Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 12 NO: 1 PAGE: 57-65 YEAR: 2023
DOI:10.17798/bitlisfen.1190134

Deep Transfer Learning to Classify Mass and Calcification Pathologies from Screen Film Mammograms

Volkan Müjdat TİRYAKİ[1*]

[1]*School of Engineering, Department of Computer Engineering, C-107, Siirt University, Kezer Campus, 56100 Türkiye*
(ORCID: 0000-0003-1824-5260)

Keywords: Breast cancer, image classification, nodule, tumor, computer-aided diagnosis.

The number of breast cancer diagnoses is the largest among all cancers among women in the world. Breast cancer treatment is possible if it is diagnosed in the early stages. Mammography is a common imaging technique to detect breast cancer abnormalities. Breast cancer symptom screening is being performed by radiologists. In the last decade, deep learning was successfully applied to big image classification databases such as the ImageNet. In this study, the breast cancer pathology classification performances of the recent deep learning models were investigated by transfer learning and fine tuning. A total of 3,360 mammogram patches were used from the Digital Database for Screening Mammography (DDSM) and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) mammogram databases for deep learning model training, validating, and testing. Transfer learning and fine tuning were applied using Resnet50, Xception, NASNet, and EfficientNet-B7 network weights. The best classification performance was achieved by transfer learning from the Xception network. The computational costs of deep learning models were considered while selecting the best one. On the original CBIS-DDSM five-way test mammogram classification problem, the mean sensitivity, specificity, F1-score, and AUC were 0.7054, 0.9264, 0.7024, and 0.9317, respectively. The results show that the proposed models may be useful for the classification of breast cancer pathologies.

1. Introduction

Since the 1970s, researchers have been studying clinical decision support systems [1]–[3]. Breast cancer is the most frequently diagnosed cancer among women, and the incidence of the disease has been increasing in recent years [4]. If breast cancer is detected early, treatment is possible [5]. Mammography is obtained by exposing breasts to low-energy X-rays [6] in order to detect breast cancer abnormalities early [7].

The mammogram interpretation is a multi-step process that is still being performed by radiologists. Mammography screening is advised for women older than 50 years [8], [9]. In Türkiye, there were 11.15 M women at this age interval in 2021, which indicates the need for the number of breast cancer screenings per year [10]. An automated high-performance mammogram screening system could reduce the workload of radiologists and the number of unnecessary screenings and biopsies. The first computer-aided mammogram interpretation model was proposed by Ackerman and Gose in 1972 [11], wherein the researchers designed a feature extraction followed by a nearest neighbor breast lesion classification system. Automatic detection of a tumor tissue from mammography is related to texture analysis, and many different approaches have been investigated to date by new textural feature definitions and classifier models [12], such as the spherical wavelet transform [13] and geometric and textural feature extraction [14]. Computer-aided breast cancer research has also been conducted by other imaging modalities, such as microwave applications [15]. Despite these and other successful results in the literature, feature extraction–based machine learning methods can be time-consuming, particularly for medical image- or video-based

*Corresponding author: tiryakiv@siirt.edu.tr
Received: 17.10.2022, Accepted: 01.03.2023

analysis. Deep learning methods enable learning features from training data, and there are example studies in the literature related to stroke [16], carotid artery [17], [18], skin cancer [19], diabetes [20], Alzheimer's disease [21], and breast cancer [22]. Deep learning methods should be useful for the solution of problems related to other biomedical imaging modalities, so cross-disciplinary studies could possibly lead to new scientific advancements.

