Image-Based Ultrasomics Classifiers: A Potential Method to Noninvasively Predict the Proliferation-Related Biomarker Ki67

anna yuan
University of Jinan https://orcid.org/0000-0001-5718-905X
Guoze Xu
University of Jinan
Guiting Fang
Jinan University First Affiliated Hospital
Tong Li
Jinan University First Affiliated Hospital
Xiaomin Lai
Jinan University First Affiliated Hospital
Xing Zhong (tzhxing@jnu.edu.cn)
https://orcid.org/0000-0003-4359-8455
Furong Huang
University of Jinan

Research article

Keywords: Breast cancer, Ki67; computer-aided diagnosis, 3D ultrasound imaging

Posted Date: August 11th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-52480/v1

(i) Title page

Title: Image-Based Ultrasomics Classifiers: A Potential Method to Noninvasively Predict the Proliferation-Related Biomarker Ki67

Author: Anna Yuan,[1#] Guoze Xu,[2#] Guiting Fang,[1] Tong Li,[1] Xiaomin Lai,[1] Furong Huang,[2] Xing Zhong[1]

[1] Medical Imaging Center, The First Affiliated Hospital of Jinan University, Guangzhou, 510627. China

[2] Guangdong Provincial Key Laboratory of Optical Fiber Sensing and Communications, Department of Optoelectronic Engineering, Jinan University, Guangzhou 510632, China

The author Anna Yuan and Guoze Xu are co-first author.

The author Xing Zhong and Furong Huang are co-corresponding authors.

Please address all correspondence to:

Dr. Xing Zhong. Medical Imaging Center, The First Affiliated Hospital of Jinan University, No. 613, Huangpu West Road, Tianhe District, Guangzhou 510627, People's Republic of China. Email: tzhxing@jnu.edu.cn; Tel.: +86-13725197597. Dr. Furong Huang. Department of Optoelectronic Engineering, Jinan University, Huangpu Road West 601, Tianhe District, Guangzhou 510632, People's Republic of China. Email: furong_huang@163.com; Tel.: +86-20-85220484

(ii) Abstract

Objectives: The proliferation-related biomarker Ki67 is excellent in predicting the prognosis of breast cancer. 3D ultrasound imaging can provide information about coronal section images that help diagnose breast cancer. Our study aims to develop an ultrasomics method that extracts information related to Ki67 by analyzing the maximum transverse/sagittal/coronal section images of breast cancer mass.

Methods: Our study retrospectively collected data on patients who had finished 3D ultrasound examinations and were pathologically diagnosed with breast cancer. Images met the criteria were segmented, and then regions of interest (ROIs) were outlined for extracting ultrasomics features, such as statistical, morphological, texture, filter, and wavelet features.

Results: The least absolute shrinkage and selection operator (LASSO) regression model selected 16 features that were closely related to the Ki67. The classification results of sensitivity, specificity, accuracy, and area under curve (AUC) of the transverse-sectional images were 0.6451, 0.8064, 0.7258, and 0.8065 (95% CI, 0.6915-0.9214), respectively; for sagittal-sectional images were 0.5806, 0.7741, 0.6774, and 0.6660 (95% CI, 0.5283-0.8037) respectively; for coronal-sectional images were 0.5806, 0.6774, 0.6290, and 0.7159 (95% CI, 0.5847-0.8471) respectively, and for a combination of three-section images were 0.7667, 0.7500, 0.7580, and 0.8510 (95% CI, 0.7537 -0.9483).

Conclusions: The model classifier based on the transverse section images performed better than that based on the sagittal/coronal section images. The model classifier based on a combination of three-section images had a better outcome than that used only single section images. Image-based ultrasomics classifiers can noninvasively predict the Ki67 of breast cancer.

Keywords: Breast cancer; Ki67; computer-aided diagnosis; 3D ultrasound imaging

Abbreviations: ROI, regions of interest; LASSO, least absolute shrinkage and selection operator; AUC, area under curve; ROC, receiver operating characteristic; CAD, computer-aided diagnosis; MRI, magnetic resonance imaging; IHC, immunohistochemistry; NAC, Neoadjuvant chemotherapy; pCR, pathological complete response.

(iii) full text

Introduction:

Breast cancer is the most prevalent cancer among women worldwide and it has continuously ranked among the leading cancers both in terms of morbidity and mortality of women for many years [1]. It is estimated that 626,679 patients die from breast cancer worldwide in 2018, which accounts for 6.6% of all cancer-related deaths, making it the cancer with the highest mortality in women [2]. Therefore, early diagnosis and therapy play a crucial role in improving the survivability of breast cancer patients.

Nowadays, breast cancer is screened by medical imaging techniques and diagnosed by immunohistochemistry (IHC) in clinical practices [3]. Medical imaging techniques, such as magnetic resonance imaging (MRI), mammography, and ultrasound, are commonly used in breast disorders because they are noninvasive and provide information about the overall anatomy of the breast tumor. MRI scans are super-sensitive to soft tissue lesions [4]. However, they cost a lot of time and money than other available tests and contain a relatively high number of contraindications. Mammography is super-sensitive to the detection of calcification, but its sensitivity decreases by 30% when scanning dense breast tissue [4]. The breast density of Asian women is generally higher than that of Western women, which slightly restricts the application of mammography in patients from Asian countries [4].

As is well known, ultrasound is quite a convenient and time-saving imaging technique with outstanding performances of non-ionizing, low expense, real-time dynamic imaging for dense breast tissue scanning. In addition, ultrasound has been shown to play an essential role in early screening, diagnosis, and follow-up of breast cancer. Currently, it is regarded as the primary detection method for breast cancer

patients [5]. Ultrasound can assess the size, morphology, orientation, internal structure, and margins of lesions from multiple sections with high resolution [6] 3D ultrasound allows the calculation of the corresponding volume. The coronal reconstruction images of the 3D volume can show the anatomical details and spatial location of a lesion [6].

