Deep learning model to predict Ki-67 expression of breast cancer using digital breast tomosynthesis

Ken Oba

kenoba@luke.ac.jp

St Luke's Hospital https://orcid.org/0000-0003-0539-0880

Maki Adachi

Tohoku University School of Medicine: Tohoku Daigaku Daigakuin Igakukei Kenkyuka Igakubu

Tomoya Kobayashi

Tohoku University School of Medicine: Tohoku Daigaku Daigakuin Igakukei Kenkyuka Igakubu

Eichi Takaya

Tohoku University School of Medicine: Tohoku Daigaku Daigakuin Igakukei Kenkyuka Igakubu

Daiki Shimokawa

Tohoku University School of Medicine: Tohoku Daigaku Daigakuin Igakukei Kenkyuka Igakubu

Toshinori Fukuda

Oregon Health & Science University Hospital

Kazuyo Yagishita

Saint Luke's International Hospital: Sei Roka Kokusai Byoin

Kengo Takahashi

Tohoku University School of Medicine: Tohoku Daigaku Daigakuin Igakukei Kenkyuka Igakubu

Takuya Ueda

Tohoku University School of Medicine: Tohoku Daigaku Daigakuin Igakukei Kenkyuka Igakubu
https://orcid.org/0000-0002-0913-5791

Hiroko Tsunoda

Saint Luke's International Hospital: Sei Roka Kokusai Byoin

Research Article

Keywords: DBT, FFDM, Ki-67, breast cancer, deep learning

Posted Date: October 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3411805/v1

Version of Record: A version of this preprint was published at Breast Cancer on March 7th, 2024. See the published version at https://doi.org/10.1007/s12282-024-01549-7.

Abstract

Background:

To develop a deep learning (DL) model for digital breast tomosynthesis (DBT) image to predict Ki-67 expression.

Methods:

The institutional review board approved this retrospective study and waived the requisite to obtain the informed consent from the patients. Initially, 499 patients (mean age of 50.5 years, ranging from 29 to 90 years) who were referred to our hospital suggestive of breast cancer were initially enrolled in this study. We selected 126 patients with pathologically confirmed breast cancer and measured Ki-67. Xception architecture was used for the DL model to predict Ki-67 expression. Diagnostic performance of the DL model was assessed by accuracy, sensitivity, specificity, and areas under the receiver operating characteristic curve (AUC). The diagnostic performance was also assessed with sub-datasets divided by radiological characteristics of breast cancer.

Results:

The average accuracy, sensitivity, specificity, and AUC were 0.856, 0.860, 0.654, 0.933, respectively. The AUC of the four sub-groups separated by radiological findings for the mass, calcification, distortion, and focal asymmetric density sub-dataset were 0.890, 0.750, 0.870, and 0.660, respectively.

Conclusions:

Our results suggest potential application of the DL model to predict the expression of Ki-67 using DBT, which may be useful in determining the treatment strategy for breast cancer preoperatively.

Introduction

Breast cancer is the most common cancer that affects women, and its incidence and mortality rates are expected to increase [1]. Determining the molecular subtype of breast cancer is crucial for its treatment [2]. Although subtypes are originally classified by gene expression, genetic testing is not clinically practical in all cases. As a result, subtypes are often classified based on an immunohistochemical (IHC) assessment of biomarkers in the clinical situation: estrogen receptor (ER); progesterone receptor (PR); human epidermal growth factor receptor 2 (HER2) [3]. Surrogate subtypes do not always correspond to molecular subtypes. Ki-67 is also a routinely-used molecular biomarker and is considered as an independent prognostic factor for breast cancer patients, which is expressed as the percentage of positively stained nuclei [4]. Ki-67 is referred to as a reference value in determining indications for chemotherapy, especially in ER-positive HER2-negative breast cancer [5, 6]. Ki-67 is typically assessed through needle biopsy specimens, which is obtained with invasive procedures [4].

Several studies investigated radiological findings associated with the expression of Ki-67 in breast cancer [7–10]. Several studies reported the application of radiomics analysis in predicting Ki-67 expression from radiological images [10–12]. Currently, many researches have emphasized the potential usefulness of applying deep learning (DL) model, especially using convolutional neural networks (CNN), for clinical imaging, such as benign and malignant classification of breast lesions using MR imaging [13].

