

Development and validation of convolutional neural network-based model to predict the risk of sentinel or non-sentinel lymph node metastasis in patients with breast cancer: a machine learning study



Mingzhen Chen,^{a,f} Chunli Kong,^{a,b,c,f} Guihan Lin,^{a,b,c} Weiyue Chen,^{a,b,c} Xinyu Guo,^a Yaning Chen,^a Xue Cheng,^{a,b,c} Minjiang Chen,^{a,b,c} Changsheng Shi,^e Min Xu,^{a,b,c} Junhui Sun,^{d,*} Chenying Lu,^{a,b,c,**} and Jiansong Ji^{a,b,c,***}



^aKey Laboratory of Imaging Diagnosis and Minimally Invasive Interventional Research of Zhejiang Province, Lishui Hospital, International Institutes of Medicine, School of Medicine, Zhejiang University, Lishui, Zhejiang 323000, China

^bInstitute of Imaging Diagnosis and Minimally Invasive Intervention Research, The Fifth Affiliated Hospital of Wenzhou Medical University, Lishui 323000, China

^cClinical College of the Affiliated Central Hospital, School of Medicine, Lishui University, Lishui 323000, China

^dDivision of Hepatobiliary and Pancreatic Surgery, Hepatobiliary and Pancreatic Interventional Treatment Centre, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

^eDepartment of Interventional Radiology, The Third Affiliated Hospital of Wenzhou Medical University, Ruian, Zhejiang, China

Summary

Background For patients with sentinel lymph node (SLN) metastasis and low risk of residual non-SLN (NSLN) metastasis, axillary lymph node (ALN) dissection could lead to overtreatment. This study aimed to develop and validate an automated preoperative deep learning-based tool to predict the risk of SLN and NSLN metastasis in patients with breast cancer (BC) using dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) images.

eClinicalMedicine

2023;63: 102176

Published Online xxx

<https://doi.org/10.1016/j.eclinm.2023.102176>

Methods In this machine learning study, we retrospectively enrolled 988 women with BC from three hospitals in Zhejiang, China between June 1, 2013 to December 31, 2021, June 1, 2017 to December 31, 2021, and January 1, 2019 to June 30, 2023, respectively. Patients were divided into the training set (n = 519), internal validation set (n = 129), external test set 1 (n = 296), and external test set 2 (n = 44). A convolutional neural network (CNN) model was proposed to predict the SLN and NSLN metastasis and was compared with clinical and radiomics approaches. The performance of different models to detect ALN metastasis was measured by the area under the curve (AUC), accuracy, sensitivity, and specificity. This study is registered at ChiCTR, ChiCTR2300070740.

Findings For SLN prediction, the top-performing model (i.e., the CNN algorithm) achieved encouraging predictive performance in the internal validation set (AUC 0.899, 95% CI, 0.887–0.911), external test set 1 (AUC 0.885, 95% CI, 0.867–0.903), and external test set 2 (AUC 0.768, 95% CI, 0.738–0.798). For NSLN prediction, the CNN-based model also exhibited satisfactory performance in the internal validation set (AUC 0.800, 95% CI, 0.783–0.817), external test set 1 (AUC 0.763, 95% CI, 0.732–0.794), and external test set 2 (AUC 0.728, 95% CI, 0.719–0.738). Based on the subgroup analysis, the CNN model performed well in tumour group smaller than 2.0 cm, with the AUC of 0.801 (internal validation set) and 0.823 (external test set 1). Of 469 patients with BC, the false positive rate of SLN prediction declined from 77.9% to 32.9% using CNN model.

Interpretation The CNN model can predict the SLN status of any detectable lesion size and condition of NSLN in patients with BC. Overall, the CNN model, employing ready DCE-MRI images could serve as a potential technique to assist surgeons in the personalized axillary treatment of in patients with BC non-invasively.

*Corresponding author. Division of Hepatobiliary and Pancreatic Surgery, Hepatobiliary and Pancreatic Interventional Treatment Centre, The First Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China.

**Corresponding author. Key Laboratory of Imaging Diagnosis and Minimally Invasive Intervention Research, Lishui Hospital of Zhejiang University, China.

***Corresponding author. Key Laboratory of Imaging Diagnosis and Minimally Invasive Intervention Research, Lishui Hospital of Zhejiang University, China.

E-mail addresses: 1307005@zju.edu.cn (J. Sun), luchenying_1981@163.com (C. Lu), jijiansong@zju.edu.cn (J. Ji).

^fThese authors contributed equally to this work.

Funding National Key Research and Development projects intergovernmental cooperation in science and technology of China, National Natural Science Foundation of China, Natural Science Foundation of Zhejiang Province, and Zhejiang Medical and Health Science Project.

Copyright © 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Breast cancer; Lymph node; Deep learning; Convolutional neural network

Research in context

Evidence before this study

Breast cancer patients with positive sentinel lymph node (SLN) but a low risk of non-sentinel lymph node (NSLN) metastasis should be exempted from axillary lymph node dissection (ALND). However, it has been reported that 43–65% of SLN-positive patients further underwent unnecessary axillary surgery.

We retrieved PubMed on June 2, 2023, for research articles that involved the terms: “(non-sentinel lymph node OR non-SLN OR NSLN) AND (deep learning OR radiomics OR artificial intelligence OR AI) AND (predict OR prediction OR identify) AND (magnetic resonance imaging OR magnetic resonance image OR MRI) AND (metastasis) AND (breast cancer OR breast carcinoma)” without date or language limits. We obtained three studies on the construction of artificial intelligence (AI)-related models for the prediction of lymph node metastasis in breast cancer. However, the establishment of these models was based on ultrasound images or still relied on time-consuming manual delineation. There was no study

utilizing magnetic resonance images to construct the AI model to identify SLN and NSLN status in breast cancer.

Added value of this study

This study developed a convolutional neural network (CNN) model for preoperative prediction of SLN and NSLN metastasis based on dynamic contrast-enhanced MRI (DCE-MRI) of breast carcinoma. For SLN prediction, the CNN model outperformed clinical and radiomics models in three datasets. It also performed well in the prediction of NSLN metastasis. Moreover, the CNN model can provide additional value to the radiologists' diagnoses.

Implications of all the available evidence

The CNN-based model is superior to radiologists' reports, the clinical model, and the traditional radiomics model in predicting the preoperative SLN and NSLN metastasis of patients with breast cancer, independent of tumour size. With the ready DCE-MRI images, this CNN model shows the possibility to serve as an imaging biomarker for personalized axillary surgical strategies.

Introduction

Breast cancer (BC) ranks first in the morbidity of female malignant tumours in the world in 2020, posing a severe threat to women's health.¹ Compared with axillary lymph node (ALN) negative patients, the 5-year overall survival rate of patients with ALN metastasis tends to decrease by up to 40%.² so precise assessment of ALN is pivotal for the treatment and prognosis of patients with BC.³ Sentinel lymph node (SLN) is the first pathway of primary tumour metastasis to the axilla. SLN dissection (SLND) is often performed to evaluate ALN status in patients with BC with clinically negative lymph node and to decide the need for subsequent ALN dissection (ALND).⁴ In routine clinical management, ALND is usually recommended if SLN is positive. According to the International Breast Cancer Study Group (IBCSG) 2301 trial, patients with limited SLN metastasis (SLNM) have a lower risk of developing non-SLN (NSLN) metastasis.⁵ For these patients, ALND does not improve survival rate compared to SLND alone, but results in more severe complications due to expanded surgical scope. Namely, patients with positive SLN but a low risk of NSLN metastasis can be exempted from ALND.⁶

Previous investigations have indicated that 43–65% of SLN-positive patients further underwent unnecessary axillary surgery, increasing surgical time and patient burden.^{7,8} Therefore, it is of great clinical benefit to seek a non-invasive approach for preoperatively identifying patients in this subgroup.

