Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
() : –
© TÜBİTAK
doi:2303-81

Automatic breast cancer detection from ultrasound images using Dirichlet distribution based deep ensemble transfer learning

Osman GÜLER*
Tusaş Şehit Hakan Gülşen Vocational and Technical Anatolian High School, Ankara, Turkiye

Received: 19.03.2023 • **Accepted/Published Online:** 09.08.2023 • **Final Version:** 2023

Abstract: Breast cancer is a common type of cancer among women. Breast ultrasound is a useful and rapid diagnostic tool for early detection of breast cancer. Thanks to expert radiologists and clinicians, suspicious situations can be detected and further examinations can be requested. Computer-aided decision systems used to assist specialist radiologists and clinicians have more reliable and faster results. Deep learning has become a valuable tool in breast cancer detection and diagnosis, and it is being increasingly used in medical imaging analysis. Therefore, in this study, a Dirichlet distribution based deep ensemble learning model using pretrained transfer learning models is proposed for breast cancer classification from ultrasound images. In the study, experiments were carried out with the Breast Ultrasound Images Dataset (Dataset BUSI). First, image preprocessing was performed on the dataset, and then the dataset was balanced with data augmentation. DenseNet121, InceptionV3, MobileNet, and ResNet50 models were used for transfer learning. With these models, repetitive trainings using the weights of the previous training as input were made. The obtained results were fit with Dirichlet distribution based ensemble learning with different variations. A total of 98.17% validation accuracy was achieved on the proposed model dataset. The results show that the proposed model is useful in breast cancer classification.

Key words: Breast cancer, breast ultrasound images, deep learning, transfer learning, ensemble learning

1. Introduction

One of the most common cancer types among women is breast cancer (BC) according to the World Health Organization and International Agency for Research on Cancer data [1-2]. As can be seen from Figure 1, the rate of BC among cancer patients is 11.7%, while the incidence of BC among women is 24.5%.

Early diagnosis and treatment process in cancer is an important step for the recovery of the patient. After late diagnosis, in the later stages of the disease, the patient cannot respond to treatment and the disease may result in death. For this reason, it is important that people go to check-ups at certain intervals. Considering the incidence of BC among women, it is very important for women to go for regular check-ups every year in terms of early diagnosis and treatment process. In the early diagnosis process of BC, firstly, manually or in a doctor's examination, manual control procedures are performed. As soon as the mass is noticed, various imaging techniques, such as ultrasound, magnetic resonance imaging (MRI), mammography, computerized tomography(CT), and positron emission tomography(PET), are used for medical diagnosis [3]. The images and analyzes obtained are ultimately examined by specialist doctors and a definitive diagnosis is made. As a result of these assays and images, the cells examined are divided into two classes as normal cells and tumor cells, while

*Correspondence: h.osmanguler@gmail.com

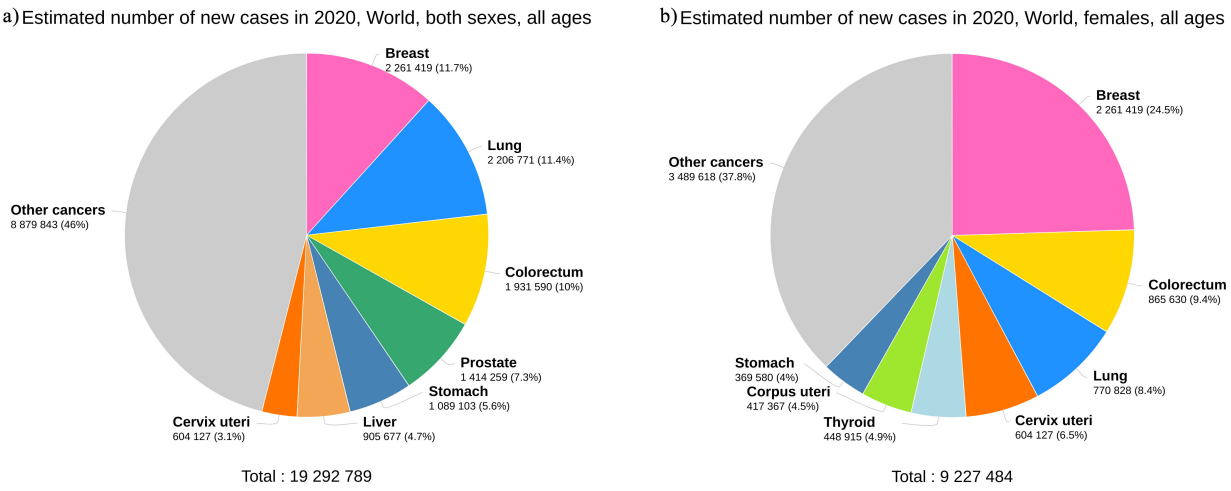

Figure 1. a) Cancer cases number both sexes, b) cancer cases number only females [2].

tumor cells are divided into two classes as benign or malignant [4]. Here, benign refers to the nondangerous tumor cell, and malignant refers to the dangerous tumor cell.

New systems, in which artificial intelligence, machine learning (ML) and deep learning techniques are used, are recommended by many researchers in recent years on disease diagnosis and diagnosis in CAD systems, since they have shown high performance and performance. Deep learning is an important technology that can be used for early diagnosis in the field of medical imaging and recognition, as it enables faster and more accurate analysis of complex datasets with deep learning. In general, using ML and deep learning techniques in many areas such as cancer and its types [9-11], Alzheimer's detection [12,13], diabetes [14,15], COVID-19 detection [16], carotid artery [17,18] CAD systems have been developed for medical diagnosis, diagnosis and classification problems.

Although machine learning and deep learning methods are widely used in classification problems, in some cases there is not enough data. In this case, the learning process is negatively affected and the generalization ability of the model is weakened, and excessive learning may occur. The hyperparameter settings and optimization methods used may also be insufficient. In cases where the amount of data is limited, data augmentation methods or previously trained models can be used to increase the effect of learning [19]. In this way, using the information learned by a deep learning model in solving a problem to solve another problem is called transfer learning. With transfer learning, time and data costs are reduced by using a previously trained model to solve a new problem.