The purpose of this study is to develop a DL model for digital breast tomosynthesis (DBT) image to predict Ki-67 expression.

Materials and Methods

The institutional review board approved this retrospective study and waived the requirement for written informed consent.

Data Collection

A total of 499 patients (mean age of 50.5 years, ranging from 29 to 90 years) who were referred to our hospital suggestive of breast cancer, admitted, and performed DBT between March 1, 2019, and August 31, 2019 were initially enrolled in this study. We reviewed clinical and pathological information on each patient from our electronic medical record system. The exclusion criteria were (a) patients who had only unilateral breast imaging, (b) male patients, (c) postoperative patients, (d) patients with metal clips after biopsy, and (e) patients whose lesions could not be detected by DBT (Fig. 1). Out of the 425 patients, we selected 126 patients with pathologically confirmed breast cancer and measured Ki-67.

Pathological evaluation and measurement of Ki-67 index

Table 1 shows the characteristics of the tumors. The pathological results and Ki-67 proliferation index were determined using the results of the preoperative biopsy. The pathological results of these 126 patients with breast cancer yielded diagnosis of invasive ductal carcinoma in 100 patients, mucinous carcinoma in 5 patients, invasive micropapillary carcinoma in 1 patient, invasive lobular carcinoma in 12 patients, microinvasive carcinoma in 2 patients, ductal carcinoma in situ (DCIS) in 10 patients.

Table 1
Patients and tumor characteristics

	High-Ki67	Low-i67
Ki-67 index	48.42	14.90
Mean (%) ± SD (Min-Max)	± 12.61 (31–74)	± 7.80 (1–29)
Breasts (n)	35	95
Patients (n)	35	91
Histopathology *		
IDC, n (%)	30 (86)	70 (74)
Mucinous Carcinoma, n (%)	1 (3)	4 (4)
Invasive micropapillary carcinoma, n (%)	0 (0)	1 (1)
ILC, n (%)	1 (3)	11 (12)
Microinvasive carcinoma, n (%)	1 (3)	1 (1)
DCIS, n (%)	2 (6)	8 (8)
High-Ki67: high Ki-67 lesion, Low-Ki67: low Ki-67 lesion, Abbreviations: IDC invasive ductal carcinoma, ILC invasive lobular carcinoma, DCIS ductal carcinoma in situ		
* No significant difference.		

The Ki-67 proliferation index was evaluated by manual counting with a counter. Each lesion was classified into two groups: high Ki-67 lesion (High-Ki67) that showed Ki-67 proliferation index over 30% (35 lesions) and low Ki-67 lesion (Low-Ki67) that showed Ki-67 proliferation less than 30% (95 lesions).

DBT Examinations

All DBT images were acquired on 3Dimentions Mammography System (Hologic, Inc., Bedford, MA). Scanning parameters of DBT images are as follows: kilovoltage peak ranges of 25–34 kV; current ranges of 10–180 mA; exposed time ranges of 51–368 msec; force ranges of 48.9-142.7 N; thick ranges of 16–92 mm; and absorbed dose ranges of 6.10–4.70 mGy. The total tomographic angler range was 15° (-7.5° – 7.5°), consisting of 15 projection views taken at 1° increments. Interslice interval was 1 mm, and resolution was 70 μm × 70 μm per pixel. This study used images from the mediolateral oblique (MLO) view.

Two radiologists with 5 and 30 years of experience annotated the lesions and classified the radiological findings into four categories: calcification, masse, distortion, and focal asymmetric density (FAD). All breasts' lesions were cropped at the center of the lesion and then resized to a grayscale image of 256 x 256 pixels.

DL model

The DL model was constructed using the Xception architecture [14], with pre-training and initialization of the CNN architecture using ImageNet [15]. RAdam was used for optimization, and binary cross-entropy was used as the loss function, with a total of 10,000 steps, a warmup proportion of 0.1, and a minimum learning rate of 1e-5 [16]. The batch size and number of epochs was set to 200 and 400, respectively.

Our computing system consisted of an Intel Core i7-7800X CPU with six cores (Intel, Santa Clara, CA, USA) and an NVIDIA Quadro RTX 8000 GPU with 48 GB of memory (Nvidia, Santa Clara, CA, USA). The operating system was Ubuntu 18.04.5 long term support (LTS): Xenial Xerus. Python 3.8.2 was used, and Keras 2.2.0 with TensorFlow 1.9.0 was used as the deep learning framework.

