

# Machine learning magnetic resonance imaging radiomics predicts axillary lymph node metastasis in invasive breast cancer

**Yunfang Yu**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University <https://orcid.org/0000-0003-2579-6220>

**Zifan He**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Jie Ouyang**

Tungwah Hospital, Sun Yat-sen University

**Yujie Tan**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Yong-Jian Chen**

Department of Medical Oncology and Guangdong Key laboratory of Liver Disease Research, the Third Affiliated Hospital of Sun Yat-sen University

**Yang Gu**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Luhui Mao**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Jue Wang**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Wei Ren**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Lili Lin**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Zhuo Wu**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Jingwen Liu**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Anlin Li**

The First Clinical Medical College, Guangdong Medical University

**Qiyun Ou**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Qiugen Hu**

Shunde Hospital, Southern Medical University

**Kai Chen**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Chenchen Li**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Nian Lu**

Sun Yat-sen University Cancer Center, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine

**Xiaohong Li**

Shunde Hospital, Southern Medical University

**Fengxi Su Su**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

**Qiang Liu**

Sun Yat-sen Memorial Hospital

**Chuan-Miao Xie**

Sun Yat-sen University Cancer Center

**Herui Yao (✉ [yaoherui@mail.sysu.edu.cn](mailto:yaoherui@mail.sysu.edu.cn))**

Sun Yat-sen Memorial Hospital, Sun Yat-sen University

---

**Article**

**Keywords:** Signature Incorporated Tumor, Molecular Subtypes, Early-stage Invasive Breast Cancer Patients, Multiomic Radiomic Signature, Tumor Microenvironment

**Posted Date:** November 21st, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-104661/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

In current clinical practice, the standard evaluation for axillary lymph node (ALN) status in breast cancer is based on the invasive procedure and many patients will suffer from operative associated complications. Hence, a novel signature incorporated tumor and lymph node magnetic resonance imaging (MRI) radiomics, clinical and pathological characteristics, and molecular subtypes based on the machine learning approach was established to accurately identify ALN metastasis in early-stage invasive breast cancer patients. Although the misjudgment of ALN status by clinicians according to preoperative MRI are common during clinical practice and even the senior radiologists make mistakes sometimes, this multiomic radiomic signature showed the superiority over clinicians and could precisely discriminate ALN metastasis among different molecular subtype patients. Furthermore, the association between MRI radiomic features and tumor-microenvironment features including immune cells, long non-coding RNAs, and types of methylated sites were found, which revealed the potential biological underpinning of MRI radiomics.

## Introduction

The importance of axillary lymph node (ALN) status in early-stage invasive breast cancer was highlighted in 10-year follow-up of the ACOSOG Z0011 and IBCSG 23-01 randomized clinical trials.<sup>1-2</sup> Since sentinel lymph node biopsy (SLNB) was included in the breast cancer standard diagnosis procedure from 2005, it have greatly improved the accuracy of ALN status judgment and changed the traditional concept that women could be away from axillary lymph nodes dissection (ALND) if they had fewer than three positive sentinel lymph nodes.<sup>3-4</sup> However, SLNB is still an invasive procedure that may bring complications such as lymphedema and upper limb numbness and its accuracy is also limited by the radiopharmaceuticals and modalities of radiocolloid injection.<sup>5-6</sup> As a consequence, it's urgent and indispensable to develop a reliable non-invasive evaluation approach of ALN status in patients who had early-stage invasive breast cancer.

Nowadays, the development of radiomics in the aspects of tumor diagnosis, treatment decision and prognosis prediction is encouraging and breast cancer is one of the pioneers in the exploration.<sup>7-9</sup> Magnetic resonance imaging (MRI) radiomics-based approaches for predicting molecular subtype, recurrence risk and survival, pathologic complete responses of breast cancer patients showed powerful predictive ability and the prospect of extended application.<sup>10-13</sup> Moreover, the association between MRI radiomics with breast tumor microenvironment have been corroborated in previous studies.<sup>14-15</sup> As MRI examination has been recommended for preoperative estimation of breast cancer patients, in this retrospective multicenter study, a new preoperative non-invasive approach based on MRI radiomics was established to predict ALN metastasis in early-stage invasive breast cancer patients, which reduces unnecessary SLNB or ALND and would decrease the incidence of associated postoperative complications to improve the life quality of patients.

# Results

## Patients' characteristics

A total of 1,161 early-stage invasive breast cancer patients were retrospectively recruited, of which 1,088 patients were eligible for this study. The study design was shown in Supplementary Fig. 1 and the study workflow was shown in Fig. 1. Table 1 showed the clinicopathological characteristics of patients in the training cohort (n=803), the prospective-retrospective validation cohort (n=106), and the external validation cohort (n=179). 389 (35.75%) of 1,088 patients were initially diagnosed as positive ALN status by radiologists through MRI, but 106 (27.25%) of 389 patients didn't have ALN metastasis (ALNM) and were confirmed as negative ALN status with pathological examination. Rather, 190 (28.27%) of 672 patients with clinical negative ALN status were found to have pathological positive ALN status.

## Distinguishing ALN metastasis by tumor and lymph node radiomic signature

The key radiomic features of ALN and tumor regions were selected from a total of 5,178 quantitative features by the random forest algorithm. Overall distribution of key radiomic features from contrast-enhanced T1-weighted imaging (T1+C), T2-weighted imaging (T2WI), and diffusion-weighted imaging quantitatively measured apparent diffusion coefficients (DWI-ADC) sequences among patients with and without ALNM in the training cohort was demonstrated in Fig. 2A. Incorporating these key features of ALN region to predict ALNM yielded AUC values of 0.85, 0.61, and 0.81 in the training cohort, the prospective-retrospective validation cohort, and the external validation cohort, respectively (Supplementary Fig. 2). Simultaneously, incorporating three-sequence key features of tumor region for ALNM prediction achieved an AUC of 0.78, 0.59, and 0.63 in the training cohort, the prospective-retrospective validation cohort, and the external validation cohort, respectively (Supplementary Fig. 3). The AUC values for ALNM prediction of combining multi-sequence features were higher than incorporating single-sequence features in the training and the validation cohorts (Supplementary Table 1).

When combined both ALN and tumor regions features, the ALN-tumor radiomic signature was constructed and illustrated good performance in detecting ALNM in the training cohort (AUC, 0.88), the prospective-retrospective validation cohort (AUC, 0.87), and the external validation cohort (AUC, 0.87), which outperformed the ALN or tumor radiomic signature alone (Fig. 2B). The detailed evaluation indicators for model performance including sensitivity and specificity were summarized in Table 2.