With favourable soft-tissue contrast and enhanced patterns of the lesion, breast dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), plays an important role in the assessment of ALN status.⁹ However, DCE-MRI depends on radiologists' interpretation and naked eye cannot accurately distinguish SLN and NSLN states. Recently, radiomics has been shown to be useful for automatic identification of ALN in patients with BC.^{10,11} For example, Yu Y et al. hypothesized the ALN-tumour radiomic signature can differentiate ALN metastasis preoperatively with an AUC of 0.87 in the external validation set. However, radiomics relies on time-consuming manual pixel-by-pixel delineation, which limits the clinical transformation of this technique.

With the advancement of computerized analysis techniques, deep learning has attracted much attention

due to its distinguished performance in image recognition and classification tasks.^{12,13} Convolutional neural network (CNN) is a prospective deep learning algorithm.¹⁴ Unlike radiomics which relies on manual annotation, CNN can automatically extract and learn complex image characteristics and thereby realizing end-to-end prediction efficiently. Therefore, CNN can achieve automatic and sufficient mining from medical images making it more feasible for clinical translation.¹⁵ Ohno et al. and Hallinan et al. demonstrated the potential of CNN for pulmonary nodules classification and lumbar disease detection, respectively.^{16,17} CNN also manifests great potential in the clinical management of tumours in diverse medical datasets.^{18,19}

Preoperative analysis and management of lymph node in patients with BC is an important but challenging procedure. The combination of DCE-MRI of primary tumours and deep learning in predicting SLN and NSLN status may yield better diagnostic results and contribute to the personalized management of axillary surgery in patients with BC. To this end, an automated preoperative CNN-based tool was developed and verified to predict SLN and NSLN metastasis in patients with BC using DCE-MRI images in this multi-institutional study.

Methods

Patients and study design

This retrospective three-centre study was reviewed and approved by the Institutional Review Boards of Lishui Hospital of Zhejiang University (centre 1), the First Affiliated Hospital of Zhejiang University (centre 2), and the Third Affiliated Hospital of Wenzhou Medical University (centre 3). The requirement for patient informed consent was waived due to the retrospective nature of the study and because all administrative procedures were part of routine care. The study was carried out in compliance with STARD-2015 guidelines (equator-network.org). Chinese Clinical Trial Registry institution approved the study with the number of ChiCTR2300070740.²⁰

Data were acquired from consecutive patients with BC at three university teaching hospitals (centre 1 (June 1, 2013 to December 31, 2021), centre 2 (June 1, 2017 to December 31, 2021), and centre 3 (January 1, 2019 to June 30, 2023)). After applying the inclusion and exclusion criteria (see [Supplementary Appendix S1](#)), a total of 988 cases were included, and the study flowchart is shown in [Fig. 1](#). Patients were divided into four cohorts: a training set ($n = 519$ from centre 1), an internal validation set ($n = 129$ from centre 1), and two external test sets ($n = 296$ from centre 2 and $n = 44$ from centre 3).

Clinical data collection and reference standards

Clinical and histopathological data were collected from medical records, including age, histological type and

grade of invasive carcinoma, molecular subtypes, BI-RADS category, tumour size, and lymph node status reported by MRI (MRI classification criterion of ALN shown in [Supplementary Appendix S2](#)).

The postoperative pathological results of SLND and ALND were collected as the reference standards for modelling. The corresponding pathological findings can be divided into the following three categories: positive SLN but negative NSLN (SLN+ & NSLN-), positive SLN and NSLN (SLN+ & NSLN+), and negative SLN (SLN-) (refer to [Supplementary Table S2](#)).

Data preprocessing

The MRI scanning scheme of each hospital was detailed in [Supplementary Appendix S3](#) and [Table S1](#). DCE-MRI cross-sectional slice with the largest lesion was selected as the input image for the deep learning model. The affected breast was then independently segmented using a rectangular shape by two experienced radiologists. [Supplementary Fig. S2](#) provides examples illustrating the cropping process. In cases where there was disagreement between the radiologists, they engaged in negotiations until reaching a consensus. The target images were uniformly resampled to a size of 70×70 pixels before being input into the model. Further data augmentation was performed through random cropping to a final size of 64×64 pixels. Finally, all images were normalized using Z-Score to ensure consistent scaling across the dataset.

ALN metastasis prediction CNN model

As the most commonly used deep learning algorithm, CNN can directly input medical images into the network for modelling. Through the calculation of each layer in the network, the prominent features are extracted, and the key information that contributes to image recognition and classification is learned automatically.¹⁵ Therefore, in this study, CNN-based models were designed to predict the metastasis of SLN and NSLN, successively.

In this paper, we used the mainstream Residual Network 18 (ResNet18) for experiments.²¹ The specific parameter settings were detailed in [Supplementary Appendix S4](#). The outline of algorithm establishment is portrayed in [Fig. 2](#). First, the ResNet18-I algorithm differentiated all the enrolled patients with BC into two subgroups. If ResNet18-I predicted the lymphatic spread risk within the axilla is low, the subjects will be directly output as SLN low-risk subgroup. When ResNet18-I recognized the patients as SLN high-risk, the SLND results are integrated to preclude SLN-negative patients in this subgroup. The ResNet18-II would sequentially differentiate NSLN status of SLN high-risk patients with positive SLN. Consequently, the ResNet18-II could classify SLN+ & NSLN+ patients as the NSLN high-risk subgroup, thus guiding clinical therapy, while patients in the NSLN low-risk subgroup (SLN+ & NSLN-) could bypass further ALND treatment.

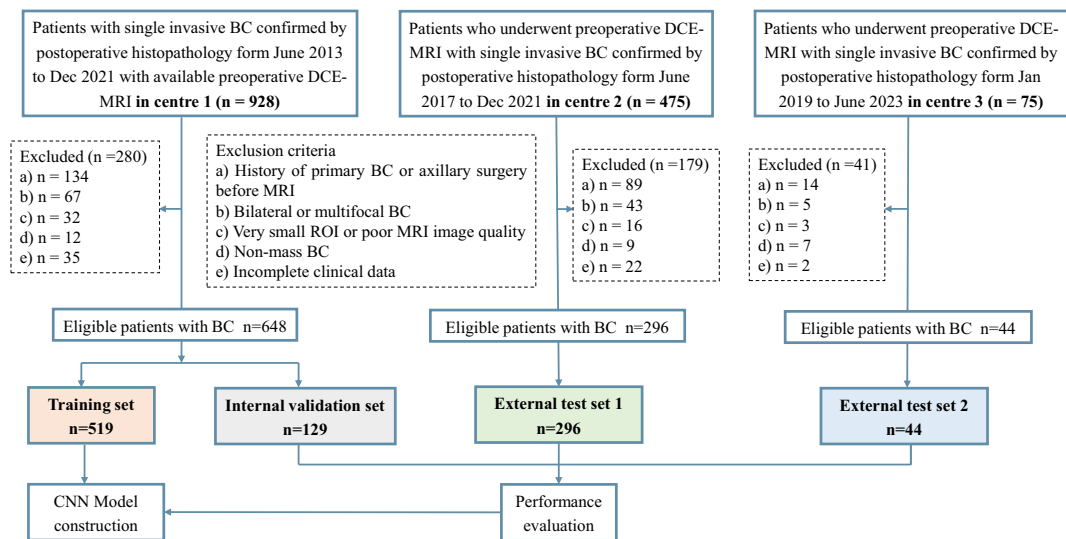


Fig. 1: Flowchart of patients and study design. BC, breast cancer; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging; ROI, region of interest; CNN, convolution neural network.