Statistical analysis

The student's t-test was used to analyze quantitative continuous variables of clinical factors, while the chi-square test was used for discrete variables. The statistical significance was set at $p < 0.05$.

To assess the performance of our DL model, 7-fold cross validation was used, where 80% of the data was used for training and 20% for validation. In cross validation assessment, the validation accuracy was calculated 7 times by changing the DBT images included in the train and validation datasets. All data were separated by patient in the cross-validation process to avoid including DBT images of the same patient in both the training and validation sets.

The performance of our DL model was assessed by accuracy, sensitivity, specificity, and the area under the curve (AUC) of the receiver operating characteristic (ROC) curve. In this study, true positive (TP) lesions were defined as the correct classification of High-Ki67, while true negative (TN) lesions were defined as the correct classification of Low-Ki67. False positive (FP) lesions were defined as the incorrect classification of Low-Ki67 as High-Ki67, and false negative (FN) lesions were defined as the incorrect classification of High-Ki67 as Low-Ki67. Sensitivity and specificity values are for a 0.5 probability cut line. Each is calculated as follows:

$$\text{accuracy} = (TP + TN) / (TP + FP + FN + TN)$$

$$\text{sensitivity} = TP / (TP + FN)$$

$$\text{specificity} = TN / (TN + FP)$$

To evaluate the diagnostic accuracy of each radiological findings, the datasets were divided into four sub-datasets: calcification dataset, mass dataset, distortion dataset, and FAD dataset. The diagnostic accuracy of our DL model for each sub-dataset was independently evaluated to examine the relationship between diagnostic performance and radiological findings.

Results

Table 1 presents the clinical characteristics of the patients included in the study. The mean age of the High-Ki67 and Low-Ki67 groups were 50.29 ± 10.53 years (range 36–81) and 54.39 ± 12.96 (range 32–88) years, respectively. There was no significant difference in age between the two groups.

Figure 2 shows the ROC curve of our DL model using 7-fold cross-validation. The AUC value for our DL model was 0.856. The accuracy, sensitivity and specificity of our DL model were 0.860, 0.654, 0.933, respectively. Table 2 shows the diagnostic accuracies performance of the four sub-groups separated by radiological findings: 0.890 for the mass sub-dataset, 0.750 for the calcification sub-dataset, 0.870 for the distortion sub-dataset, and 0.660 for the FAD sub-dataset. Figure 3 illustrates representative images of TP, TN, FP and FN lesions.

Table 2
Diagnostic accuracies of 4 groups

	Accuracy	Sensitivity	Specificity
Mass	0.89	0.72	0.95
Calcification	0.75	0.64	0.82
Distortion	0.87	0.64	0.94
FAD	0.66	-	0.66
Abbreviations: FAD focal asymmetric density			

Discussion

Our research indicates potential clinical application of DL model to predict expression of Ki-67 using breast cancer imaging before surgery. As Ki-67 expression is an important indicator, which also influences breast cancer subtype classification, early prediction of the Ki-67 expression through clinical images before obtaining biopsy results may have potential benefits by enabling early decision-making to determine initial treatment strategy.

Previous research reported the usefulness of radiomics analysis to predict Ki67 values of breast cancer [12, 17]. In spite of the different setting of the cut-off values for positive Ki-67 status, our DL model showed higher AUC values than previous radiomics analyses [12, 17]. Previous radiomics studies suggested that the results of radiomics analysis may vary significantly depending on the setting of the ROI due to tumor heterogeneity of breast cancer, such as the periphery or inside of the tumor [12, 17] and the robustness of radiomics analysis is highly affected by the ROI setting [17]. Whereas, DL model using CNN, as a process to convolute morphological information of image, allows to capture comprehensive information throughout the entire tumor. Therefore, our DL model inherently offers potentially more consistent and reliable insights.

Although image-based prediction of Ki-67 expression is also investigated by various imaging modalities such as MRI [10], our DL model used DBP for the imaging modality. DBT, as an advanced imaging modality of FFDM, has become a widely used and commonly employed breast imaging technique. Our DL models using widely available DBP imaging have potential advantage of clinical application when compared to MRI. In addition, as DBT provides more detailed information about morphological information of tumor when compared to FFDM, this may accurately reflect tumor heterogeneity [19].