The first step in the transfer learning process is the selection of the pretrained model [20]. In the selection of the pretrained model, a model that has been successful on the similar dataset or on the general datasets is selected. After the model is selected, the pretrained model is adapted for problem solving. In this step, the last classification layer of the model needs to be retrained, but any layer of the new model can be fine-tuned if desired [21]. Finally, trainings are carried out on the dataset. Thanks to the use of transfer learning with pretrained models, the training time is shortened and generalization errors are also reduced [22]. Because it increases generalization ability, a pretrained model learns common patterns between different datasets and provides better performance in a new task. It achieves successful results on fewer datasets and reduces the

risk of overfitting. In transfer learning, deepening and abstraction techniques are used to better transfer the extracted features in different data sets and to achieve high performance. Deepening is done by sending features from the middle or upper layers of a pretrained model to more hidden layers. The first layers represent lower-level features, while the last layers represent higher-level features. In this way, the model can learn higher-level features. Abstraction is done so that the extracted features represent more general patterns or attributes. This increases the generalization of the features learned by the model and its usability in other tasks.

In this study, we proposed three approaches using Dirichlet distribution based ensemble learning including pretrained architectures to classify the BC from ultrasound images. The main contributions of this study are:

- Reproducing images with data augmentation, balancing the dataset and applying image preprocessing techniques,
- Implementation of fine-tuned transfer learning methodology on BUSI dataset with pretrained architectures,
- Selection of successful models with 3 different approaches of community learning structure based on Dirichlet distribution,
- Thanks to 3 different approaches and combinations, more reliable and more stable results are obtained,
- The data and test results of the study support deep learning-based CAD systems.

The structure of the paper continues as follows: The literature review for the research subject is stated in Section 2. In Section 3, the dataset, materials and suggested approaches used in the study are explained. In Section 4, the experimental results of the tests performed with the proposed approaches are shared. In Section 5, the results of the research are discussed and the study is concluded.

2. Releated works

Breast ultrasound images are a useful diagnostic tool for early detection and diagnosis of BC. Deep learning techniques are widely used to help radiologists and clinicians analyze breast ultrasound images and assist in decision making. Different CAD systems have been proposed by many researchers on BC classification from ultrasound images with deep learning methods and techniques. For this reason, in this study, previous studies with ML and deep learning techniques for BC detection from ultrasound images were examined and the results are summarized below.

In [3], transfer learning and deep convolutional neural networks (DCNN) were used to classify thyroid and breast lesions from ultrasound images. TNet architectures for thyroid cancer and BNet architectures for Breast cancer were created. A dataset consisting of 719 thyroid and 672 breast images taken between 2016 and 2018 was used for classification. Average accuracy rate was 86.5 % for TNet and 89% for BNet. To compare the results, the performance of the radiologists was compared and it was seen that the proposed model achieved a higher level of accuracy than the radiologists. In [6], a model consisting of a combination of unsupervised learning algorithm (fuzzy c-means clustering) and supervised back-propagation artificial neural network(ANN) learning algorithm is proposed. In the study, experiments were performed on a total of 178 breast ultrasound images, including 88 benign and 90 malignant. A total of 457 features were extracted from the ultrasound images and the most important features were selected. An accuracy rate of 95.86% was obtained in experiments using 457 features, and an accuracy rate of 94.13% in experiments using 19 features. In [7], classification was made using B-mode ultrasound (B-mode) and strain elastography ultrasound (SE) images obtained from 85 patients, 42 of whom had benign lesions and 43 of whom were malignant. Firstly, trainings were made with AlexNet and ResNet models. Then, tests were conducted using the ensemble learning method with these two

models. The ensemble learning model showed a better result than the AlexNet and RexNet models and achieved 90.00% accuracy. In [23], a recurrent wavelet Elman neural network (RWENN) was used to classify BC images in the Wisconsin BC dataset. The RWENN classifier achieved an accuracy rate of 97.78% for the ten-fold cross-validation test. Although the proposed RWENN model has more computational costs, the increased complexity is acceptable. The simulation results were compared with the studies performed on the Wisconsin BC dataset and it was stated that the proposed RWENN model performed better than other methods. In [24], a specially created convolutional neural network (CNN) for classification of breast lesions and transfer learning methods for comparison were used on a dataset consisting of a total of 641 cases, 413 benign and 228 malignant lesions. Five-fold cross-validation was used in the study. The result was an overall accuracy of 85.98% and area under the curve (AUC) equal to 0.94 for classification, with the accuracy and AUC increasing to 92.05% and 0.97, respectively, after image augmentation and L2 regularization were applied. With transfer learning, 87.07% accuracy and 0.96 AUC were obtained. In [25], breast cancer classification was studied on a dataset consisting of 1328 images, 707 benign and 621 malignant, by combining four different datasets. For image preprocessing, segmentation was performed using fuzzy enhancement and two-sided filtering algorithms. Then, the fine-tuned VGG11, VGG16, VGG19, ResNet101, DensNet121 and DensNet161 transfer learning models were used for classification. As a result, it was shown that the accuracy, precision, specificity, F1 score, and area under the curve of the proposed method were 95.48%, 98.11%, 98.33%, 95.71%, and 98.83, respectively. In [26], a 4-stage system is proposed, namely preprocessing, segmentation, feature extraction, and classification. Preprocessing was done using speckle reduction anisotropic diffusion, active contour-based segmentation was used and images were classified with three classifiers: K-nearest neighbors (KNN) algorithm, decision tree algorithm and random forest classifier. KNN algorithm achived 83% accuracy, decision tree algorithm achieved 85% accuracy and random forest classifier outperformed the other classifiser and achieved 88% accuracy for classification of breast ultrasound images. In [27], different breast ultrasound images datasets were combined and classified with an ensemble structure using VGGNet, ResNet, and DenseNet models. The created dataset consists of a total of 1687 images, 953 of which are benign and 734 are malignant. As a result of the study, the accuracy, sensitivity, specificity, precision, F1 score, and AUC values of the proposed method were 91.10%, 85.14%, 95.77%, 94.03%, 89.36%, and 0.97, respectively. In experiments performed on BUSI, which is an open dataset, with the proposed model, the accuracy, sensitivity, specificity, precision, F1 score, and AUC values are 94.62%, 92.31%, 95.60%, 90.00%, 91.14%, and 0.9711, respectively. In [28], an approach that combines feature selection and parameter tuning simultaneously has been proposed to improve classification accuracy and reduce feature extraction time for breast cancer detection. In this study images were automatically segmented using the level set method, and the tumor was classified as benign or malignant using a genetic algorithm and a support vector machine (SVM). The accuracy of the proposed system is 95.24% for classification of breast tumours. It is stated that the feature calculation time of the proposed system is 8% of a system without feature selection. In [29], ensemble CNN is proposed by using 3 different stacked feature extractors and Gaussian dropout for classification problem of breast tumours as benign, malignant, and normal. The proposed ensemble model achieved 92.15%, 92.21%, 92.26%, and 92.17% accuracy, F1 score, precision, and recall rates, respectively. In [30], a model consisting of three recurrent neural networks (RNN) constructs trained as classifiers with ResNet-18 and spatial attention mechanism is proposed named as SAFNet. First, the image features were extracted by fine-tuning the ResNet-18 model. Next, three RNNs are trained using extreme learning machine (ELM), random vector functional link network (RVFL) and Schmidt neural network (SNN). Finally, the predictions of the proposed model were