## Distinguishing ALN metastasis by multiomic radiomic signature

In the univariate analysis, which was presented in Supplementary Table 2, five differentially expressed clinical characteristics were found to be associated with ALN status in the training cohort, including age

( $P = 0.012$ ), clinical T stage ( $P < 0.001$ ), clinical N stage ( $P < 0.001$ ), Ki67 expression ( $P = 0.010$ ), and molecular subtype ( $P = 0.033$ ). To develop a more precise and clinically applicable method that could predict an individual's ALN status, the multiomic radiomic signature incorporated all key radiomic features of ALN and tumor region with clinical characteristics, pathological characteristics and molecular subtype that significantly associated with ALNM was built and showed better performance of ALNM prediction, which achieved the higher AUC (0.90) in the training cohort, the prospective-retrospective validation cohort (AUC, 0.93), and the external validation cohort (AUC, 0.91), respectively (Fig. 2C). The sensitivity and specificity of the multiomic radiomic signature were list in Table 2. As for combination of these clinicopathological characteristics and molecular subtype, it achieved an AUC of 0.74, 0.68, and 0.72 for predicting ALNM in the training cohort, the prospective-retrospective validation cohort, and the external validation cohort, respectively (Supplementary Fig. 4).

The multiomic radiomic signature also presented the ability of discriminating ALNM patients with 1, 2, and 3 positive nodes (AUC of 0.88, 0.89 and 0.92 in the training cohort; AUC of 0.79, 1.00 and 0.93 in the prospective-retrospective validation cohort; AUC of 0.97, 0.93 and 0.87 in the external validation cohort; Supplementary Table 3).

Besides, to further assess the added value of the multiomic radiomic signature to the ALN status, we conducted subgroup analysis within patients with different molecular subtypes. Encouragingly, the multiomic radiomic signature could identify ALNM patients in the subgroups of Luminal A (AUC, 0.91, 0.91, respectively), Luminal B (AUC, 0.89, 0.92, respectively), human epidermal growth factor receptor 2 (Her-2) positive (AUC, 0.90, 0.76, respectively), and triple-negative breast cancer (TNBC) (AUC, 1.00, 1.00, respectively) patients in the training and external validation cohorts (Supplementary Table 3). In addition, when stratified by other factors such as age, Ki67 expression level, and clinical T stage, the AUC value of the multiomic radiomic signature remained at 0.86-1.00 in these subgroups (Supplementary Table 3).

Decision curve analysis (DCA) was conducted to determine the clinical usefulness of the multiomic radiomic signature by quantifying the net benefits at different threshold probabilities. The decision curve showed that if the threshold probability is  $> 5\%$ , using the multiomic radiomic signature to predict ALNM adds more benefit than either the ALN-tumor radiomic signature or the radiologists' diagnosis (Fig. 2D).

According to the radiomic score of the multiomic radiomic signature, an optimal cutoff value (0.334) was generated to classify patients into high- and low-score groups in the training cohort. High-score patients with high risk of ALNM had significantly shorter DFS compared with the low-score group (HR 0.43, 95% CI 0.21-0.86,  $P=0.014$ ; Supplementary Fig. 5).

During clinical practice, patients' clinical ALN status judged by the preoperative MRI are commonly inconsistent with the pathological one and even the senior radiologists make mistakes sometimes. The multiomic radiomic signature could precisely recognize ALNM among patients with different clinical tumor stage in the training and the validation cohorts (Supplementary Table 3). As shown in Fig. 3, patient 1 had pathological positive ALN status but was considered as a non-ALNM patient by radiologists though MRI before surgery. On the contrary, patient 2 was initially diagnosed as ALNM by radiologists but

found to be a non-ALNM patient by pathological examination. The ALN status of two patients could be accurately assessed through the multiomic radiomic signature by the cutoff value of 0.334.

### **Radiomics associated with tumor microenvironment**

According to the cutoff values of radiomic score from T1+C and T2WI sequences signatures of tumor region in the training cohort, 91 breast cancer patients from The Cancer Genome Atlas (TCGA) and The Cancer Imaging Archive (TCIA) with T1+C and T2WI sequences MRI were classified into two group. In total, 1381 T1+C and T2WI sequences-based differentially expressed genes (DEGs) were obtained among low-score and high-score patients. Next, the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed to further examine the biological functions of the identified radiomic-based genes. The GO enrichment analysis indicated that the radiomic-based genes were enriched in various physiological metabolic processes, such as affection of transmembrane transporter activity, NADP binding and ATPase complex (Supplementary Table 4 and Supplementary Fig. 6). The KEGG pathway enrichment analysis found these genes were involved in the oxidative phosphorylation signaling pathway (Supplementary Table 4).

We further explored the association between MRI radiomic features and tumor microenvironment including 22 immune cells, long non-coding RNAs, and types of methylated sites in these patients. The expression level of 395 long non-coding RNAs and the enrichment level of 1784 types of methylated sites were found significantly different among ALNM and non-ALNM patients and the top 30 one were selected via random forest algorithm for further analysis. Overall distribution of key tumor-microenvironment features among patients with and without ALNM was shown in Fig. 4A. The key radiomic features of ALN and tumor region were found to be remarkably correlated linearly with immune cells like the M0 macrophages, B naïve cells and neutrophils, long non-coding RNAs like P11.563P16.1 and RP11.888D10.3, and types of methylated sites like cg14681629 and cg02784848 (Fig. 4B-D).

## **Discussion**

Unlike previous studies just focusing on the tumor region for predicting ALNM,<sup>16-18</sup> in this study, ALN-tumor radiomic signature that combined multi-sequence key radiomic features of ALN and tumor regions showed high predictive ability and could be applied to predict the ALN status precisely and non-invasively. The multiomic radiomic signature integrated radiomic features above and ALNM-associated clinicopathologic characteristics, molecular subtype demonstrated better performance for predicting ALNM. In addition, the multiomic radiomic signature contributed to improve the accuracy of judgement for ALN status in patients who had inconsistent clinical ALN status with pathological ALN status, and it could accurately recognize ALNM in patients with different number of positive lymph nodes, different clinical T stage, different age, different Ki67 expression level, and even different molecule subtypes. Besides, the key radiomic features were also found to be associated with tuomr microenvironment.

Therefore, the multiomic radiomic signature was an effective tool for ALNM prediction that provide useful message for diagnosis and treatment decision preoperatively and meanwhile could preventing unnecessary ALND and even SLNB in early-stage invasive breast cancer patients.

Breast cancer patients have distinct ALN status when they are newly diagnosed, and the treatment plans are so as different depend on their ALN status. It has been demonstrated that patients with positive ALN status had poorer outcome than patients with negative ALN status.<sup>19</sup> It matters clinician's decision-making of therapy that patients with positive ALN status are consider as high-risk patients and need to undergo adjuvant chemotherapy according to NCCN guideline.<sup>20</sup> Parts of TNBC patients with lager tumor burden and clinical positive ALN status are consider receiving neoadjuvant therapy. It should be noted that the misjudgment of the ALN status is a cause of patients' overtreatment and waste in health care. As for the patients with negative ALN status and without other risk factors, they could be treated by endocrine therapy alone, which means lower cost and more comfortable treatment experience. Thus, it's very important to accurately and clearly distinguish the ALN status for patients with early-stage breast cancer.