Our result suggested that the predictive accuracy of Ki-67 expression varies among sub-dataset of radiological characteristics of breast cancer. The accuracy in the mass sub-dataset was higher compared to the other sub-datasets, whereas that in the calcification sub-dataset was lower compared to the other sub-datasets. This pattern is consistent with the prior research, where a lower accuracy of calcification compared to other findings was observed to predict the presence of stromal invasion of breast cancer [20]. In the paper, they suggested that the DL model did not represent the relationship between calcification and invasion because of the donwnsampling image processing for DBT image. We assumed the same reason for the lower accuracy to predict Ki-67 expression in the calcification sub-dataset. The accuracy in the distortion sub-dataset was lower than that in the mass sub-dataset, and lower than that in the calcification sub-dataset. Although image interpretation is sometimes difficult, presence of distortion, presence of distortion considered to be pathognomonic for malignancy, when once detected. Therefore, it is reasonable that the accuracy for distortion would fall between that of mass and calcification. FAD sub-dataset did not show any clear trends, which may be associated with the small number of cases.

The clinical utility of Ki-67 has undergone historical transitions. It was previously used for the classification of luminal breast cancer. Based on the St. Gallen International Consensus Guidelines [3] previous radiomics studies set the threshold of Ki-67 expression 14% [21]. Recently, many studies recommended clinically significant threshold of Ki-67 expression as 30% in association with determination of adjuvant chemotherapy in hormone-positive, HER2-negative early breast cancer [21]. This is the reason why 30% was used as the threshold for Ki-67 classification in this research.

Our research has several limitations. First, this research conducted at a single institution using a single vendor, has a small sample size. This could impact the accuracy of the predictions and limit the generalizability of the findings. In the future, it will be necessary to conduct studies across multiple institutions and with larger sample sizes to ensure the robustness of our DL model and its applicability to broader populations. Second, we used cropped images of lesions for our DL model. For clinical application, targeted breast lesions need to automatically detect targeted breast lesions. Third, the expression of Ki-67 was obtained from a preoperative biopsy sample. Localized biopsy samples may not represent the Ki-67 expression of entire tumors due to heterogeneity.

In conclusion, the DL model utilizing DBT has the potential to accurately predict the expression of Ki-67, which can serve as a valuable non-invasive tool in determining the treatment strategy for breast cancer preoperatively.

Declarations

Author Contributions

Conceptualization: Takuya Ueda and Hiroko Tsunoda; Methodology: Ken Oba, Takuya Ueda and Hiroko Tsunoda; Investigation: Ken Oba, Toshinori Fukuda, Kazuyo Yagishita and Hiroko Tsunoda; Formal analysis: Ken Oba; Validation: Maki Adachi, Tomoya Kobayashi, Daiki Shimokawa and Kengo Takahashi; Data Curation: Ken Oba, Maki Adachi, Daiki Shimokawa, Toshinori Fukuda, Kengo Takahashi and Hiroko Tsunoda; Software: Maki Adachi, Eichi Takaya, Daiki Shimokawa and Kengo Takahashi, Resources: Ken Oba, Toshinori Fukuda, Kazuyo Yagishita and Hiroko Tsunoda; Writing - original draft preparation: Ken Oba, Takuya Ueda and Hiroko Tsunoda; Writing - review and editing: Maki Adachi, Daiki Shimokawa, Toshinori Fukuda, Kengo Takahashi and Kazuyo Yagishita; Supervision: Takuya Ueda and Hiroko Tsunoda; Funding acquisition: Takuya Ueda.

Acknowledgements

Funding

This work was supported by JST, CREST JPMJCR15D1, Japan and JSPS KAKENHI 20H03637, Japan.

Potential Conflict of Interest

Takuya Ueda received two research grants: Kakenhi, MEXT/JSPS WISE Program from MEXT, Grant Number 20H03738; CREST from JST, Grant Number JPMJCR15D1.

