Texture Analysis of Breast DCE-MRI to Predict Benign and Malignant Tumors and HER-2 Status Using Machine Learning and Deep Learning

Jia You (2292264387@qq.com)
Shengjing Hospital of China Medical University
Bin Zhang
School of Medicine and Bioinformatics Engineering, Northeastern University
Jiandong yin
Shengjing Hospital of China Medical University Department of Radiology

Research Article

Keywords: Breast cancer, Convolutional neural network, DCE-MRI, HER-2, Machine learning

Posted Date: March 10th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1010490/v1

Texture Analysis of Breast DCE-MRI to Predict Benign and Malignant Tumors and HER-2 Status Using Machine Learning and Deep Learning

Author Names and Degrees: Jia You, MS,[1] Bin Zhang, MS,[2] and Jiandong Yin, PhD[1*]

Author Affiliation:

[1]Department of Radiology, Shengjing Hospital of China Medical University

[2]School of Medicine and Bioinformatics Engineering, Northeastern University

Corresponding Author Info: Jiandong Yin; Department of Radiology, Shengjing Hospital of China Medical University, 110004, No. 36, Sanhao Street, Heping District, Shenyang, Liaoning, China; Tel: +86 18940254851; no fax numbers; Email: jiandongyin@sina.com

ABSTRACT

Objectives: The purpose of this study was to evaluate the diagnostic performance of three classification models that utilize logistic regression analysis (LRA), support vector machine (SVM), and convolutional neural network (CNN) in distinguishing breast tumor properties and HER-2 status using quantitative analysis of breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI).

Methods: A total of 354 patients who underwent breast DCE-MRI were included in this retrospective study. All enrolled patients were divided into benign and malignant groups. Patients with breast cancer were divided into positive and negative HER-2 groups. Each group was then divided into training and validation cohorts using different MRI scanners. Machine learning texture analysis was realized using LRA and SVM. Least absolute shrinkage and selection operator (LASSO) were used to filter texture features. Deep learning was implemented using CNN. The classification model performance was evaluated using receiver operating characteristic (ROC) analysis.

Results: When distinguishing breast tumor properties, three classifiers based on LASSO_LRA, LASSO_SVM, and CNN in the training cohorts achieved the areas under ROC curves (AUCs) of 0.937, 0.946, and 0.995 (accuracy = 88.75%, 89.37%, and 95.63%), respectively. Three classifiers in the corresponding validation cohorts achieved AUCs of 0.896, 0.914, and 0.965 (accuracy = 83.30%, 85.56%, and 91.11%), respectively. When identifying the HER-2 status, three classifiers in the training cohorts reached AUCs of 0.907, 0.914, and 0.954 (accuracy = 82.61%, 86.09%, and 92.17%), respectively. Three classifiers in the corresponding validation cohorts reached AUCs of 0.819, 0.869, and 0.923 (accuracy = 81.43%, 84.29%, and 87.14%), respectively.

Conclusion: The performance of the deep learning approach with CNN is superior to that of machine learning approaches employing LRA and SVM. These results suggest that the classifier based on CNN can be used as an efficient tool for preoperative evaluation of breast tumor properties and HER-2 status.

Keywords: Breast cancer; Convolutional neural network; DCE-MRI; HER-2; Machine learning.

Abbreviations

LRA: Logistic regression analysis; SVM: Support vector machine; CNN: Convolutional neural network; HER-2: Human epidermal growth factor receptor 2; DCE-MRI: Dynamic contrast-enhanced magnetic resonance imaging; LASSO: Least absolute shrinkage and selection operator; ROC: Receiver operating characteristic; AUC: Area under the ROC curve; IHC: Immunochemistry; FISH: Fluorescence in situ hybridization; TA: Texture analysis; TIC: Time-intensity curve; ROI: Region of interest; HIS: First-order histogram; GLCM: Gray-level co-occurrence matrix; GRLM: Gray-level run-length matrix; BN: Batch normalization; ReLU: Rectified Linear Units; LOOCV: Leave-one-out cross-validation; ICC: Intraclass correlation coefficient.

Introduction

Breast cancer is the most frequently diagnosed cancer among women worldwide. It remains the second leading cause of death in women[1-2]. Early diagnosis, personalized treatment selection, and disease management are increasingly needed to better evaluate breast cancer prognosis[3]. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is one of the best imaging techniques that provides temporal information about contrast agent kinetics in suspicious lesions along with acceptable spatial resolution[4]. It has been shown to have great potential in breast tumor assessment[5].

Human epidermal growth factor receptor 2 (HER-2) amplification or overexpression was first discovered in 1987 and demonstrated in about 20–30% of human breast cancers[6]. Two approved methods for detecting the amplification and overexpression status of HER-2 in breast cancer are immunochemistry (IHC) and fluorescence in situ hybridization (FISH)[7]. Currently, HER-2 status has been used to guide breast cancer treatment and to assess prognosis[8]. However, due to the high tumor heterogeneity and the relatively small size of tissue specimens, it may not be possible to accurately assess HER-2 status in the entire tumor based on needle biopsy specimens[9].

Texture analysis (TA) is a part of radiomics analysis that serves as a tool for high-throughput extraction and evaluation of quantitative features from images[10-11]. Over the course of the last 5–10 years, TA has been increasingly become a tool for maximizing the amount of information extracted from virtually any medical imaging modality[12-14]. Least absolute shrinkage and selection operator (LASSO)[15], logistic regression analysis (LRA)[16], and support vector machine (SVM)[17] have become very popular machine learning tools for analyzing complex data that often arise in radiological research. LASSO is a novel feature selection method that can use a penalty function to filter data with a large number of covariates[18-19]. SVM and LRA are common tools for data classification and aim to create a decision boundary between two classes that allows the prediction of labels from one or more feature vectors[20-22].

Recently, deep learning has emerged as a powerful technology for defining network architecture that connects multiple neural-like processing layers with multiple abstraction layers[23-24]. A convolutional neural network (CNN) is a supervised deep learning model that incorporates concatenated convolutional layers, pooling layers, and fully connected layers[25]. After a series of data processing steps, deep features can be learned from the original data. Each point in the obtained depth feature map corresponds to a region in the original image. At present, many studies

have applied the CNN model to the field of medical image analysis[26-27].
Nevertheless, little is known about the differences in performance between machine learning and deep learning in the differentiation of benign and malignant breast tissue. There have also been few comparative studies on the differentiation of HER-2 gene status between the two methods. Therefore, the aim of this work was to evaluate the classification performance using a machine learning method that utilized SVM and LRA and deep learning method that was implemented using CNN in distinguishing breast tumor properties and HER-2 status.

Materials and methods

This study was approved by the Ethics Review Board of Shengjing Hospital of China Medical University (2020PS011K). Requirements for written informed consent were waived because of the retrospective nature of the study.