IHC is extensively applied to estimate biomarkers and differentiate molecular subtypes [3]. Specifically, Ki67, is present during all active phases of the cell cycle (G(1), S, G(2), and mitosis) and is recognized as a biomarker closely connected to cell proliferation [7]. The higher the value of Ki67, the more cells are undergoing cell division, which indicates the breast tumor is more aggressive. Several studies have well-attested to the predictive and the prognosis value of Ki67 in breast cancer [8-10]. The level of the proliferation-associated biomarker Ki67 serves as a delegate of the intrinsic molecular subtypes in clinical practice [11]. It differentiates HR+breast tumors into two distinct molecular subtypes: Luminal A (representing low Ki67 levels) and Luminal B (representing high Ki67 levels) [11]. Also, Ki67 is one of the most important prognostic factors for NAC, patients with high Ki67 tend to be more sensitive to NAC. Many studies have recognized a high Ki67 proliferation rate to be a predictive factor for a higher percentage of pathological complete response (pCR) after NAC [12-15]. Some scholars have proposed combining the Ki67 index into AJCC 2018 staging to enhance prognostic information in breast tumor candidates undergoing genomic profiling [16] to help identify patients who would most likely benefit from chemotherapy [17]. Regarding the patients without pCR, the Ki67 level is also important, without a decrease in the Ki67 index of residual tumors after NAC will have a poorer prognosis [18]. Until now, the Ki67 index of breast cancer has depended on histopathology. However, histopathology examination still has some

limitations. First, a biopsy of tumor tissue is an invasive procedure with the uncertainty of tissue sampling [19]. Due to the heterogeneity of tumor tissue, Ki67 results of different tissue sampling can be distinctive. In addition, the visual diagnostic analysis of physicians is generally subjective, and the reproducibility among observers of the visually derived Ki67 score is weak, which may lead to misjudgment at cutoff values [20].

Recent studies have shown that the hereditary and molecular features of a tumor can be captured inside medical images with the aid of computer-aided diagnosis (CAD) [3]. CAD can help analyze the tumor's overall features, excavate and extract hundreds or even thousands of quantitative features from medical images that are indistinguishable with the naked eye, and unveil the predictive and prognostic relevance between images and medical outcomes [3,21]. Numerous studies have explored the association between breast ultrasound images and some biological features [3,22-24]. The results prove that ultrasound imaging has a promising prospect of application in the evaluation of tumor heterogeneity. Coronal images of breast ultrasound could provide information beyond the conventional two-dimensional images that have reached widespread availability [6,25-27]. Still, our investigation discovered that there is no advanced research on incorporating the coronal images' information into a CAD system. Meanwhile, predicting the Ki67 proliferative index based on ultrasound images has not yet been studied.

Our research aims to develop a machine learning–based ultrasomics analysis system to extract quantitative information from the Ki67 proliferative index by analyzing the maximum transverse/sagittal/coronal section images of breast cancer mass segmented from a 3D video and help clinicians predict the prognosis of patients in determining the most suitable therapeutic options.

Materials and Methods

Patient Data

The institutional review board of our hospital approved this retrospective study and waived the informed consent requirement. Our study retrospectively collected data on patients who had finished 3D ultrasound examinations and were pathologically diagnosed with breast cancer in our hospital between July 2017 and December 2019.

Inclusion criteria were as follows: (1) Underwent ultrasound examination within two weeks before surgery; (2) single breast mass that could be recognized by ultrasound; (3) pathologically confirmed primary breast cancer; (4) precise IHC outcome of Ki67 proliferative index; and (5) no biopsy before surgery. Exclusion criteria were as follows: (1) The 3D probe could not cover the entire tumor; (2) metastatic breast disease; (3) received radiation therapy, chemotherapy, and neoadjuvant chemotherapy before ultrasound examination; and (4) incomplete clinical information.

Ultimately, a total of 70 breast cancer patients met the criteria, of which 50 were Ki67 positive (Ki67 $\geq$ 14%) and 20 were Ki67 negative (Ki67 < 14%).

US Data Acquisition

In order to ensure the consistent quality of the images, we coherently used a single ultrasound device, the Philips iU elite (Philips, Netherlands), VL13-5 probe, SmPrt AdBrst with default parameter settings to perform the breast ultrasound scanning, minimizing the image variations generated by machine conditions to the best extent. Steps in the scanning procedure were as follows. First, the patient was maintained at a supine position, hands behind the head to fully expose the bilateral breast. The gel was applied evenly, then acoustic beam emission surface of the 3D

probe was placed above the breast mass. When the maximum transverse section of the mass was displayed on the screen, the sonographers used the 3D mode of the machine scanning the mass layer by layer automatically to generate a dynamic 3D video and then saved it in the image archiving and communication system. Finally, they used the built-in tool of the ultrasonic device to segment the 3D video in a 1 mm layer thickness and obtain transverse/sagittal/coronal section images in the Digital Imaging and Communications in Medicine (DICOM) format.

Ki67 Assessment

Pathological tissue was acquired after surgery. The pathologists assessed the Ki67 proliferation index by manually calculating the percentage of positively stained cells among the total invasive cells in the stained area. In this study, we considered Ki67 $\geq$ 14% to be a high proliferation level, denoted as a positive expression, while Ki67 < 14% was a low proliferation level, denoted as a negative expression [28].

Image Segmentation

DICOM format ultrasound images were selected from the image archiving and communication system. Senior sonographers respectively selected three images with maximum transverse section/sagittal section (perpendicular to transverse section)/coronal section (perpendicular to both transverse and sagittal section), and manually sketched the ROI according to the outline of the lesion for extracting features. ROIs of 30 patient (outlined by sonographer 1, with at least 10 years of work experience) and ROIs of all patient (outlined by sonographer 2, with at least 8 years of work experience) were chosen to evaluate the inter-observer reproducibility of the extracted features. The two sonographers were utterly ignorant about the pathological diagnoses of these patients.

Features Extraction and Selection

Maximum and minimum normalized transformation was performed for ultrasound image standardization (Equation 1). Mathworks, a program coded by Matlab 2017 (Natick, MA, USA) was used to extract ultrasomics features from the ROIs of the standardized images, namely statistical (17), morphological (10), texture (including GLCM (20), GLRLM (11), GLSZM (11), and NGTDM (5)), filter (408), and wavelet features (306) [29,30].

$$I_{\text{norm}} = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \tag{1}$$

The ultrasomics features were selected by the LASSO regression method. The LASSO regression method is a shrinkage estimation method with the idea of reducing the variable set (order reduction). It can compress the coefficients of variables and make some regression coefficients become zero through constructing a penalty function, achieving the purpose of variable selection [31].

The LASSO regression optimization problem can be expressed as:

$$\arg\min_{\beta} \|y - X\beta\|^2, \|\beta\| \le s \tag{2}$$

The corresponding Lagrangian can be expressed as:

$$\arg\min_{\beta} \|y - X\beta\|^2 + \lambda\|\beta\| \tag{3}$$

Where y is the target variable or the corresponding variable, β is the regression coefficient vector, X is the data matrix corresponding to the explanatory variable, λ is the penalty parameter, and s is a constant greater than 0 and exists certain corresponding relations with λ.