calculated by fusion of the predictions from these three RNNs based on majority voting. The proposed model achieved 94.10%, 94.93%, 98.14%, and 96.50% accuracy, sensivity, precision, and F1 score with 5-fold cross-validation, respectively. In [31], a Visual Geometry Group Attention - Vision Transformer (VGGA-ViT) network consisting of a VGG attention module and a frequent and alert block was proposed for the spine. Local features were extracted with CNN, global relations between different regions were determined with ViT and relevant features were extracted. In the study, tests were carried out with the proposed model and transfer learning models on two different breast cancer datasets, one consisting of 974 and the other 163 samples. The proposed model was more successful than other models, achieving 88.71% accuracy on the first dataset and 81.72% accuracy on the second dataset. In [32], a method on singleton filter banks, local configuration pattern features and classification is proposed for the detection of breast lesions on ultrasound images. In this study, a dataset consisting of 448 nodule images, 147 normal, 210 benign and 91 malignant, was used. In the proposed system, LM, Schmid, MR8 and MR4 singleton filter banks were applied on the images. Then, the features of the images were extracted using the native binary pattern. Finally, using the LSDA feature reduction approach, the number of features for each filter bank was reduced to 30 and classification was performed. The proposed method obtained accuracy, sensitivity, specificity, and pedictive value of 96.10%, 96.50%, 96.50%, 95.30%, and 97.90%, respectively. In [33], the temporal sequence dual-branch network model is proposed to classify breast tumor using B-mode ultrasound and contrast-enhanced ultrasound (CEUS) video. A Temporal Sequence Regression Mechanism (TSRM) was designed to extract CEUS video features by using Gram matrix and a Shuffle Temporal Sequence Mechanism (STSM) was designed to shuffle temporal sequences. With ResNext18 as backbone network, image features were extracted from B-mode ultrasound. The developed method obtained 90.20%, 91.40%, 95.20%, and 93.20% accuracy, recall, precision, and F1 score values, respectively. In [34], a new residual DCNN model is proposed based on the idea of larger residual blocks, benefit assessment of tumor localization, and fine-tuning of optimizer settings. In the study, a CNN model designed from 5 residual blocks is proposed. In this study, Adam, momentum stochastic gradient descent (SGD), Nesterov momentum SGD and AdaGrad optimization algorithms were used with 3 different learning rates (0.01, 0.001, and 0.0001). With the momentum SGD (learning rate: 0.0001) based optimizer, accuracy, sensitivity, specificity, and precision values were obtained as 96.24%, 97.37%, 93.95%, 95.30%, and 97.14%, respectively. In [38], two datasets were combined and the number of data was increased with data augmentation. The combined dataset, consisting of a total of 943 images, was increased to 9430 images with 10-fold data augmentation. A CNN model was proposed in the study. At the same time, trainings were made with VGG19 and Yolov3 models for comparison. With the proposed CNN model, fine-tuned training for multiple iterations extracted features from ultrasound images. The results show that the proposed method achieved an accuracy, sensitivity, and specificity of 96.31%, 92.63% and 96.71%, respectively. In [36], a shallow custom CNN was proposed and BC classification was performed from ultrasound images with eight different fine-tuned pretrained models and the results were compared. In the study, trainings were made using the five-fold cross-validation technique and the Grad-CAM heat map visualization technique was used. The ResNet50 model, with the highest accuracy score of 92.04% with Adam optimizer and 92.73% with the RMSprop optimizer, performed best from the pretrained models. The proposed CNN structure, on the other hand, reached 100% accuracy and 1.0 AUC score. In [37], the features of Alexnet, MobilenetV2, DenseNet121, and Resnet50 models were obtained separately and combined to develop a hybrid-based CNN system. In this way, the number of features of the proposed hybrid model has been increased. The most valuable features were selected using the minimum redundancy maximum relevance (mRMR) feature selection method. SVM and KNN machine learning classifiers were used as classifiers. With 95.60% accuracy, the