Besides MRI examination, ultrasound is also favored by clinicians in clinical practice due to the wide application and easily accessible data for study. Previous research demonstrated that an approach based on ultrasound radiomics also showed acceptable performance for ALN status prediction in early-stage breast cancer patients.<sup>7,21</sup> However, compared with MRI examination, operators' experience and heterogeneity from the angle of incidence of the ultrasound beam often affect the standardization of ultrasound image data, which result in the lack of reproducibility of model. As MRI examination possesses the advantage of showing the three-dimensional spatial position of each organ and tissue with high resolution, it could provide more comprehensive information for diagnosis and TNM staging, and therefore has been recommended for lesions difficult for the ultrasound and mammography and for preoperative estimation of breast cancer patients.<sup>22</sup>

Additionally, the ALN-tumor signature based on MR radiomic in this study illustrated the superiority over general radiologists in detecting ALNM. Hence, it may be beneficial to combine MR radiomics with ultrasound radiomics, which might achieve better performance for ALNM prediction, especially improve the diagnosis and treatment level of hospitals in rural area and make up the disadvantages of areas with insufficient medical resources in the future.

Previous studies demonstrated that the tumor microenvironment harbored variety of immune cells, blood vessels and extracellular matrix, and the changes in the distribution of immune cells and blood vessel formation could facilitate the development and metastasis of tumors.<sup>23</sup> Therefore, many tumor microenvironment related factors like cancer-associated fibroblasts, myeloid-derived suppressor cells, and tumor associated macrophages have been used to predict tumor metastasis.<sup>24-26</sup> It has been reported that radiomic features could provide large amount of information of tumor microenvironment.<sup>27</sup> In this study, the ALN-tumor radiomic signature based on the integration of radiomic features from the tumor and the ALN regions could effectively predict ALNM, which also confirmed the correlation between

MRI radiomics and the breast tumor microenvironment. Combination of radiomics with tumor microenvironment is promising to improve the predictive ability of ALNM.

Several limitations still need to be addressed in our study. First, the clinical information and MRI sequences of some patients were missing due to retrospective nature, which led to a decrease in sample size. Second, we failed to construct a ALNM prediction model with the combination of tumor-microenvironment features, MR radiomics and ultrasound radiomics, but further exploration could be conducted in the future to realize more precisely prediction for ALNM.

In conclusion, this study presented a multiomic radiomic signature that incorporated MRI multi-sequence key radiomic features of ALN and tumor regions with clinicopathological characteristics and molecular subtype, which could be conveniently used for identifying ALNM patients among different molecular subtypes in early-stage invasive breast cancer and be promising to eventually result in a noninvasive approach to guide future clinical practice. Furthermore, the association between MRI radiomic features and tumor-microenvironment features including immune cells, long non-coding RNAs, and types of methylated sites were found, which revealed the potential biological underpinning of MRI radiomics.

## Method

### Patients and study design

This multicenter study was conducted in accordance with the Declaration of Helsinki. The study's protocol was approved by the ethics committee of each participating hospital (Sun Yat-sen Memorial Hospital of Sun Yat-sen University, SYSEC-KY-KS-2019-054-001; Sun Yat-sen University Cancer Center, B2020-114-01; Shunde Hospital of Southern Medical University, KYLS-20190579; Tungwah Hospital of Sun Yat-sen University, 2020DHLL018). The requirement for informed consent in retrospective cohorts was waived. Participants from the prospective phase III clinical trials [NCT01503905]<sup>28</sup> have signed informed consent; the trial was also approved by the ethics committee with number [2011] EC # (12).

A total of 1,161 early-stage invasive breast cancer patients were recruited from four institutions in China, of which 1,088 patients passed quality control. The inclusion criteria included the followings: (a) female patients aged at least 18 years, who had histologically confirmed staged I-III invasive breast cancer;<sup>29</sup> (b) patients who had been treated with surgery and SLNB or ALND, and had been pathologically confirmed ALN status; (c) preoperative MRI scan of breast tumor and/or ALN were conducted, including contrast-enhanced T1-weighted imaging (T1+C), T2-weighted imaging (T2WI), or diffusion-weighted imaging quantitatively measured apparent diffusion coefficients (DWI-ADC). The exclusion criteria included the followings: (a) patients underwent biopsy at an external institution and pathological results were not available; (b) patients suffering from other tumor diseases before or at the same time. The primary outcome was ALN metastasis status.

The patients were divided into three cohorts: the training cohort (n= 803 from Sun Yat-sen Memorial Hospital of Sun Yat-sen University, a national hospital [Guangzhou, China] and Sun Yat-sen University

Cancer center, a national hospital [Guangzhou, China]), the prospective-retrospective validation cohort (n= 106 from a prospective phase 3 clinical trials [NCT01503905] collected from Sun Yat-sen University, a national hospital [Guangzhou, China]), and the external validation cohort (n= 179 from the Shunde Hospital of Southern Medical University [Foshan, China] and Tungwah Hospital of Sun Yat-sen University [Dongguan, China]).

Breast tumors were classified into four molecular subtypes according to St. Gallen Consensus Conference 2013.<sup>30</sup> Patients with estrogen receptor (ER) and progesterone receptor (PR) positive, human epidermal growth factor receptor 2 (HER2) negative and Ki-67's level of < 14% breast tumors were defined as Luminal A subtype patients. ER positive and HER2 over-expressed or amplified patients, or ER positive and HER2 negative patients with at least one of situations: Ki-67's level of > 14%, PR negative or low, were considered as Luminal B subtype patients. While ER and PR absent, HER2-positive subtype patients were distinguished by HER2 over-expressed or amplified, and HER2 negative breast tumors were classified as triple negative subtype. These biomarkers were measured with immunohistochemical methods or in situ hybridization.