Patient Cohort

A total of 449 female patients with breast disease were evaluated for inclusion in the study between January 2018 and December 2020. Inclusion criteria were as follows: 1) patient underwent a DCE-MRI examination within one month before surgery or biopsy; 2) no anti-cancer treatment, such as radiotherapy and chemotherapy, was administered before DCE-MRI; and 3) benign or malignant primary breast disease confirmed by surgery or pathological biopsy. Exclusion criteria included the following: 1) patients with breast cancer without HER-2 status (n = 24); 2) patients with incomplete clinical data (n = 39); and 3) poor image quality due to artifacts (n = 32). The expression of HER-2 gene was considered according to the result of IHC and FISH[7]. The recruitment pathway for patients in this study is shown in Figure 1. A total of 354 patients were included in the study, of which 113 were benign and 241 (122 were HER-2-positive, 119 were negative) were malignant cases. All patients were divided into training (n = 239) and validation (n = 115) cohorts using different MRI scanners. Patients' clinicopathologic characteristics are shown in Table 1.

TABLE 1. Clinicopathologic characteristics of patients undergoing DCE-MRI examinations

Characteristic	Number (%)	
	Training Cohort (n = 239)	**Validation Cohort (n = 115)**
Age (year)[a]	50.54 ± 10.674	52.02 ± 9.769
Pathology		
Benign	68	45
Adenosis	40 (16.7)	22 (19.1)
Fibroadenoma	23 (9.62)	17 (14.8)
Papilloma	5 (2.09)	6 (5.22)

Malignant	171	70
Tis	11 (4.60)	6 (5.22)
Invasive ductal carcinoma	145 (60.7)	53 (46.1)
Invasive lobular carcinoma	7 (2.93)	5 (4.35)
Invasive papillary carcinoma	2 (0.837)	2 (1.74)
Mucinous carcinoma	6 (2.51)	4 (3.48)
Molecular subtypes		
HER-2		
Positive	89 (73.0)	33 (27.0)
Negative	82 (68.9)	37 (31.1)
Ki-67		
≥ 14%	120 (72.3)	46 (27.7)
＜ 14%	51 (68.0)	24 (32.0)
Lymph metastasis		
N_0	109 (72.2)	42 (27.8)
N_+	62 (68.9)	28 (31.1)

Tis: carcinoma in situ; HER-2: human epidermal growth factor receptor 2.

[a]Values are mean±standard deviation.

MR Image Acquisition

Two different 3.0T MRI scanners were utilized, including a GE 3.0T MR scanner (Signa HDxt, GE Healthcare, USA) with a dedicated eight-channel bilateral breast coil and a Philips 3.0 T MR scanner (Ingenia, Philips Medical System, Best, Netherlands) equipped with a dedicated seven-channel bilateral breast coil. Two groups of patients underwent breast DCE-MRI examination and were enrolled in training and validation cohorts, respectively. Each patient was placed on the scanning bed in a prone position with established venous access. The orientation of slice images was transverse. The VIBRANT-VX sequence T1-weighted 3D images were initially acquired as mask images. Eight phases were selected for the MRI scan after an intravenous injection of the contrast agent (0.5 mmol/mL, gadodiamide, Magnevist, Bayer Healthcare Pharmaceuticals, Berlin, Germany) at 4 mL/s (0.15 mmol/kg body weight) and an equal volume of saline flush at the same injection speed. The corresponding acquisition parameters were as follows: repetition time = 7.42 ms, echo time = 4.25 ms, flip angle = 15°, slice thickness = 2.20 mm, spacing between slices = 2.20 mm, matrix = 1024 × 1024; repetition time = 4.14 ms, echo time = 2.10 ms, flip angle = 12°, slice thickness = 2.00 mm, spacing between slices = 1.00 mm, and matrix = 380 × 380. Eight subtraction sequences were obtained by subtracting the pre-contrast scan from each of the eight post-contrast scans.

Image Section

Two radiologists with more than 10 years of breast MRI diagnosis experience pointed out the

location and slice range of the tumor by referring to clinical, radiological, and pathological reports. The image was downloaded corresponding to the display layer of the largest tumor lesion[28]. The image corresponding to the phase with the highest lesion signal intensity displayed on the time-intensity curve (TIC) was selected[29].

Machine Learning

Lesion Segmentation

The regions of interest (ROIs) were semi-automatically segmented on the selected DCE-MRI images according to lesion locations provided by radiologists using Maximum Between-Class Variance Method (Otus Method). This operation was realized using MATLAB R2018b (Mathworks, Natick, MA, United States). The effect of lesion segmentation is shown in Figure 2.

Texture Extraction and Selection

MATLAB R2018b was used to calculate ROI texture parameters. The extracted features were classified into four categories[30]: 1) first-order histogram (HIS)-based (n = 4), 2) gray-level co-occurrence matrix (GLCM)-based (n = 19), 3) gray-level run-length matrix (GRLM)-based (n = 11), and 4) wavelet transform-based features (n = 21). A total of 55 features were derived from each image(Table 2).

TABLE 2 . Texture features extracted from machine learning DCE-MRI

Methods	Texture Features	Number
HIS	Mean, variance, skewness, kurtosis	4
GLCM[a]	Autocorrelation (ACOR), contrast (CON), correlation (COR), cluster prominence (CP), cluster shade (CS), dissimilarity (DIS), energy (ENE), entropy (ENT), homogeneity (HOM), maximum probability (MP), sum of squares (SOS), sum average (SA), sum variance (SV), sum entropy (SE), difference variance (DV), difference entropy (DE), information measure of correlation (IMC), inverse difference normalized (IDN), inverse difference moment normalized (IDMN),	19
GRLM[b]	run-length non-uniformity (RLN), gray level non-uniformity (GLN), long run emphasis (LRE), short run emphasis (SRE), fraction of image in runs (FIR), low gray-level run emphasis (LGRE), high gray-level run emphasis (HGRE), short run low gray-level emphasis (SRLGE), short run high gray-level emphasis (SRHGE), long run low gray-level emphasis (LRGE), long run high gray-level emphasis (LRGE)	11
Wavelet transform[c]	Harr parameters	7
	Deubechies2 parameters	7
	Symlet4 parameters	7
Total		55

HIS: histogram; GLCM: gray level co-occurrence matrix; GRLM: gray level run-length matrix.

[a]GLCM parameters were calculated for four distances (1, 2, 3, and 4 pixels) and four angles (0, 45, 90, and 135°). (d, 0), (d, d), (0, d), and (-d, -d) represent 0, 45, 90, and 135°, respectively, where d stands for distance. For example, CON (1, 1) represents the contrast feature calculated for the distance of 1 and the angle of 45°.

[b]GRLM parameters were calculated for four angles (0, 45, 90, and 135°).

[c]Wavelet transform parameters were calculated for four layers and three directions (horizontal, vertical, and diagonal) to produce low- and high-frequency components. For example, Haar HD_1 represents a diagonal high-frequency component of the first layer using the Haar wavelet.

