The European Research Journal 2025;11(2):279-288 Original Article
DOI: https://doi.org/10.18621/eurj.1600293 *Radiology*

Differentiating types of breast cancer from digital mammography images with artificial intelligence methods

Ela Kaplan[1], Orhan Yaman[2], Hacı Taner Bulut[1], Mehmet Şirik[1], Türker Tuncer[2], Şengül Doğan[2]

[1]Department of Radiology, Faculty of Medicine, Adıyaman University, Adıyaman, Türkiye; [2]Department of Digital Forensics Engineering, Faculty of Technology, Fırat University, Elazığ, Türkiye

ABSTRACT

Objectives: Breast cancer (BCA) is one of the world's most prevalent cancer and the top cause of mortality. For many decades, mammography has been used routinely for screening of early breast cancer and diagnosing symptomatic patients. The main purpose of this work is to investigate the usefulness of machine learning techniques using mammography images.

Methods: A total of 194 patients who underwent ultrasound examination after observing suspicious lesions on mammography images and were diagnosed with BCA by ultrasound-guided core needle biopsy were included in the study. A set of mammography images with complete cancer subtypes was used. A transfer learning-based computer vision method was adopted in this study. AlexNet was to extract the features and select the most significant features using a feature selection function. Our deep learning-based model attained more than 80% accuracy in classifying malignant and benign cancers. However, the employed deep learning model cannot classify subtypes accurately.

Results: Per the results, the commonly used image classification model is highly accurate in distinguishing malignant and benign changes, however unable to classify cancer subtypes.

Conclusions: In conclusion, machine learning can still not simulate conventional immunohistochemistry subtyping using tissue biopsy.

Keywords: Artificial intelligence, breast cancer, deep learning, mammography

Breast cancer (BCA) is the most common malignancy in women [1]. For this reason, screening programs have been developed for women for breast cancer and efforts are being made to detect cancer at an early stage [2]. The most effective, most frequently used and easily accessible method for screening breast cancer is mammography. Although the success of mammography in detecting breast cancer depends on breast density, its specificity varies between 75-90% [3, 4]. BCA tumor morphology is diverse and results from different markers in the intratumoral heterogeneity structure [5, 6]. Determining BCA types is very important in treatment planning. Although histopathological sampling is a minimally invasive procedure, it is very important to be able to diagnose and determine breast cancer types with non-

Corresponding author: Ela Kaplan, MD., Assist. Prof.,
Phone: +90 416 223 16 90, E-mail: elakaplan15@gmail.com

How to cite this article: Kaplan E, Yaman O, Bulut HT, Şirik M, Tuncer T, Doğan Ş. Differentiating types of breast cancer from digital mammography images with artificial intelligence methods. Eur Res J. 2025;11(2):279-288. doi: 10.18621/eurj.1600293

Received: December 12, 2024
Accepted: December 28, 2025
Published Online: February 10, 2025

Copyright © 2025 by Prusa Medical Publishing
Available at https://dergipark.org.tr/en/pub/eurj

This is an open access article distributed under the terms of Creative CommonAttribution-NonCommercial-NoDerivatives 4.0 International License

invasive methods [7]. In recent years, artificial intelligence (AI) methods have been increasingly adapted to assist in cancer detection, reducing false negative or false positive findings, reducing recall rate, and at a more advanced level, assisting in identifying cancer subtypes [8]. There are studies investigating the detectability of tumoral lesions by analyzing digital mammographic images used in BCA screening with AI methods [9, 10]. There is also increasing interest in exploring the possibility of AI to assist in cancer subtyping and predicting treatment response [11, 12].

The aim of this study is to investigate the detectability of malignant breast lesions and BCA subtypes in mammography images using deep learning methods.

METHODS

This study is a cross-sectional study conducted by retrospective recruitment of patients who underwent mammography in a tertiary center between 2017 and 2021. The study was carried out according to the Declaration of Helsinki and local ethics committee approval was obtained.

Table 1. Characteristic of the BCA included in the selected patient dataset.

	Data
Age (years)	42.22±11.25
Breast Density, n (%)	
C	102 (52.6%)
D	92 (47.4%)
Breast cancer molecular subtypes, n (%)	
Luminal A	72 (37.1%)
Luminal B HER –	40 (20.6%)
Luminal B HER +	54 (27.8%)
HER+	18 (9.3%)
Triple negative	10 (5.2%)
Breast Cancer Types, n (%)	
Invasive ductal carcinoma	182 (93.8%)
Invasive lobular carcinoma	12 (6.2%)

Data are shown as mean± standard deviation or n (%). HER2=Human epidermal growth factor receptor 2

Study Population

Patients with breast density C and D presenting BIRADS 4A-4B-4C and 5 lesions on mammography images according to ACR BIRADS (Breast Imaging-Reporting and Data System) were recruited. Patients with inconclusive tissue biopsy or benign lesions were excluded. A total of 194 patients who underwent ultrasound examination by observing suspicious lesions on mammography images and were diagnosed with BCA by ultrasound-guided core needle biopsy were included.