Model Establishment and Evaluation

The ultrasomics features selected by LASSO regression were evaluated by a binomial logistic regression model. As a classification model, a binomial logistical

regression model is delegated by a conditional probability table $P(Y \mid X)$ in the form of parameterized logistic distribution. In this model, the random variable X is a real number, and the random variable Y is either 1 or 0. This model can be expressed by conditional probability distributions as follows:

$$P(Y=1 \mid x)=\frac{\exp(w \bullet x+b)}{1+\exp(w \bullet x+b)} \tag{4}$$

$$P(Y=0 \mid x)=\frac{1}{1+\exp(w \bullet x+b)} \tag{5}$$

In the above formula, $x \in R^n$ is the input, $Y \in \{0,1\}$ is the output, $w \in R^n$ and $b \in R$ are parameters, w is the weight vector, b is the offset, $w \bullet x$ is the inner product of w and x.

The receiver operating characteristic (ROC) curve was employed to evaluate the diagnostic efficacy of the logistic regression model and to calculate the corresponding AUC, sensitivity, specificity, and accuracy.

Statistical Analysis Method

The reproducibility of the extracted ultrasomics features between inter-observers was analyzed using the intraclass correlation coefficient (ICC). The statistical difference among the ultrasomics features was analyzed using the Mann–Whitney U test. All statistical analysis methods, model establishment, and diagnostic efficacy assessment were performed by R statistical software. (version 3.6.2 , https://www.r-project.org/).

Results

Features Extraction and Selection of Images

A total of 2376 ultrasomics features, evaluated for inter-group consistency, were extracted from the maximum transverse/sagittal/coronal section images of breast

cancer mass (about 792 ultrasomics features in each section). Finally, 2,152 features were selected. (ICC>0.75)

The LASSO regression model further selected 16 features that were closely related to the classification of positive Ki67 and negative Ki67 from the above 2,152 ultrasomics features, including 8 transverse-sectional, 4 sagittal-sectional, and 4 coronal-sectional image features, which somewhat reflects that transverse-sectional images contain more proliferative information of Ki67 than other sections; the results are shown in Figure 1. For the features, the numerical values of Ki67 positive were greater than those of Ki67 negative, which was obtained by comparing the features of the transverse-sectional images. The 16 features selected by the LASSO regression model showed statistical differences between positive Ki67 and negative Ki67 after analysis by the Mann–Whitney U test, ($p < 0.05$), as shown in Table 1.

Model Classification Results

The classification results for different section models are listed in Table 2. The classification results based on the ultrasomics features of the transverse section are the best. Concerning the classification results of sagittal section and coronal section, the sensitivity, specificity, and accuracy are both worse than those for the transverse section. The AUC results for different section models are shown in Figure 2.

The datasets containing all sectional ultrasomics features were divided into training and validation datasets in a ratio of 7:3 and then both the training and validation datasets were trained with the logistic regression model. The classification results of the model are shown in Table 3. Compared with the results in a single section, the model classifier combining three-section images has the best classification outcome. The ROC curves of the training dataset and validation dataset are shown in Figure 3.

Discussion

In our investigation, the relationship between the quantitative sonographic features of breast cancer in a single or combined transverse/sagittal/coronal section image and the proliferation-related biomarker Ki67 was objectively assessed by using ultrasomics method. Compared with traditional statistical methods, which usually consider and evaluate limited assumptions, machine learning methods are superior in generating models to analyze images for prediction by extensively searching models and parameter spaces [32]. Currently, many definitive diagnoses have been made for breast disease by using the computational algorithms of ultrasomics, including identifying benign and malignant breast lesions based on the texture features of ultrasound images [33], predicting axillary lymph node metastasis and associated potential biomarkers in breast cancer [34], and analyzing the biological behavior of infiltrating ductal carcinoma of the breast [3].

However, the lack of a common standard dataset in breast cancer ultrasound research limits the fair evaluation of algorithm performance. Simultaneously, the quality of ultrasonic breast images is highly dependent on the acquisition process, and the images vary significantly between different ultrasound systems, which will affect the results obtained by the algorithm. In order to solve this problem, we used the same mammary gland scanning mode of a 3D ultrasound machine to automatically scan the entire breast mass and obtain the video images, which can avoid the errors caused due to using different machines, parameters, and operators. In addition, 3D ultrasound reduces the difficulty of collection by simplifying the operation method and is characterized by the reliable function of post-processing software.

Coronal section images (the surface parallel to the probe's surface) of breast cancer have been documented to provide valuable information in differentiating

benign and malignant breast masses and assessing the prognosis of breast cancer [6,25,35], but conventional 2D ultrasound scanning cannot gain the images of the coronal section, we believe that transverse/sagittal/coronal section images have their own heterogeneity and contain essential information. Therefore, our study carried out separate and integrated calculations on these three-section images, which was different from the previous research that adopted a simple 2D image of the maximum transverse section of the mass. The spread rate of the tumor can vary in all directions. Generally, a faster growth rate in a section represents a more extensive infiltration range, which indicates that this section of the tumor grows rapidly and is the most aggressive compared with other sections. Therefore, it is meaningful for us to segment the 2D images of the maximum transverse/sagittal/coronal section images to assess the prediction of proliferation-related biomarker Ki67. Our sonographers have also demonstrated that the computational model, which uses these three maximum sectional images segmented by the 3D video for calculation, possesses the advantages of excellent repeatability and small selection error among different personnel.

Biomarkers of breast cancer play an essential role in the determination of treatment and prognosis of patients. Some biomarkers are prone to be differentiated within and between observers, which hinders the repeatability and feasibility of their application, and Ki67 is particularly vulnerable in this regard [36]. The St Gallen International Expert Consensus 2015 pointed out that image analysis could help reduce the variation in Ki67 scores to reveal better potential biomarkers [36], which was also supported by our analysis. The maximum transverse section of a breast mass generally represents the largest section of the entire mass, indicating the quickest growing and the comparatively most aggressive. Besides, the direction of the transverse section is consistent with the direction of sound beam dispersion, which

means the image quality of this section is the highest and may contain more information than other sections. These two points were reflected in our results as the features extracted from the transverse section were the most detailed, and the prediction of Ki67 was the best.