Regarding the applications of deep learning for breast cancer research, convolutional neural networks (CNN) were used for benign versus malignant breast mass classification [23]. In another study based on more than one million mammograms, the breast cancer screening performance of radiologists was shown to be improved when CNN-based diagnosis was used together with radiologists [24]. An end-to-end mammogram classifier based on a five-class mammogram patch classification like the one in the present study was investigated using Resnet50 and VGG16 [25]. Mass detection and classification have been performed by Momminet-V2 using multi-view mammograms [26]. The computational cost of deep learning model training is becoming a concern, and recently breast cancer mass pathology was classified by implementing pre-trained deep neural networks without transfer learning [27]. NASNet is a high performing CNN for image classification on the ImageNet [28], [29]. Recently, EfficientNets were proposed by considering the computational expense and the classification performance together [30]. Among EfficientNets, the EfficientNet-B7 is the best performing CNN model on the ImageNet.

In this study, transfer learning and fine tuning of recently proposed pre-trained deep neural networks, including Resnet50 [31], Xception [32], NASNetMobile [28], NASNetLarge [28], and EfficientNet-B7 [30] were investigated for the classification of normal, benign, and malignant masses, and benign and malignant calcification patches. These CNN models are among the best performing ones on ImageNet and the successful implementations of these networks for the breast cancer pathology classification problem are demonstrated in the present study. The Digital Database for Screening Mammography (DDSM) and the Curated Breast Imaging Subset of DDSM (CBIS-DDSM) datasets were used for deep learning model training [33]–[35]. This study is organized as follows: Section 2 includes materials and methods that describe the mammogram dataset, preprocessing, deep learning model training, and performance evaluation. Experimental results and discussion are presented in section 3, and conclusions and suggestions are given in section 4.

2. Material and Method

Significant amounts of data are necessary for deep learning model training. This study was conducted using publicly available datasets.

2.1. Dataset and preprocessing

The CBIS-DDSM and DDSM mammograms were selected as data sources since these are large and publicly available mammogram databases. The CBIS-DDSM database includes mammogram patches with the abnormality and pathology information in the form of mass versus calcification and benign versus malignant. The mammogram patches were resampled to 331×331 and saved as 8-bit gray level images. The normal tissue patches were randomly selected from the DDSM dataset craniocaudal (CC) views of mammogram regions which does not include any abnormalities [36], [37]. Breast cancer abnormalities have a great variety of shape, texture, and intensity, and two representative training patches from each class are given in Figure 1.

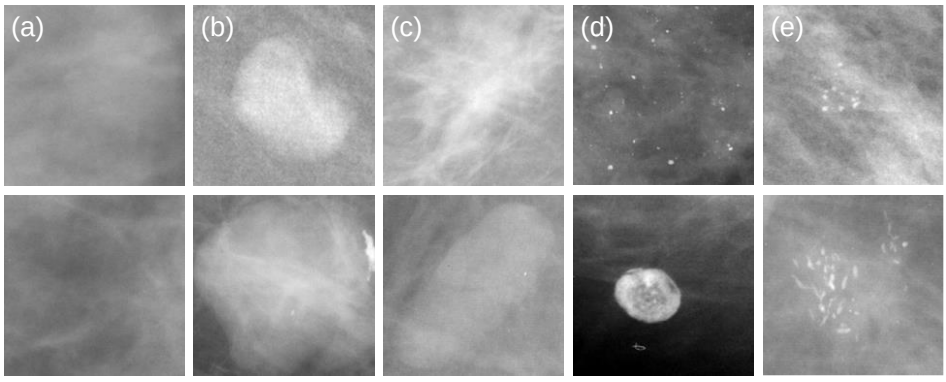

Figure 1. Two representative training patches belong to normal (a), benign mass (b), malignant mass (c), benign calcification (d), and malignant calcification (e).

Mammogram patches were classified into normal, benign mass, malignant mass, benign calcification, and malignant calcification. Patches that have a "benign without callback" annotation were considered "benign." Normal patches that do not include any breast cancer abnormalities were selected from the DDSM dataset because the CBIS-DDSM includes mammograms with abnormalities. Normal patches were extracted from the middle of the mammogram so that they do not include background pixels. Test mammograms defined in the CBIS-DDSM database were used as test data in the present study, and none of them were used for training or validating. The dataset was balanced by deleting excess files. Randomly selected 20% of the CBIS-DDSM training images were used as validation and rest of the mammograms were used for training. The percentages of the number of training, validation, and test patches were 65%, 16%, and 19%, respectively. The number of training, validation, and test patches is given in Table 1.