Deep Learning

Deep learning was implemented using a CNN model[23]. Batch normalization (BN) was first used to preprocess images. The CNN model had three convolutional layers. A convolutional layer contained two convolution operations using a 3×3 convolution kernel[31]. The Rectified Linear Units (ReLU) activation function was used for nonlinear transformation after each operation. A max pooling layer with a step size of 2 was used to realize global feature extraction and image size reduction. The two fully connected layers appeared at the end of the network. The first fully connected layer compressed the feature channels into a size of 1×1 with a total of 120 features. The second contained a one-feature channel, which compressed all features into a single feature. The output layer of the sigmoid activation function outputted the prediction results. The structure of the CNN model is shown in Figure 3.

The images were batch-normalized and converted into a size of 128 × 128 × 1 before being input to the CNN. This does not only accelerate the convergence speed of the network and improves its performance, but also makes the extracted feature distribution more in line with the true data distribution, ensuring the nonlinear model expression.

Two activation functions were used: ReLU and Sigmoid functions (Figure 4). The ReLU function is commonly used in image processing and is defined as:

$$R\mathrm{eLU}(x)=\begin{cases} x & if x>0 \\ 0 & if x\le 0 \end{cases} \tag{1}$$

If the input value of the function is ≤0, the output result is 0. If the input value is >0, the result is equal to itself.

The Sigmoid function, also known as the logistic function, is more commonly used in binary classification problems. This function is defined as:

$$\delta(x)=\frac{1}{1+\mathrm{e}^{-\mathrm{x}}} \tag{2}$$

It converts the continuous real input value into a value between 0 and 1. In particular, if the input value is a large negative value, the output is 0. If the input value is a large positive value, the result is 1.

Statistical Analysis

LASSO was applied to select features. LRA and SVM were used to establish classifiers. To avoid classifier overfitting, a leave-one-out cross-validation method was utilized[29]. The statistical analysis was performed using SPSS 19.0 (IBM, Armonk, NY, United States). Independent sample t-test and Mann-Whitney U test were performed for statistical analysis. Spearman's correlation analysis was employed to assess the correlation between features and tumor nature or HER-2 status.

The performance of the three classification models was evaluated using receiver operating characteristic (ROC) analysis on Medcalc (version 14.10.20, http://www.medcalc.org/) by measuring the area under the ROC curves (AUCs). Accuracy, sensitivity, and specificity were also calculated and compared.

A random set of images was used to extract features based on the locations provided by two radiologists. The texture features were compared and the inter-observer variability was evaluated using the intraclass correlation coefficients (ICCs)(0–0.4, poor agreement; 0.41–0.6, moderate agreement; 0.61–0.8, good agreement; and 0.81–1, excellent agreement)

Results

Benign and Malignant Groups

In the training cohort, there were 171 (71.5%) were malignant and 68 (28.5%) were benign. There was no significant difference between benign and malignant lesions in terms of age (p = 0.558). Table 3 shows the results of nine features that were extracted from the images processed by LASSO. Mean, kurtosis, RLN (1, 0), Deubechies2_2_HH, Deubechies2_2VH, Symlet4_2HH, Symlet4_2VH, and Deubechies2_3DH were significantly different. Among them, RLN (1, 0) had the greatest correlation with tumor properties.

TABLE 3. Features extracted from images of benign and malignant lesions

Features	Benign	Malignant	AUCs	R_S	p value
Mean[a]	129.397±32.342	157.870±39.012	0.718	0.339	<0.001
kurtosis[a]	2.379±0.625	2.682±0.419	0.660	0.250	0.004
RLN (1, 0)[b]	434.411±194.899	295.742±52.344	0.808	-0.480	<0.001
Deubechies2_2_HH[b]	0.996±0.540	1.493±0.501	0.707	0.323	0.002
Deubechies2_2VH[b]	1.073±0.635	1.760±0.547	0.741	0.375	0.002
Deubechies2_1DH[b]	0.146±0.103	0.150±0.0381	0.506	0.010	0.725
Symlet4_2HH[b]	0.801±0.233	1.302±0.609	0.758	0.402	<0.001
Symlet4_2VH[b]	0.870±0.403	1.574±0.676	0.777	0.431	<0.001
Deubechies2_3DH[b]	0.506±0.290	0.856±0.300	0.727	0.354	<0.001

[a]mean ± standard deviation; [b]median ± range

RLN: run-length non-uniformity.

HH: horizontal high-frequency; VH: vertical high-frequency; DH: diagonal high-frequency.

The ROC curves are shown in Figure 5. The AUC, sensitivity (sen), specificity (spe) and accuracy (acc) were obtained (Table 4). In the training cohorts, AUC = 0.937 for LRA (Acc = 88.75%), AUC = 0.946 for SVM (Acc = 89.37%), and AUC = 0.995 for CNN (Acc = 95.63%). There was no statistical difference between LRA and SVM (p = 0.359). ROC analysis indicated a statistical difference between LRA and CNN (p = 0.012) and SVM and CNN (p = 0.014). In the validation cohorts, AUC = 0.896 for LRA (Acc = 83.30%), AUC = 0.914 for SVM (Acc = 85.56%), and AUC = 0.965 for CNN (Acc = 91.11%). There was no statistical difference between LRA and SVM (p = 0.462). There was a statistical difference between LRA and CNN (p = 0.047) and SVM and CNN (p = 0.032).

TABLE 4. Performance of three classifiers when distinguishing benign and malignant lesions

Methods	Cohorts	AUCs	95% CI	Sensitivity (%)	Specificity (%)	Accuracy (%)
LASSO_LRA	training	0.937	0.887 - 0.969	79.13	95.56	88.75
	validation	0.896	0.813 - 0.950	81.82	88.57	83.30
LASSO_SVM	training	0.946	0.898 - 0.975	88.70	97.78	89.37
	validation	0.914	0.836 - 0.963	83.67	91.43	85.56
CNN	training	0.995	0.968 - 0.981	94.78	98.34	95.63
	validation	0.965	0.903 - 0.992	96.33	80.00	91.11

LASSO: least absolute shrinkage and selection operator. LRA: logistic regression analysis;
SVM: support vector machine; CNN: convolutional neural network; AUC: area under curve.

Positive and Negative HER-2 Status Groups

In the training cohorts, there were 89 (52.05%) patients with positive HER-2 status and 82 (47.95%) with negative. The results in Table 5 show no significant associations in age (p = 0.784), progesterone receptor expression (p = 0.569), Ki-67 (p = 0.171), and pathology types (p = 0.446) between the positive and negative HER-2 status groups. Twelve features were extracted from the images processed by LASSO (Table 6). Mean, variance, GLN (1, 0), and Haar_4VH were significantly different. Mean had the strongest correlation w
ith HER-2 status.