The presence of estrogen receptors (ER), progesterone receptor (PR), human epidermal growth factor (HER2) gene expression, and Ki-67 immunohistochemical staining were obtained from the pathological evaluation results of each patient and was presented in Table 1.

Identification and Classification of Tumoral Subtypes

The presence of tumoral ER, PR, HER2, and immunohistochemical staining with Ki-67 as proliferation marker and epidermal growth factor (EGF) play a role in determining BCA subtypes [5]. BCA was divided into subtypes according to its molecular properties.

Molecular subtypes of invasive breast tumors are divided into 4 basic groups: 1- Luminal A like: ER-positive (+) PR+ HER2 negative (-) and Ki-67 index <15%, Luminal B like: This group is divided into two: Luminal B like (HER2 -) ER+ HER2- and at least one; P- or weak Ki-67:>30%, Luminal B like (HER 2 +): ER: +, HER2: overexpressed or amplified, Ki-67: any result, PR: any result, 3) HER 2 +: ER-, PR:-, HER2: overexpressed or amplified, 4) Triple-negative: ER+, PR+, Ki-67: negative [13].

Artificial Intelligence Method

Dataset

This dataset had 806 mammography images with two categories. These categories were named Malignant Group and Control Group. A total of 382 images belonged to malignant and there were 424 images in the normal category. These images were stored in JPEG format. Images of 382 lesions in the malignant group were categorized according to their subtypes. Craniocaudal (CC) and Mediolateraloblique (MLO) images of the breast with the detected lesion were ob-

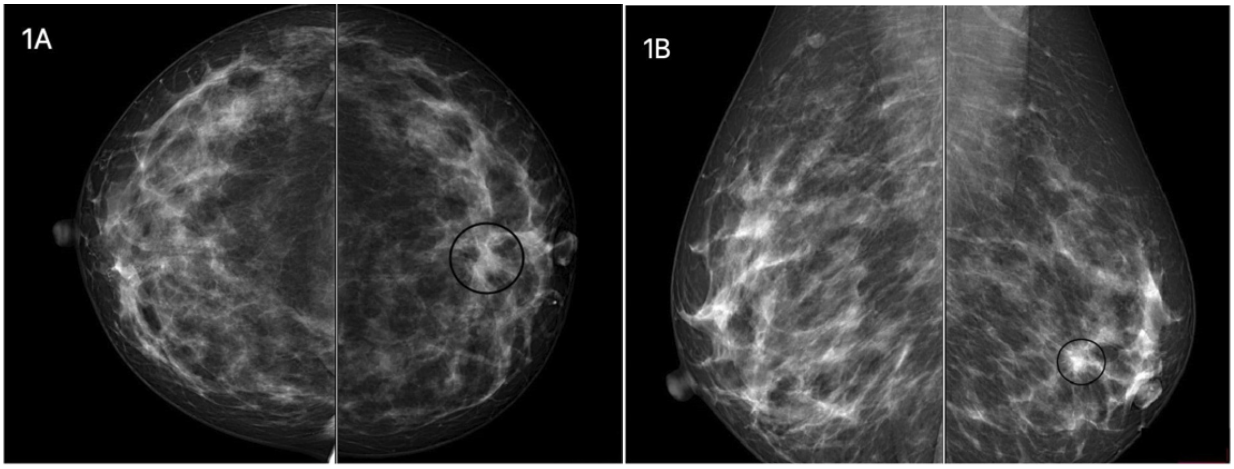


Fig. 1. Mamography images. (a) CC mammogram of a 12×10 mm, irregularly shaped, spiculated contoured mass lesion in the retroareolar area of the left breast (circle) in a 52-year-old female patient. (b) MLO mammogram of a 12×10 mm, irregularly shaped, spiculated contoured mass lesion in the retroareolar area of the left breast (circle) in a 52-year-old female patient.

tained from 188 patients and 6 patients in whom the lesion could be detected only in CC mammograms. Mammography images are presented in Fig. 1.

First Aim

The proposed model has four sections and these sections are breast image segmentation (breast area localization), deep feature extraction using pre-trained AlexNet, NCA-based feature selection, and classification with Cubic SVM classifier. An overview of the proposed intelligent AlexNet-NCA-SVM-based model is shown in Fig. 2.