Table 1. The number of training, validating, and test patches in the dataset.

Folder	Normal	Benign mass	Malignant mass	Benign calcification	Malignant calcification	Total
Train	434	434	434	434	434	2170
Validation	109	109	109	109	109	545
Test	129	129	129	129	129	645
Total	672	672	672	672	672	3360

2.2. Hardware configuration

Deep learning models were trained on a workstation computer that has an NVIDIA RTX 3060 GPU, double Xeon E5-2630 2.6 GHz CPU, and a 16 GB RAM.

2.3. Deep transfer learning and fine tuning

Deep transfer learning classification models were implemented since they are known to reduce the training time, reduce the requirements for the amount of training data, and improve the performance [25], [38]. Data augmentation was applied to the patches by setting the width shift, height shift, zoom, and shear ranges to 0.2, rotation range to 90°, and enabling the vertical and horizontal random flips. The brightness range was not used since intensity is a feature that is used by radiologists for mass and calcification detection. Deep learning models were implemented by importing the weights of Resnet50, Xception, NASNetMobile, NASNetLarge, and EfficientNet-B7 [28] networks, which have 138 M, 23 M, 5 M, 88,9 M, and 66 M parameters, respectively. The models were named as backbone followed by TL (transfer learning). Adam optimizer was used [39]. Transfer learning and fine tuning were implemented by using the three-step approach described in Shen *et al.* [25]. First, all of the layers except the final layer were frozen, and the model was trained for three epochs with a 10^{-3} learning rate. Second, 33% of layers from the end were unfrozen and the model was trained for ten epochs with a 10^{-4} learning rate. Third, all layers were unfrozen and trained with a learning rate of 10^{-5} until one of the stopping criteria was met. Early stopping criteria were: (1) the maximum number of epochs was 150, and (2) the maximum number of consecutive epochs when the loss function did not decrease was ten. (3) The minimum learning rate was 10^{-8}. The transfer learning and fine tuning codes are accessible at https://github.com/tiryakiv/breast-cancer-pathology-classification.

2.4. Performance evaluation

Patch classification accuracy, sensitivity, specificity, and F-measure were defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} \quad (3)$$

$$F-measure = \frac{2TP}{2TP+FP+FN} \quad (4)$$

where TP, TN, FP, and FN are abbreviations for true positive, true negative, false positive, and false negative, respectively. Sensitivity, specificity, and F-measure of each class were calculated. The area under the receiver operating curve (AUC) was analyzed, and a confusion matrix was constructed to evaluate the classification performance [40]. Accuracy, specificity, sensitivity, and F-measure show the classification performance, but they depend on the threshold level. Sensitivity shows the ratio of

correctly selected relevant items, and specificity shows the ratio of negatively selected elements to the true negative elements. AUC shows the overall performance of the binary classifier by considering all possible thresholds. In the present study, a one-versus-rest classifier was implemented for AUC calculations.

3. Results and Discussion

The classification results of normal, benign, and malignant masses, and benign and malignant calcification patches using ResNet50, Xception, NASNetMobile, NASNetLarge, and EfficientNet-B7 transfer learning methods on the validation data are shown in Table 2.

Table 2. The breast cancer pathology classification results on the validation data.

TL backbone	Accuracy	AUC	Train time
Resnet50	0.7321	0.9392	37 min
Xception	0.7193	**0.9407**	**29 min**
NASNetMobile	0.6807	0.9183	47 min
NASNetLarge	**0.7376**	0.9357	114 min
EfficientNet-B7	0.7156	**0.9407**	259 min

The Xception and EfficientNet-B7 TL models had the highest AUC, and the NASNetLarge TL model had the highest accuracy on the validation data. To compare the performances of all models, the confusion matrices on the validation data are given in Table 3.

