ, et al. High resolution in vivo characterization of apparent diffusion coefficient at the tumor-stromal boundary of breast carcinomas: a pilot study to assess treatment response using proximity-dependent diffusion-weighted imaging. *J Magn Reson Imaging* 2014; 39:1308-1313.

17. Singer L, Wilmes LJ, Saritas EU, et al. High-resolution diffusion-weighted magnetic resonance imaging in patients with locally advanced breast cancer. *Acad Radiol* 2012; 19:526-534.

18. Shin HJ, Park JY, Shin KC, et al. Characterization of tumor and adjacent peritumoral stroma in patients with breast cancer using high-resolution diffusion-weighted imaging: Correlation with pathologic biomarkers. *Eur J Radiol* 2016; 85:1004-1011.

19. Dalca AV, Bouman KL, Freeman WT, Rost NS, Sabuncu MR, Golland P. Medical Image Imputation from Image Collections. *IEEE Trans Med Imaging* 2018.

20. Dong C, Loy CC, He K, Tang X. Learning a Deep Convolutional Network for Image Super-Resolution. 2014; Cham: Springer International Publishing; 2014. p. 184-199.
21. Dong C, Loy CC, He K, Tang X. Image Super-Resolution Using Deep Convolutional Networks. IEEE Trans Pattern Anal Mach Intell 2016; 38:295-307.
22. Ledig C, Theis L, Huszar F, et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. 30th Ieee Conference on Computer Vision and Pattern Recognition (Cvpr 2017) 2017:105-114.
23. Lim B, Son S, Kim H, Nah S, Mu Lee K. Enhanced deep residual networks for single image super-resolution. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2017; 2017. p. 136-144.
24. Oktay O, Bai W, Lee M, et al. Multi-input Cardiac Image Super-Resolution Using Convolutional Neural Networks. 2016; Cham: Springer International Publishing; 2016. p. 246-254.
25. Pham C, Ducournau A, Fablet R, Rousseau F. Brain MRI super-resolution using deep 3D convolutional networks. 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017); 2017 18-21 April 2017; 2017. p. 197-200.
26. McDonagh S, Hou B, Alansary A, et al. Context-Sensitive Super-Resolution for Fast Fetal Magnetic Resonance Imaging. 2017; Cham: Springer International Publishing; 2017. p. 116-126.
27. Zeng K, Zheng H, Cai C, Yang Y, Zhang K, Chen Z. Simultaneous single- and multi-contrast super-resolution for brain MRI images based on a convolutional neural network. Computers in Biology and Medicine 2018; 99:133-141.
28. Li Y, Xu F, Zhang F, et al. DLBI: deep learning guided Bayesian inference for structure reconstruction of super-resolution fluorescence microscopy. Bioinformatics 2018; 34:i284-i294.

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
29. Chaudhari AS, Fang Z, Kogan F, et al. Super-resolution musculoskeletal MRI using deep learning. *Magn Reson Med* 2018; 80:2139-2154.
30. Rakha EA, Reis-Filho JS, Baehner F, et al. Breast cancer prognostic classification in the molecular era: the role of histological grade. *Breast Cancer Res* 2010; 12:207.
31. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. *Lancet Oncol* 2010; 11:174-183.
32. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology* 1991; 19:403-410.
33. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thurlimann B, Senn HJ. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. *Ann Oncol* 2011; 22:1736-1747.
34. Yang Q, Li L, Zhang J, Shao G, Zhang C, Zheng B. Computer-aided diagnosis of breast DCE-MRI images using bilateral asymmetry of contrast enhancement between two breasts. *J Digit Imaging* 2014; 27:152-160.
35. Fan M, Zhang P, Wang Y, et al. Radiomic analysis of imaging heterogeneity in tumours and the surrounding parenchyma based on unsupervised decomposition of DCE-MRI for predicting molecular subtypes of breast cancer. *Eur Radiol* 2019; 29:4456-4467.
36. Normalization B. Accelerating deep network training by reducing internal covariate shift. *CoRR-2015-Vol abs/150203167-URL: <http://arxiv.org/abs/150203167> 2015.*
37. Tensorflow. Accessed: Jun. 2019. [Online]. Available: <https://github.com/tensorflow/tensorflow>.
38. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 2017; 77:e104-e107.