## **Radiomic Features extraction**

The protocol of multi-sequence MRI acquisition across four institutions and parameters of MR scanners for patients were shown in Supplementary Table 5. For all cohorts, multi-sequence MRI images from all cohorts were retrieved from Picture Archiving and Communication System, and radiomic features corresponding to the quantitative data obtained after computational translation of images were extracted from T1+C, T2WI, and DWI-ADC sequences imaging. Regions of interest of the breast tumor area (ROI-1), and ALN area (ROI-2) were semi-automatically delineated on each slice obtained via T1+C, T2WI and DWI-ADC (delineated with b value of 800 or 1000 s/mm<sup>2</sup> and then copied to the corresponding ADC maps) sequences imaging by 3D Slicer software method (<https://www.slicer.org/>, version 4.10.2).<sup>31</sup> N4ITK Bias Field Correction were applied to obtained standard normal distribution of image intensities in ROI-1 and ROI-2, then the voxel-based features of ROIs were extracted using the SlicerRadiomics extension in 3D Slicer software, the in-house texture extraction platform developed based on the python package "PyRadiomics". A total of 5,718 quantitative radiomic features were extracted from ROI-1s and ROI-2s separately, including six groups of radiomic features: shape, first-order, the gray-level co-occurrence matrix (GLCM), the gray-level size zone matrix (GLSZM), the gray-level dependence matrix (GLDM), and the neighbouring gray tone difference matrix (NGTDM). All patients were separately reassessed by two radiologists (Lu N and Li XH) blinded to the patients' clinical outcomes, MRI reassessed under the guidance of two senior radiologists (Xie CM and Wu Z) who major in MRI interpretation more than 20 years.

All ROIs of each MRI scan for each patient was normalization separately using Z-score to obtain a standard normal distribution of image intensities before radiomic features selection. The Random forest

algorithm<sup>32</sup> was used to select the top 30 radiomic features from T1+C, T2WI, and DWI-ADC sequence of ROI-1, and ROI-2, respectively, in the training cohort.

### **Tumor and lymph node radiomic signature construction and validation**

The signature were built based on the training cohort with MRI radiomic features that extracted from the breast primary tumor and the ALN regions of T1+C, T2WI, and DWI-ADC sequences. First, t-test was used to detect the associations between each feature and patients' ALN status. Next, to reduce the risk of bias and potential overfitting, MRI radiomic features that achieved significance at  $P < 0.05$  were entered into the further selection and top 30 features identified by random forest algorithm were selected from the each T1+C, T2WI, and DWI-ADC sequences for the construction of the signature via the support vector machine (SVM) algorithm.<sup>33</sup> The predictive accuracy of the ALN-tumor radiomic signature for predicting ALNM was initially assessed in the training cohort and then validated in the prospective-retrospective validation and external validation cohorts using receiver operating characteristic (ROC) curve analysis.

### **Multiomic radiomic signature construction and validation**

The independent-sample t-test was used to assess the association between clinical characteristics, pathological characteristics, molecular subtype and ALN status in the training cohort. To provide the clinician a quantitative tool to predict individual probability of ALNM, all these characteristics with  $P < 0.05$  and key MRI radiomic features were used for the construction of the multiomic radiomic signature via the SVM algorithm. ROC curve analysis was also performed to evaluate the performance of the multiomic radiomic signature in the training cohort and was validated in the prospective-retrospective and the external validation cohorts.

### **MRI radiomics features associated with tumor microenvironment**

The t-test was utilized through to identify differentially expressed genes<sup>34</sup> associated with the radiomic score in 91 patients with T1+C and T2WI sequences MRI from TCGA and TCIA. The GO and KEGG analysis were performed using the clusterProfiler R package<sup>35</sup>. The GO terms and KEGG pathways were considered statistically significant with  $P$  values and false discovery rates less than 0.05.

The CIBERSORT algorithm<sup>36</sup> and the LM22 gene signature were used for highly sensitive and specific discrimination of 22 human immune cell phenotypes. CIBERSORT is a deconvolution algorithm that uses a set of reference gene expression values (a signature with 547 genes) that is considered a minimal representation for each cell type. Based on those values, CIBERSORT infers cell type proportions in data from bulk tumor samples with mixed cell types using support vector regression. Gene expression profiles

were prepared using standard annotation files, the data were uploaded to the CIBERSORT web portal (<http://cibersort.stanford.edu/>), and the algorithm was run using the LM22 signature at 1,000 permutations.

12,578 long non-coding RNAs data of transcriptome RNA sequencing based on Illumina platform and 23,381 types of methylated sites data from both Illumina Human Methylation 27 and Illumina Human Methylation 450 platform were used for analysis. The top 30 ALNM-associated long non-coding RNAs and types of methylated sites were selected using the random forest algorithm. Correlation analysis and weighted linear regression models were used to estimate the strength of the correlations with Pearson  $\rho$ .

### Statistical analysis.

The  $\chi^2$  test was performed to examine the differences in categorical variables, and the independent t-test was applied to compare the differences in continuous variables between two groups. The predictive accuracy of the signatures was assessed by using ROC analysis. The area under ROC curve (AUC) was used to evaluate the performance of sensitivity and specificity in each signature. Additionally, DCA was performed to assess the clinical utility of the prediction model by quantifying the net benefits when different threshold probabilities were considered.<sup>37</sup> Patients were categorized into high- and low-score groups with the optimal cutoff values defined by the R package `cutpointr`. Heatmaps were generated to show the distribution and expression levels of radiomic features, immune cells, long non-coding RNAs and types of methylated sites. Survival was calculated using the Kaplan-Meier method and the log-rank test, and hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated using a Cox regression analysis. For all the analyses, two-sided P-values less than 0.05 were considered statistically significant. Statistical analyses were performed using R software (version 4.0.2). This study is registered with ClinicalTrials.gov, number NCT04003558.

## References

1. Armando E Giuliano, Karla V Ballman, Linda McCall, et al. Effect of Axillary Dissection vs No Axillary Dissection on 10-Year Overall Survival Among Women With Invasive Breast Cancer and Sentinel Node Metastasis: The ACOSOG Z0011 (Alliance) Randomized Clinical Trial. *JAMA*. **318**(10):918-926. (2017).
2. Viviana Galimberti, Bernard F Cole, Giuseppe Viale, 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*. **19**(10):1385-1393. (2018).
3. Lyman GH, Giuliano AE, Somerfield MR, et al. American Society of Clinical Oncology guideline recommendations for sentinel lymph node biopsy in early-stage breast cancer. *J Clin Oncol*. **23**:7703–7720. (2005).