Comparison with other models

We developed two different predictive models for comparison: a clinical model, combining pathologic and radiologic factors, and a radiomics model based on the hand-crafted features. The traditional radiomics process and related algorithms explanation were described concretely in the [Supplementary Appendix S5](#). As shown in [Supplementary Fig. S2](#), this technique mainly included precise regions of interest (ROI) segmentation, feature extraction, algorithm analysis, and model construction.

Statistical analysis

Statistical tests used in this study were performed using SPSS, R, and Python v3.7.6, and a bilateral P value less

than 0.05 was considered significant. Continuous variables are recorded as medians and IQRs, and categorical variables are recorded as numbers and percentages. Univariate and multivariable logistic regression were conducted to identify potential clinical risk factors. Once developed, the performance of the CNN algorithm was compared to clinical and radiomics models using the following metrics: the area under the receiver operating characteristics (ROC) curve (AUC), accuracy, sensitivity, and specificity. In this study, the probability threshold used for the accuracy results was consistent with the conventional one, which was 0.5. The ROCs of these models are compared by the DeLong test.²² The percentage of true positive, false positive, true negative, and

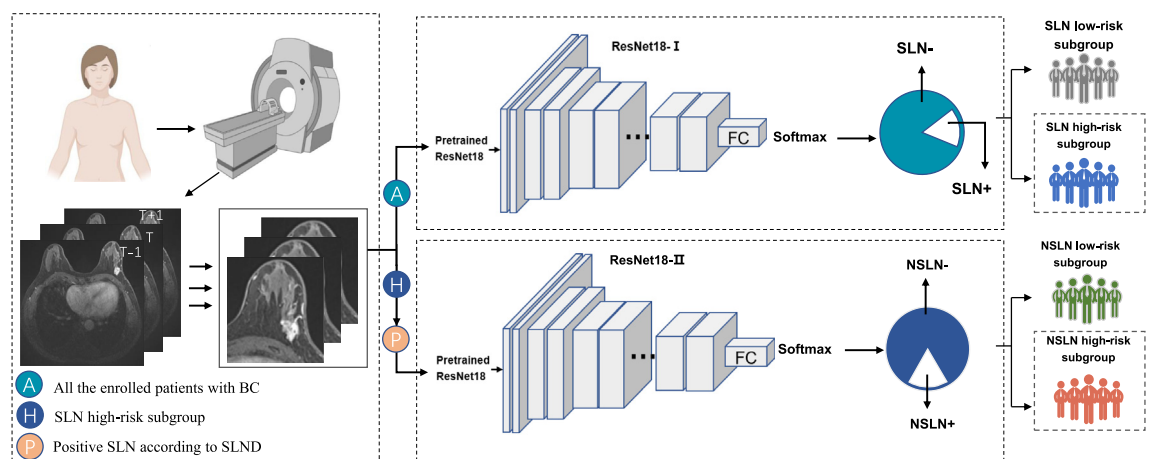


Fig. 2: Overview of the convolutional neural network model development process. ResNet18, Residual Network 18; BC, breast cancer; SLN, sentinel lymph node; NSLN, non-SLN; SLND, sentinel lymph node dissection.

false negative results of the method is determined according to the reference standard of pathological results, and the results are displayed in the form of a confusion matrix diagram. By reckoning the net benefits under different threshold probabilities, decision curve analysis is used to evaluate the clinical usefulness of each model. Inference code and illustrative examples of the proposed CNN model are publicly available on GitHub (https://github.com/wen-alan/BC_ALN).

Role of the funding source

The funder of the study had a role in study design, data collection, data analysis, and writing of the report. All authors had access to dataset and JSJ had the final responsibility for the decision to submit for publication.

Results

Clinical characteristics of the involved patients

The median age of all the included patients is 51 years (IQR, 46–58 years). The histological type in most patients is invasive ductal carcinoma (907 of 988 [91.8%]). Most patients had molecular type of Luminal B (585 of

988 [59.2%]). Patient characteristics are depicted in Table 1. Of all the included patients, 234 (23.7%) patients had sentinel lymph node metastasis (refer to Supplementary Table S2). Univariate and multivariate logistic regression analysis results were shown in the Supplementary Appendix S6 and Table S3.

Algorithm performance of SLN metastasis identification

In radiomics modelling, the performance of four machine learning classifiers for SLN prediction is listed in Supplementary Table S4. The AUC and accuracy of DNN were not the best among the four classifiers in the training set. However, DNN model is more generalizable to other datasets. Therefore, the DNN model was output as a traditional radiomics model and further compared with the clinical or CNN model.

The top-performing model, CNN model, showed AUCs of 0.986 (95% CI, 0.962–1.000), 0.899 (95% CI, 0.887–0.911), 0.885 (95% CI, 0.867–0.903), and 0.768 (95% CI, 0.738–0.798) in the training (n = 519), internal validation (n = 129), external test 1 (n = 296), and

Variables	Training set N = 519	Internal validation set N = 129	External test set 1 N = 296	External test set 2 N = 44
Median age (IQR) ^a	51 (46–57)	49 (44.5–56)	52 (45–59)	57 (48–62.75)
Tumour size				
≤2 cm	326 (62.8)	84 (65.1)	183 (61.8)	15 (34.1)
>2 cm	193 (37.2)	45 (34.9)	113 (38.2)	29 (65.9)
Menopausal status (%)				
Premenopause	242 (46.6)	69 (53.5)	159 (53.7)	12 (27.3)
Postmenopause	277 (53.4)	60 (46.5)	137 (46.3)	32 (72.7)
Histological type (%)				
Invasive ductal carcinoma	486 (93.6)	121 (93.8)	265 (89.5)	35 (79.5)
Invasive lobular carcinoma	15 (2.9)	5 (3.9)	10 (3.4)	2 (4.6)
Others	18 (3.5)	3 (2.3)	21 (7.1)	7 (15.9)
Histological grade (%)				
I	42 (8.1)	15 (11.6)	12 (4.1)	7 (15.9)
II	262 (50.5)	62 (48.1)	157 (53.0)	31 (70.5)
III	215 (41.4)	52 (40.3)	127 (42.9)	6 (13.6)
Molecular subtype (%)				
Luminal A	93 (17.9)	24 (18.6)	59 (19.9)	9 (20.5)
Luminal B	307 (59.2)	74 (57.4)	180 (60.8)	24 (54.5)
HER2 positive	41 (7.9)	10 (7.7)	22 (7.4)	10 (22.7)
Triple negative	78 (15.0)	21 (16.3)	35 (11.8)	1 (2.3)
BI-RADS category (%)				
4A	10 (1.9)	1 (0.8)	24 (8.1)	1 (2.3)
4B	68 (13.1)	14 (10.8)	26 (8.8)	6 (13.6)
4C	205 (39.5)	61 (47.3)	124 (41.9)	9 (20.5)
5	236 (45.5)	53 (41.1)	122 (41.2)	28 (63.6)
MRI-reported ALN status (%)				
Negative	433 (83.4)	105 (81.4)	151 (51)	23 (52.3)
Positive	86 (16.6)	24 (18.6)	145 (49)	21 (47.7)

Data are n (%), unless otherwise indicated. HER2, human epidermal growth factor receptor-2; MRI, magnetic resonance imaging; ALN, axillary lymph node. ^aData are medians, with IQRs in parentheses.