- 1
2
3 39. Witten I, Frank E, Hall M. Data mining: practical machine learning tools and techniques, 3rd
4 ed. Amsterdam: Elsevier, <http://www.cs.waikao.ac.nz/ml/weka/>.
5
6
7 40. Chaudhari AS, Stevens KJ, Wood JP, et al. Utility of deep learning super-resolution in the
8 context of osteoarthritis MRI biomarkers. J Magn Reson Imaging 2019.
9
10 41. Bahrami K, Shi F, Rekik I, Gao Y, Shen D. 7T-guided super-resolution of 3T MRI. Med Phys
11
12 2017; 44:1661-1677.
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Tables

Table 1. Tumour characteristics

Characteristic	All (n=322)	Development set (n=222)	Validation set (n=100)	P-value ^a
Ki-67 ^b				0.2748
High	181	108/222 (33.5%)	73/100 (22.7%)	
Low	82	55/222 (17.1%)	27/100 (8.4%)	
NA	59	59/222 (18.3%)	0/100 (0%)	
Histological grade ^b				0.7833
I/II	94	49/222 (15.2%)	45/100 (14.0%)	
III	121	66/222 (20.5%)	55/100 (17.1%)	
NA	107	107/222 (33.2%)	0/100 (0%)	
Menopausal status ^b				0.9026
Premenopausal	150	99/222 (30.7%)	51/100 (15.8%)	
Postmenopausal	150	101/222 (31.5%)	49/100 (15.2%)	
NA	22	22/222 (6.8%)	0/100 (0%)	
Maximum diameter ^c (mm)	23.68(6-120)	24.21(6-120)	22.64(8-50)	0.2836
Age ^c (y)	52.37(22-84)	52.02 (22-84)	53.09 (31-83)	0.4252

^aData were tested using the chi-square test.

^bData were tested using analysis of variance (ANOVA).

The data in parentheses indicate percentages.

Table 2. Image features

Feature class	Feature description
Morphologic (n=14)	Elongation, Flatness, Least Axis Length, Major Axis Length, Maximum 2D diameter (Column), Maximum 2D diameter (Row), Maximum 2D diameter (Slice), Maximum 3D diameter, Mesh Volume, Minor Axis Length, Sphericity, Surface Area, Surface Area to Volume ratio, Voxel Volume
Histogram distribution (n=18)	10th Percentile, 90th Percentile, Energy, Entropy, Interquartile Range, Kurtosis, Maximum, Mean Absolute Deviation, Mean, Median, Minimum, Range, Robust Mean Absolute Deviation, Root Mean Squared, Skewness, Total Energy, Uniformity, Variance
Texture GLCM (n=24)	Autocorrelation, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Inverse Difference, Inverse Difference Moment, Inverse Difference Moment Normalized, Inverse Difference Normalized, Informational Measure of Correlation1, Informational Measure of Correlation2, Inverse Variance, Joint Average, Joint Energy, Joint Entropy, Maximal Correlation Coefficient, Maximum Probability, Sum Average, Sum Entropy, Sum of Squares
Texture GLRLM (n=16)	Grey Level Non-Uniformity, Grey Level Non-Uniformity Normalized, Grey Level Variance, High Grey Level Run Emphasis, Long Run Emphasis, Long Run High Grey Level Emphasis, Long Run Low Grey Level Emphasis, Low Grey Level Run Emphasis, Run Entropy, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Grey Level Emphasis, Short Run Low Grey Level Emphasis
Texture GLSZM (n=16)	Grey Level Non-Uniformity, Grey Level Non-Uniformity Normalized, Grey Level Variance, High Grey Level Zone Emphasis, Large Area Emphasis, Large Area High Grey Level Emphasis, Large Area Low Grey Level Emphasis, Low Grey Level Zone Emphasis, Size-Zone Non-Uniformity, Size-Zone Non-Uniformity Normalized, Small Area Emphasis, Small Area High Grey Level Emphasis, Small Area Low Grey Level Emphasis, Zone Entropy, Zone Percentage, Zone Variance
Texture GLDM (n=14)	Dependence Entropy, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Dependence Variance, Grey Level Non-Uniformity, Grey Level Variance, High Grey Level Emphasis, Large Dependence Emphasis, Large Dependence High Grey Level Emphasis, Large Dependence Low Grey Level Emphasis, Low Grey Level Emphasis, Small Dependence Emphasis, Small Dependence High Grey Level Emphasis, Small Dependence Low Grey Level Emphasis
Texture NGTDM (n=5)	Busyness, Coarseness, Complexity, Contrast, Strength