TABLE 5. Characteristics with HER-2 status breast cancer

Characteristics	Training Cohorts		Validation Cohorts		p value
	Positive (n = 89)(%)	Negative (n = 82)(%)	Positive (n = 37)(%)	Negative (n = 33)(%)	
Age (year)[ab]					0.784
> 50 years	52 (61.9)	32 (38.1)	21 (53.8)	18 (46.2)	
≤ 50 years	37 (42.5)	50 (57.5)	16 (51.6)	15 (48.4)	

Progesterone receptor[b]					0.569
Positive	71 (51.4)	67 (48.6)	27 (65.9)	14 (34.1)	
Negative	18 (54.5)	15 (45.5)	10 (52.6)	9 (47.4)	
Ki-67[b]					0.171
≥ 14%	67 (55.8)	53 (44.2)	28 (60.9)	18 (39.1)	
< 14%	22 (43.1)	29 (56.9)	9 (37.5)	15 (62.5)	
Pathology[c]					0.446
Tis	5 (45.5)	6 (54.5)	2 (33.3)	4 (66.7)	
Invasive ductal carcinoma	82 (56.6)	63 (43.4)	34 (64.2)	19 (35.8)	
Invasive lobular carcinoma	2 (28.6)	5 (71.4)	3 (60.0)	2 (40.0)	
Invasive papillary carcinoma	1 (50.0)	1 (50.0)	2 (100.0)	0 (0.00)	
Mucinous carcinoma	2 (33.3)	4 (66.7)	1 (25.0)	3 (75.0)	

Values are no. (%); Tis: carcinoma in situ.

[a]*p* value is for positive-HER-2 vs. negative-HER-2 comparison.

[b]Variables are tested using x^2 test.

[c]Variables are tested using Fisher's exact test.

TABLE 6. Features extracted from images with HER-2 status

Features	Negative	Positive	AUCs	*Rs*	*p* value
Mean[a]	126±37.5	175±27.8	0.846	0.571	<0.001
variance[a]	32.4±8.63	41.6±7.02	0.824	0.535	<0.001
GLN (1, 0)[b]	793.30± 275.910	521.983±80.898	0.701	-0.331	<0.001
RLN (1, 0)[a]	323±107	299±65.8	0.535	-0.058	0.205
Haar_2HH[b]	1.929±0.518	2.483±0.705	0.610	0.181	0.053
Deubechies2_2HH[b]	1.470±0.403	1.52±0.368	0.536	0.060	0.522
Haar_4VH[b]	6.599±2.096	8.401±2.492	0.641	0.233	0.013
Deubechies2_3HH[b]	2.377±0.745	2.681±0.638	0.607	0.177	0.059
Deubechies2_3VH[b]	2.759±0.593	3.349±1.009	0.575	0.124	0.184
Deubechies2_4DH[b]	1.347±0.511	1.471±0.400	0.571	0.117	0.211
Symlet4_3VH[b]	2.605±0.937	2.663±0.467	0.560	0.100	0.288
Symlet4_4DH[b]	1.630±0.638	1.516±0.411	0.529	-0.048	0.609

[a]mean ± standard deviation; [b]median ± range

GLN: gray level non-uniformity; RLN: run-length non-uniformity.

HH: horizontal high-frequency; VH: vertical high-frequency; DH: diagonal high-frequency.

The ROC curves are shown in Figure 6. The results for AUC, Sen, Spe and Acc are shown in Table 7. In the training cohorts, AUC = 0.907 for LRA (Acc = 82.61%), AUC = 0.914 for SVM (Acc = 86.09%), and AUC = 0.954 for CNN (Acc = 92.17%). There was no statistical difference between LRA and SVM (p = 0.171). Statistical difference was observed between LRA and CNN

(p = 0.025) and SVM and CNN (p = 0.013). In validation cohorts, AUC = 0.819 for LRA (Acc = 81.43%), AUC = 0.869 for SVM (Acc = 84.29%), and AUC = 0.923 for CNN (Acc = 87.14%). There was no statistical difference between LRA and SVM (p = 0.096). Statistical difference was observed between LRA and CNN (p = 0.043) and SVM and CNN (p = 0.034).

TABLE 7. Performance of three classifiers when distinguishing HER-2 status

Methods	Cohorts	AUCs	95% CI	Sensitivity (%)	Specificity (%)	Accuracy (%)
LSAAO_LRA	training	0.907	0.839 - 0.953	82.67	82.50	82.61
	validation	0.819	0.709 - 0.901	85.00	76.67	81.43
LASSO_SVM	training	0.914	0.847 - 0.958	78.67	95.00	86.09
	validation	0.869	0.767 - 0.938	79.38	86.67	84.29
CNN	training	0.954	0.898 - 0.984	90.67	97.50	92.17
	validation	0.923	0.833 - 0.973	80.00	96.67	87.14

LASSO: least absolute shrinkage and selection operator. LRA: logistic regression analysis;
SVM: support vector machine; CNN: convolutional neural network; AUC: area under curve.

Interobserver Agreement Evaluation

Features derived from the ROIs segmented separately by two radiologists showed excellent agreement, while the ICCs ranged from 0.817 to 0.926.

Discussion

The present study evaluated the classification performance of texture analysis of DCE-MRI when distinguishing between benign and malignant breast tissue, as well as HER-2 status. Three classifiers were involved, two of which were established by LASSO_SVM and LASSO_LRA, while the remaining one was established using a CNN model. The results showed that the three classification models had good performance in evaluating tumor properties and HER-2 status. The CNN model had the highest diagnostic accuracy. Therefore, the classifier based on the CNN model can better serve as a complementary aid in the radiological evaluation of breast cancer and HER-2 status.

At present, awareness and attention to breast cancer are increasing because it has the second most fatal rate among women[1-2]. With the development of modern examination techniques, patients have a higher expectation for non-invasive diagnosis of breast tumors. In addition, serves as a breast tumor biomarker, the overexpression and/or amplification of the HER-2 receptor has been shown to be associated with more aggressive tumor phenotypes, malignant cell transformation, and poor clinical outcome[32]. Therefore, early diagnosis of breast cancer and HER-2 status are crucial for guiding therapy.

Huang et al.[33] and Vidić et al.[34] indicated that binary LRA and SVM classifiers showed good

performance in distinguishing benign and malignant breast tumors, respectively. The results of the 3D texture feature analysis from Huang et al.[33] (Acc = 88.42%, AUC = 0.8926) were similar to those for the present study, showing that features from the largest cross-sectional tumor views had a similar diagnostic power compared to those from the whole lesions. In addition, our results showed that RLN (1, 0) had the strongest correlation with tumor malignancy and can be used as an independent predictor of breast cancer.

Chang et al.[35] and Agner et al.[36] demonstrated that breast MRI texture features classified using LRA and SVM can be used to predict HER-2 status (AUC = 0.8458, AUC = 0.74). In addition, our results showed that the mean had the strongest correlation with HER-2 status and can be used as an independent predictor of HER-2 positivity.

For each LASSO_LRA and LASSO_SVM model, the AUC for a single feature was lower than that for combined features, which was similar to those by Vidić et al.[34]. The joint application of multiple parameters may improve the accuracy of texture analysis.

The basic premise behind the use of deep learning is that while human experts are often good at visual recognition of image characteristics, crafting image metrics for recognition tasks is difficult. Meanwhile, a recent study has shown that deep learning with CNN is a powerful technique for medical imaging classification[37].