Finally, we verified the ability of the model classifier (combined three-section images) in predicting Ki67. Our results show that the model classifier analyzing all maximum transverse/sagittal/coronal section images can take full advantage of the heterogeneity information of each section and ranks first in predictive capacity, which demonstrates that ultrasomics has the potential to be an excellent noninvasive predictive method for Ki67. Ultrasomics can investigate the quantitative relationship among breast cancer sonographic features and the proliferation-related biomarker Ki67. It can also improve clinicians' understanding of the relationship between target image features and biological features, enabling them to carry out early medical management for breast cancer patients, including therapeutic planning, therapeutic-response evaluation, and prognosis prediction. The use of ultrasound in breast diseases is universal and applicable, consequently, we believe that predict the proliferation-related biomarker Ki67 by ultrasound images will be more practical and valuable than that by other type of images.

Our research had some limitations. First, our study was a retrospective study where the data were only from a single-center, and the amount of data is relatively small. Second, this study excluded tumors with a maximum diameter beyond the scanning range of the 3D probe, which may lead to selection bias in the study cohort. Third, there may be some sample bias and participation bias due to the number of Ki67 negative cases being relatively low. Fourth, our study only extracted the features from the maximum transverse/sagittal/coronal 2D section images of breast cancer

mass and used these features to represent the entire mass features. Some studies have proved that 2D texture analysis of a single section is adequate in computed tomography images, and no significant difference was found between 2D and 3D texture analysis [37]. However, we did not determine whether this applies to our ultrasomics study.

Prospects: We first need to try our best to increase the data volume by using multi-center research. Another step would be to design a 3D algorithm to reconstruct the volume of interest （VOI） of the video directly for the sake of analyzing the whole mass features and evaluate the prediction distinction between 3D and 2D texture analysis. Furthermore, it may be innovative to combine breast ultrasound, mammography, and MR image features to comprehensively predict the proliferation-related biomarker Ki67.

Conclusions

Our study determined that ultrasound image-based ultrasomics classifiers can noninvasively predict the proliferation-related breast cancer biomarker Ki67. Ultrasomics can evaluate the quantitative relationship among sonographic features and the proliferation-related breast cancer biomarker Ki67, providing the needed prognostic and predictive information for breast cancer patients.

(iv) References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394-424.

2. Ahmad A. Breast Cancer Statistics: Recent Trends. Adv Exp Med Biol 2019;1152:1-7.

346 3. Guo Y, Hu Y, Qiao M, Wang Y, Yu J, Li J, et al. Radiomics Analysis on Ultrasound
347 for Prediction of Biologic Behavior in Breast Invasive Ductal Carcinoma. Clin Breast
348 Cancer 2018;18(3):e335-e44.
349 4. Chiao J-Y, Chen K-Y, Liao KY-K, Hsieh P-H, Zhang G, Huang T-C. Detection and
350 classification the breast tumors using mask R-CNN on sonograms. Medicine
351 (Baltimore) 2019;98(19):e15200.
352 5. Zhang Q, Song S, Xiao Y, Chen S, Shi J, Zheng H. Dual-mode
353 artificially-intelligent diagnosis of breast tumours in shear-wave elastography and
354 B-mode ultrasound using deep polynomial networks. Med Eng Phys 2019;64:1-6.
355 6. Guo R, Lu G, Qin B, Fei B. Ultrasound Imaging Technologies for Breast Cancer
356 Detection and Management: A Review. Ultrasound Med Biol 2018;44(1):37-70.
357 7. Scholzen T, Gerdes J. The Ki-67 protein: from the known and the unknown. J Cell
358 Physiol 2000;182(3):311-22.
359 8. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast
360 cancer: prognostic and predictive potential. Lancet Oncol 2010;11(2):174-83.
361 9. Viale G, Regan MM, Dell'Orto P, Mastropasqua MG, Maiorano E, Rasmussen BB,
362 et al. Which patients benefit most from adjuvant aromatase inhibitors? Results using a
363 composite measure of prognostic risk in the BIG 1-98 randomized trial. Ann Oncol
364 2011;22(10):2201-7.
365 10. Hugh J, Hanson J, Cheang MCU, Nielsen TO, Perou CM, Dumontet C, et al.
366 Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use

of an immunohistochemical definition in the BCIRG 001 trial. J Clin Oncol 2009;27(8):1168-76.

11. Dieci MV, Orvieto E, Dominici M, Conte P, Guarneri V. Rare breast cancer subtypes: histological, molecular, and clinical peculiarities. Oncologist 2014;19(8):805-13.

12. Alba E, Lluch A, Ribelles N, Anton-Torres A, Sanchez-Rovira P, Albanell J, et al. High Proliferation Predicts Pathological Complete Response to Neoadjuvant Chemotherapy in Early Breast Cancer. Oncologist 2016;21(2):150-5.

13. Fasching PA, Heusinger K, Haeberle L, Niklos M, Hein A, Bayer CM, et al. Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment. BMC Cancer 2011;11:486.

14. Colleoni M, Bagnardi V, Rotmensz N, Viale G, Mastropasqua M, Veronesi P, et al. A nomogram based on the expression of Ki-67, steroid hormone receptors status and number of chemotherapy courses to predict pathological complete remission after preoperative chemotherapy for breast cancer. Eur J Cancer 2010;46(12):2216-24.

15. Fernández-Sánchez M, Gamboa-Dominguez A, Uribe N, García-Ulloa AC, Flores-Estrada D, Candelaria M, et al. Clinical and pathological predictors of the response to neoadjuvant anthracycline chemotherapy in locally advanced breast cancer. Med Oncol 2006;23(2):171-83.