SVM classifier was the most successful classifier. In [38], a ML approach with a radiomic-based classification pipeline is proposed. Images were masked using the original masks for feature extraction on the dataset. Histogram oriented gradient, Hu-moment, shape and texture features of masked images are extracted. The dataset is balanced with SMOTE oversampling, which is a synthetic data replication method. In the study, it was stated that ensemble learners (such as random forest, gradient boosting and AdaBoost classifiers) were able to achieve successful results according to multiple evaluation metrics. The proposed approach achieved accuracy, AUC, F1 score, and Mathews correlation coefficient values of 97.40%, 0.97, 94.00%, and 95.90%, respectively, in the BUSI dataset. In [39], a two-step, fully automated pipeline approach is proposed, the first step of which is to classify the ultrasound image, and the second step to generate the predictions in parallel with three CNN ensemble methods to segment it. Firstly, trainings were made using ResNet50, Xception, InceptionV3, InceptionResNetV2, and DenseNet121 architectures and classification prediction was made with ensemble learning. Secondly, segmentation of the lesioned regions was performed on the image with double labeling. As a result, experiments with 5-fold cross-validation on the BUSI dataset show 91.14% in terms of accuracy (classification) and 82.59% in terms of Dice coefficient (segmentation). In [40], image decomposition method is proposed to obtain fuzzy enhanced and bilateral filtered images as input to a pretrained model. Image enhancement and bilateral filtering methods are used to extend the original image data and create a multi-channel feature fusion model. The original ultrasound images were improved to remove edge features and noise was removed by bilateral filtering. In order not to lose useful information in the images, the original images were used as input alongside the bileteral filtered images that were expanded in the training. In the results, 93.00% accuracy, 95.00% sensitivity, 88.00% specificity, 93.00% F1 score, and 0.97 AUC values were obtained with the VGG16 model.

3. Material and methods

In this section, the dataset used in the study, image preprocessing steps and suggested approaches are explained.

3.1. Dataset and image preprocessing

In our study, Breast Ultrasound Images (BUSI) Dataset [41] was used to classify ultrasound images taken on BC as normal, benign, and malignant. The BUSI dataset was published by the Egyptian Baheya Hospital in 2018. It consists of 780 images, including 437 benign cases, 210 malignant cases, and 133 normal cases, from a total of 600 female patients aged 25 to 75 years. Images are in png format and 500 × 500 pixels. The LOGIQ E9 ultrasound system and the LOGIQ E9 Agile ultrasound system were used for scanning. The transducers' frequencies are 1–5 MHz on the ML6-15-D Matrix linear probe. Sample images of the BUSI dataset are given in Figure 2.

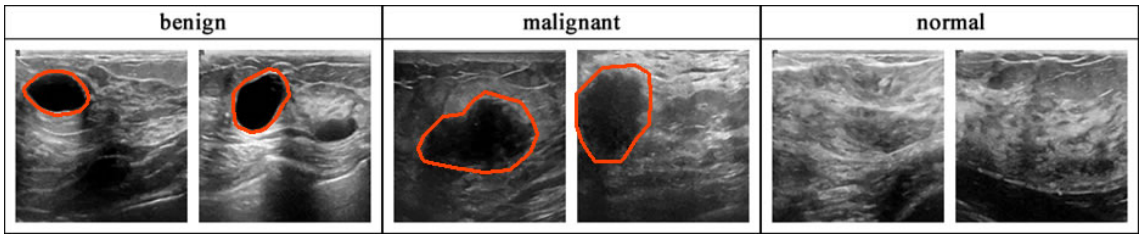

Figure 2. The BUSI dataset sample images.

The preprocessing steps first involve resizing the image dimensions to 224 × 224 pixels. For image

enhancement on the dataset, the Gaussian Blur method was applied and the images were converted to RGB format. Secondly, since the dataset has an unbalanced structure, the data augmentation method is applied to balance the dataset and reproduce new data. In the data augmentation method, cropping, translating, rotating, flipping, filling pixels, brighten, darken and recoloring techniques were applied to the images. A total of 3750 images, 1250 in each class, were created with data augmentation. Sample images of the dataset obtained after data augmentation are given in Figure 3.

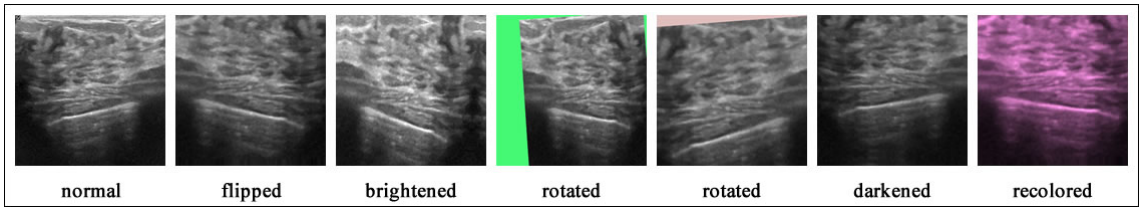

Figure 3. Sample images after data augmentation.

3.2. Transfer learning models

In this study, feature extraction was performed using pretrained models for BC detection from ultrasound images. Within the scope of the study, trainings were made with DenseNet121, InceptionV3, MobileNet, and ResNet50 transfer learning models, which were trained on ImageNet dataset and used successfully in image classification tasks in general. Information on these models used is given below.

3.2.1. DenseNet

In the DenseNet model, the outputs of each convolutional layer are connected to the inputs of all subsequent layers [42]. Thanks to these dense connections, information transfer is better, gradient loss is prevented even in the deeper parts of the network, and more effective learning takes place. The DenseNet121 version used in the study is a DenseNet model with a total of 121 layers.

3.2.2. Inception

The Inception model has a depth of 22 layers and consists of 144 layers [43]. With the Inception block inside, the convolution operation is performed with filter sizes of 1×1, 3×3, 5×5, and maximum pooling is performed with the size of 3×3. The InceptionV3 model is a version developed to increase efficiency [44]. Since the InceptionV3 model has a more complex and deep structure than other models, it performs better on larger data sets and more powerful hardware resources.

3.2.3. MobileNet

MobileNet is a modern neural network architecture developed for more efficient operation of real-time image processing and artificial intelligence applications in mobile devices and embedded systems [45]. MobileNet is lightweight because it uses a convolution operation called depthwise separable convolution. In the Depthwise separable convolution process, convolutions are performed for each channel separately and then combined. In this way, fewer parameters are used and the computational load is reduced.