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;68:394–424.
2. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.
3. Goldhirsch A, Wood WC, Coates AS, Gelber RD, Thürlimann B, Senn H-J, et al. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol. 2011;22:1736–47.
4. Lakhani SR, International Agency for Research on Cancer. WHO classification of breast tumours. 2nd ed. Who Classification of Tumours Editorial Board, editor. IARC; 2019.
5. de Azambuja E, Cardoso F, de Castro G Jr, Colozza M, Mano MS, Durbecq V, et al. Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. Br J Cancer. 2007;96:1504–13.
6. Thomssen C, Balic M, Harbeck N, Gnant M. St. Gallen/Vienna 2021: A Brief Summary of the Consensus Discussion on Customizing Therapies for Women with Early Breast Cancer. Breast Care .

2021;16:135–43.

7. Jiang L, Ma T, Moran MS, Kong X, Li X, Haffty BG, et al. Mammographic features are associated with clinicopathological characteristics in invasive breast cancer. Anticancer Res. 2011;31:2327–34.
8. Cheng C, Zhao H, Tian W, Hu C, Zhao H. Predicting the expression level of Ki-67 in breast cancer using multi-modal ultrasound parameters. BMC Med Imaging. 2021;21:150.
9. Fang J, Zhao W, Li Q, Zhang B, Pu C, Wang H. Correlation Analysis of Conventional Ultrasound Characteristics and Strain Elastography with Ki-67 Status in Breast Cancer. Ultrasound Med Biol. 2020;46:2972–8.
10. Surov A, Clauser P, Chang Y-W, Li L, Martincich L, Partridge SC, et al. Can diffusion-weighted imaging predict tumor grade and expression of Ki-67 in breast cancer? A multicenter analysis. Breast Cancer Res. 2018;20:58.
11. Ma W, Ji Y, Qi L, Guo X, Jian X, Liu P. Breast cancer Ki67 expression prediction by DCE-MRI radiomics features. Clin Radiol. 2018;73:909.e1-909.e5.
12. Tagliafico AS, Bignotti B, Rossi F, Matos J, Calabrese M, Valdora F, et al. Breast cancer Ki-67 expression prediction by digital breast tomosynthesis radiomics features. Eur Radiol Exp. 2019;3:36.
13. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic versus Convolutional Neural Networks Analysis for Classification of Contrast-enhancing Lesions at Multiparametric Breast MRI. Radiology. 2019;290:290–7.
14. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. p. 1800–7.
15. Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. 2009;248–55.
16. Liu L, Jiang H, He P, Chen W, Liu X, Gao J, et al. On the Variance of the Adaptive Learning Rate and Beyond [Internet]. arXiv [cs.LG]. 2019. Available from: http://arxiv.org/abs/1908.03265
17. Jiang T, Jiang W, Chang S, Wang H, Niu S. Intratumoral analysis of digital breast tomosynthesis for predicting the Ki-67 level in breast cancer: A multi-center radiomics study. Medical [Internet]. 2022; Available from: https://aapm.onlinelibrary.wiley.com/doi/abs/10.1002/mp.15392
18. Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, et al. Assessing robustness of radiomic features by image perturbation. Sci Rep. 2019;9:614.
19. Amer HA, Schmitzberger F, Ingold-Heppner B, Kussmaul J, El Tohamy MF, Tantawy HI, et al. Digital breast tomosynthesis versus full-field digital mammography-Which modality provides more accurate prediction of margin status in specimen radiography? Eur J Radiol. 2017;93:258–64.
20. Shimokawa D, Takahashi K, Oba K, Takaya E, Usuzaki T, Kadowaki M, et al. Deep learning model for predicting the presence of stromal invasion of breast cancer on digital breast tomosynthesis. Radiol Phys Technol [Internet]. 2023; Available from: http://dx.doi.org/10.1007/s12194-023-00731-4
21. Burstein HJ, Curigliano G, Thürlimann B, Weber WP, Poortmans P, Regan MM, et al. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus

Guidelines for treatment of early breast cancer 2021. Ann Oncol. 2021;32:1216–35.