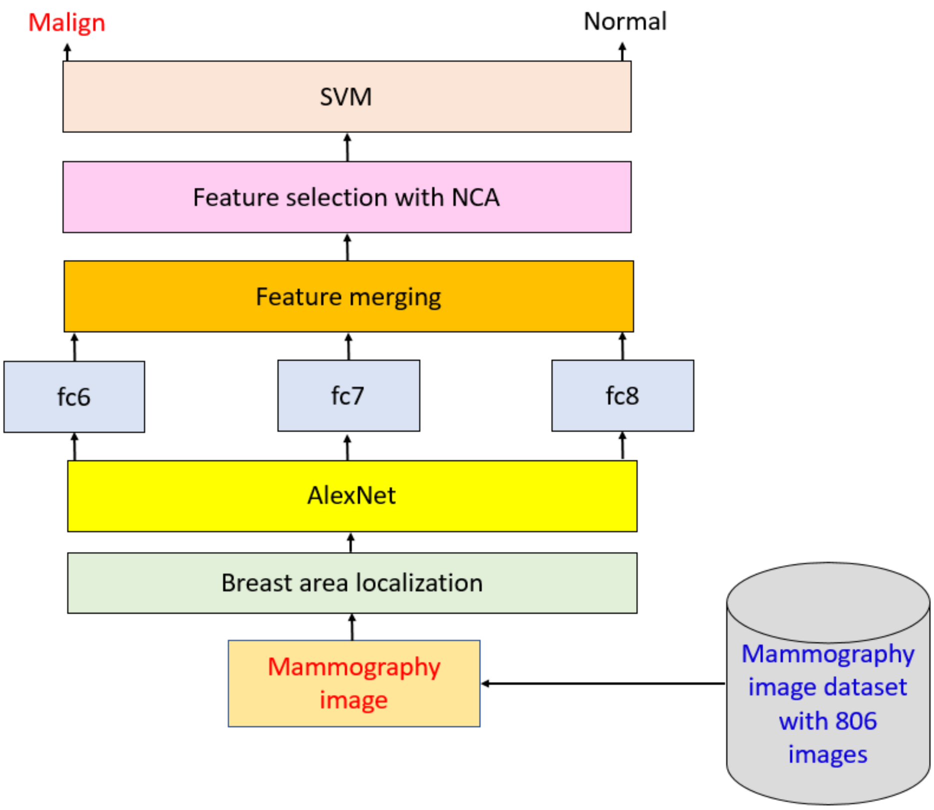


Fig. 2. Overview of the proposed AlexNet-NCA-SVM based breast cancer detection model. .

The steps of the presented model are given as follows.

Step 1: Load the mammography image dataset.

Step 2: Read each mammography image.

Step 3: Segment the breast using binary image conversion.

Step 4: Apply pre-trained AlexNet [14] on these images. The used AlexNet was trained on the ImageNet [15] dataset and the ImageNet dataset contains a few million images with 1000 classes. We transferred this information (optimized weights) to solve our BCA detection problem.

Step 5: The used pre-trained AlexNet has three fully connected layers and the names of them are fc6, fc7, and fc8. By using the fc6, fc7, and fc8, 4096, 4096, and 1000 features are generated from an image respectively.

$$f^1 = fc6(M) \quad (1)$$

$$f^2 = fc7(M) \quad (2)$$

$$f^3 = fc8(M) \quad (3)$$

In Equations (1)-(3), we used three deep feature creation functions and these functions are $fc6(.)$,$fc7(.)$ and $fc8(.)$. These functions are called names of the fully connected layers. By employing these function, first (f^1), second (f^2) and third (f^3) feature vectors with a length of 4096, 4096 and 1000 are generated.

Step 6: Merge the generated features and obtain 4096+4096+1000=9192 features.

$$X=merge(f^1,f^2,f^3) \quad (4)$$

Herein, merge(.,.,.) is concatenation function and X defines merged feature vector with a length of 9192.

Step 7: Select the most informative 1000 features from the generated 9192 features deploying NCA. NCA is a feature selection version of the k nearest neighbor (kNN) and uses the distance metric (Manhattan distance) to calculate the weights of the features. The most valuable features are selected using weight values. The big weight values assign discriminative features and small weight values denote redundant features. We selected the top 1000 features by using indexes of the weights.

Step 8: Classify the best 1000 features deploying the SVM classifier. The attributes of the used SVM are; Kernel: 3rd degree polynomial (Cubic), Box-constraint: 2, Standardize: True, and Validation: 10-fold CV.