Table 3. Breast cancer pathology classification confusion matrix on validation data. BC, MC, BM, MM, and N are abbreviations for benign calcification, malignant calcification, benign mass, malignant mass, and normal patches, respectively. TL: transfer learning. The highest *TP* for each class is given in bold.

Radiologist	Resnet50 TL predictions					Xception TL predictions					NASNetMobile TL predictions				
	BC	MC	BM	MM	N	BC	MC	BM	MM	N	BC	MC	BM	MM	N
BC	77	29	1	2	0	77	22	2	7	1	58	24	5	15	7
MC	14	74	0	21	0	17	68	6	18	0	4	65	10	28	2
BM	3	2	48	56	0	3	1	52	52	1	1	3	41	56	8
MM	2	4	7	96	0	3	3	13	90	0	1	1	8	**99**	0
N	0	1	2	2	104	2	0	1	1	105	0	0	0	1	**108**

Radiologist	NASNetLarge TL predictions					EfficientNet-B7 TL predictions				
	BC	MC	BM	MM	N	BC	MC	BM	MM	N
BC	**79**	22	2	4	2	65	39	1	3	1
MC	22	62	4	20	1	5	**85**	2	16	1
BM	2	0	**58**	46	3	0	2	48	56	3
MM	0	0	10	**99**	0	2	7	14	86	0
N	2	0	0	3	104	0	0	0	3	106

The confusion matrices in Table 3 showed that all models performed well in discriminating the normal patches. The highest misclassification rates were observed for the benign mass patches. *TP*s of the NASNet TL malignant mass were equal. EfficientNet-B7 TL model had the highest *TP* for malignant calcification patches. Based on overall accuracy and confusion matrices, the best breast pathology classification model was selected as NASNetLarge TL because it has the highest accuracy and the three *TP*s of the model were the highest among all of the models.

3.1. Test results

The performance of the NASNetLarge TL model trained by the breast cancer pathology patches was evaluated. The AUC and accuracy on the test data were 0.9404 and 0.7318 respectively. The performance difference between the validation and test data was close. To analyze the classification errors, the NASNetLarge TL model confusion matrix on the test data is given in Table 4.

Table 4. Breast cancer pathology classification confusion matrix on the test data

Radiologist	BC	76	33	8	10	2
	MC	35	84	0	9	1
	BM	6	2	88	29	4
	MM	4	3	26	95	1
	N	0	0	0	0	129
		BC	MC	BM	MM	N
		NASNetLarge TL predictions				

The confusion matrix in Table 4 showed that the malignant calcification patch misclassification probability as benign calcification was the highest. The misclassification probability of benign calcifications as malignant calcifications and benign masses as malignant masses was high. These results showed that the model has better discrimination for abnormality types than the pathology type. The patch classification performance of mass was better than calcification. The NASNetLarge TL model's sensitivity, specificity, and F-measure for each class are listed in Table 5.

The test results in Table 5 showed that the highest and lowest sensitivity were obtained on normal and benign calcification patches, respectively. Among the abnormalities, including patches, the highest specificity was obtained for malignant masses and the lowest specificity was obtained for benign calcification. The multi-class receiver operating characteristic (ROC) analysis of the NASNetLarge TL model is given in Figure 2.

Table 5. NASNetLarge TL model classification performance on the test patches. The best result for each metric is shown in bold.