16. Vissio E, Metovic J, Osella-Abate S, Bertero L, Migliaretti G, Borella F, et al. Integration of Ki-67 index into AJCC 2018 staging provides additional prognostic

388 information in breast tumours candidate for genomic profiling. Br J Cancer
389 2020;122(3):382-7.
390 17. Jones RL, Salter J, A'Hern R, Nerurkar A, Parton M, Reis-Filho JS, et al.
391 Relationship between oestrogen receptor status and proliferation in predicting
392 response and long-term outcome to neoadjuvant chemotherapy for breast cancer.
393 Breast Cancer Res Treat 2010;119(2):315-23.
394 18. Cabrera-Galeana P, Muñoz-Montaño W, Lara-Medina F, Alvarado-Miranda A,
395 Pérez-Sánchez V, Villarreal-Garza C, et al. Ki67 Changes Identify Worse Outcomes in
396 Residual Breast Cancer Tumors After Neoadjuvant Chemotherapy. Oncologist
397 2018;23(6):670-8.
398 19. Aleskandarany MA, Green AR, Ashankyty I, Elmouna A, Diez-Rodriguez M,
399 Nolan CC, et al. Impact of intratumoural heterogeneity on the assessment of Ki67
400 expression in breast cancer. Breast Cancer Res Treat 2016;158(2):287-95.
401 20. Abubakar M, Orr N, Daley F, Coulson P, Ali HR, Blows F, et al. Prognostic value
402 of automated KI67 scoring in breast cancer: a centralised evaluation of 8088 patients
403 from 10 study groups. Breast Cancer Res 2016;18(1):104.
404 21. Lee SE, Han K, Kwak JY, Lee E, Kim E-K. Radiomics of US texture features in
405 differential diagnosis between triple-negative breast cancer and fibroadenoma. Sci
406 Rep 2018;8(1):13546.
407 22. Çelebi F, Pilancı KN, Ordu Ç, Ağacayak F, Alço G, İlgün S, et al. The role of
408 ultrasonographic findings to predict molecular subtype, histologic grade, and hormone
409 receptor status of breast cancer. Diagn Interv Radiol 2015;21(6):448-53.

410 23. Stein RG, Wollschläger D, Kreienberg R, Janni W, Wischnewsky M, Diessner J, et
411 al. The impact of breast cancer biological subtyping on tumor size assessment by
412 ultrasound and mammography - a retrospective multicenter cohort study of 6543
413 primary breast cancer patients. BMC Cancer 2016;16:459.
414 24. Costantini M, Belli P, Bufi E, Asunis AM, Ferra E, Bitti GT. Association between
415 sonographic appearances of breast cancers and their histopathologic features and
416 biomarkers. J Clin Ultrasound 2016;44(1):26-33.
417 25. Watermann DO, Földi M, Hanjalic-Beck A, Hasenburg A, Lüghausen A,
418 Prömpeler H, et al. Three-dimensional ultrasound for the assessment of breast lesions.
419 Ultrasound Obstet Gynecol 2005;25(6):592-8.
420 26. Vourtsis A, Kachulis A. The performance of 3D ABUS versus HHUS in the
421 visualisation and BI-RADS characterisation of breast lesions in a large cohort of
422 1,886 women. Eur Radiol 2018;28(2):592-601.
423 27. Tang G, An X, Xiang H, Liu L, Li A, Lin X. Automated Breast Ultrasound:
424 Interobserver Agreement, Diagnostic Value, and Associated Clinical Factors of
425 Coronal-Plane Image Features. Korean J Radiol 2020;21(5):550-60.
426 28. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn HJ.
427 Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the
428 St. Gallen International Expert Consensus on the Primary Therapy of Early Breast
429 Cancer 2011. Ann Oncol 2011;22(8):1736-47.

29. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun 2014;5:4006.

30. Huang Y-Q, Liang C-H, He L, Tian J, Liang C-S, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. J Clin Oncol 2016;34(18):2157-64.

31. Tibshirani R. Regression Shrinkage and Selection Via the Lasso 1996;58(1):267-88.

32. Tahmassebi A, Wengert GJ, Helbich TH, Bago-Horvath Z, Alaei S, Bartsch R, et al. Impact of Machine Learning With Multiparametric Magnetic Resonance Imaging of the Breast for Early Prediction of Response to Neoadjuvant Chemotherapy and Survival Outcomes in Breast Cancer Patients. Invest Radiol 2019;54(2):110-7.

33. Alvarenga AV, Pereira WCA, Infantosi AFC, Azevedo CM. Complexity curve and grey level co-occurrence matrix in the texture evaluation of breast tumor on ultrasound images. Med Phys 2007;34(2):379-87.

34. Yu F-H, Wang J-X, Ye X-H, Deng J, Hang J, Yang B. Ultrasound-based radiomics nomogram: A potential biomarker to predict axillary lymph node metastasis in early-stage invasive breast cancer. Eur J Radiol 2019;119:108658.

35. Jiang J, Chen Y-q, Xu Y-z, Chen M-l, Zhu Y-k, Guan W-b, et al. Correlation between three-dimensional ultrasound features and pathological prognostic factors in breast cancer. Eur Radiol 2014;24(6):1186-96.

36. Robertson S, Azizpour H, Smith K, Hartman J. Digital image analysis in breast pathology-from image processing techniques to artificial intelligence. Transl Res 2018;194:19-35.

37. Lubner MG, Stabo N, Lubner SJ, del Rio AM, Song C, Halberg RB, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. Abdom Imaging 2015;40(7):2331-7.

(v) tables

Table1. The characteristic parameters of ki67 selected by LASSO regression

Feature parameters		Mean±Std		p-value
		ki-67(-)	ki-67(+)	
Transverse section	perimeter	905.105±309.875	1191.476±310.091	<0.0001
	RLNT3_log	617.751±452.342	808.569±552.033	<0.0001
	HGRET3_log	6948.540±5623.447	11452.133±8892.17	0.0011
	HGRET6_laplace	6817.058±5650.132	11189.810±8862.65	<0.0001
	LRHGE_bior11	12937.849±8467.782	19314.229±10968.8	<0.0001
	GLS_LZHGE_dmey	121546331.974±12552918	294402819.256±430	0.0113
	Correlation_rbio11	0.879±0.064	0.890±0.062	<0.0001
	RLN_sym2	481.689±296.056	625.501±345.733	0.0039
Sagittal section	Contrast_dmey	0.212±0.222	0.1119±0.147	0.0020
	IMC2_dmey	0.228±0.133	0.159±0.093	0.0023
	Variance_dmey	4.277±5.508	2.128±1.650	0.0003
	LRHGE_dmey	16083.752±8987.511	23621.054±13.1635	0.0164
Coronal section	Autocorrelation_sym2	1.553±1.798	0.799±1.067	0.0016
	Standard uniformity	2.374E-09±9.463E-09	4.639E-09±4.196E-0	0.0183
	LRHGET3_log	15374.526±12527.388	20885.975±16496.8	<0.0001
	NGTDM_StrengthT3_	1.037±1.372	0.526±0.697	<0.0001