4. Gary H Lyman et al. Sentinel Lymph Node Biopsy for Patients With Early-Stage Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update J Clin Oncol. **35**(5):561-564. (2017).
5. Giammarile F, Alazraki N, Aarsvold JN, et al. The EANM and SNMMI practice guideline for lymphoscintigraphy and sentinel node localization in breast cancer. Eur J Nucl Med Mol Imaging. **40**:1932–1947. (2013).
6. George E Naoum et al. Quantifying the Impact of Axillary Surgery and Nodal Irradiation on Breast Cancer-Related Lymphedema and Local Tumor Control: Long-Term Results From a Prospective Screening Trial J Clin Oncol. **38**(29):3430-3438. (2020).
7. Xueyi Zheng et al. Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer. Nat Commun. **11**(1):1236. (2020).
8. Yan-Qi Huang et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol. **34**(18):2157-64. (2016).
9. Zhenyu Liu et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal. Cancer Clin Cancer Res. **23**(23):7253-7262. (2017).
10. Hyunjin Park et al. Radiomics Signature on Magnetic Resonance Imaging: Association with Disease-Free Survival in Patients with Invasive Breast Cancer. Clin Cancer Res. **24**(19):4705-4714. (2018).
11. Hui Li et al. MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays. Radiology. **281**(2):382-391. (2016).
12. Zhenyu Liu et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. Clin Cancer Res. **25**(12):3538-3547. (2019).
13. Tianwen Xie et al. Machine Learning-Based Analysis of MR Multiparametric Radiomics for the Subtype Classification of Breast Cancer. Front Oncol. **9**:505. (2019).
14. Shota Yamamoto, Wonshik Han, Youngwoo Kim, et al. Breast Cancer: Radiogenomic Biomarker Reveals Associations among Dynamic Contrast-enhanced MR Imaging, Long Noncoding RNA, and Metastasis. Radiology. **275**(2):384-92. (2015).
15. Nathaniel Braman, Prateek Prasanna, Jon Whitney, et al. Association of Peritumoral Radiomics With Tumor Biology and Pathologic Response to Preoperative Targeted Therapy for HER2 (ERBB2)-Positive Breast Cancer. JAMA Netw Open. **2**(4):e192561. (2019).
16. Jia Liu, Dong Sun, Linli Chen, et al. Radiomics Analysis of Dynamic Contrast-Enhanced Magnetic Resonance Imaging for the Prediction of Sentinel Lymph Node Metastasis in Breast Cancer. Front Oncol. **9**:980. (2019).
17. Lu Han, Yongbei Zhu, Zhenyu Liu, et al. Radiomic nomogram for prediction of axillary lymph node metastasis in breast cancer. Eur Radiol. **29**(7):3820-3829. (2019).

18. Xiaoyu Cui, Nian Wang, Yue Zhao, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer using Radiomics Features of DCE-MRI. *Sci Rep.* **9**(1):2240. (2019).
19. Christoph Tausch, Susanne Taucher, Peter Dubsy, et al. Prognostic value of number of removed lymph nodes, number of involved lymph nodes, and lymph node ratio in 7502 breast cancer patients enrolled onto trials of the Austrian Breast and Colorectal Cancer Study Group (ABCSG). *Ann Surg Oncol.* **19**(6):1808-17. (2012).
20. NCCN Clinical Practice Guidelines in Oncology Breast Cancer Version 4.2020 – May 8, 2020.
21. Ohuchi N, Suzuki A, Sobue T, et al. JSTART investigator groups. Sensitivity and specificity of mammography and adjunctive ultrasonography to screen for breast cancer in the Japan Strategic Anti-cancer Randomized Trial (J-START): a randomised controlled trial. *Lancet.* **387**:341–348. (2016).
22. Min Sun Bae et al. Preoperative MR Imaging in Women with Breast Cancer Detected at Screening US. *Radiology.* **282**(3):681-689. (2017).
23. Arneth B. Tumor Microenvironment. *Medicina (Kaunas).* 56(1):15. (2019).
24. J M Houthuijzen, J Jonkers, et al. Cancer-associated fibroblasts as key regulators of the breast cancer tumor microenvironment. *Cancer Metastasis Rev.* **37**(4):577-597. (2018).
25. Elham Safaradz, et al. Myeloid-derived suppressor cells: Important contributors to tumor progression and metastasis. *J Cell Physiol.* **233**(4):3024-3036. (2018).
26. YuxinLin, JianxinXu, et al. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *J Hematol Oncol.* **12**(1):76. (2019).
27. Sun R, Limkin EJ, Vakalopoulou M, Derclé L, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* **19**(9):1180-1191. (2018).
28. Armando E Giuliano, Stephen B Edge, Gabriel N Hortobagyi. Eighth Edition of the AJCC Cancer Staging Manual: Breast Cancer. *Ann Surg Oncol.* **25**(7):1783-1785. (2018).
29. Fengxi Su. Neoadjuvant chemotherapy for operable premenopausal breast cancer patients. Updated May 24, 2106. Accessed November 4, 2020.

<https://www.clinicaltrials.gov/ct2/show/NCT01503905?cond=NCT01503905&draw=2&rank=1>.

30. A Goldhirsch, E P Winer et al. The treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. *Ann Oncol.* **24**(9):2206-23. (2013).
31. Fedorov, A. et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn Reson Imaging.* **30**(9):1323-41 (2012).
32. Breiman, L. Random forests. *Mach Learn.* **45**:5-32 (2001).
33. William S Noble. What is a support vector machine? *Nat Biotechnol.* **24**:1565-7 (2006).

34. Ritchie, M.E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**(7):e47. (2015).
35. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* **16**(5):284-7. (2012).
36. Newman, A.M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods.* **12**(5):453-7. (2015).
37. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol.* **34**(21):2534-40. (2016)
38. Carstensen B, Plummer M, Laara E, Hills M. Epi: A Package for Statistical Analysis in Epidemiology. R package version 2.41. 2020. <http://CRAN.R-project.org/package=Epi>.

## Declarations

### Article information

**Author Contributions:** All authors had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Yunfang Yu, Zifan He, Jie Ouyang, Yujie Tan are co-first authors. Herui Yao, Chuanmiao Xie are co-corresponding authors.

**Concept and design:** All authors.

**Acquisition, analysis, or interpretation of data:** Herui Yao, Yunfang Yu, Yujie Tan, Zifan He, Chuanmiao Xie.

**Drafting of the manuscript:** All authors.

**Critical revision of the manuscript for important intellectual content:** All authors.

**Statistical analysis:** All authors.

**Obtained funding:** Herui Yao.

**Administrative, technical, or material support:** Herui Yao, Yunfang Yu, Yujie Tan, Zifan He, Chuanmiao Xie.

**Supervision:** Herui Yao.

**Conflict of Interest Disclosures:** The authors have no conflicts of interest to declare.

**Funding:** This study was supported by grant 2020ZX09201021 from the National Science and Technology Major Project, grant YXRGZN201902 from the Medical Artificial Intelligence Project of Sun Yat-Sen Memorial Hospital, grants 81572596, 81972471, and U1601223 from the National Natural Science Foundation of China, grant 2017A030313828 from the Natural Science Foundation of

Guangdong Province, grant 201704020131 from the Guangzhou Science and Technology Major Program, grant 2017B030314026 from the Guangdong Science and Technology Department, grant 2018007 from the Sun Yat-Sen University Clinical Research 5010 Program, grant SYS-C-201801 from the Sun Yat-Sen Clinical Research Cultivating Program.