Class	Sensitivity	Specificity	F-measure
Benign calcification	0.6281	0.8989	0.6080
Malign calcification	0.6885	0.9140	0.6693
Benign mass	0.7213	0.9216	0.7012
Malign mass	0.6643	0.9323	0.6985
Normal	**0.9416**	**1.0000**	**0.9699**

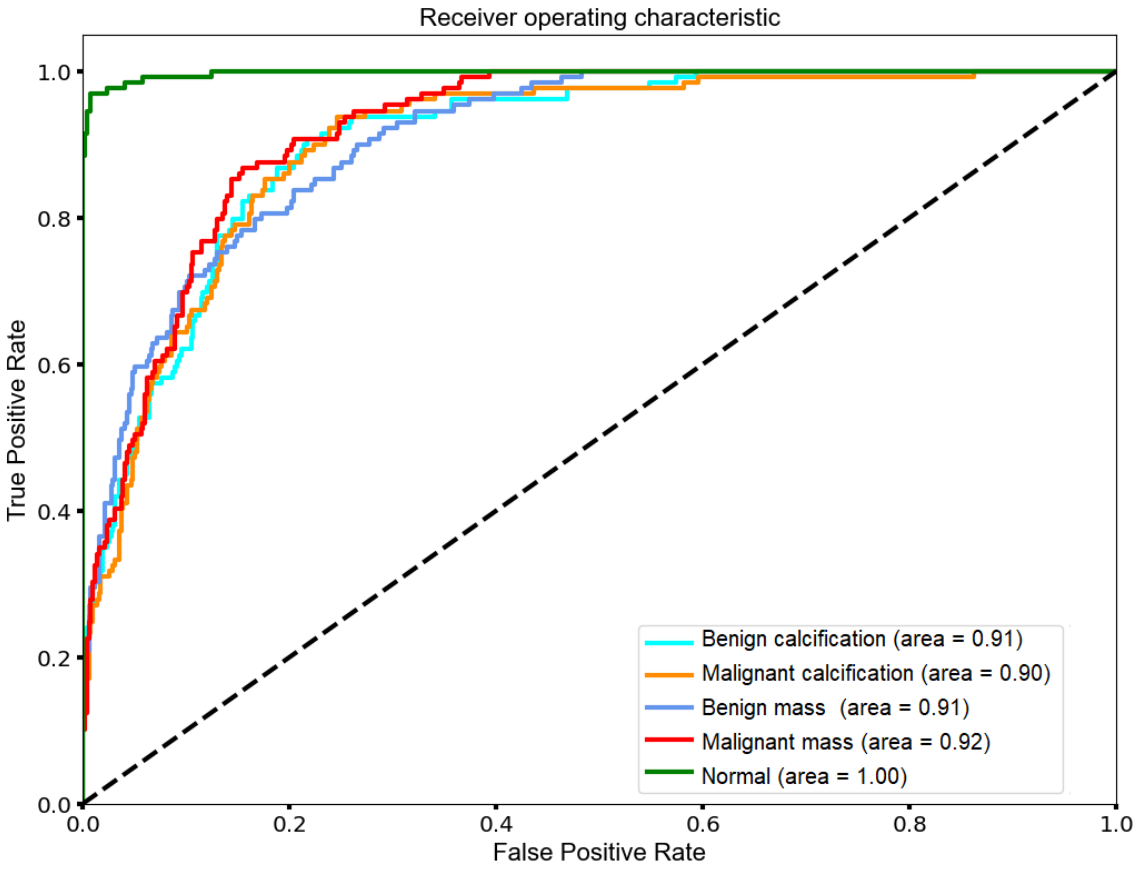

Figure 2. ROC analysis of a five-way breast cancer pathology classification. (One versus rest classifier)