Figures

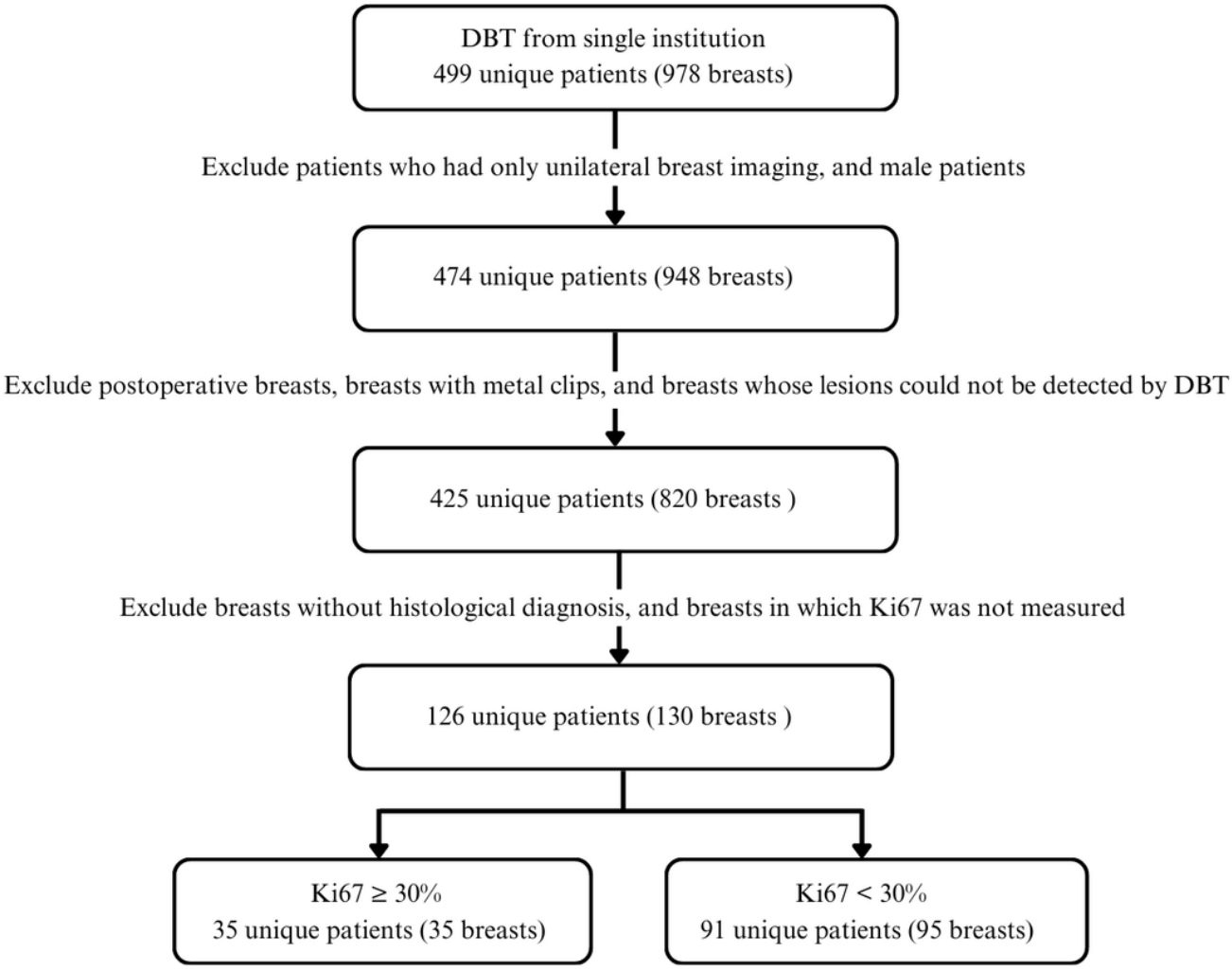

Figure 1

Flowchart of inclusion and exclusion criteria.

In total, 126 patients were analyzed in this study. DBT = digital breast tomosynthesis