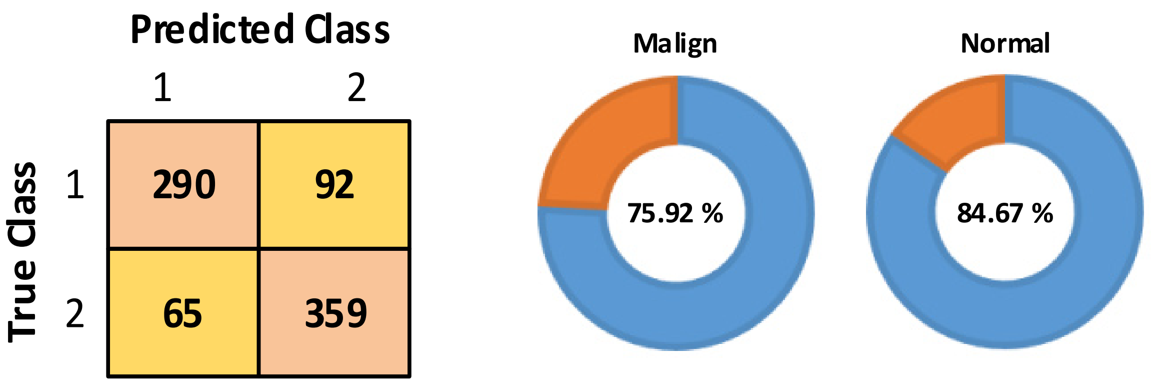
Second Aim

Deep learning is the flagship of machine learning and various automated diagnosis models. In this work, we have used a deep learning model to classify the mammogram according to tumor subtype. To classify our images, DarkNet53 [15] which is a widely preferred computer vision model in the literature. These categories are: (1) Luminal B HER2+, (2) HER 2+, (3) Luminal A, (4) Luminal B, HER- (5) Control Group, and (6) Triple Negative Subtype.

RESULTS

Detection of BCA with Deep Learning Method

In this study, MATLAB 2020a programming environment and Classification Learner Toolbox have been used and the results were calculated. In the collected dataset, there are 382 “Malignant” and 424 “Normal” mammography images, and the proposed AlexNet-



Confusion Matrix   Class by Class Results

Fig. 3. Confusion matrix and class-wise results of the proposed AlexNet-NCA-SVM based model using Cubic SVM classifier.

Table 2. Accuracy, sensitivity, specificity, geometric mean, balanced accuracy, precision and F1-score results of our proposal

Parameters	Results (%)
Accuracy	80.52
Sensitivity	75.92
Specificity	84.67
Geometric mean	80.17
Balanced Accuracy	80.29
Precision	79.60
F1-score	82.06

F1 Score is the harmonic mean of Precision and Recall scores. It calculated in AI.

NCA-SVM model was developed. Cubic SVM is chosen as a classifier. The calculated confusion matrix and class-wise results are also shown in Fig. 3.

As seen in Figure 3, 75.92% accuracy for "Malign class" and 84.67% accuracy for "Normal class" were calculated. Accuracy, Sensitivity, Specificity, Geometric mean, Balanced Accuracy, Precision, and F1-Score values were calculated by running 1000 iterations of the Cubic SVM classifier, and the calculated best results are listed in Table 2 [16-18].

The results calculated in Fig. 3 were obtained by performing a 10-Fold CV [19]. The calculated fold-wise accuracy results are shown in Fig. 4.

Fig. 4 denotes that, the highest result was obtained with fold-7 and the lowest result with fold-3 in the calculated fold-wise accuracies. The cubic SVM algorithm was used to classify the calculated features in the proposed method. The reason why this classifier is preferred is that it achieves higher performance results than other used classifiers. 12 classifiers have been used for classification and they are Fine Decision Tree (DT), Medium DT, Coarse DT, Linear Discriminant, Gaussian Naïve Bayes, Linear SVM, Cubic SVM, Quadratic SVM, Fine k nearest neighbor (KNN), Medium KNN, Cosine KNN, and Ensemble Boosted Trees [18]. The accuracy results calculated for 12 classifiers are shown in Fig. 5.

As can be seen in Fig. 5, Cubic SVM was preferred among these classifications with high accuracy. The calculated unit annulus rate (UAR) rates are calculated as 0.78 and 0.85 respectively for these groups. Therefore, the average UAR rate is equal to 81.50%. (Fig. 6).

The calculated UAR rates are calculated as 0.78 and 0.85 respectively. Therefore, the average UAR rate is equal to 81.50%.

Differentiation of Breast Cancer Subtypes with Deep Learning Methods

We could not obtain successful results on our dataset. By applying DarkNet53, the reckoned confusion matrix is denoted in Fig. 7. In this confusion ma-

Fig. 4. The calculated fold-wise accuracies employing Cubic SVM classifiers.