## **Tables**

<b>Table 1.</b> Clinicopathologic characteristics between axillary lymph node positive and negative groups in the training, prospective-retrospective validation and external validation cohorts						
	<b>Training cohort</b>		<b>Prospective-retrospective validation cohort</b>		<b>External validation cohort</b>	
	(No. of patients[n]=803)		(n=106)		(n=179)	
	<b>Axillary lymph node status</b>		<b>Axillary lymph node status</b>		<b>Axillary lymph node status</b>	
<b>Characteristic</b>	<b>Negative</b>	<b>Positive</b>	<b>Negative</b>	<b>Positive</b>	<b>Negative</b>	<b>Positive</b>
	(n=475)	(n=328)	(n=28)	(n=78)	(n=94)	(n=85)
<b>Age, years (median [IQR])</b>	48 [43, 56]	48 [41, 56]	42 [39, 46]	41 [38, 45]	49 [43, 56]	46 [42, 56]
<b>Number of tumors (%)</b>						
1	421 (88.6)	284 (86.9)	21 (75.0)	62 (79.5)	76 (80.9)	72 (84.7)
>1	54 (11.4)	43 (13.1)	7 (25.0)	16 (20.5)	18 (19.1)	13 (15.3)
<b>Tumor size, cm (median [IQR])</b>	2.0 [1.6, 2.7]	2.5 [2.0, 3.2]	3.6 [2.6, 5.0]	3.1 [2.5, 5.1]	2.3 [1.7, 2.8]	2.8 [2.1, 3.7]
<b>Clinical T stage (%)</b>						
T1	241 (50.7)	97 (29.6)	0 (0.0)	12 (15.4)	39 (41.5)	18 (21.2)
T2	213 (44.8)	209 (63.7)	22 (78.6)	44 (56.4)	50 (53.2)	58 (68.2)
T3	16 (3.4)	15 (4.6)	4 (14.3)	18 (23.1)	3 (3.2)	7 (8.2)
T4	5 (1.1)	7 (2.1)	2 (7.1)	4 (5.1)	2 (2.1)	2 (2.4)
<b>Clinical N stage (%)</b>						
N0	382 (80.4)	132 (40.2)	15 (60.0)	21 (29.6)	85 (96.6)	37 (50.0)
N1	90 (18.9)	178 (54.3)	10 (40.0)	45 (63.4)	3 (3.4)	20 (27.0)
N2	2 (0.4)	18 (5.5)	0 (0.0)	4 (5.6)	0 (0.0)	16 (21.6)
N3	1 (0.2)	0 (0.0)	0 (0.0)	1 (1.4)	0 (0.0)	1 (1.4)
<b>Clinical TNM stage (%)</b>						
I	207	51	0 (0.0)	4 (5.6)	37 (42.0)	12

	(43.6)	(15.5)				(16.2)
II	253 (53.3)	241 (73.5)	22 (88.0)	51 (71.8)	48 (54.5)	42 (56.8)
III	15 (3.2)	36 (11.0)	3 (12.0)	16 (22.5)	3 (3.4)	20 (27.0)
<b>Histological grade (%)</b>						
Grade 1 (low)	22 (5.0)	2 (0.6)	0 (0.0)	0 (0.0)	4 (4.9)	5 (6.7)
Grade 2 (intermediate)	235 (53.0)	149 (46.1)	10 (50.0)	43 (72.9)	46 (56.1)	33 (44.0)
Grade 3 (high)	186 (42.0)	172 (53.3)	10 (50.0)	16 (27.1)	32 (39.0)	37 (49.3)
<b>Pathological T stage (%)</b>						
T1	272 (57.3)	140 (42.7)	19 (67.9)	30 (40.0)	45 (47.9)	23 (27.1)
T2	188 (39.6)	171 (52.1)	7 (25.0)	32 (42.7)	45 (47.9)	55 (64.7)
T3	13 (2.7)	16 (4.9)	1 (3.6)	7 (9.3)	2 (2.1)	6 (7.1)
T4	2 (0.4)	1 (0.3)	1 (3.6)	6 (8.0)	2 (2.1)	1 (1.2)
<b>Pathological TNM stage (%)</b>						
I	271 (57.1)	0 (0.0)	19 (67.9)	0 (0.0)	45 (47.9)	0 (0.0)
II	202 (42.5)	206 (62.8)	8 (28.6)	28 (37.3)	47 (50.0)	44 (51.8)
III	2 (0.4)	122 (37.2)	1 (3.6)	47 (62.7)	2 (2.1)	41 (48.2)
<b>ER status (%)</b>						
Negative	64 (13.6)	51 (15.5)	5 (17.9)	8 (10.3)	20 (21.3)	25 (29.8)
Positive	407 (86.4)	277 (84.5)	23 (82.1)	70 (89.7)	74 (78.7)	59 (70.2)
<b>PR status (%)</b>						
Negative	128 (27.2)	91 (27.7)	14 (50.0)	23 (29.5)	31 (33.0)	29 (34.5)
Positive	343 (72.8)	237 (72.3)	14 (50.0)	55 (70.5)	63 (67.0)	55 (65.5)

<b>HER2 status (%)</b>						
Negative	326 (72.0)	208 (65.4)	15 (68.2)	43 (65.2)	59 (71.1)	44 (62.9)
Positive	127 (28.0)	110 (34.6)	7 (31.8)	23 (34.8)	24 (28.9)	26 (37.1)
<b>Ki67 status (%)</b>						
<30	259 (55.0)	150 (45.7)	12 (42.9)	63 (80.8)	52 (57.1)	49 (59.8)
≥30	212 (45.0)	178 (54.3)	16 (57.1)	15 (19.2)	39 (42.9)	33 (40.2)
<b>Molecular subtypes (%)</b>						
Luminal A	92 (20.0)	30 (9.2)	4 (16.0)	21 (30.0)	20 (23.0)	15 (19.7)
Luminal B	306 (66.7)	249 (76.6)	16 (64.0)	42 (60.0)	50 (57.5)	40 (52.6)
HER2-positive	31 (6.8)	28 (8.6)	1 (4.0)	3 (4.3)	6 (6.9)	12 (15.8)
Triple negative	30 (6.5)	18 (5.5)	4 (16.0)	4 (5.7)	11 (12.6)	9 (11.8)
<b>Type of surgery (%)</b>						
Breast-conserving surgery	250 (52.7)	127 (38.7)	14 (50.0)	28 (35.9)	22 (23.4)	18 (21.2)
Others	224 (47.3)	201 (61.3)	14 (50.0)	50 (64.1)	72 (76.6)	67 (78.8)
<b>Follow-up time, months (median [IQR])</b>	23.4 [15.8, 35.6]	21.7 [15.3, 35.1]	45.5 [21.5, 65.2]	44.6 [25.5, 57.1]	22.6 [10.6, 41.5]	24.3 [9.4, 55.1]
Abbreviations: IQR, interquartile range; TNM, tumor–node–metastasis; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; Ki67, proliferation marker protein Ki-67.						