Table 1: Baseline characteristics of patients in the training, internal validation and external test sets.

external test 2 ($n = 44$) sets, respectively. Notably, it achieved a higher AUC than clinical model (AUC, 0.885 vs. 0.486, $P < 0.001$) and radiomics model (AUC, 0.885 vs. 0.561, $P < 0.001$) in the external test set 1 according to the DeLong test. In addition, in the external test set 1, the accuracy of the CNN model was 89.2%, which outperformed the clinical model (43.9%) and radiomics model (82.6%). Fig. 3a demonstrates the confusion matrices that CNN model attained the highest sensitivity of 75.5%, and specificity of 88.3%, when compared with clinical (sensitivity of 48.8%, and specificity of 43.1%), and radiomics model (sensitivity of 60%, and specificity of 54.4%) in the external test set 1. ROCs for SLN prediction according to clinical, radiomics, and CNN models are shown in Fig. 3b and the detailed performance of the three models is summarized in Table 2. According to the results of the decision curves analysis, if the threshold probability was greater than 0.2, using the CNN model to predict SLN metastasis gains more benefits than treat-all or treat-none tactics and more benefits than the clinical and radiomics models (Fig. 3c).

Furthermore, the false positive rate of SLN status evaluation will be decreased from 77.9% (determined by radiologist on MRI images) to 32.9% by utilizing the CNN model on SLND decision in validation and test datasets (Fig. 4). To some extent, this can reduce the rate of unnecessary SLND. Simultaneously, the false negative rate of SLN status predicted by the CNN model is also lower than that of radiologists (10.3% vs. 18.6%). This means that high-risk patients requiring SLND would not be easily omitted by the ResNet18-I frame.

Algorithm performance of NSLN metastasis identification

Due to the orderly lymphatic drainage of the mammary gland, it is necessary to further identify the NSLN status if the patient is in SLN high-risk subgroup (Fig. 2). In this part, the CNN model achieved outstanding performance with an AUC of 0.998 (95% CI, 0.994–1), and an accuracy of 100% in the training set. The performance slightly decreased in the internal validation (AUC, 0.800 (95% CI, 0.783–0.817), accuracy, 79%), external test set 1 (AUC, 0.763 (95% CI, 0.732–0.794), accuracy, 77.8%), and external test set 2 (AUC, 0.728 (95% CI, 0.719–0.738), accuracy, 73%). The relevant performance details are demonstrated in Table 2. As stated in confusion matrices (Fig. 5a), the sensitivity, and specificity values for the training set are 100%, and 100%, respectively. Data for the internal validation set are 72.7%, and 86.1%, respectively, data for the external test set 1 are 90.5%, and 66.7%, respectively, and data for the external test set 2 are 66.7%, and 88.9%, respectively. ROCs for NSLN prediction of the CNN model in different datasets are exhibited in Fig. 5b. Moreover, utilizing CNN model to predict NSLN metastasis yields more net benefits than treat-all or treat-none strategies (Fig. 5c).

Model performance of SLN metastasis according to tumour size

In multivariable analysis, tumour size is one of the prognostic factors in the clinical models (Supplementary Table S3). Therefore, the performance of CNN model was further compared according to the classification of tumour size, shown in Table 3. The CNN model showed the highest discrimination (AUC, 0.823, 95% CI, 0.781–0.865) of SLN metastasis for patients with lesions measuring 2.0 cm or smaller in the external test set 1. The corresponding accuracy, sensitivity, and specificity were 88.8%, 90%, and 68.2%, respectively. Whereas the correlated performance of clinical (AUC, 0.505, 95% CI, 0.406–0.605) and radiomics (AUC, 0.571, 95% CI, 0.565–0.577) models were approximate to random in the external test set 1. Additionally, the CNN model still outperformed other models in predicting SLN metastasis of breast lesions larger than 2.0 cm, with an AUC of 0.858 (95% CI, 0.824–0.891) in the external test set 1. The associated accuracy, sensitivity, and specificity were 86.5%, 87%, and 79.1%, respectively.

Activation heat maps from DCE-MRI images

To better understand the model, we employed class activation mapping (CAM) to visualize the internal features of the neural network. CAM can generate a rough localization map that highlights the input area of the classification target, which can help us understand whether the CNN focuses on the suitable tumour area.²³ In this way, CAM can improve the interpretability of the CNN model. We found that the high signal inside the lesion and the area around the tumour is constructive in predicting SLN or NSLN status, as shown in Supplementary Fig. S3. Patient examples for the actual use of the established CNN model are displayed in Fig. 6.

Discussion

In this study, we constructed and tested a CNN model to predict preoperative SLN and NSLN metastasis in patients with BC using datasets from three institutions. The CNN model exhibited satisfactory discrimination for SLN prediction in the internal validation set (AUC, 0.899) and the external test set 1 (AUC, 0.885) and surpassed a clinical model, and a radiomics model. Based on the result of subset analysis, the model can predict SLN of any detectable lesion size effectively. For patients with positive SLN, the CNN model could further differentiate NSLN condition, with an AUC of 0.800 in the internal validation set and an AUC of 0.763 in the external test set 1. Moreover, this model can add value to radiologists' diagnoses, which may decrease the rate of unnecessary SLND in patients with BC.

Several studies have explored risk stratification methods for ALN metastasis in patients with BC, including lymphatic or vascular structure involvement,