Havaei et al.[31] have presented an automatic tumor segmentation method and have achieved a high performance level by stacking two CNNs. Zhou et al.[28] found that deep learning achieved a higher accuracy rate compared to ROI-based radiology when evaluating the accuracy of mass lesions detected on breast DCE-MRI. Sentana et al.[38] indicated that CNN had a higher accuracy than SVM on the detection of lung nodules. Nyflot et al.[37] have used CNN and four machine learning classifiers to evaluate quality of images from patients who underwent intensity-modulated radiotherapy. And the performance of deep learning networks showed the best potential.

Compared to the above researches, two commonly used machine learning models, LASSO_LRA and LASSO_SVM, and deep learning with CNN have been comprehensively studied on breast DCE-MRI images in the present work. In addition to analyzing their performance in diagnosing benign and malignant breast tumors, the predictive abilities of HER-2 status were also evaluated.

There are two possible reasons for the best performance of the classifiers based on CNN. The first one is that the CNN is a deep learning model used in algorithms to identify, extract, and quantify analysis features without human intervention. Using the repeated CNN model training, some small differences that were invisible to the naked eye were extracted. The second reason might be the slight differences in the intensity of the blood vessel signal. Although, in order to minimize the error, image selection was performed according to the peak signal intensity of TIC, there might also be differences in the maximum concentration of contrast media in different patients. In contrast, BN performed before image convolution in a CNN model might reduce the error caused by individual factors.

Limitations

This study has several limitations. First, it was a small-sample study. Second, one deep learning model and two machine learning methods were utilized. Third, differential diagnosis was performed on tumor properties, while analysis of breast cancer with different data system

(BI-RADS) scores was not carried out. Fourth, among the common molecular subtypes of breast cancer patients, only the study of HER-2 gene was performed. In addition, CNN lacks a feature interpretation ability, which is a common problem of a deep learning model[37]. Therefore, it is necessary to explore the features' meaning of CNN before applied in a clinical setting.

In the future, the sample size needs to be increased and the performance of classifications based on other deep learning models needs to be investigated. The predictive ability of deep learning models in different BI-RADS classifications and expression intensity of ER, PR, and Ki-67 in breast diseases should be explored.

Conclusion

Three classifiers based on LRA, SVM and CNN showed good performance in classifying the benign and malignant lesions and HER-2 status. The results suggest that a classifier based on CNN can be an efficient tool for the preoperative evaluation of breast tumor property and HER-2 status.

Declarations

Ethics approval and consent to participate The protocol was approved by Shengjing Hospital of China Medical University Institutional Review Board. Written informed consent was waived by the Institutional Review Board because of the retrospective of this study.

Consent for publication Not applicable.

Availability of data and materials The data that support the findings of this study are available from Shengjing Hospital of China Medical University, but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Shengjing Hospital of China Medical University.

Competing interests The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

Funding This study has received funding from Research and Development Foundation for Major Science and Technology from Shenyang (No. 19-112-4-105), Natural Fund Guidance Plan from Liaoning (No. 2019-ZD-0743) and Big Data Foundation for Health Care from China Medical University (No. HMB201902105).

Authors' contributions J You and B Zhang contributed to the study design. B Zhang completed the image processing and analysis. B Zhang and J You completed the data processing. J You conducted the manuscript writing. JD Yin made the final revision and polishing of the article. All authors read and approved the final manuscript.

Acknowledgements We thank Shengjing Hospital of China Medical University for the data support. We thank Northeastern University for the technical support.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. CA Cancer J Clin. 2019;69(1):7-34.

doi:10.3322/caac.21551
2. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Akinyemiju TF, et al. Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2016: A Systematic Analysis for the Global Burden of Disease Study. JAMA Oncol. 2018;4(11):1553-1568. doi:10.1001/jamaoncol.2018.2706
3. Sun X, He B, Luo X, et al. Preliminary Study on Molecular Subtypes of Breast Cancer Based on Magnetic Resonance Imaging Texture Analysis. J Comput Assist Tomogr. 2018;42(4):531-535. doi:10.1097/RCT.0000000000000738
4. Li W, Yu K, Feng C, Zhao D. Molecular Subtypes Recognition of Breast Cancer in Dynamic Contrast-Enhanced Breast Magnetic Resonance Imaging Phenotypes from Radiomics Data. Comput Math Methods Med. 2019;2019:6978650. doi:10.1155/2019/6978650
5. Fan M, Wu G, Cheng H, et al. Radiomic Analysis of DCE-MRI for Prediction of Response to Neoadjuvant Chemotherapy in Breast Cancer Patients. Eur J Radiol. 2017;94:140-147. doi:10.1016/j.ejrad.2017.06.019
6. Burstein HJ. The Distinctive Nature of HER2-positive Breast Cancers. N Engl J Med. 2005;353(16):1652-1654. doi:10.1056/NEJMp058197
7. Wolff AC, Hammond ME, Hicks DG, et al. Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. J Clin Oncol. 2013;31(31):3997-4013. doi:10.1200/JCO.2013.50.9984
8. Fiordelisi MF, Auletta L, Meomartino L, et al. Preclinical Molecular Imaging for Precision Medicine in Breast Cancer Mouse Models. Contrast Media Mol Imaging. 2019;2019:8946729. doi:10.1155/2019/8946729
9. Zhou J, Tan H, Bai Y, et al. Evaluating the HER-2 Status of Breast Cancer Using Mammography Radiomics Features. Eur J Radiol. 2019;121:108718. doi:10.1016/j.ejrad.2019.108718
10. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: Extracting More Information from Medical Images Using Advanced Feature Analysis. Eur J Cancer. 2012;48(4):441-446. doi:10.1016/j.ejca.2011.11.036
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-577. doi:10.1148/radiol.2015151169
12. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in Breast Cancer Classification and Prediction. Semin Cancer Biol. 2021;72:238-250. doi:10.1016/j.semcancer.2020.04.002
13. He B, Ji T, Zhang H, et al. MRI-based Radiomics Signature for Tumor Grading of Rectal Carcinoma Using Random Forest Model. J Cell Physiol. 2019;234(11):20501-20509. doi:10.1002/jcp.28650
14. Zhang Y, Zhang B, Liang F, et al. Radiomics Features on Non-contrast-enhanced CT Scan Can Precisely Classify AVM-related Hematomas From Other Spontaneous Intraparenchymal Hematoma Types. Eur Radiol. 2019;29(4):2157-2165. doi:10.1007/s00330-018-5747-x
15. Lu H, Yin J. Texture Analysis of Breast DCE-MRI Based on Intratumoral Subregions for Predicting HER2 2+ Status. Front Oncol. 2020;10:543. doi:10.3389/fonc.2020.00543
16. Song L, Yin J. Application of Texture Analysis Based on Sagittal Fat-Suppression and Oblique