Figure 2 showed that the mass abnormality classification performed better than the calcification abnormality classification. The performance of discriminating the malignant calcification patches was lower than benign calcification, on the other hand, the performance of discriminating the malignant mass was higher than benign mass. As expected, AUC of normal patches was better than the abnormality including patches. Computer-aided breast cancer diagnosis research involves a number of types of problems. Some studies in the literature focus on mass detection and classification [41], [42], and others focus on calcification detection and classification [43], [44]. A five-way classification of breast cancer abnormalities and the malignant behavior of tissues from mammograms were investigated in the present study. The closest work in the literature to the present one is Shen *et al.* [25], wherein the researchers used 2,478 CBIS-DDSM patches and achieved an accuracy of 0.99 on the test data when they applied the ResNet50 transfer learning method. The performance difference with the present study may be because of: 1) the difference in the number of patches (3,360 patches were used in the present study); 2) training, validating, and test data included DDSM normal mammograms in the present study; and 3) original CBIS-DDSM test images were used in the present study. The other study that is related to the present one is Hekal *et al.*, where the researchers investigated the four-way classification performance of benign calcification, malignant calcification, benign mass, and malignant mass [45]. They achieved an accuracy of 0.91 using AlexNet by training with 2800 CBIS-DDSM mammograms. The relatively higher accuracy than the present study can be explained by the lower number of classes (n=4) and the performance evaluation on the validation data. The five-way classification of breast cancer pathology patches similar to the present work was investigated by Chun-ming *et al.* [46]. They achieved an accuracy of 0.91 using the Deep Cooperation CNN model on the CBIS-DDSM dataset. Their model performance is higher than the present study, possibly because they used a smaller test dataset (%10) and their dataset was randomly distributed. In the present study, the classification performance of state-of-the-art transfer learning methods was demonstrated on the original CBIS-DDSM test dataset.

4. Conclusion and Suggestions

Transfer learning and fine tuning pre-trained deep neural networks have great potential for the classification of breast cancer abnormalities and pathologies. Pre-trained deep neural networks reduce the computational cost and yield high performance when they are used for different classification domains. In this study, successful transfer learning applications have been shown using the recent deep neural networks trained with the ImageNet. By applying transfer learning and fine tuning, the useful features learned from the ImageNet database were adopted for the current breast pathology classification problem. The best classification performance has been obtained by using the NASNetLarge network, which resulted in a 0.9404 AUC and 0.7318 accuracy. The NASNetLarge network can be used for the breast cancer pathology classification at the reported performance. The proposed model may be useful for detecting breast cancer abnormalities and classifying them as malignant versus benign.

The performance ranking of CNNs on the ImageNet was not the same for the breast cancer pathology classification models. This could be due to the contextual difference between the ImageNet images and the mammograms. The sensitivity of benign mass and benign calcification showed that further performance improvement can be investigated. Novel machine learning techniques and deep ensemble models will be investigated in the future to increase the system performance. The CBIS-DDSM and other databases have lesions that sometimes have mass and calcification abnormalities overlap. The classification of such tumors with double abnormalities will be investigated in the future.

Acknowledgments

The author thanks the Siirt University Scientific Research Projects Directorate for providing the NVIDIA GeForce RTX 3060 GPU under Grant No. 2021-SİÜMÜH-01. The author thanks the owners of the DDSM and CBIS-DDSM for publicly sharing the datasets for research.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] H. L. Bleich, "The computer as a consultant," *N. Engl. J. Med.*, vol. 284, no. 3, pp. 141–7, 1971.

[2] M. E. Cohen and D. Le Hudson, "A hybrid system for diagnosis involving biosignals," *Proc. 2005 IEEE Int. Conf. Comput. Intell. Meas. Syst. Appl. CIMSA 2005*, vol. 2005, no. July, pp. 312–315, 2005.

[3] I. Gökbay, S. Karaman, S. Yarman, and B. Yarman, "An intelligent decision support tool for early diagnosis of functional pituitary adenomas," *TWMS J. Appl. Eng. Math.*, vol. 5, no. 2, p. 169, 2015.

[4] V. Ozmen, "Breast Cancer in Turkey: Clinical and Histopathological Characteristics (Analysis of 13.240 Patients)," *J. Breast Heal.*, vol. 10, no. 2, pp. 98–105, 2014.

[5] S. W. Duffy *et al.*, "Mammography screening reduces rates of advanced and fatal breast cancers: Results in 549,091 women," *Cancer*, vol. 126, no. 13, pp. 2971–2979, 2020.