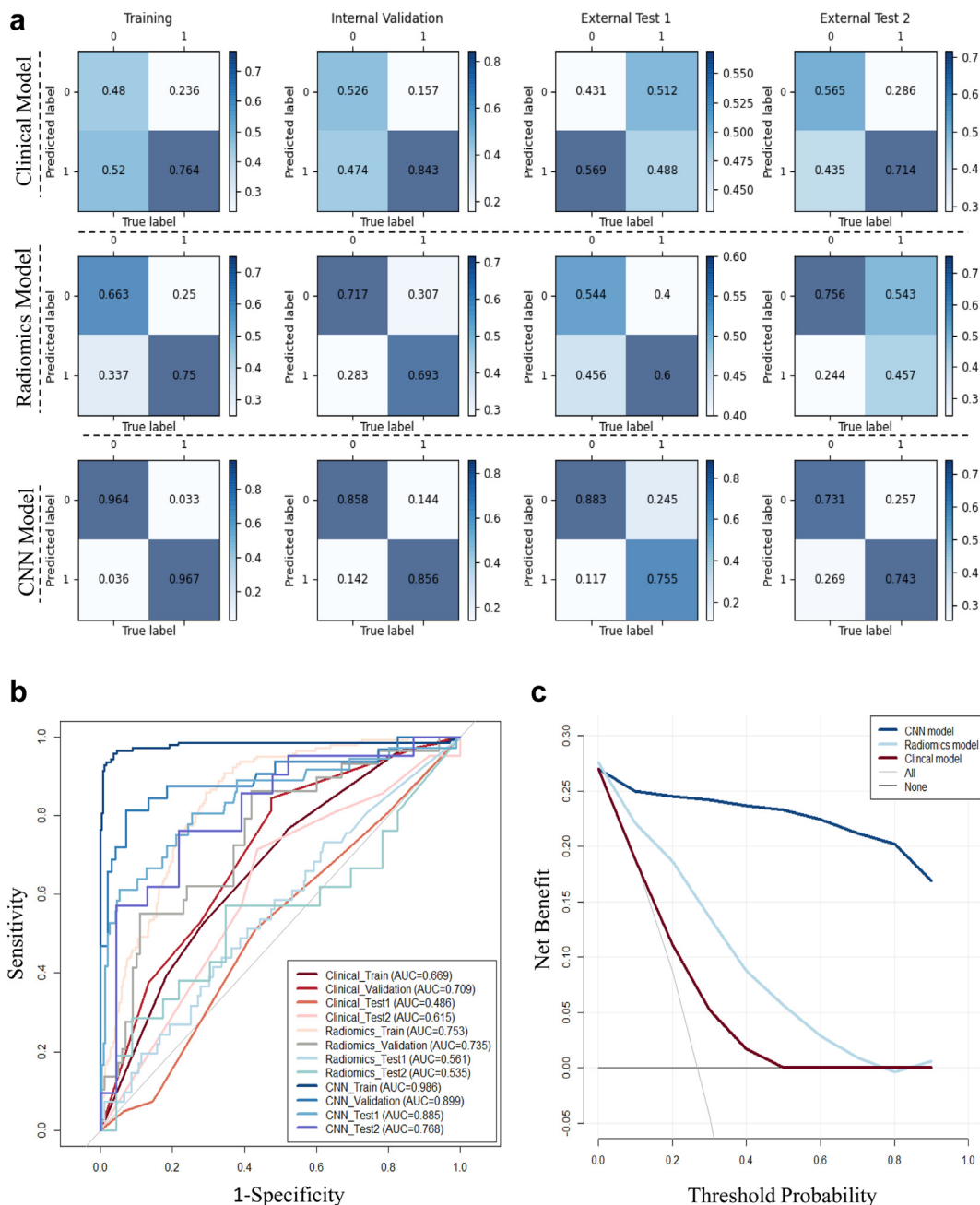


Fig. 3: Confusion matrices, receiver operating characteristic curves (ROC) and decision curve analysis of sentinel lymph node (SLN) status classification for different models. (a) Confusion matrices of the clinical, radiomics, and CNN model in training set, internal validation set, external test set 1, and external test set 2. The colour depends on the number inside the square: the higher the number, the darker the colour. (b) ROC results and (c) decision curve analysis of different models for predicting SLN status.

histological type, and oestrogen/progesterone receptor status.^{24–26} However, some histopathological data, such as lymph-vascular invasion, could only be assessed after the surgery, thereby failing to pilot ALND decisions preoperatively. In this study, we only included clinical factors that could be obtained before breast surgery.

Finally, MRI-reported tumour size and BI-RADS category were incorporated into clinical model. This accords with the previous conclusion that SLN metastasis was highly correlated with tumour size (>2.0 cm).²⁷ However, the performance of the clinical model was not satisfactory.

	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)	P value ^a
SLN					
Training set (n = 519)					
Clinical model	0.669 (0.619–0.718)	55.7 (55.6–55.8)	76.4 (69.4–83.5)	48.0 (43.0–53.1)	<0.0001
Radiomics model	0.753 (0.715–0.790)	74.0 (72.7–75.3)	75.0 (65.1–84.9)	66.3 (60.1–72.5)	<0.0001
CNN model	0.986 (0.962–1.000)	96.8 (92.6–100)	96.7 (91.9–100)	96.4 (91.1–100)	–
Internal validation set (n = 129)					
Clinical model	0.709 (0.612–0.807)	60.5 (60.1–60.8)	84.4 (71.8–97.0)	52.6 (42.6–62.5)	0.0031
Radiomics model	0.735 (0.727–0.742)	78.6 (77.4–79.8)	69.3 (57.0–81.6)	71.7 (58.6–84.8)	0.024
CNN model	0.899 (0.887–0.911)	84.2 (80.9–87.4)	85.6 (81.9–89.3)	85.8 (84.2–92.9)	–
External test set 1 (n = 296)					
Clinical model	0.486 (0.395–0.577)	43.9 (38.2–49.5)	48.8 (33.2–64.6)	43.1 (37.0–49.4)	<0.0001
Radiomics model	0.561 (0.559–0.563)	82.6 (80.2–84.9)	60.0 (46.2–73.8)	54.4 (41.6–67.3)	<0.0001
CNN model	0.885 (0.867–0.903)	89.2 (85.5–90.0)	75.5 (69.4–81.6)	88.3 (83.7–92.9)	–
External test set 2 (n = 44)					
Clinical model	0.615 (0.454–0.775)	63.6 (62.6–64.7)	71.4 (52.1–90.8)	56.5 (36.3–76.8)	0.043
Radiomics model	0.535 (0.520–0.551)	54.5 (52.1–57.0)	45.7 (38.1–53.3)	75.6 (65.7–85.6)	0.021
CNN model	0.768 (0.738–0.798)	70.9 (67.7–74.2)	74.3 (60.9–87.7)	73.1 (63.2–82.9)	–
NSLN					
CNN model					
Training set (n = 139)	0.998 (0.994–1.000)	100 (100–100)	100 (100–100)	100 (100–100)	–
Internal validation set (n = 32)	0.800 (0.783–0.817)	79.0 (77.6–80.5)	72.7 (64.7–80.7)	86.1 (76.9–95.3)	–
External test set 1 (n = 36)	0.763 (0.732–0.794)	77.8 (75.3–80.3)	90.5 (76.0–100)	66.7 (59.0–74.3)	–
External test set 2 (n = 21)	0.728 (0.719–0.738)	73.0 (68.2–77.8)	66.7 (57.0–76.4)	88.9 (84.6–93.2)	–

Note: Data in parentheses are the 95% confidence interval. All the results in the primary cohort were evaluated by five-fold cross-validation. SLN, sentinel lymph node; NSLN, non-SLN; AUC, area under the receiver operating characteristic curve. ^aComparison between CNN model and other models using DeLong test.

Table 2: Predictive performance of various approaches in the training, internal validation and external test sets for SLN and NSLN status.

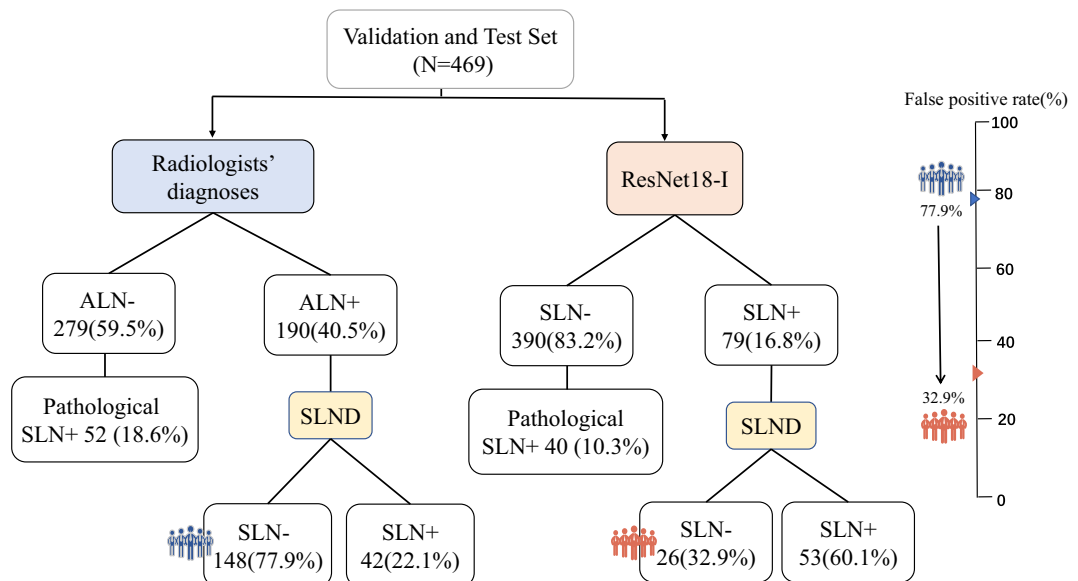


Fig. 4: Recommendation for SLND according to radiologists' diagnoses or the CNN model in patients with BC in the validation and test set. ResNet18, Residual Network 18; ALN, axillary lymph node; SLN, sentinel lymph node; SLND, sentinel lymph node dissection.