Axial T2-Weighted Magnetic Resonance Imaging to Identify Lymph Node Invasion Status of Rectal Cancer. Front Oncol. 2020;10:1364. doi:10.3389/fonc.2020.01364
17. Qian Z, Li Y, Wang Y, et al. Differentiation of Glioblastoma from Solitary Brain Metastases Using Radiomic Machine-learning Classifiers. Cancer Lett. 2019;451:128-135. doi:10.1016/j.canlet.2019.02.054
18. McEligot AJ, Poynor V, Sharma R, Panangadan A. Logistic LASSO Regression for Dietary Intakes and Breast Cancer. Nutrients. 2020;12(9):2652. doi:10.3390/nu12092652
19. Lee S, Kwon S, Kim Y. A Modified Local Quadratic Approximation Algorithm for Penalized Optimization Problems[J]. Computational Statistics and Data Analysis. 2016;94:275-286. doi:10.1016/j.csda.2015.08.019
20. Reynolds E, Callaghan B, Banerjee M. SVM-CART for Disease Classification. J Appl Stat. 2019;46(16):2987-3007. doi:10.1080/02664763.2019.1625876
21. Yin JD, Song LR, Lu HC, Zheng X. Prediction of Different Stages of Rectal Cancer: Texture Analysis Based on Diffusion-weighted Images and Apparent Diffusion Coefficient Maps. World J Gastroenterol. 2020;26(17):2082-2096. doi:10.3748/wjg.v26.i17.2082
22. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W. Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. Cancer Genomics Proteomics. 2018;15(1):41-51. doi:10.21873/cgp.20063
23. LeCun Y, Bengio Y, Hinton G. Deep Learning. Nature. 2015;521(7553):436-444. doi:10.1038/nature14539
24. Lee H, Kwon H. Going Deeper With Contextual CNN for Hyperspectral Image Classification. IEEE Trans Image Process. 2017;26(10):4843-4855. doi:10.1109/TIP.2017.2725580
25. Szegedy C, Liu W, Jia Y, et al. Going Deeper with Convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015:1-9. doi: 10.1109/CVPR.2015.7298594.
26. Setio AA, Ciompi F, Litjens G, et al. Pulmonary Nodule Detection in CT Images: False Positive Reduction Using Multi-View Convolutional Networks. IEEE Trans Med Imaging. 2016;35(5):1160-1169. doi:10.1109/TMI.2016.2536809
27. Thaha MM, Kumar KPM, Murugan BS, Dhanasekeran S, Vijayakarthick P, Selvi AS. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. J Med Syst. 2019;43(9):294. doi:10.1007/s10916-019-1416-0
28. Zhou J, Zhang Y, Chang KT, et al. Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning With Consideration of Peritumor Tissue. J Magn Reson Imaging. 2020;51(3):798-809. doi:10.1002/jmri.26981
29. Li H, Sun H, Liu S, et al. Assessing the Performance of Benign and Malignant Breast Lesion Classification with Bilateral TIC Differentiation and Other Effective Features in DCE-MRI. J Magn Reson Imaging. 2019;50(2):465-473. doi:10.1002/jmri.26646
30. Rizzo S, Botta F, Raimondi S, et al. Radiomics: The Facts and The Challenges of Image Analysis. Eur Radiol Exp. 2018;2(1):36. doi:10.1186/s41747-018-0068-z
31. Havaei M, Davy A, Warde-Farley D, et al. Brain Tumor Segmentation with Deep Neural Networks. Med Image Anal. 2017;35:18-31. doi:10.1016/j.media.2016.05.004
32. Tai W, Mahato R, Cheng K. The Role of HER2 in Cancer Therapy and Targeted Drug Delivery. J Control Release. 2010;146(3):264-275. doi:10.1016/j.jconrel.2010.04.009
33. Huang YH, Chang YC, Huang CS, Wu TJ, Chen JH, Chang RF. Computer-aided Diagnosis of Mass-like Lesion in Breast MRI: Differential Analysis of The 3-D Morphology Between Benign

and Malignant Tumors. Comput Methods Programs Biomed. 2013;112(3):508-517. doi:10.1016/j.cmpb.2013.08.016

34. Vidić 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. 2018;47(5):1205-1216. doi:10.1002/jmri.25873

35. Chang RF, Chen HH, Chang YC, Huang CS, Chen JH, Lo CM. Quantification of Breast Tumor Heterogeneity for ER Status, HER2 Status, and TN Molecular Subtype Evaluation on DCE-MRI. Magn Reson Imaging. 2016;34(6):809-819. doi:10.1016/j.mri.2016.03.001

36. Agner SC, Rosen MA, Englander S, et al. Computerized Image Analysis for Identifying Triple-negative Breast Cancers and Differentiating Them from Other Molecular Subtypes of Breast Cancer on Dynamic Contrast-enhanced MR Images: a Feasibility Study. Radiology. 2014;272(1):91-99. doi:10.1148/radiol.14121031

37. Nyflot MJ, Thammasorn P, Wootton LS, Ford EC, Chaovalitwongse WA. Deep Learning for Patient-specific Quality Assurance: Identifying Errors in Radiotherapy Delivery by Radiomic Analysis of Gamma Images with Convolutional Neural Networks. Med Phys. 2019;46(2):456-464. doi:10.1002/mp.13338

38. Sentana IWB, Asri SA, Jawas N, Wardani AE. CNN and SVM Based Classifier Comparation to Detect Lung Nodule In Computed Tomography Images. Paper presented at Proceedings of the International Conference on Science and Technology (ICST 2018) Bali, Indonesia. 2018. doi:10.2991/icst-18.2018.7