Std: standard deviation; p-value: Mann-Whitney U test

461 Table2. Different section models' classification results

Section	AUC (95% CI)	Sensitivity	Specificity	Accuracy
transverse section	0.8065(0.6915-0.9214)	0.6451	0.8064	0.7258
sagittal section	0.6660(0.5283-0.8037)	0.5806	0.7741	0.6774
coronal section	0.7159(0.5847-0.8471)	0.5806	0.6774	0.6290

462

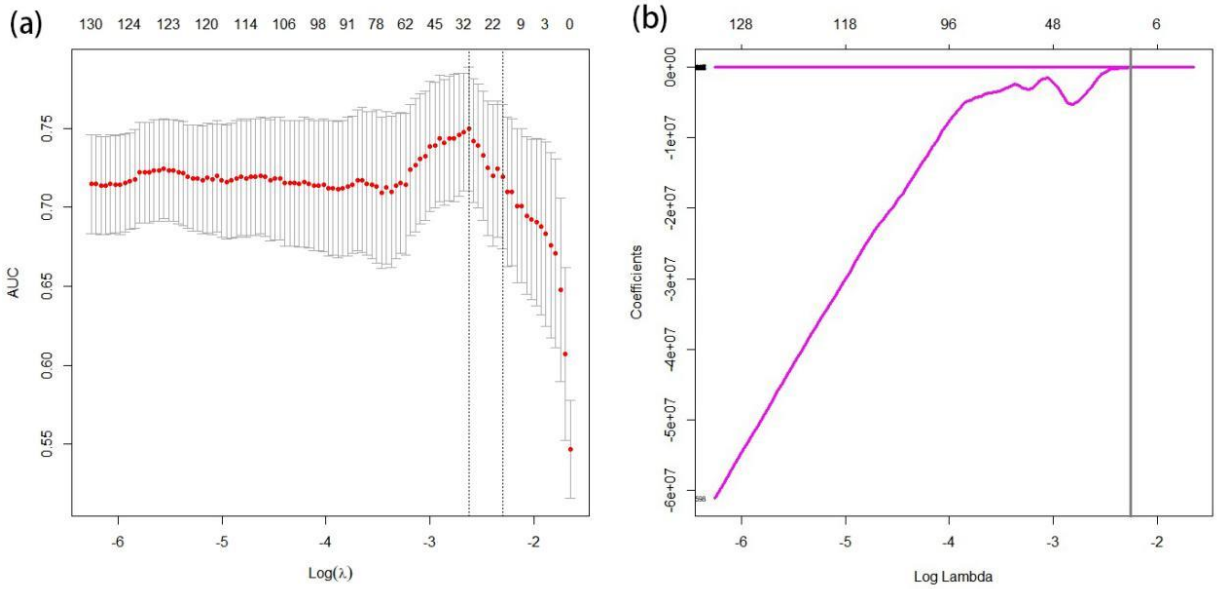
463 Table3. The classification results of the model

Dataset	AUC (95% CI)	Sensitivity	Specificity	Accuracy
Training	0.8383(0.7734-0.9030)	0.7746	0.7464	0.7606
Testing	0.8510(0.7537-0.9483)	0.7667	0.7500	0.7580

464

465 **(vi) figures**

466 Figure 1.



467

468 Figure 2.

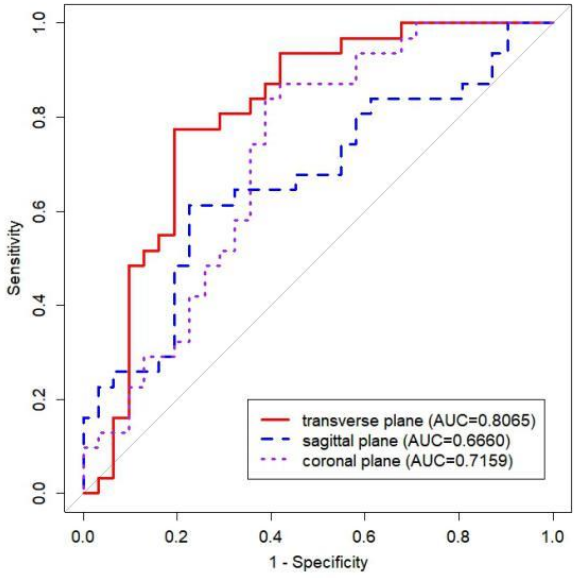
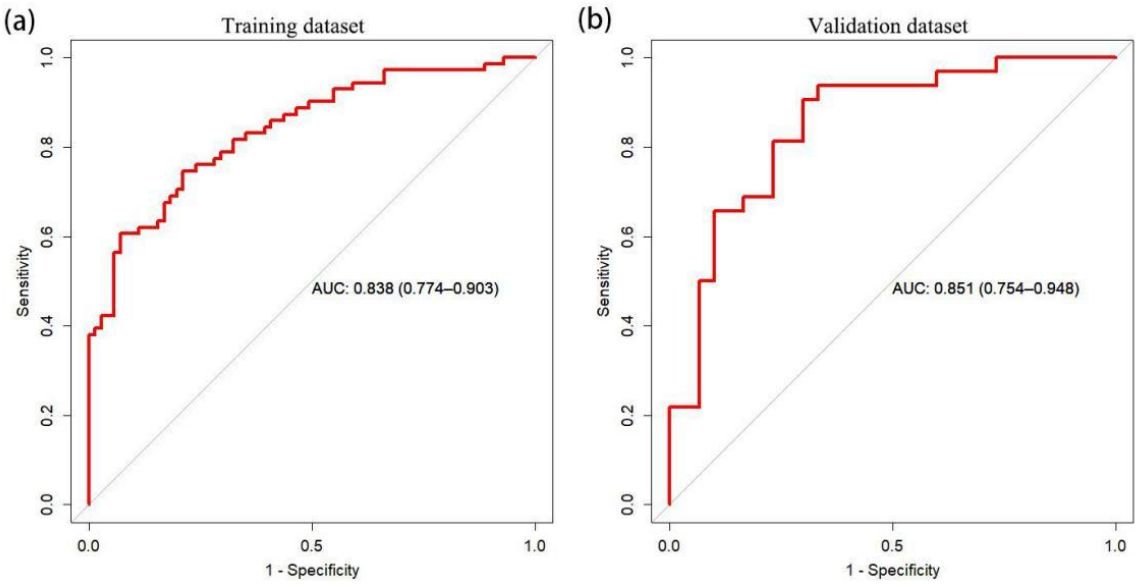


Figure 3.