GLSZM=Grey-Level Size Zone Matrix; GLRLM=Grey-Level Run-Length Matrix; GLCM=Grey-Level Co-Occurrence Matrix; GLDM=Grey-Level Spatial Dependence Matrix; NGTDM=Neighbourhood Grey Tone Difference Matrix

Table 3. Performance comparison between the EDSR analysis and bicubic method in terms of PSNR and SSIM

Methods	PSNR		SSIM	
	Mean	95% CI	Mean	95% CI
Bicubic	34.334	[34.187,34.481]	0.956	[0.955,0.957]
SRGAN	35.565	[35.439,35.692]	0.964	[0.962,0.965]
EDSR	37.577	[37.420,37.734]	0.974	[0.974,0.975]

PSNR=peak signal-to-noise ratio; SSIM=structural similarity. CI=confidence interval

Peer Review Only

Table 4. Individual feature performance (in terms of AUC values) and the ICC of radiomics from the original ADC image and SR-ADC images based on bicubic, SRGAN and EDSR methods in predicting the tumour grade and Ki-67 expression level

Feature	Original		BICUBIC		SRGAN		EDSR		ICC
	Grade	Ki-67	Grade	Ki-67	Grade	Ki-67	Grade	Ki-67	
Uniformity	0.012	0.111	0.564	0.607	0.632	0.679	0.644	0.631	0.863
Grey Level Variance	0.474	0.015	0.539	0.512	0.648	0.632	0.632	0.490	0.876
Mean Absolute Deviation	0.179	0.324	0.468	0.518	0.618	0.629	0.593	0.484	0.894
Long Run High Grey Level Emphasis	0.301	0.473	0.618	0.690	0.605	0.669	0.690	0.708	0.771
Grey Level Variance-GLDM	0.453	0.210	0.509	0.485	0.648	0.632	0.626	0.467	0.875
10th percentile	0.280	0.625	0.613	0.730	0.642	0.705	0.656	0.637	0.907
Inverse Difference Normalized	0.496	0.212	0.403	0.503	0.658	0.594	0.629	0.577	0.858
Median	0.223	0.634	0.554	0.691	0.516	0.649	0.543	0.657	0.932
Contrast-GLCM	0.531	0.304	0.535	0.515	0.653	0.599	0.611	0.502	0.858
Maximum Probability	0.493	0.413	0.548	0.574	0.619	0.579	0.634	0.643	0.685
Joint Energy	0.535	0.416	0.514	0.582	0.634	0.612	0.641	0.618	0.497
Difference Average	0.557	0.413	0.583	0.558	0.677	0.613	0.641	0.558	0.871
Grey Level Variance-GLSZM	0.573	0.340	0.570	0.469	0.629	0.579	0.669	0.492	0.810
Inverse Difference Moment	0.581	0.502	0.630	0.596	0.709	0.627	0.698	0.635	0.847
Large Dependence									
High Grey Level Emphasis	0.509	0.594	0.583	0.675	0.632	0.659	0.683	0.701	0.860

ICC=intra-class correlation coefficient; GLDM=Grey-Level Spatial Dependence Matrix; GLSZM=Grey-Level Size Zone Matrix; GLCM=Grey-level co-occurrence matrix

Figure legends

Figure 1. Study overview. SR=super-resolution; LR=low resolution; HR=high resolution.

Figure 2. Performance comparison between the EDSR analysis and bicubic method in terms of a) PSNR and b) SSIM.

Figure 3. A case study of the DCE-MRI and ADC images of the tumour using bicubic-, SRGAN- and EDSR-based methods to improve image resolution. a) Tumour area on DCE-MRI. The resolution of the image was improved with the a) original DCE-MRI b) bicubic-, c) SRGAN- and d) EDSR-based methods. e) The original ADC image and the corresponding tumour. The resolution was improved by the f) bicubic-, the g) SRGAN- and h) EDSR-based methods.