3.2.4. ResNet

The ResNet model consists of resudial blocks, and in this model, the calculated weight values of one layer are transferred directly to the next layer [46]. Additional connections called "skip connection" or "shortcut connection" are used in the ResNet model. These connections allow information to be transferred without loss by jumping directly from one layer to another. The Resnet50 model is a version of ResNet that uses 150 layers.

3.3. Proposed method

In this study, an ensemble learning approach based on Dirichlet distribution [47], which is a probability density function on distributions, is proposed. The Dirichlet ensemble is obtained by combining Dirichlet distributions with different parameter values and can be used to estimate a probability distribution. The Dirichlet ensemble can take any value where the weights are not zero. Therefore, a more flexible model can be created compared to other probability modeling methods. This is especially useful for getting more accurate results when working with categorical data.

The formula for Dirichlet ensemble is based on the formulae of the Dirichlet distributions that are combined. The formula for the probability density function of the Dirichlet ensemble is defined below.

The probability density function of a Dirichlet distribution is

$$f(x|\alpha) = \left(\frac{1}{B(\alpha)}\right) * \prod_{d=1}^{K} x_d^{a_d - 1}. \tag{1}$$

Here, $x = (x_1, x_2, , x_k)$ is a K-dimensional probability vector, and $\alpha = (\alpha_1, \alpha_2, , \alpha_k)$ is a K-dimensional positive parameter vector. $B(\alpha)$ is the normalization constant known as the beta function.

Dirichlet ensemble arranges the weights $w = (w_1, w_2, , w_M)$ and the parameters $\alpha_1, \alpha_2, , \alpha_M$ of the combined Dirichlet distributions. In this case, the probability density function of the Dirichlet ensemble is

$$f(x|w, \alpha_1, \alpha_2, ..., \alpha_M) = \sum_{j=1}^{M} w_j * f_{j(x|a_j)}. \tag{2}$$

Here, $f_(j(x|\alpha_j))$ is the probability density function of the jth Dirichlet distribution. w_j is the weight of the jth Dirichlet distribution, and the $\sum$symbol represents the sum of the product of $f_(j(x|\alpha_j))$ with w_j for all j from 1 to M.

With the Dirichlet distribution, in the ensemble learning application, the dataset to be trained first is divided into training and test data. Secondly, the learning models used in ensemble learning are determined. In this study, pretrained models in the Keras library were used. Third, with each selected learning model, training is carried out on the training data set. Each weight obtained as a result of the trainings is recorded. Fourth, the ensemble model is created using the weights obtained as a result of the trainings and the α parameters of the Dirichlet distribution are determined. These parameters determine the weight of each learning model. By using different α values, the weighting of each model and its contribution to the model can be adjusted. Finally, the prediction performance of the created ensemble model is measured on the test dataset.

The parameter α determines the shape and density of the distribution in the Dirichlet distribution. The parameter α is usually represented by positive real numbers. If the value of α is large, the density of the distribution will be concentrated and narrow, and if the value of α is small, the density of the distribution will be spread and wide. For example, a Dirichlet distribution of $\alpha = (1, 1, 1)$ indicates that all variables have

equal weight, while a Dirichlet distribution of $\alpha = (0.5, 0.5, 0.5)$ indicates that the distribution is more spread out. The Dirichlet ensemble is a method in community learning that combines predictions using the weighted vote method. In this method, weights are assigned to the model estimates using the Dirichlet distribution and the final estimate is made using these weights. We have previously stated that the α parameter controls the distribution of these weights. We can determine the α parameter ourselves, or we can have the Dirichlet distribution determine it for statistically better performance. As a result of different tests, we obtained the best performance where the α parameter was determined by the Dirichlet distribution. The Dirichlet distribution distributed the α parameter with a total value of 1 according to the performance and complexity of the models.

In this study, we propose three approaches using deep ensemble transfer learning based on Dirichlet distribution with pretrained DenseNet121, InceptionV3, MobileNet, and ResNet50 transfer learning models to classify BC from ultrasound images. The approach named as Dirichlet based deep ensemble transfer learning (DB-DETL). Transfer learning is based on the idea that the features learned by one model are used to help another model learn. In our approach, after balancing the dataset with image processing and data augmentation, image features were extracted with DenseNet121, InceptionV3, MobileNet, and ResNet50 models and the weights obtained as a result of 30 epoch trainings were saved. With a model, repetitive trainings were carried out on the same dataset, each consisting of 30 epochs and in which the weights gained as a result of the previous training were given as an introduction to the new training.

The dataset used in the study consists of 224 $\times$ 224 pixel RGB images. A total of 3750 images were used in 3 different classes. Adaptive moment estimation (Adam) optimizer was used for the transfer learning methods used in the study. Because Adam optimizer combines the benefits of both momentum and RMSprop, accelerating the optimizer, balancing the updating of parameters at different scales, and improving the training process. The learning rate is set to 0.001. The categorical cross-entropy loss function was used as the loss function. Since the dataset used in our study was unbalanced and the number of data was insufficient, it was balanced by performing data augmentation and the number of data was increased. When using deep ensemble transfer learning based on the data augmentation process and Dirichlet distribution, overfitting and underfitting is a problem to consider. Overfitting can occur if the model architecture is too complex, if regularization operations are performed poorly or incorrectly, and if data is generated too close to the original data during the data augmentation process. Underfitting occurs when the proposed model is insufficient for the dataset, the hyperparameter settings are not appropriate, and the data is not suitable for training, resulting in poor performance. Therefore, in our proposed model, the model architecture is arranged in an appropriate structure, not too complex, to address overfitting and underfitting problems. The regularization techniques have carefully tuned the hyperparameter and data augmentation parameters.