(vii) figure legends

Figure1. Extraction of ultrasomics features by LASSO regression model. (a) Performed a 10-fold interactive verification of the LASSO model to select the best adjustment parameters. (b) Sixteen ultrasomics features were selected from the model with the optimal adjustment parameters.

Figure2. The AUC results of different section models.

Figure3. The ROC curves of the datasets. (a) Training dataset; (b)Validation dataset.

(viii) Abbreviations

ROI, regions of interest; LASSO, least absolute shrinkage and selection operator; AUC, area under curve; ROC, receiver operating characteristic; CAD, computer-aided diagnosis; MRI, magnetic resonance imaging; IHC, immunohistochemistry; NAC, Neoadjuvant chemotherapy; pCR, pathological complete response.

(ix) Acknowledgments

None

(x) Authors' contributions

Anna Yuan: Conceptualization, Methodology,Validation, Investigation, Writing-Original Draft, Writing-Review & Editing. Guoze Xu: Conceptualization, Software, Formal analysis, Data Curation, Writing - Original Draft. Guiting Fang: Methodology, Investigation, Visualization. Tong Li: Investigation, Visualization. Xiaomin Lai: Investigation, Visualization. Furong Huang: Formal analysis, Resources, Data Curation, Supervision. Xing Zhong: Formal analysis, Investigation, Resources, Supervision, Project administration.

(xi) Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

(xii) Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

(xiii) Ethics approval and consent to participate

The institutional review board of our hospital approved this retrospective study and waived the informed consent requirement.

(xiv) Consent for publication

Not applicable.

(xv) Competing interests

The authors declare that they have no competing interests.

Figures

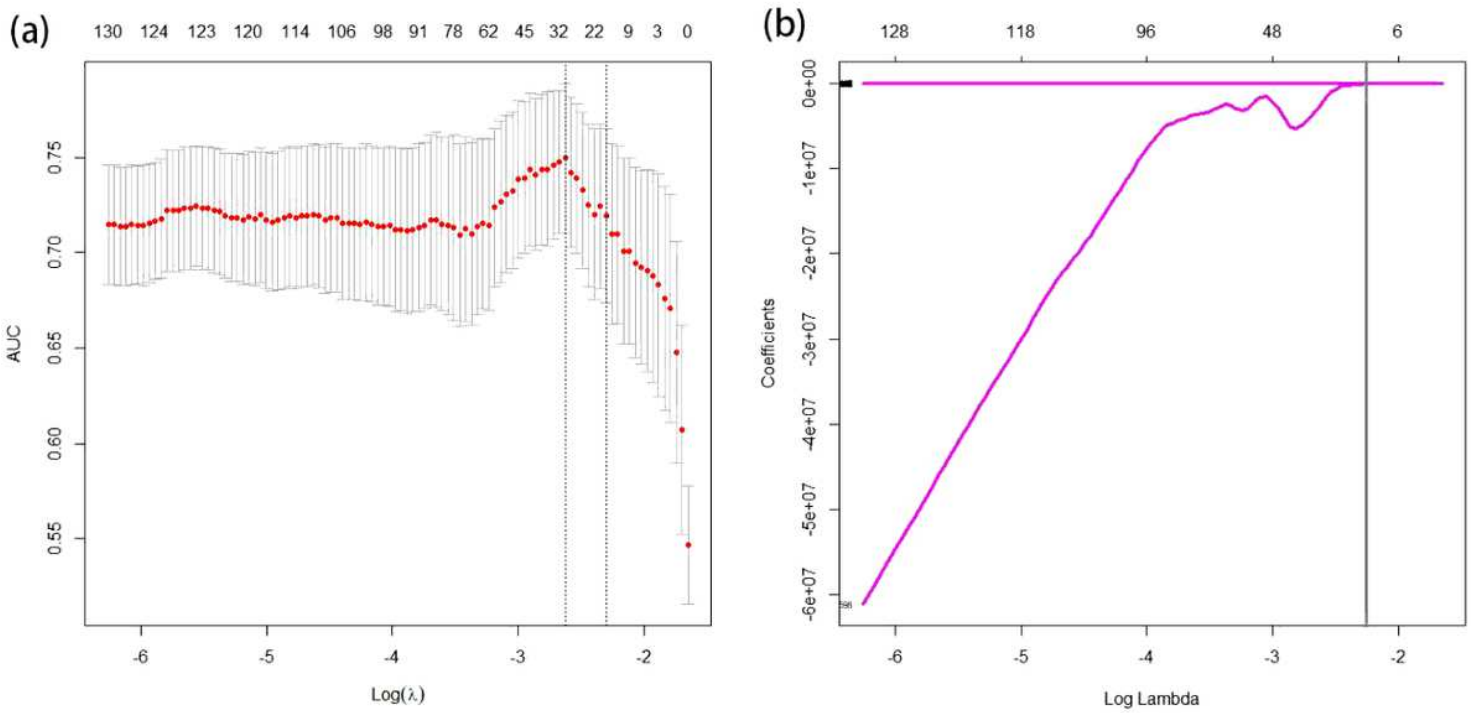


Figure 1

Extraction of ultrasomics features by LASSO regression model. (a) Performed a 10-fold interactive verification of the LASSO model to select the best adjustment parameters. (b) Sixteen ultrasomics features were selected from the model with the optimal adjustment parameters.