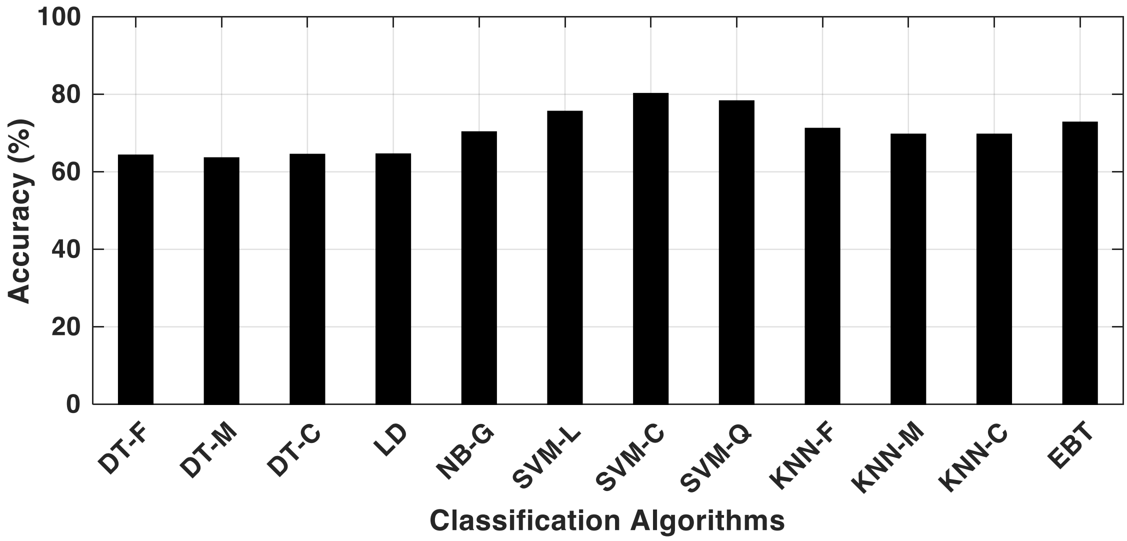


Fig. 5. The accuracy results calculated for Fine DT, Medium DT, Coarse DT, Linear Discriminant, Gaussian Naïve Bayes, Linear SVM, Cubic SVM, Quadratic SVM, Fine KNN, Medium KNN, Cosine KNN, and Ensemble Boosted Trees. DT=decision tree, KNN=k nearest neighbor.

trix, true positive rates (TPR) and false negative rates (FNR) have been shown using blue and pink color tones. According to this matrix, 49.49% classification accuracy has been calculated (Fig. 7).

Moreover, the class-wise area under curves (AUC) has been reckoned and the results are illustrated in Fig. 8.

DISCUSSION

In this work, a set of mammograms with biopsy-proven cancer was collected and an automatic malignant lesion detection model was presented using transfer learning. By using transfer learning, the pre-

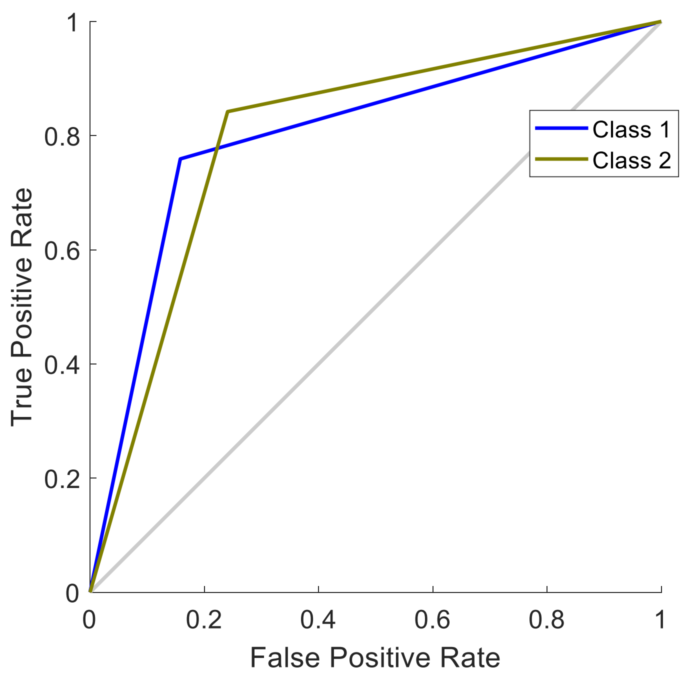


Fig. 6. These classes are named Normal (Class 1) and Malign (Class 2).