In this study, 3 different approaches have been proposed with the proposed DB-DETL model structure. In the first approach, the results obtained as a result of each training with 4 models were fitted with ensemble learning. In the proposed approach, each training cycle is expressed with t. t1 is the first training result, t2 is the second training result, and t3 is the third training result. DenseNet121-t1, InceptionV3-t1, MobileNet-t1, and ResNet50-t1 results, DenseNet121-t2, InceptionV3-t2, MobileNet-t2, and ResNet50-t2 results, DenseNet121-t3, InceptionV3-t3, MobileNet-t3, and ResNet50-t3 results were fitted with Dirichlet ensemble and 3 different outputs were obtained. The recommended architecture for the first approach is shown in Figure 4. In the second approach, the weights obtained as a result of transfer learning repeated 3 times with the same model were fitted with Dirichlet ensemble learning. DenseNet121-t1, DenseNet121-t2 and DenseNet121-t3 results, InceptionV3-t1, InceptionV3-t2 and InceptionV3-t3 results, MobileNet-t1, MobileNet-t2 and MobileNet-t3 results, ResNet50-

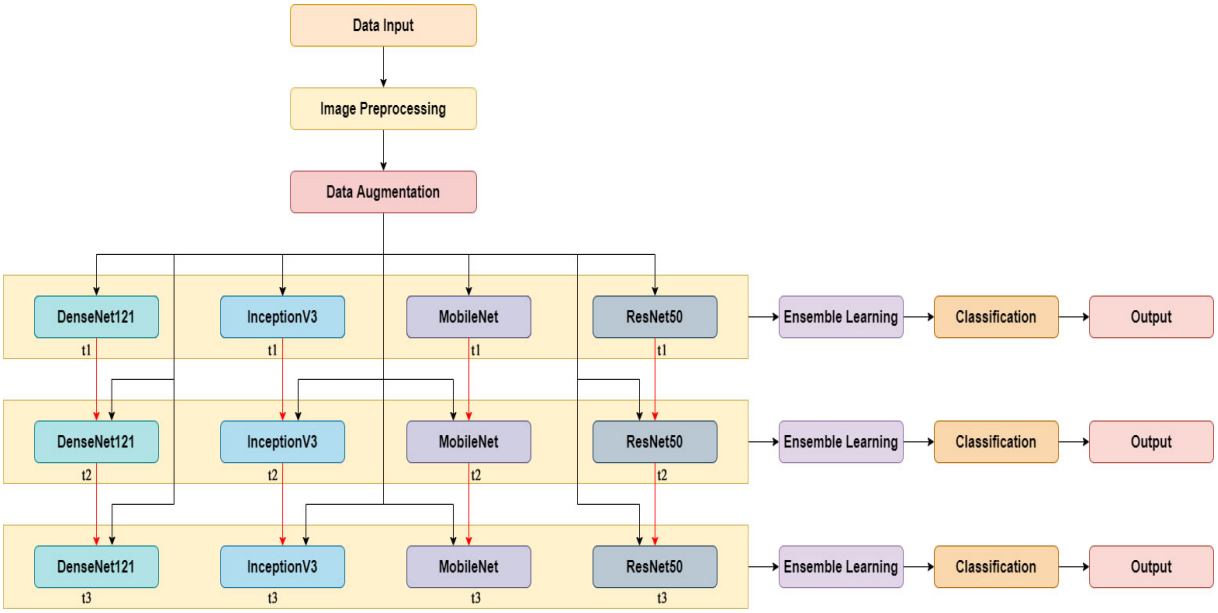

Figure 4. DB-DETL1 approach.

t1, ResNet50-t2, and ResNet50-t3 results were fitted with Dirichlet ensemble and 4 different outputs were obtained. The proposed architecture is shown in Figure 5. In the third approach, all weights obtained as a result of repetitive trainings, DenseNet121-t1, DenseNet121-t2, DenseNet121-t3, InceptionV3-t1, InceptionV3-t2, InceptionV3-t3, MobileNet-t1, MobileNet-t2, MobileNet-t3, ResNet50-t1, ResNet50-t2. and ResNet50-t3 results were fitted with Dirichlet ensemble and 1 output was obtained. The proposed architecture is shown in Figure 6. In this study, accuracy, loss, precision, recall, and F1 score metric values were calculated as a result of the trainings. Equations for accuracy, precision, recall, and F1 score are given below, respectively.

$$Accuracy = \left(\frac{TP + TN}{TP + TN + FP + FN}\right) \tag{3}$$

$$Precision = \left(\frac{TP}{TP + FP}\right) \tag{4}$$

$$Recall(Sensivity) = \left(\frac{TP}{TP + FN}\right) \tag{5}$$

$$F1score = 2x\left(\frac{PrecisionxRecall}{Precision + Recall}\right), \tag{6}$$

where TP is the number of true positives, TN is the number of true negatives, FP is the number of false positives, and FN is the number of false negatives.

4. Experimental results

In this study, using the previously trained DenseNet121, InceptionV3, MobileNet, and ResNet50 models, the weights obtained as a result of the previous training were reused and self-training was performed with the same