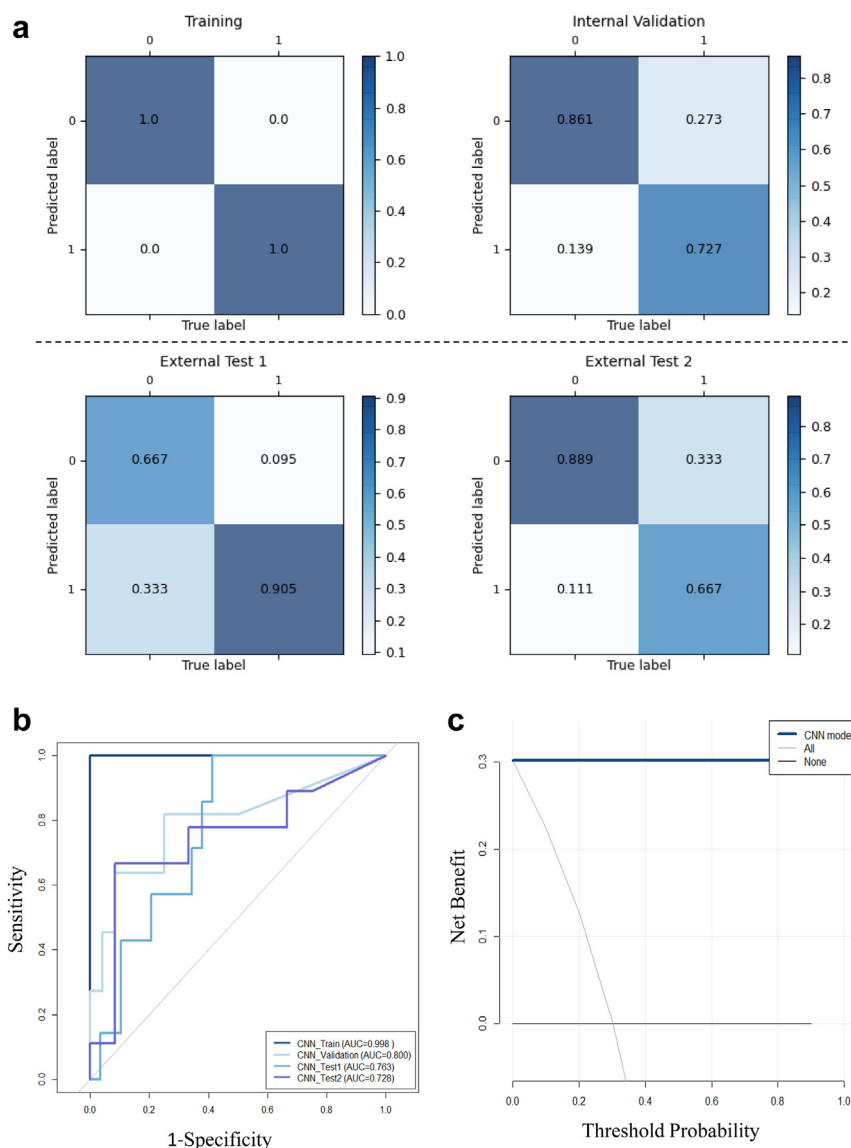


Fig. 5: Confusion matrices (a), receiver operating characteristic curves (b) and decision curve analysis (c) of non-sentinel lymph node status classification in training set, internal validation set, external test set 1, and external test set 2.

Providing a complete scope of each organ with a high soft-tissue resolution, MRI is advocated for breast lesions that are tricky to be assessed by mammography or ultrasound.²⁸ Evaluating axillary lymph node via ultrasound is definitely more convenient and cost-effective. However, the quality of ultrasound images can vary depending on the skill and experience of the physician performing the examination. This variability can affect the accuracy and reliability of the ultrasound evaluation. On the contrary, MRI scanning of breast has been standardized and is less affected by operators, so the image quality is more stable. Specifically, there are various MRI sequences available for breast scans,

among which DCE-MRI holds significant importance. By utilizing a contrast agent, DCE-MRI can effectively display the enhancement of lesions and the characteristics of the blood supply. As a result, DCE-MRI offers superior lesion conspicuity and plays a crucial role in achieving high diagnostic efficiency. Therefore, in this study, we utilized DCE-MRI images to develop model for the evaluation of axillary lymph node. Although MRI is very helpful in breast cancer staging and preoperative evaluation, the diagnosis process by radiologists is relatively subjective, and MRI is usually inconsistent with pathological ALN status. Several investigations have evaluated ALN metastasis prediction using

	AUC	ACC(%)	Sensitivity (%)	Specificity (%)
Lesions ≤ 2.0 cm				
Training set				
Clinical model	0.617 (0.550–0.684)	64.7 (64.6–64.9)	47.6 (35.5–60.0)	68.8 (63.2–74.4)
Radiomics model	0.830 (0.694–0.965)	79.0 (69.1–88.8)	79.0 (70.8–87.3)	77.2 (60.5–94.0)
CNN model	0.997 (0.995–1.000)	98.1 (97.0–99.2)	98.3 (97.2–99.3)	98.4 (96.9–99.9)
Internal validation set				
Clinical model	0.693 (0.549–0.838)	71.4 (71.0–71.9)	64.3 (39.2–89.4)	72.9 (62.4–83.3)
Radiomics model	0.687 (0.678–0.695)	67.1 (64.2–70.0)	75.0 (67.3–82.7)	64.2 (55.8–72.6)
CNN model	0.801 (0.770–0.831)	81.0 (79.0–82.9)	84.7 (74.7–94.7)	67.4 (55.3–79.4)
External test set 1				
Clinical model	0.505 (0.406–0.605)	32.8 (32.6–33.0)	88.5 (76.2–100.0)	23.6 (16.9–30.2)
Radiomics model	0.571 (0.565–0.577)	84.1 (83.8–84.3)	53.8 (49.7–58.0)	64.5 (61.4–67.6)
CNN model	0.823 (0.781–0.865)	88.8 (87.7–89.9)	90.0 (82.0–98.0)	68.2 (60.8–75.7)
External test set 2				
Clinical model	0.440 (0.118–0.762)	22.7 (4.3–49.0)	80.0 (29.9–98.9)	0 (0–34.5)
Radiomics model	0.490 (0.438–0.542)	60.0 (55.8–64.2)	52.0 (36.3–67.7)	66.0 (52.7–79.3)
CNN model	0.812 (0.731–0.893)	72.0 (67.1–76.9)	76.0 (53.1–98.9)	90.0 (79.3–100)
Lesions > 2.0 cm				
Training set				
Clinical model	0.561 (0.492–0.629)	52.8 (52.6–53.1)	71.4 (61.3–81.5)	40.5 (31.6–49.5)
Radiomics model	0.709 (0.608–0.810)	75.6 (71.8–79.5)	75.5 (63.2–87.7)	59.4 (48.2–70.5)
CNN model	0.999 (0.999–1.000)	99.3 (98.6–99.9)	99.4 (98.7–100)	99.6 (98.7–100)
Internal validation set				
Clinical model	0.578 (0.424–0.732)	57.8 (56.7–58.8)	66.7 (44.9–88.4)	51.9 (33.0–70.7)
Radiomics model	0.759 (0.736–0.783)	79.7 (78.2–81.3)	68.6 (61.4–75.7)	78.0 (68.8–87.2)
CNN model	0.847 (0.791–0.903)	77.9 (72.6–83.2)	91.3 (79.9–100)	73.8 (69.2–78.3)
External test set 1				
Clinical model	0.489 (0.338–0.640)	47.8 (47.4–48.2)	53.3 (28.1–78.6)	46.9 (37.1–56.8)
Radiomics model	0.543 (0.516–0.569)	85.3 (84.0–86.6)	72.0 (55.8–88.2)	44.9 (25.6–64.2)
CNN model	0.858 (0.824–0.891)	86.5 (84.6–88.4)	87.0 (83.1–90.9)	79.1 (71.3–86.9)
External test set 2				
Clinical model	0.548 (0.337–0.723)	62.1 (60.5–63.7)	93.8 (81.9–100.0)	23.1 (0.2–46.0)
Radiomics model	0.436 (0.338–0.534)	46.9 (42.8–50.9)	58.8 (21.7–95.8))	55.4 (16.5–94.2)
CNN model	0.699 (0.662–0.736)	66.2 (62.3–70.1)	68.7 (53.3–84.2)	76.9 (67.4–86.4)

Note: Data in parentheses are the 95% confidence interval. All the results in the primary cohort were evaluated by five-fold cross-validation. SLN, sentinel lymph node. AUC: area under the receiver operating characteristic curve.