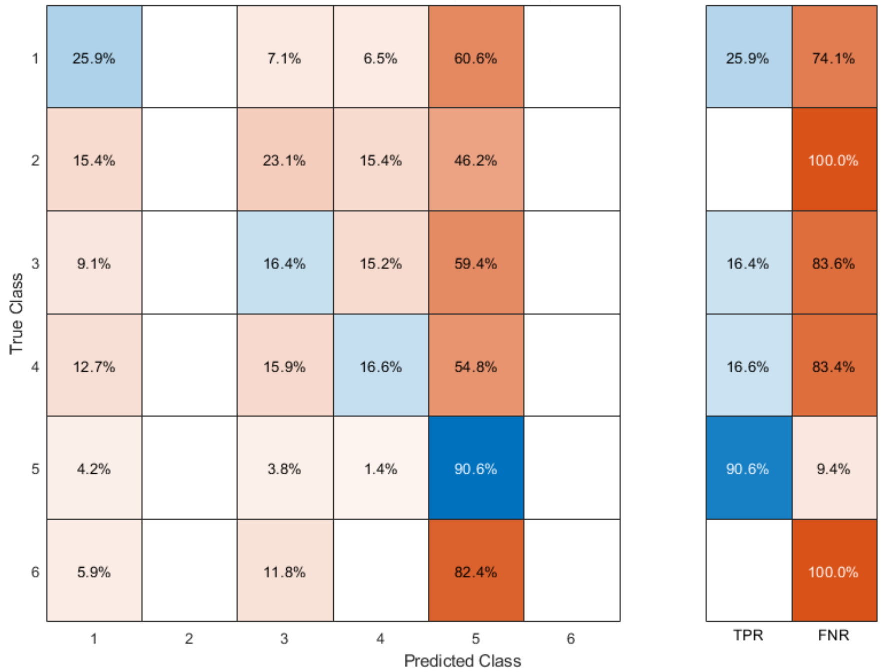


Fig. 7. Confusion matrix and class-wise results of the proposed AlexNet-NCA-SVM based model using Cubic SVM classifier for BCA subtypes.

trained networks have been used and the optimal weight has been used for feature extraction and there is no need to use back propagation. Therefore, we used AlexNet in transfer learning mode to extract deep features. All fully connected layers of the AlexNet have been utilized as feature generation functions and 9192 features are generated from each mammography image. NCA has chosen 1000 features and Cubic SVM has been used to classify these features. The presented AlexNet-NCA-SVM-based mammography classification method attained 80.52% classification accuracy, 75.92% sensitivity, and 84.67% specificity rates. In this view, our model is successful in detecting BCA images in the evaluation and varies depending on breast density.

Classification of mammographic images as benign and malignant is the most important step in the evaluation and varies depending on breast density [19]. In a multicenter study, the performance of screening mammograms in the evaluation of radiologists was investigated, and this sensitivity was reported as 73% and specificity as 96% in the initial evaluation [20]. This highlights the necessity of robust AI methods to assist radiologists in the initial evaluation. There are multiple studies on the use of AI methods in the evaluation of screening mammograms and their results are variable. In these studies, conducted by evaluating the images in screening mammograms by different radiologists and AI, it has been shown that AI contributes to the evaluation of mammographic images and increases the detection of breast cancer [21-23].

In this study, the effectiveness of the deep learning method was investigated in a dataset that included mammographic images of patients with C and D breast patterns with BCA detected and normal mammograms. According to the results of our study, the sensitivity and specificity of BCA detection by deep learning method in mammographic images of patients with dense breast patterns is very high. In other words, the use of artificial intelligence methods in the initial evaluation of screening mammography may increase the evaluation performance of radiologists, reduce reading time, and reduce recall rate.

In the curves in which the AI method was used to determine the pathological subtypes of BCA detected in mammography images, the AUC value was 0.75 in determining the HER2+ subtype, while the AUC value in the detection of other subtypes varied between 0.64-0.69. These low values may be due to the small number of patients in the groups, as well as the lack of