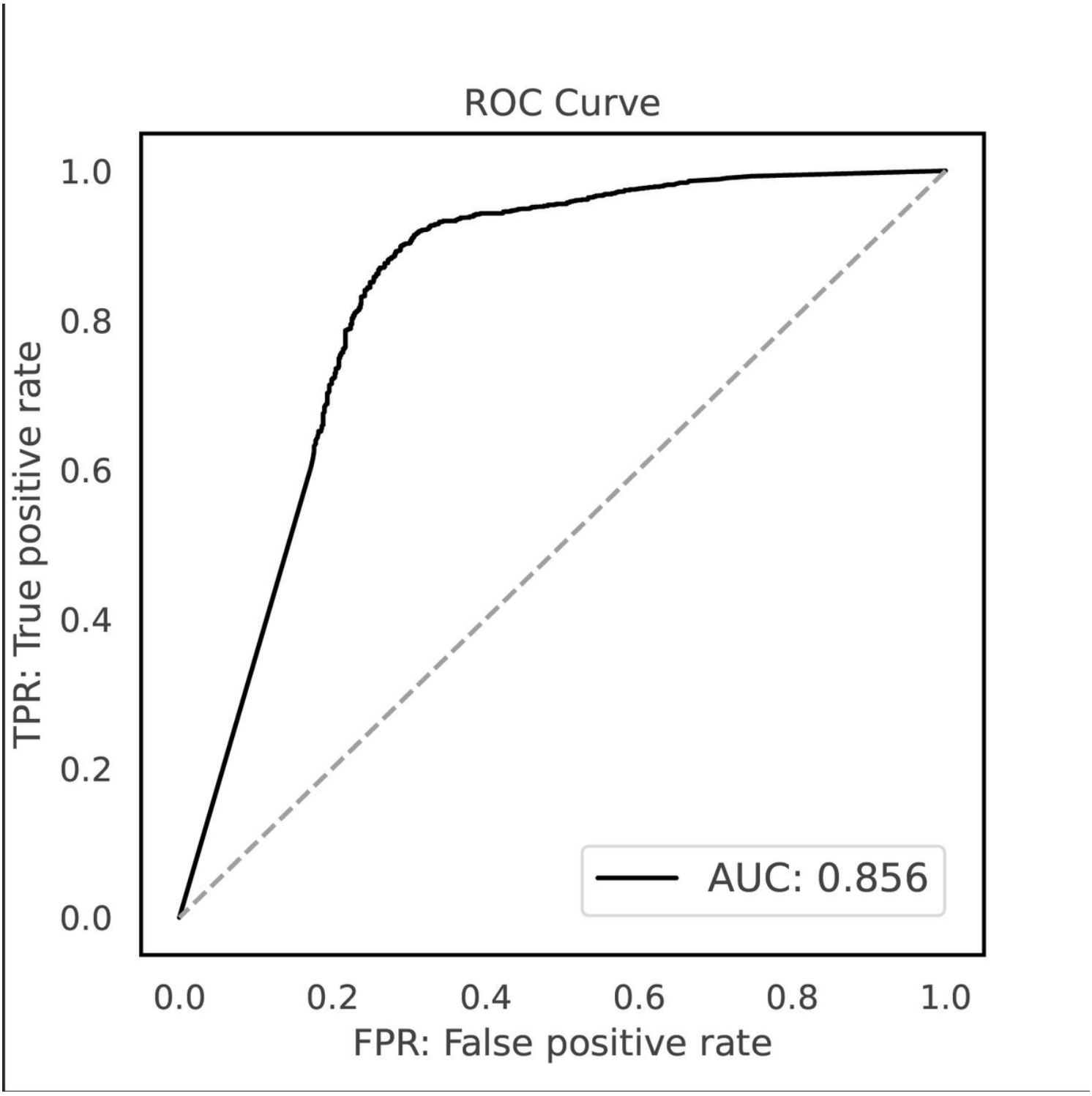

Figure 2

Receiver operating characteristic curves (ROC) for testing, based on the Ki-67 cut-off value of 30%. (AUC: area under the curve).