Table 3: Comparison of the performance for SLN prediction between the clinical, radiomics and CNN models according to the tumour size.

radiomics.^{11,29} A limitation of these studies was that the radiomics features in the prediction model differ from one another, which posed challenges for reproducibility. In this research, the AUC of the top-performing radiomics model in the external test set 1 was only 0.561.

Deep learning-aided lymph node prediction has been enrolled in some solid tumours.^{30,31} In the present study, we built a CNN model based on DCE-MRI images of primary lesions to predict SLN and NSLN metastasis successively. Firstly, for SLN prediction, the CNN model wound up with an AUC of 0.885 in the external test set 1. According to Fig. 4, CNN model reduced the false positive rate of SLN evaluation compared with radiologists' reports. In the second stage model, the CNN achieved an AUC of 0.763 when applied on the external

test set 1. The corresponding sensitivity was 90.5%, indicating a false negative rate of 9.5%. This suggests that the CNN model can effectively determine the NSLN status and assist in deciding whether ALND is needed for patients with positive SLN. By identifying SLN+ & NSLN– patients, unnecessary ALND procedures can be avoided, thus reducing overtreatment. Several researchers carried out CNN analysis used MRI of ALN region to predict ALN status.^{32–34} For example, Gao et al. constructed a deep-learning model for predicting overall ALN metastasis based on primary lesion and lymph node images and obtained an AUC of 0.852 in the external test cohort.³² However, there are some gaps in their study. First, conventional DCE-MRI scans may not cover all ALN, so predicting ALN metastasis based on

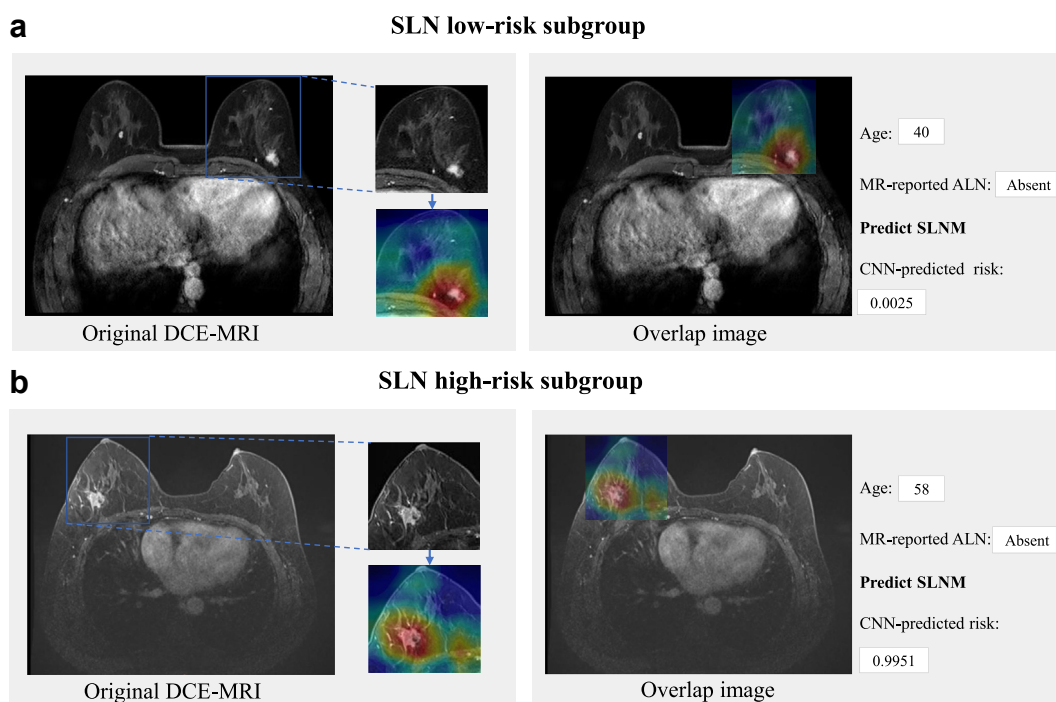


Fig. 6: Patient instances for the application of the proposed convolution neural network (CNN) model. (a) A 40-year-old female had breast cancer (BC) with negative SLN according to pathological results. She was identified to be in the low-risk subgroup by the CNN model, and the MR-reported ALN was absent. (b) A 58-year-old female had BC with positive SLN according to pathological results. She was identified to be in the high-risk subgroup by the CNN model while the MR-reported ALN was absent. ALN, axillary lymph node; SLN, sentinel lymph node; SLNM, sentinel lymph node metastasis; DCE-MRI, dynamic contrast-enhanced magnetic resonance imaging.

ALN region may not be generalizable. In addition, their model predicts gross ALN metastasis without classification, which is inconsistent with clinical routine. In our study, we predict SLN and NSLN status successively by following the clinical practice, thereby offering vital information for clinical trajectory: patients with SLN⁻ do not need SLND or ALND, patients with SLN⁺ & NSLN⁻ call for SLND but not ALND, and ALND only for patients with SLN⁺ & NSLN⁺. In daily clinical practice, breast surgeons frame a rectangle bounding box (including the full tumour region of the affected breast), and then the CNN model can provide the details of ALN status.

Previous studies demonstrated that patients with BC with tumours greater than 2.0 cm are at higher risk of developing SLN metastasis.²⁷ However, for cases with relatively small lesions (≤ 2.0 cm), it is difficult to predict whether they will suffer from adverse prognostic events, such as ALN metastasis. Breast surgeons would be more intrigued in whether the model could precisely differentiate SLN status in this group. Hence, we performed a subgroup analysis based on tumour size. The results indicated that the CNN model performs well when the lesion is less than 2.0 cm, with an AUC of 0.823 (external test set 1). This suggested that the model

performs classification tasks not only based on tumour size, shape, or location, but also integrates more crucial latent features. According to the heat map, the CNN algorithm can focus on the lesion and its surrounding important information through an attention mechanism. In general, the model can effectively predict SLN condition of any detectable lesion size.

The current study has some limitations. First, the nature of the retrospective study may cause selection bias. Second, the rectangular segmentation of the DCE-MRI images still requires manual intervention, which is relatively laborious. Third, there was only one sequence (i.e., DCE-MRI) involved in the modelling. The performance of the CNN model indicated that using a single image can initially achieve satisfactory outcomes. However, a single image may provide relatively limited information for predicting ALN. To address this limitation and enhance the prediction performance, we plan to incorporate more MRI sequences and explore the development of 2.5D or 3D models for ALN prediction in future research. Furthermore, to enhance the robustness and generalizability of the proposed CNN model, we will collect a larger and more diverse dataset from various medical institutions in future studies.