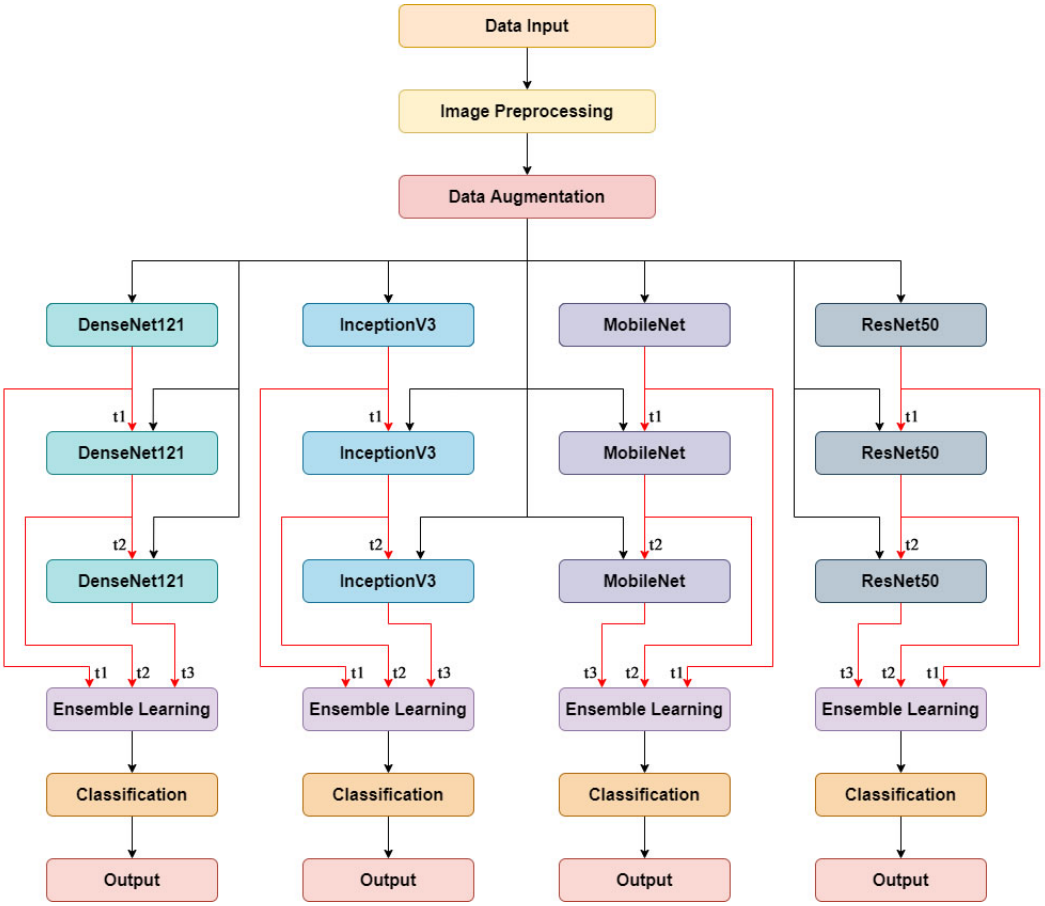

Figure 5. DB-DETL2 approach.

model with their own weights 3 times. The expression t represents the training number here. The results of train, validation and test accuracy with DenseNet121, InceptionV3, MobileNet, and ResNet50 transfer learning models are given in Table 1.

Table 1. Accuracy rates of pretrained models.

Pretrained model	Train number	Training accuracy	Validation accuracy	Test accuracy
DenseNet121	t1	96.21%	89.50%	88.80%
	t2	96.96%	90.83%	88.53%
	t3	91.46%	86.00%	86.67%
InceptionV3	t1	96.88%	89.00%	90.00%
	t2	98.12%	92.33%	91.20%
	t3	99.17%	91.00%	91.87%
MobileNet	t1	96.08%	85.00%	87.07%
	t2	99.00%	94.17%	93.47%
	t3	100.00%	96.33%	95.47%
ResNet50	t1	98.58%	84.83%	84.53%
	t2	92.54%	79.33%	78.93%
	t3	83.21%	70.17%	70.67%

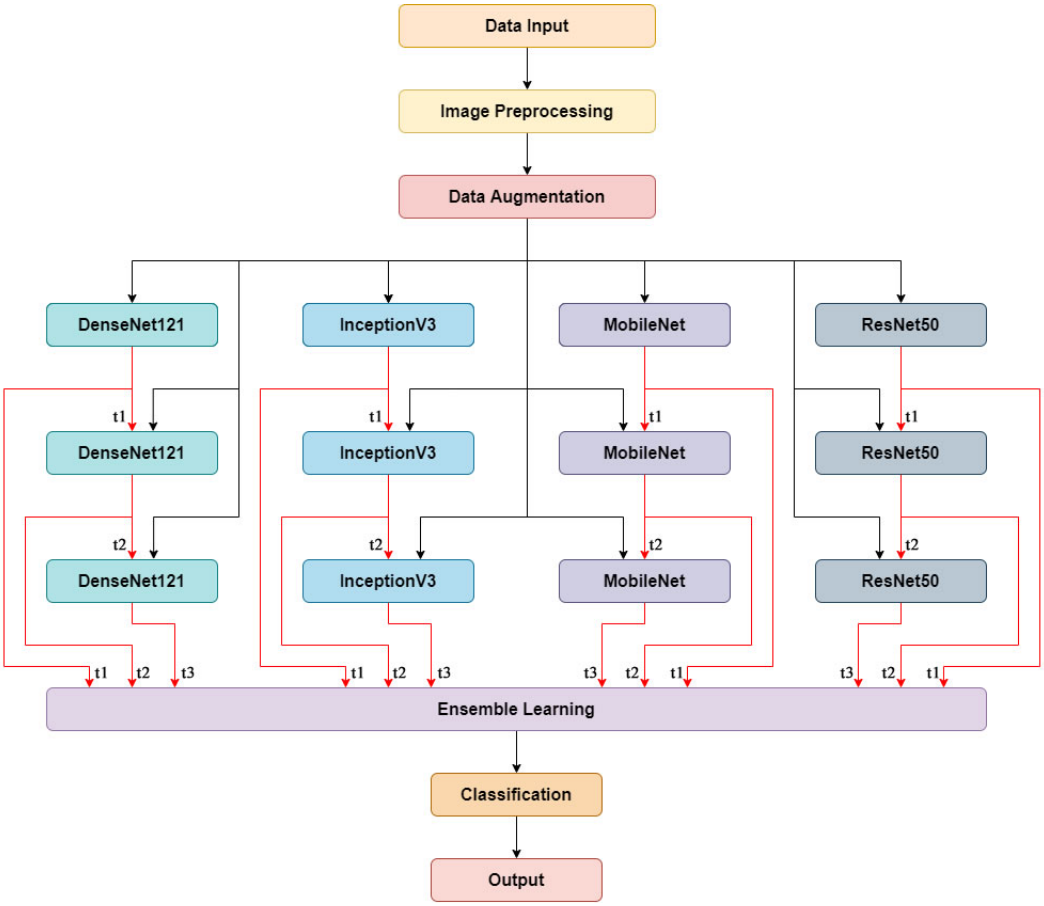

Figure 6. DB-DETL3 approach.