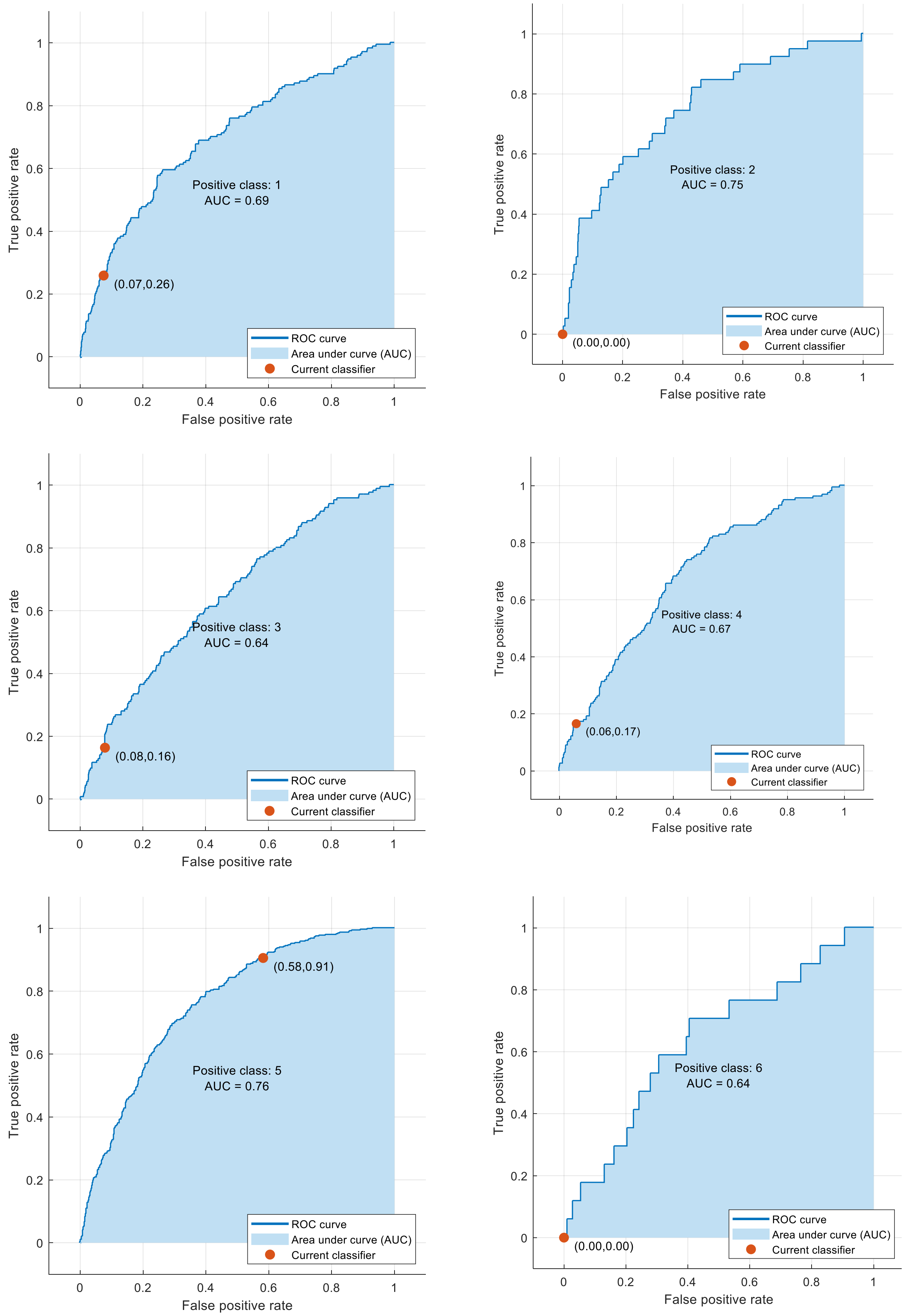


Fig. 8. ROC curves of BCA subtypes classes.

specific imaging characteristics of subtypes in mammographic imaging features. However, studies using larger databases can still be conducted to determine the effectiveness of AI methods in detecting the HER2+ subtype.

Limitations

There are some limitations of our study. First of all, this study is a single-center study, and evaluation was made using a single AI method. In addition, unlike other studies investigating the diagnostic effectiveness of AI in mammography, the evaluation performance of radiologists was not taken into account in this study. Based on the idea that radiologists can detect asymmetric breast tissue by comparing both breasts in the first reading, it was tried to determine the diagnosis rates of AI methods compared to normal breast tissue. This finding may have led to the detection of a lower rate compared to the higher AUC rates in other studies and may have made the detection of BCA subtypes unsuccessful. However, from another perspective, it may also represent a strong aspect of this study. Undoubtedly, the purpose of using AI methods in the evaluation of screening mammograms is to obtain positive results independent of radiologists, and to ensure that the radiologists waive their workforce in the initial evaluation.

CONCLUSION

The use of AI-based screening methods in the field of mammography can be used as auxiliary methods for radiologists by including them in the medical diagnosis stages. The most important concept in radiomics is to see beyond what a radiologist can see, and subsequently aid in diagnosing, reducing 'missed' lesions, and finally, assisting in treatment planning and prognostication. However, the detection of radiomic features of mammography with AI is insufficient, and its clinical use should be supported by studies with larger patient populations. Therefore, in future work, we aim to collect bigger mammography datasets to detect more classes of the BCA and new generation deep learning models can be proposed to detect type of the breast accurately. Although AI detection of mammographic features and classification of tumor subtyping will not replace biopsies, it will have an important place in the diagnosis and treatment of breast cancer.

Ethical statement

The study was approved by Adıyaman University Non-Interventional Clinical Research Ethics Committee (Date: 26.10.2021, number: 08).

Authors' Contribution

Study Conception: EK; Study Design: EK, TT, ŞD; Supervision: EK, TT, ŞD; Materials: EK, HTB, MŞ; Data Collection and/or Processing: EK, MŞ; Statistical Analysis and/or Data Interpretation: EK, HTB, MŞ; Literature Review: EK, TT, ŞD; Manuscript Preparation: EK, TT, ŞD, OY and Critical Review: TT, OY.

Conflict of interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during conduction or writing of this study.