In conclusion, the CNN-based model is superior to the clinical and the traditional radiomics models in predicting the preoperative SLN metastasis of patients with BC, independent of tumour size. Additionally, the model can also effectively predict the metastasis of the next station drainage lymph node, namely, NSLN tumour involvement. With the ready DCE-MRI images, this CNN model shows the possibility to serve as an imaging biomarker for personalized axillary surgical strategies.

Contributors

Scientific guarantor, JSJ, CYL, JHS; study concepts or data acquisition or data analysis, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, MZC, CLK, GHL; clinical studies, XYG, YNC; experimental studies, XC, MJC, CSS, MX; statistical analysis, WYC; and manuscript editing, MZC, CL K; all authors have accessed and verified the study data.

Data sharing statement

Data generated or analysed during the study are available from the corresponding author by request. Inference code and illustrative examples of the proposed CNN model are publicly available on GitHub (https://github.com/wen-alan/BC_ALN).

Declaration of interests

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

Acknowledgements

This study was supported by the National Key Research and Development projects intergovernmental cooperation in science and technology of China (2018YFE0126900), National Natural Science Foundation of China (82102162 and 82072026), Natural Science Foundation of Zhejiang Province (LGF21H180002), Zhejiang Medical and Health Science Project (2022RC087).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.eclinm.2023.102176>.

References

- Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209–249.
- Danko ME, Bennett KM, Zhai J, et al. Improved staging in node-positive breast cancer patients using lymph node ratio: results in 1,788 patients with long-term follow-up. *J Am Coll Surg*. 2010;210:797–805.
- Ahmed M, Purushotham AD, Douek M. Novel techniques for sentinel lymph node biopsy in breast cancer: a systematic review. *Lancet Oncol*. 2014;15:e351–e362.
- DeSantis CE, Ma J, Gaudet MM, et al. Breast cancer statistics, 2019. *CA Cancer J Clin*. 2019;69:438–451.
- Galimberti V, Cole BF, Viale G, et al. Axillary dissection versus no axillary dissection in patients with breast cancer and sentinel-node micrometastases (IBCSG 23-01): 10-year follow-up of a randomised, controlled phase 3 trial. *Lancet Oncol*. 2018;19:1385–1393.
- National Comprehensive Cancer Network. (NCCN) Clinical practice guidelines in oncology. Breast Cancer Version 3. Natl compr cancer network clin pract guidel oncol https://www.nccn.org/professionals/physician_gls/default.aspx#breast; 2019. Accessed September 6, 2019.

- Chu KU, Turner RR, Hansen NM, et al. Sentinel node metastasis in patients with breast carcinoma accurately predicts immunohistochemically detectable nonsentinel node metastasis. *Ann Surg Oncol*. 1999;6(8):756–761.
- Kamath VJ, Giuliano R, Dauway EL, et al. Characteristics of the sentinel lymph node in breast cancer predict further involvement of higher-echelon nodes in the axilla: a study to evaluate the need for complete axillary lymph node dissection. *Arch Surg*. 2001;136(6):688–692.
- Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. *Radiology*. 2004;233:830–849.
- Yu Y, He Z, Ouyang J, et al. Magnetic resonance imaging radiomics predicts preoperative axillary lymph node metastasis to support surgical decisions and is associated with tumor microenvironment in invasive breast cancer: a machine learning, multicenter study. *EBioMedicine*. 2021;69:103460.
- Yu YF, Tan YJ, Xie CM, et al. Development and validation of a preoperative magnetic resonance imaging radiomics-based signature to predict axillary lymph node metastasis and disease-free survival in patients with early-stage breast cancer. *JAMA New Open*. 2020;3(12):e2028086.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436–444.
- Bi WL, Hosny A, Schabath MB, et al. Artificial intelligence in cancer imaging: clinical challenges and applications. *CA Cancer J Clin*. 2019;69:127–157.
- Soffer S, Ben-Cohen A, Shimon O, et al. Convolutional neural networks for radiologic images: a radiologist's guide. *Radiology*. 2019;290(3):590–606.
- Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. *Med Image Anal*. 2017;42:60–88.
- Ohno Y, Aoyagi K, Yaguchi A, et al. Differentiation of benign from malignant pulmonary nodules by using a convolutional neural network to determine volume change at chest CT. *Radiology*. 2020;296(2):432–443.
- Hallinan JTPD, Zhu L, Yang K, et al. Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology*. 2021;300(1):130–138.
- Zhou LQ, Wu XL, Huang SY, et al. Lymph node metastasis prediction from primary breast cancer US images using deep learning. *Radiology*. 2020;294(1):19–28.
- Fu S, Lai H, Huang M, et al. Multi-task deep learning network to predict future macrovascular invasion in hepatocellular carcinoma. *EClinicalMedicine*. 2021;42:101201.
- Ji JS. Intelligent assessment of lymph nodes involvement in breast cancer based on CNN algorithm. <https://www.chictr.org.cn/showproj.html?proj=193209>.
- He K, ZhangX, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016:770–778.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837–845.
- Selvaraju RR, Cogswell M, Das A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Int J Comput Vis*. 2020;128:336–359.
- Bevilacqua JL, Kattan MW, Fey JV, et al. Doctor, what are my chances of having a positive sentinel node? A validated nomogram for risk estimation. *J Clin Oncol*. 2007;25(24):3670–3679.
- Kim GR, Choi JS, Han BK, et al. Preoperative axillary US in early-stage breast cancer: potential to prevent unnecessary axillary lymph node dissection. *Radiology*. 2018;288(1):55–63.
- Yajima R, Fujii T, Yanagita Y, et al. Prognostic value of extracapsular invasion of axillary lymph nodes combined with peritumoral vascular invasion in patients with breast cancer. *Ann Surg Oncol*. 2015;22(1):52–58.
- Viale G, Zurrida S, Maiorano E, et al. Predicting the status of axillary sentinel lymph nodes in 4351 patients with invasive breast carcinoma treated in a single institution. *Cancer*. 2005;103(3):492–500.
- Bae MS, Lee SH, Chu AJ, et al. Preoperative MR imaging in women with breast cancer detected at screening US. *Radiology*. 2017;282(3):681–689.
- Han L, Zhu Y, Liu Z, et al. Radiomic nomogram for prediction of axillary lymph node metastasis in breast cancer. *Eur Radiol*. 2019;29(7):3820–3829.

-
- 30 Bian Y, Zheng Z, Fang X, et al. Artificial intelligence to predict lymph node metastasis at CT in pancreatic ductal adenocarcinoma. *Radiology*. 2023;306(1):160–169.
 - 31 Xu X, Xi L, Wei L, et al. Deep learning assisted contrast-enhanced CT-based diagnosis of cervical lymph node metastasis of oral cancer: a retrospective study of 1466 cases. *Eur Radiol*. 2022;33:4303–4312.
 - 32 Gao J, Zhong X, Li W, et al. Attention-based deep learning for the preoperative differentiation of axillary lymph node metastasis in breast cancer on DCE-MRI. *J Magn Reson Imaging*. 2022;57(6):1842–1853. <https://doi.org/10.1002/jmri.28464>.
 - 33 Ha R, Chang P, Karcich J, et al. Axillary lymph node evaluation utilizing convolutional neural networks using MRI dataset. *J Digit Imaging*. 2018;31(6):851–856.
 - 34 Ren T, Lin S, Huang P, et al. Convolutional neural network of multiparametric MRI accurately detects axillary lymph node metastasis in breast cancer patients with PreNeoadjuvant chemotherapy. *Clin Breast Cancer*. 2022;22(2):170–177.