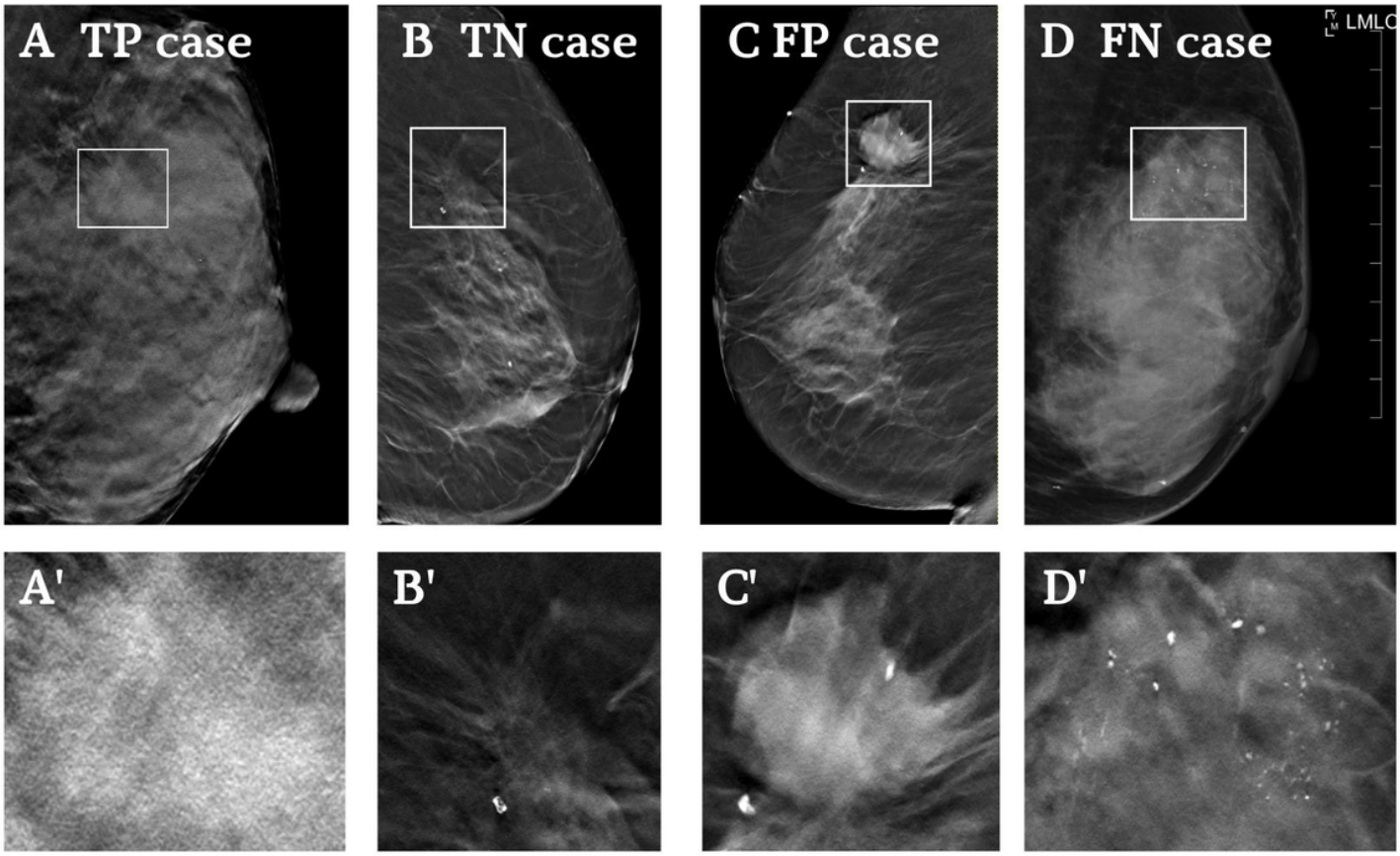

Figure 3

Representative image from the dataset.

(A) DBT and cropped image (A') of true positive case (TP).

A 46-year-old woman with high-Ki-67 expression (43%). Spiculated mass was observed in digital breast tomosynthesis (DBT). The entire tumor is included in the area of interest. Our DL model accurately classified cases as High-Ki67.

(B) DBT and cropped image (B’) of true negative case (TN).

A 75-year-old woman with low-Ki-67 expression (5%). Distortion is the only observable finding in DBT, and obvious calcification or mass cannot be identified. Our DL model accurately classified cases as Low-Ki67.
(C) DBT and cropped image (C’) of false positive case (FP).

A 75-year-old woman with high-Ki-67 expression (22%). DBT shows high-density tumor with distortion on the dorsal side. Our DL model mistakenly classified cases as High-Ki67.

(D) DBT and cropped image (D’) of false positive case (FN).

A 50-year-old woman with high-Ki-67 expression (51%). DBT shows grouped fine pleomorphic and amorphous calcifications spanning 3.0 cm. The cropped ROI does not contain some calcification. Our DL model mistakenly classified cases as Low-Ki67.