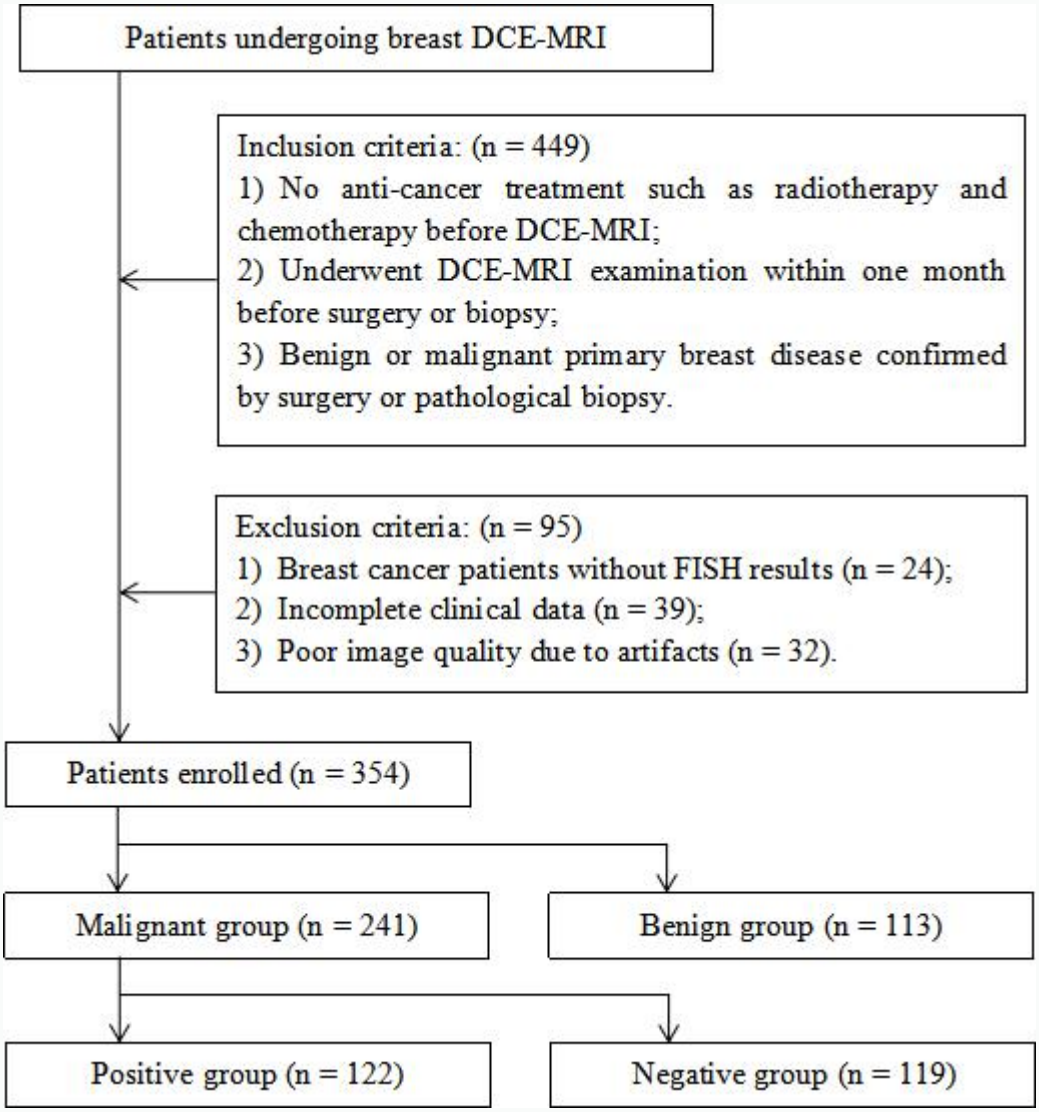

FIGURE 1: Recruitment pathway for study patients. DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging.

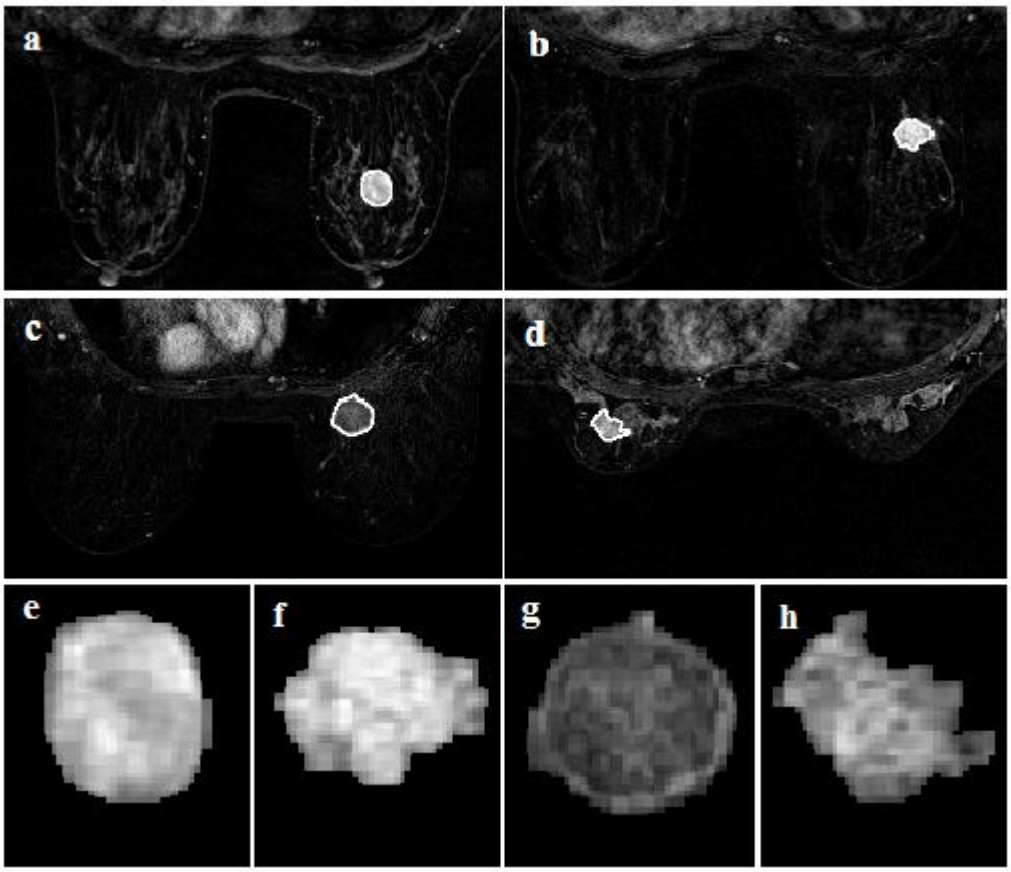

FIGURE 2: Lesion extraction results in randomly chosen cases: a and b) benign; c) malignant with negative HER-2 status, and d) malignant with positive HER-2 status. Lesions in e–h) correspond to a–d) results after segmentation.

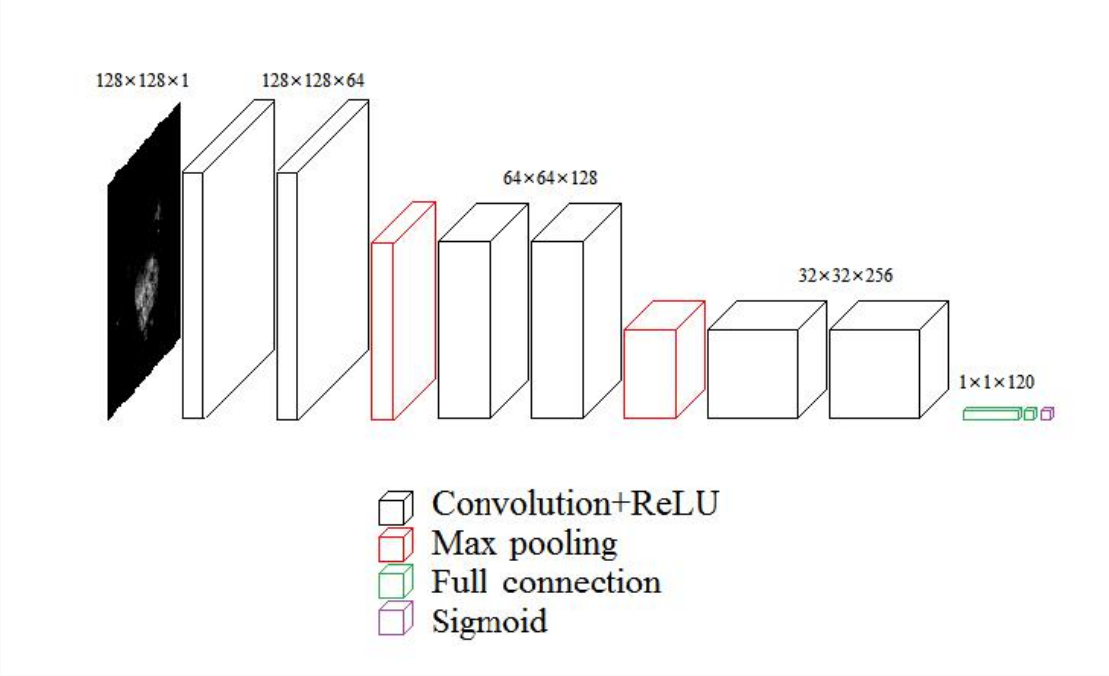

FIGURE 3: CNN model structure.