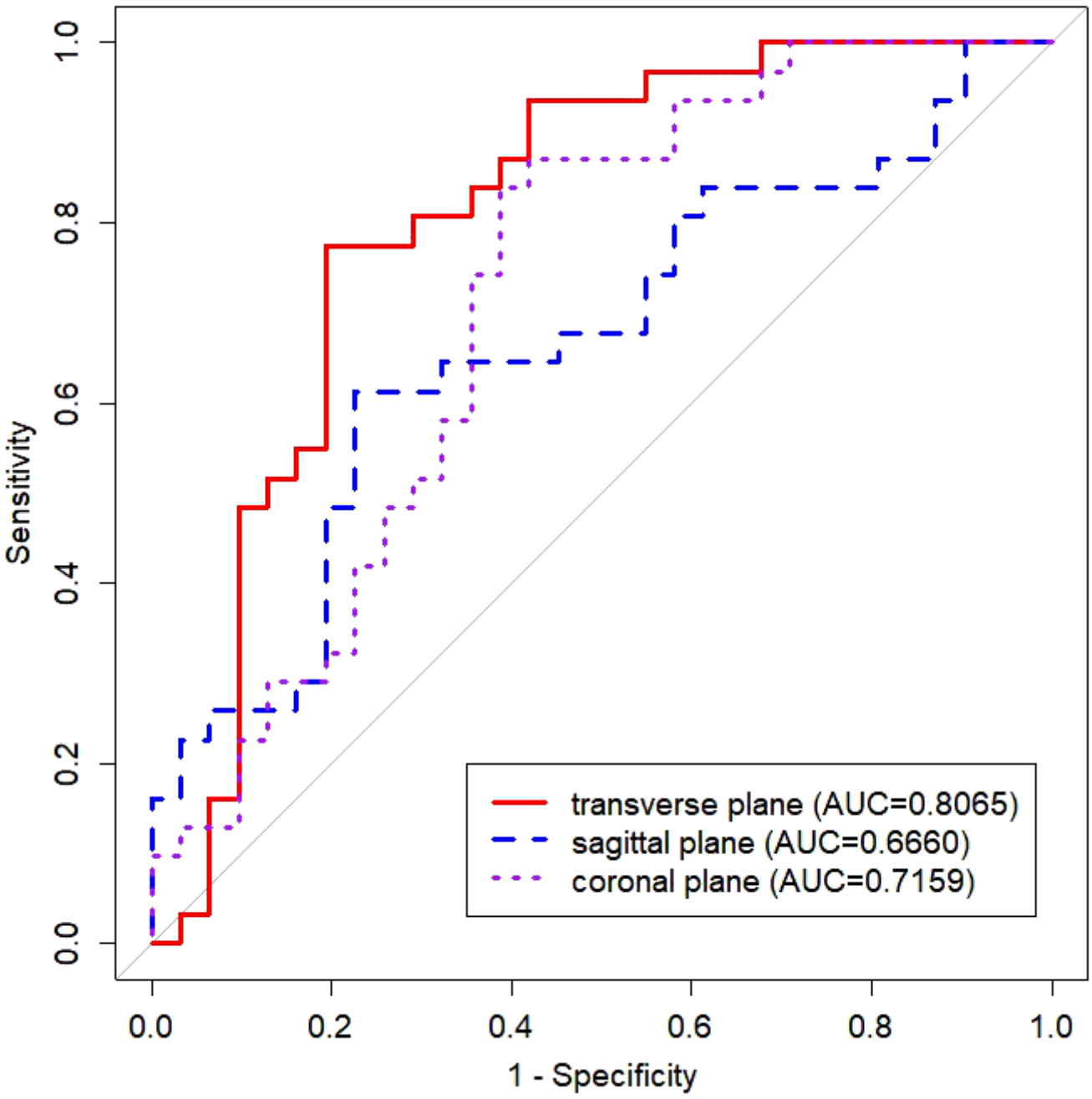


Figure 2

The AUC results of different section models

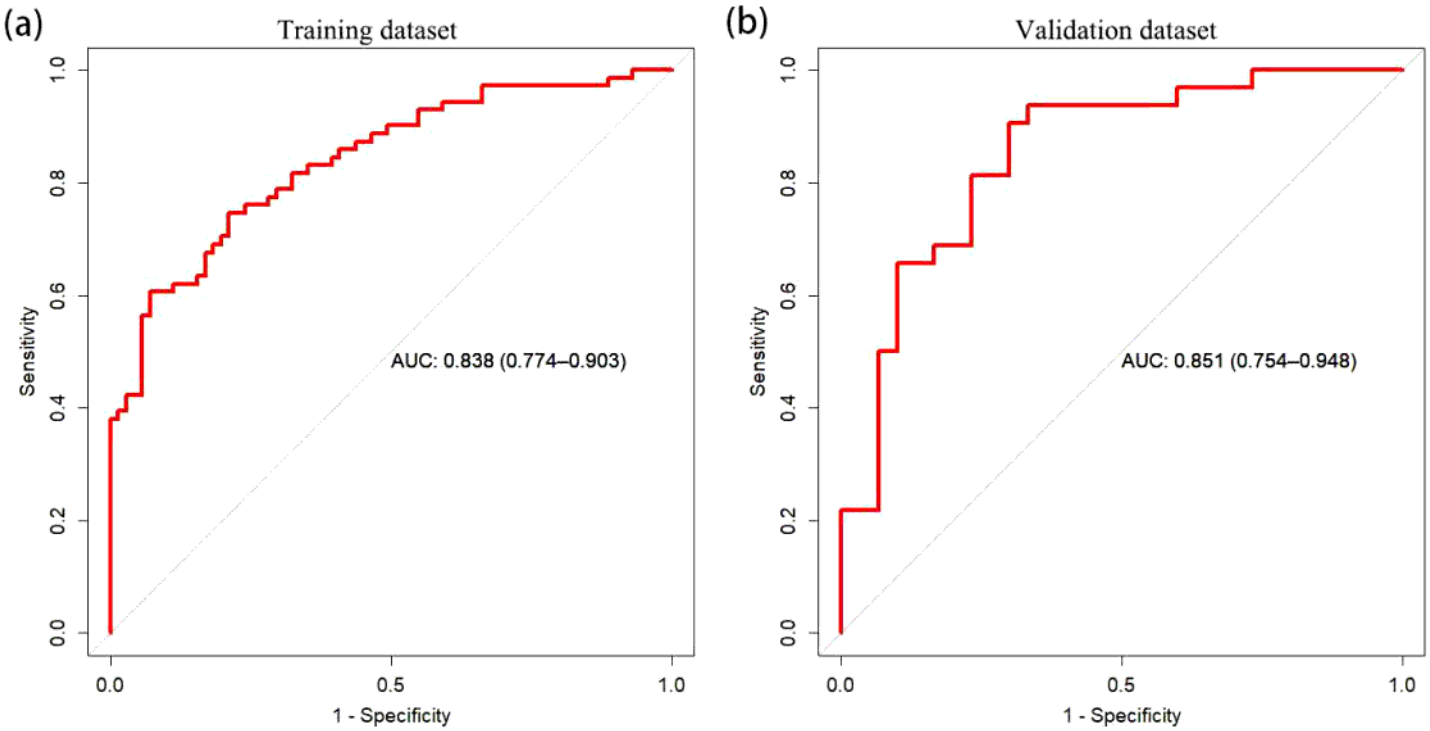


Figure 3

The ROC curves of the datasets. (a) Training dataset; (b)Validation dataset