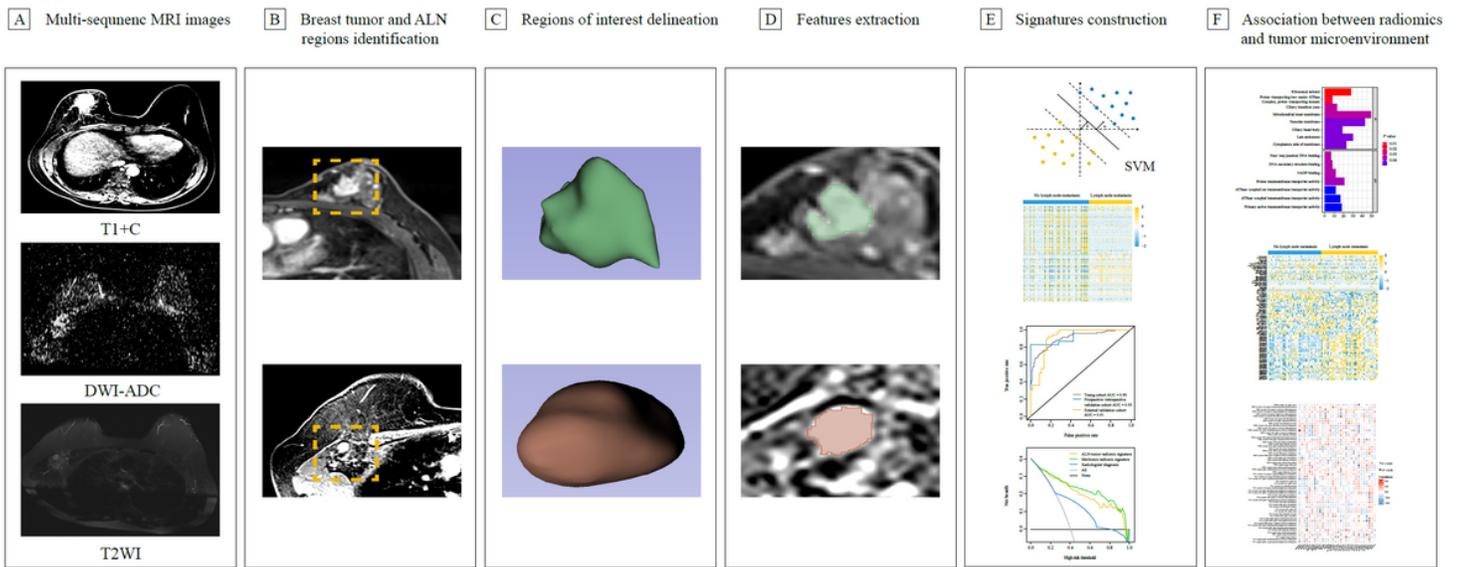
**Table 2.** The performance of each signature for axillary lymph node metastasis prediction

Cohort	Signature	Signature performance					
		Sensitivity	Specificity	Accuracy	PPV	NPV	AUC (95%CI)
Training cohort	Tumor radiomic signature	0.83	0.63	0.71	0.59	0.86	0.78 (0.74-0.83)
	ALN radiomic signature	0.89	0.70	0.77	0.66	0.91	0.85 (0.81-0.89)
	ALN-tumor radiomic signature	0.83	0.79	0.81	0.72	0.88	0.88 (0.84-0.91)
	Clinical signature	0.65	0.80	0.74	0.69	0.76	0.74 (0.70-0.77)
	Multiomic radiomic signature	0.85	0.89	0.81	0.72	0.89	0.90 (0.87-0.93)
Prospective-retrospective validation cohort	Tumor radiomic signature	0.34	1.00	0.48	1.00	0.28	0.59 (0.45-0.75)
	ALN radiomic signature	0.62	0.75	0.65	0.88	0.39	0.61 (0.44-0.77)
	ALN-tumor radiomic signature	0.82	0.91	0.84	0.97	0.63	0.87 (0.76-0.98)
	Clinical signature	0.97	0.41	0.82	0.82	0.82	0.68 (0.53-0.83)
	Multiomic radiomic signature	0.83	1.00	0.87	1.00	0.64	0.93 (0.84-1.00)
External validation cohort	Tumor radiomic signature	0.59	0.70	0.64	0.70	0.59	0.63 (0.54-0.73)
	ALN radiomic signature	0.86	0.69	0.78	0.74	0.83	0.81 (0.72-0.89)
	ALN-tumor radiomic signature	0.86	0.75	0.80	0.78	0.84	0.87 (0.80-0.93)

Clinical Signature	0.57	0.88	0.74	0.78	0.72	0.74 (0.63-0.81)
Multiomic radiomic signature	0.92	0.82	0.86	0.80	0.93	0.91 (0.84-0.97)