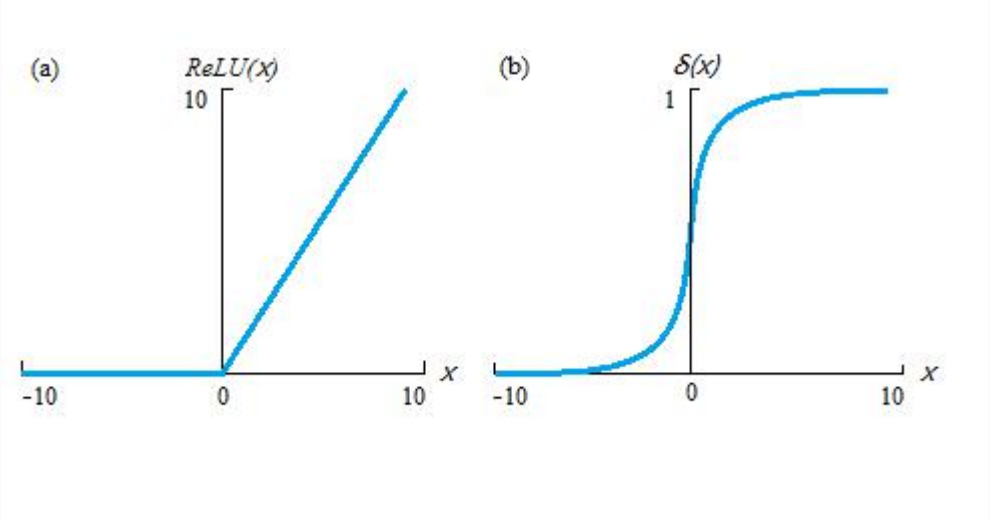

FIGURE 4: Functions used in CNN model: (a) ReLU function and (b) Sigmoid function.

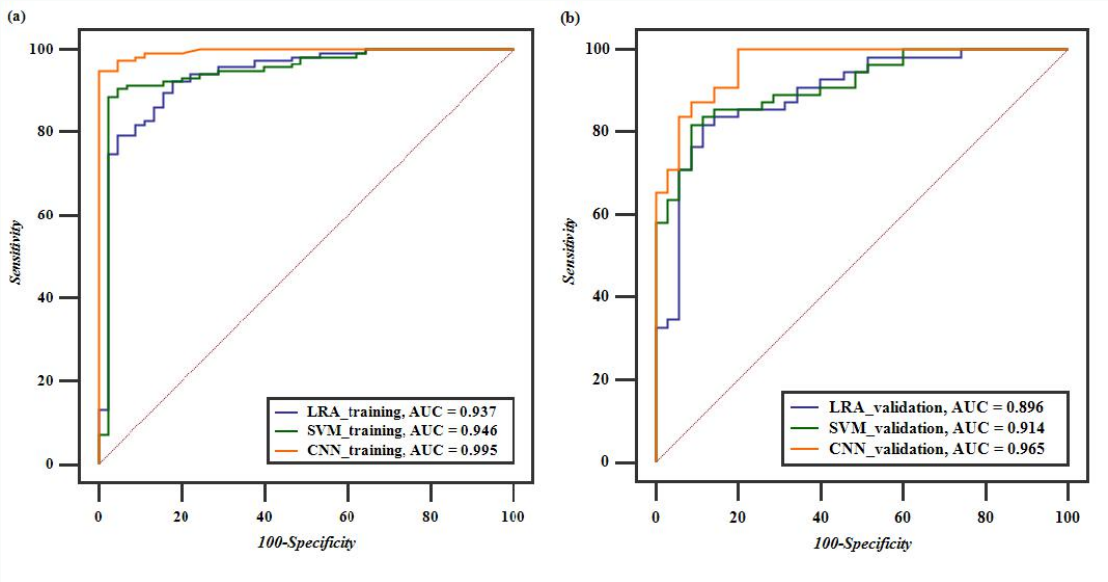

FIGURE 5: Comparison of ROC curves among three classification models for benign and malignant lesions. ROC curves for three classifiers on (a) training cohorts and (b) validation cohorts. LRA: logistic regression analysis; SVM: support vector machine; CNN: convolutional neural network.

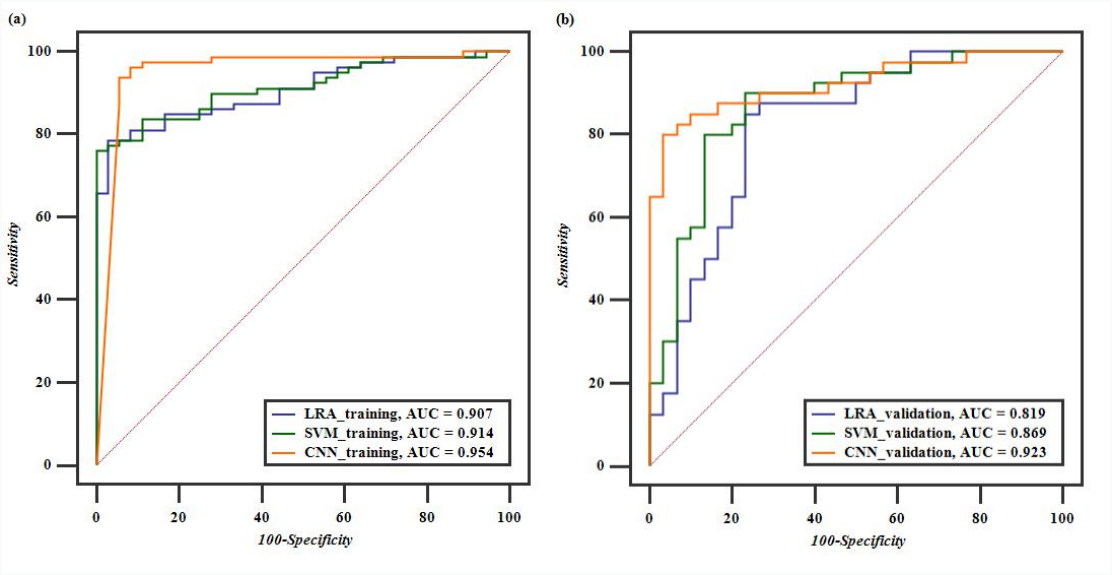

FIGURE 6: Comparison of ROC curves among three classification models for HER-2 status: ROC curves of three classifiers in (a) training cohorts and (b) validation cohorts. LRA: logistic regression analysis; SVM: support vector machine; CNN: convolutional neural network.