[6] P. Xi, C. Shu, and R. Goubran, "Abnormality Detection in Mammography using Deep Convolutional Neural Networks," in *2018 IEEE International Symposium on Medical Measurements and Applications, Proceedings*, 2018, pp. 1–6.

[7] J. Tang, R. M. Rangayyan, J. Xu, I. E. El Naqa, and Y. Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 2, pp. 236–251, 2009.

[8] "What Is Breast Cancer Screening?," *Center for disease control and prevention*, 2022. [Online]. Available: https://www.cdc.gov/cancer/breast/basic_info/screening.htm#:~:text=Breast Cancer Screening Recommendations&text=The USPSTF recommends that women,often to get a mammogram.

[9] L. Tabar *et al.*, "Efficacy of breast cancer screening by age. New results swedish two-county trial," *Cancer*, vol. 75, no. 10, pp. 2507–2517, 1995.

[10] TÜİK, "İstatistiklerle Kadın, 2021," *TÜİK veri portalı*, 2022. [Online]. Available: https://data.tuik.gov.tr/Bulten/Index?p=Istatistiklerle-Kadin-2021-45635. [Accessed: 06-Dec-2022].

[11] L. Ackerman and E. Gose, "Breast lesion classification by computer and xeroradiograph," *Cancer*, vol. 30, no. 4, pp. 1025–1035, 1972.

[12] Y. Kaya, "A new intelligent classifier for breast cancer diagnosis based on a rough set and extreme learning machine: Rs + elm," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 21, no. SUPPL. 1, pp. 2079–2091, 2013.

[13] P. Görgel, A. Sertbas, and O. N. Uçan, "Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines," *Expert Syst.*, vol. 32, no. 1, pp. 155–164, 2015.

[14] H. Singh, V. Sharma, and D. Singh, "Machine learning based computer aided diagnosis system for classification of breast masses in mammograms," *J. Phys. Conf. Ser.*, vol. 2267, no. 1, 2022.

[15] E. Onemli *et al.*, "Classification of rat mammary carcinoma with large scale in vivo microwave measurements," *Sci. Rep.*, vol. 12, no. 1, pp. 1–11, 2022.

[16] R. A. J. Alhatemi and S. Savaş, "Journal of Computer Science," *J. Comput. Sci.*, vol. IDAP-2022, pp. 192–201, 2022.

[17] S. Savaş, N. Topaloğlu, Ö. Kazcı, and P. N. Koşar, "Comparison of Deep Learning Models in Carotid Artery Intima-Media Thickness Ultrasound Images : CAIMTUSNet," *Bilişim Teknol. Derg.*, vol. 15, no. 1, pp. 1–12, 2022.

[18] S. Savaş, N. Topaloğlu, Ö. Kazcı, and P. N. Koşar, "Performance Comparison of Carotid Artery Intima Media Thickness Classification by Deep Learning Methods," in *International Congress on Human-Computer Interaction, Optimization and Robotic Applications*, 2019, pp. 125–131.

[19] D. K. A. Al-saedi and S. Savaş, "Journal of Computer Science," *J. Comput. Sci.*, vol. IDAP-2022, pp. 202–210, 2022.

[20] S. Buyrukoglu and A. Akbas, "Machine Learning based Early Prediction of Type 2 Diabetes : A New Hybrid Feature Selection Approach using Correlation Matrix with Heatmap and SFS," *Balk. J. Electr. Comput. Eng.*, vol. 10, no. 2, pp. 110–117, 2022.

[21] S. Buyrukoğlu, "Improvement of Machine Learning Models' Performances based on Ensemble Learning for the detection of Alzheimer Disease," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 2021, pp. 102–106.