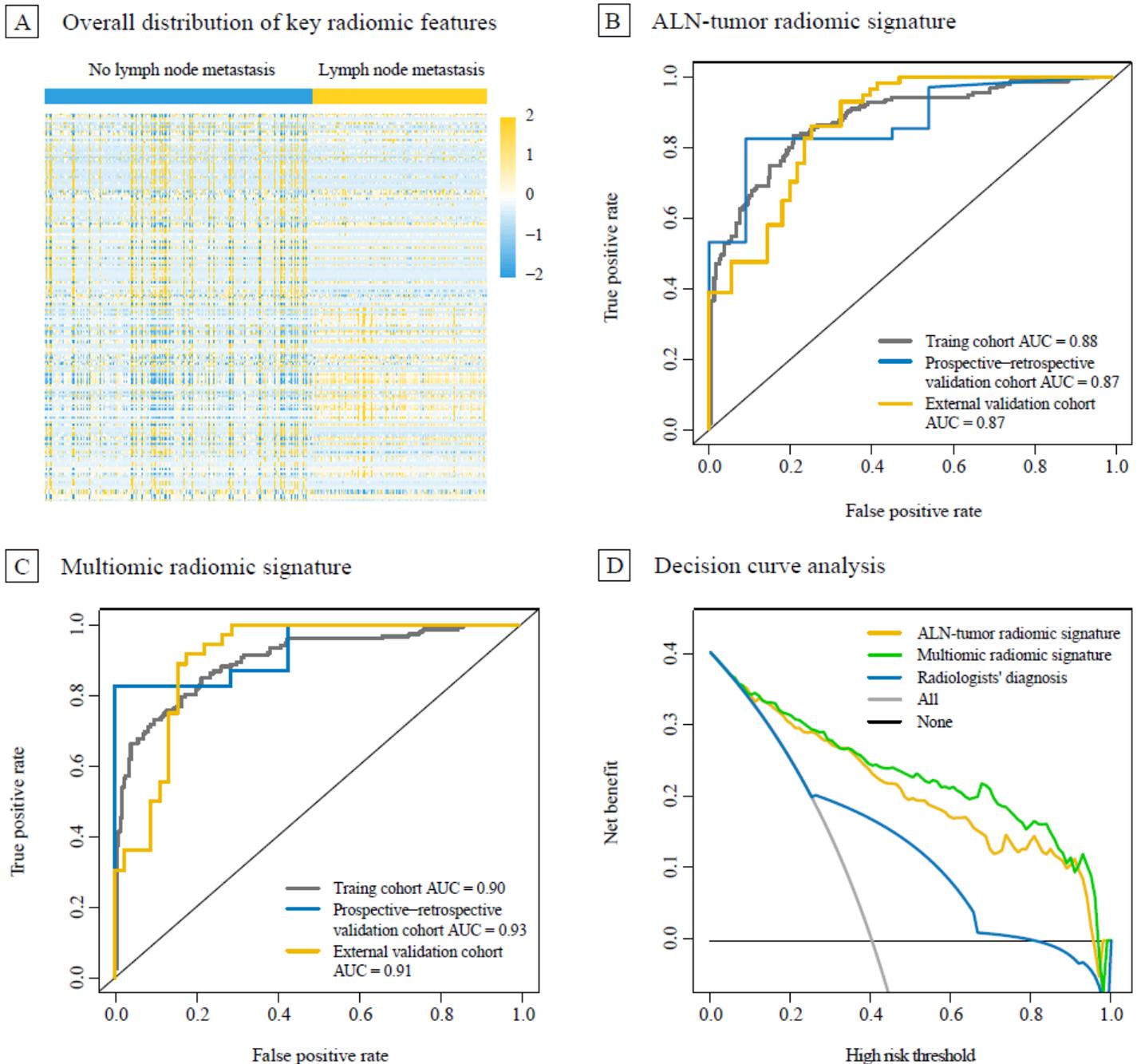
Abbreviations: ALN, axillary lymph node; PPV, positive predictive values; NPV, negative predictive values; CI, confidence interval; AUC, area under the receiver operating characteristics curve.

## Figures



**Figure 1**

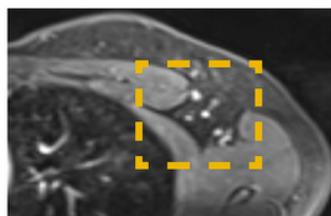
Study workflow (A) Multi-sequence MRI images were used for (B) breast tumor and ALN regions identification and (C) delineation of regions of interest, then (D) features were extracted using 3D Slicer software for (E) signatures construction and analysis of (F) the association between MRI radiomics and tumor microenvironment. MRI, Magnetic resonance imaging; ALN, Axillary lymph node; T1+C, contrast-enhanced T1-weighted imaging; T2WI, T2-weighted imaging; DWI-ADC, diffusion-weighted imaging; SVM, support vector machine.



**Figure 2**

Performance and clinical value of magnetic resonance imaging radiomic signatures (A) Overall distribution of key radiomic features from T1+C, T2WI, and DWI-ADC sequences among patients with and without ALN metastasis in the training cohort. Performance of (B) the ALN-tumor radiomic signature and (C) the multiomic radiomic signature for predicting ALN metastasis in the training, the prospective-retrospective validation and the external validation cohorts. (D) Decision curve analysis for the ALN-tumor radiomic signature and the multiomic radiomic signature. ALN, Axillary lymph node; AUC, area under the receiver operating characteristics curve; T1+C, contrast-enhanced T1-weighted imaging; T2WI, T2-weighted imaging; DWI-ADC, diffusion-weighted imaging.

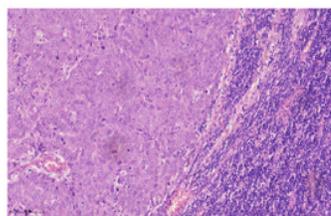
**A** Axillary lymph node-positive patient 1



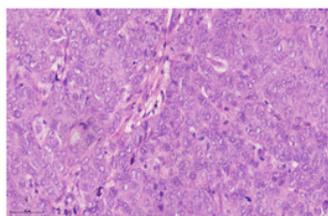
T1+C ALN



T1+C ALN

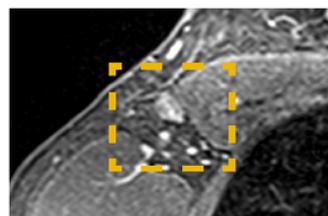


20\*Microscope field

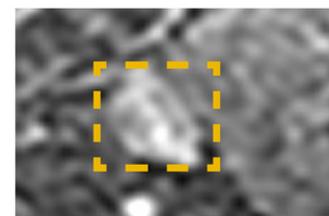


40\*Microscope field

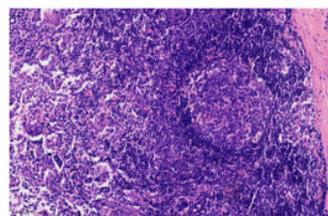
**B** Axillary lymph node-negative patient 2



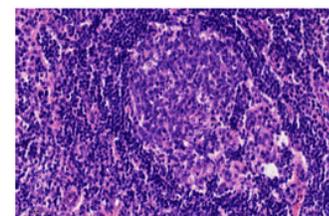
T1+C ALN



T1+C ALN



20\*Microscope field



40\*Microscope field

Age: 38 years old  
MRI-ALN status: false-negative  
Pathology-ALN status: positive  
(metastatic breast cancer cells can be seen in the axillary lymph node)  
Radiomic score of the multiomic radiomic signature: 0.873, true-positive

Age: 37 years old  
MRI-ALN status: false-positive  
Pathology-ALN status: negative  
(the morphology of the axillary lymph nodes is normal, no metastatic cancer cells are seen)  
Radiomic score of the multiomic radiomic signature: 0.326, true-negative

**Figure 3**

Patients with inconsistent MRI-ALN status and pathology-ALN status (A) Patient 1 had pathological positive ALN status but the MRI-status was considered as negative by radiologists before surgery. On the contrary, (B) the MRI-status of patient 2 was initially diagnosed as positive by radiologists but found to be negative by pathologically examination. The ALN status of two patients were accurately assessed through the multiomic radiomic signature by the cutoff value of 0.334. MRI, Magnetic resonance imaging; ALN, Axillary lymph node; T1+C, contrast-enhanced T1-weighted imaging.



Archive. T1+C, contrast-enhanced T1-weighted imaging; T2WI, T2-weighted imaging; lncRNA, long non-coding RNA; MRI, magnetic resonance imaging.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplement.docx](#)