Figure 4. Distribution of the ICC values of the statistical, morphologic features and the GLCM, GLRLM, GLSZM, GLDM, NGTDM based texture features.

Figure 5. Distribution of the AUC values using the original ADC image and SR-ADC images based on the bicubic interpolation, SRGAN and EDSR methods for predicting a) Ki-67 and b) tumour grade.

Figure 6. Distribution of the radiomic feature values of the original and SR images based on the EDSR-based image. In the prediction of tumour grade, the a) grey level variance and b) inverse difference features are shown. In the prediction of the Ki-67 expression status, the c) high grey-level zone emphasis and the d) short run high grey-level emphasis features are shown. The pink lines indicate an interaction between the feature values extracted from the original ADC images and those from the EDSR-based SR-ADC image.

Figure 7. ROC curve for the prediction of a) Ki-67 expression and b) tumour grade using the original ADC images and the bicubic-, SRGAN- and EDSR-based resolution enhancement methods.

Figure 1. Study overview. SR=super-resolution; LR=low resolution; HR=high resolution.

175x152mm (300 x 300 DPI)

Figure 2. Performance comparison between the EDSR analysis and bicubic method in terms of a) PSNR and b) SSIM.

254x104mm (300 x 300 DPI)

Figure 3. A case study of the DCE-MRI and ADC images of the tumour using bicubic-, SRGAN- and EDSR-based methods to improve image resolution. a) Tumour area on DCE-MRI. The resolution of the image was improved with the a) original DCE-MRI b) bicubic-, c) SRGAN- and d) EDSR-based methods. e) The original ADC image and the corresponding tumour. The resolution was improved by the f) bicubic-, the g) SRGAN- and h) EDSR-based methods.

285x134mm (300 x 300 DPI)

Figure 4. Distribution of the ICC values of the statistical, morphologic features and the GLCM, GLRLM, GLSZM, GLDM, NGTDM based texture features.

138x83mm (300 x 300 DPI)

250x98mm (300 x 300 DPI)

Figure 6. Distribution of the radiomic feature values of the original and SR images based on the EDSR-based image. In the prediction of tumour grade, the a) grey level variance and b) inverse difference features are shown. In the prediction of the Ki-67 expression status, the c) high grey-level zone emphasis and the d) short run high grey-level emphasis features are shown. The pink lines indicate an interaction between the feature values extracted from the original ADC images and those from the EDSR-based SR-ADC image.

260x185mm (300 x 300 DPI)

Figure 7. ROC curve for the prediction of a) Ki-67 expression and b) tumour grade using the original ADC images and the bicubic-, SRGAN- and EDSR-based resolution enhancement methods.

253x104mm (300 x 300 DPI)

Generative adversarial networks-based super-resolution of diffusion-weighted imaging: application to tumour radiomics in breast cancer

Ming Fan¹, Zuhui Liu¹, Maosheng Xu², Shiwei Wang², Tiejong Zeng³, Xin Gao⁴, Lihua Li^{1*}

¹Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou, China

²Department of Radiology, First Affiliated Hospital of Zhejiang Chinese Medical University, Hangzhou, Zhejiang, China

³Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong, China

⁴King Abdullah University of Science and Technology (KAUST), Computational Bioscience Research Center (CBRC), Computer, Electrical and Mathematical Sciences and Engineering Division (CEMSE), Thuwal, 23955-6900, Saudi Arabia.

***Corresponding Author**

Lihua Li,

Institute of Biomedical Engineering and Instrumentation, Hangzhou Dianzi University, Hangzhou, China, 310018

Phone: +86-571-86878587

Fax: +86-571-87713528

Email: lilh@hdu.edu.cn

Accurate tumour characterization with DWI and the corresponding apparent diffusion coefficient (ADC) maps is challenging due to their limited resolution. This study aimed to produce super-resolution (SR) ADC images from its low-resolution counterparts and to assess the clinical utility of these SR images by performing a radiomic analysis for predicting the histologic grade and Ki-67 expression status of breast cancer. In addition to improving the resolution of ADC images, deep-learning-based SR methods also improved tumour image-based diagnosis in breast cancer.