[22] A. Hamidinekoo, E. Denton, and R. Zwiggelaar, "Automated Mammogram Analysis with a Deep

Learning Pipeline," pp. 1–5, 2019.
[23] A. Jain and D. Levy, "Breast Mass Classification from Mammograms using Deep Convolutional Neural Networks," *arXiv Prepr. arXiv1612.00542*, 2016.
[24] N. Wu *et al.*, "Deep Neural Networks Improve Radiologists' Performance in Breast Cancer Screening," *IEEE Trans. Med. Imaging*, vol. 39, no. 4, pp. 1184–1194, 2020.
[25] L. Shen, L. R. Margoiles, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to improve Breast cancer Detection on Screening Mammography," *Sci. Rep.*, vol. 9, no. 12495, pp. 1–12, 2019.
[26] Z. Yang *et al.*, "MommiNet-v2: Mammographic multi-view mass identification networks," *Med. Image Anal.*, vol. 73, p. 102204, 2021.
[27] E. Al-Mansour, M. Hussain, H. A. Aboalsamh, and Fazal-e-Amin, "An Efficient Method for Breast Mass Classification Using Pre-Trained Deep Convolutional Networks," *Mathematics*, vol. 10, no. 14, pp. 1–19, 2022.
[28] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning Transferable Architectures for Scalable Image Recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 8697–8710, 2018.
[29] Jia Deng, Wei Dong, R. Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
[30] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," *36th Int. Conf. Mach. Learn. ICML 2019*, vol. 2019-June, pp. 10691–10700, 2019.
[31] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-Decem, pp. 770–778.
[32] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 1800–1807, 2017.
[33] R. S. Lee, F. Gimenez, A. Hoogi, and D. Rubin, "Curated Breast Imaging Subset of DDSM [Dataset]," 2016. [Online]. Available: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM#22516629cf2ec23796854d91bc86c4ae2e499baa.
[34] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, pp. 1–9, 2017.
[35] K. Clark *et al.*, "The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.," *J. Digit. Imaging*, vol. 26, no. 6, pp. 1045–1057, Dec. 2013.
[36] M. Heath *et al.*, "Current status of the Digital Database for Screening Mammography," in *Proceedings of the Fourth International Workshop on Digital Mammography*, 1998, pp. 457–460.
[37] M. Heath, K. Bowyer, D. Kopans, R. Moore, and W. P. Kegelmeyer, "The Digital Database for Screening Mammography," in *Proceedings of the Fifth International Workshop on Digital Mammography*, 2001, pp. 212–218.
[38] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, 2015.
[39] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.
[40] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.
[41] D. Abdelhafiz, J. Bi, R. Ammar, C. Yang, and S. Nabavi, "Convolutional neural network for automated mass segmentation in mammography," *BMC Bioinformatics*, vol. 21, no. Dl, pp. 1–16, 2020.
[42] Z. Assari, A. Mahloojifar, and N. Ahmadinejad, "Discrimination of benign and malignant solid breast masses using deep residual learning-based bimodal computer-aided diagnosis system," *Biomed. Signal Process. Control*, vol. 73, no. March 2021, p. 103453, 2022.
[43] R. Zamir *et al.*, "Segmenting microcalcifications in mammograms and its applications," *https://doi.org/10.1117/12.2580398*, vol. 11596, pp. 788–795, Feb. 2021.
[44] R. Hou *et al.*, "Anomaly Detection of Calcifications in Mammography Based on 11,000 Negative Cases," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 5, pp. 1639–1650, 2022.

[45] A. A. Hekal, A. Elnakib, and H. E. Moustafa, "Automated early breast cancer detection and classification system," *Signal, Image Video Process.*, vol. 15, no. 7, pp. 1497–1505, 2021.

[46] T. Chun-ming, C. U. I. Xiao-mei, Y. U. Xiang, and Y. Fan, "Five Classifications of Mammography Images Based on Deep Cooperation Convolutional Neural Network," *Am. Sci. Res. J. Eng. Technol. Sci.*, vol. 57, no. 1, pp. 10–21, 2019.