REFERENCES

1. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M. The benefits and harms of breast cancer screening: an independent review. Br J Cancer. 2013;108(11):2205-2240. doi: 10.1038/bjc.2013.177.
2. Li X, Qin G, He Q, et al. Digital breast tomosynthesis versus digital mammography: integration of image modalities enhances deep learning-based breast mass classification. Eur Radiol. 2020;30(2):778-788. doi: 10.1007/s00330-019-06457-5.
3. Conti A, Duggento A, Indovina I, Guerrisi M, Toschi N. Radiomics in breast cancer classification and prediction. Semin Cancer Biol. 2021;72:238-250. doi: 10.1016/j.semcancer.2020.04.002.
4. Hogg P, Kelly J, Mercer C. Digital Mammography: Springer; 2015.
5. Wu M, Ma J. Association Between Imaging Characteristics and Different Molecular Subtypes of Breast Cancer. Acad Radiol. 2017;24(4):426-434. doi: 10.1016/j.acra.2016.11.012.
6. Senkus E, Kyriakides S, Penault-Llorca F, et al; ESMO Guidelines Working Group. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2013;24 Suppl 6:vi7-23. doi: 10.1093/annonc/mdt284.
7. Lin F, Wang Z, Zhang K, et al. Contrast-Enhanced Spectral Mammography-Based Radiomics Nomogram for Identifying Benign and Malignant Breast Lesions of Sub-1 cm. Front Oncol.

2020;10:573630. doi: 10.3389/fonc.2020.573630.
8. Pesapane F, Suter MB, Rotili A, et al. Will traditional biopsy be substituted by radiomics and liquid biopsy for breast cancer diagnosis and characterisation? Med Oncol. 2020;37(4):29. doi: 10.1007/s12032-020-01353-1.
9. Watanabe AT, Lim V, Vu HX, et al. Improved Cancer Detection Using Artificial Intelligence: a Retrospective Evaluation of Missed Cancers on Mammography. J Digit Imaging. 2019;32(4):625-637. doi: 10.1007/s10278-019-00192-5.
10. Dembrower K, Wåhlin E, Liu Y, et al. Effect of artificial intelligence-based triaging of breast cancer screening mammograms on cancer detection and radiologist workload: a retrospective simulation study. Lancet Digit Health. 2020;2(9):e468-e474. doi: 10.1016/S2589-7500(20)30185-0.
11. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid review: radiomics and breast cancer. Breast Cancer Res Treat. 2018;169(2):217-229. doi: 10.1007/s10549-018-4675-4.
12. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to Radiomics. J Nucl Med. 2020;61(4):488-495. doi: 10.2967/jnumed.118.222893.
13. WHO. World Health Organization classification of tumors 5th edition Breast Tumors. 2019:95-99.
14. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Adv Neural Inf Process Syst. 2012;25:1097-105.
15. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: p. 7263-7271.
16. Botchkarev A. Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. arXiv:180903006. 2018.
17. Warrens MJ. On the equivalence of Cohen's kappa and the Hubert-Arabie adjusted Rand index. Journal of classification. 2008;25(2):177-183. doi: 10.1007/s00357-008-9023-7
18. Pokorny T, Tesarik J. Microwave Stroke Detection and Classification Using Different Methods from MATLAB's Classification Learner Toolbox. 2019 European Microwave Conference in Central Europe (EuMCE): IEEE; 2019. p. 500-3.
19. Cui Y, Li Y, Xing D, Bai T, Dong J, Zhu J. Improving the Prediction of Benign or Malignant Breast Masses Using a Combination of Image Biomarkers and Clinical Parameters. Front Oncol. 2021;11:629321. doi: 10.3389/fonc.2021.629321. doi: 10.3389/fonc.2021.694094.
20. Salim M, Dembrower K, Eklund M, Lindholm P, Strand F. Range of Radiologist Performance in a Population-based Screening Cohort of 1 Million Digital Mammography Examinations. Radiology. 2020;297(1):33-39. doi: 10.1148/radiol.2020192212.
21. Pacilè S, Lopez J, Chone P, Bertinotti T, Grouin JM, Fillard P. Improving Breast Cancer Detection Accuracy of Mammography with the Concurrent Use of an Artificial Intelligence Tool. Radiol Artif Intell. 2020;2(6):e190208. doi: 10.1148/ryai.2020190208.
22. Kim HE, Kim HH, Han BK, et al. Changes in cancer detection and false-positive recall in mammography using artificial intelligence: a retrospective, multireader study. Lancet Digit Health. 2020;2(3):e138-e148. doi: 10.1016/S2589-7500(20)30003-0.
23. McKinney SM, Sieniek M, Godbole V, et al. International evaluation of an AI system for breast cancer screening. Nature. 2020;577(7788):89-94. doi: 10.1038/s41586-019-1799-6.