As seen in Table 1, it is the MobileNet model that performs well among the pretrained models. The MobileNet model achieved 100.00% training, 96.33% validation, and 95.47% test accuracy scores as a result of the third training. The second successful model is the MobileNet model again. As a result of the second training, the MobileNet model achieved 99.00% training, 94.17% validation, and 93.47% test accuracy scores. The third successful model is the InceptionV3 model. InceptionV3 model achieved 99.17% training, 91.00% validation, and 91.87% test accuracy scores as a result of the third training. When Table 1 is examined in general, the repetition of the model and the use of the previous training results as input increased the performance rate in other models except the ResNet50 model.

In this study, accuracy and loss graphs of the trainings made with each model of the models were obtained in order to evaluate and compare the performance of each model. Figure 7 shows the accuracy graphs of t1 training results, Figure 8 shows the accuracy graphs of t1 training results and Figure 9 shows the accuracy graphs of t3 training results of DenseNet121, InceptionV3, MobileNet and ResNet50 models. As seen in Figures 7–9, it is understood that each model went through a fluctuating learning process and experienced sudden drops during training. The MobileNet model increased the learning process in the second training and showed a more stable learning process in the third training.

Loss is a performance measure that indicates that the model's performance is successful as its value ap-

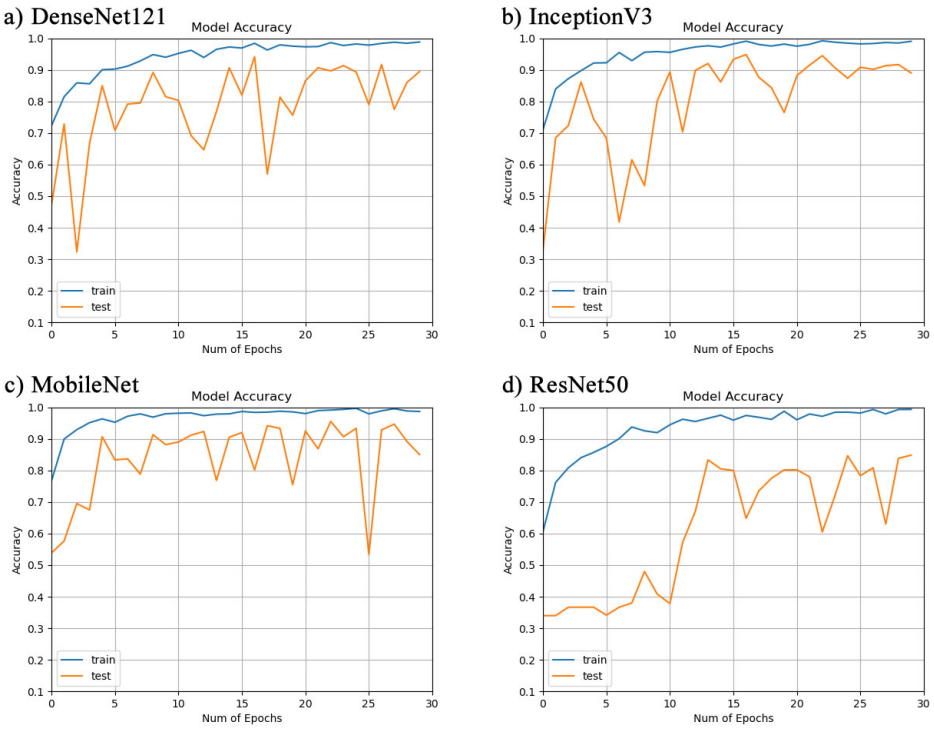

Figure 7. Accuracy graphs of a) DenseNet121-t1, b) InceptionV3-t1, c) MobileNet-t1, d) ResNet50-t1.

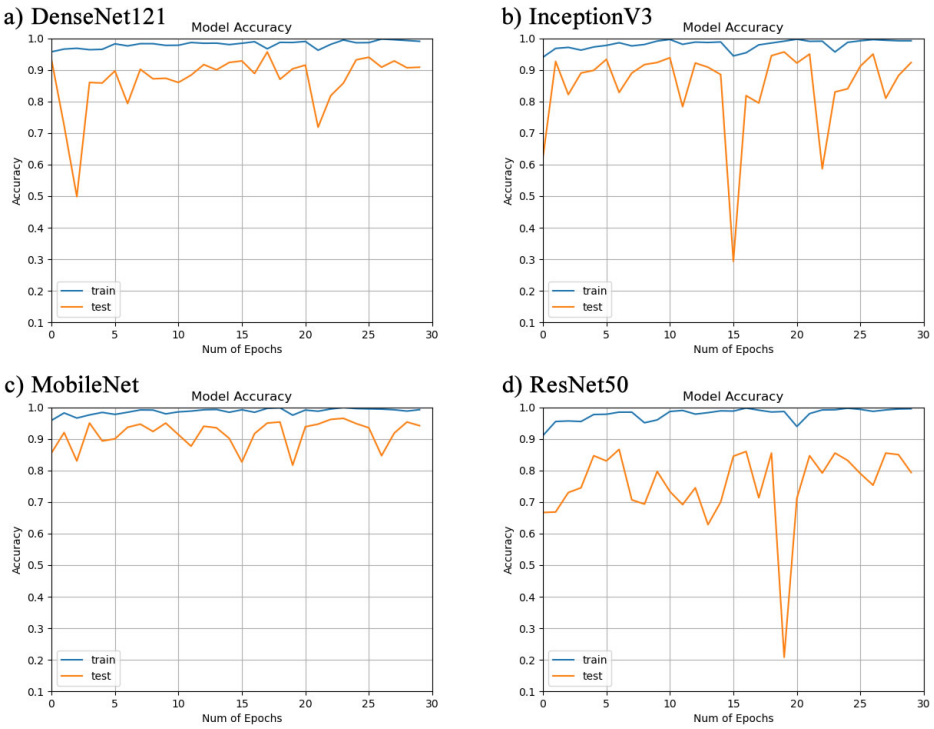

Figure 8. Accuracy graphs of a) DenseNet121-t2, b) InceptionV3-t2, c) MobileNet-t2, d) ResNet50-t2.

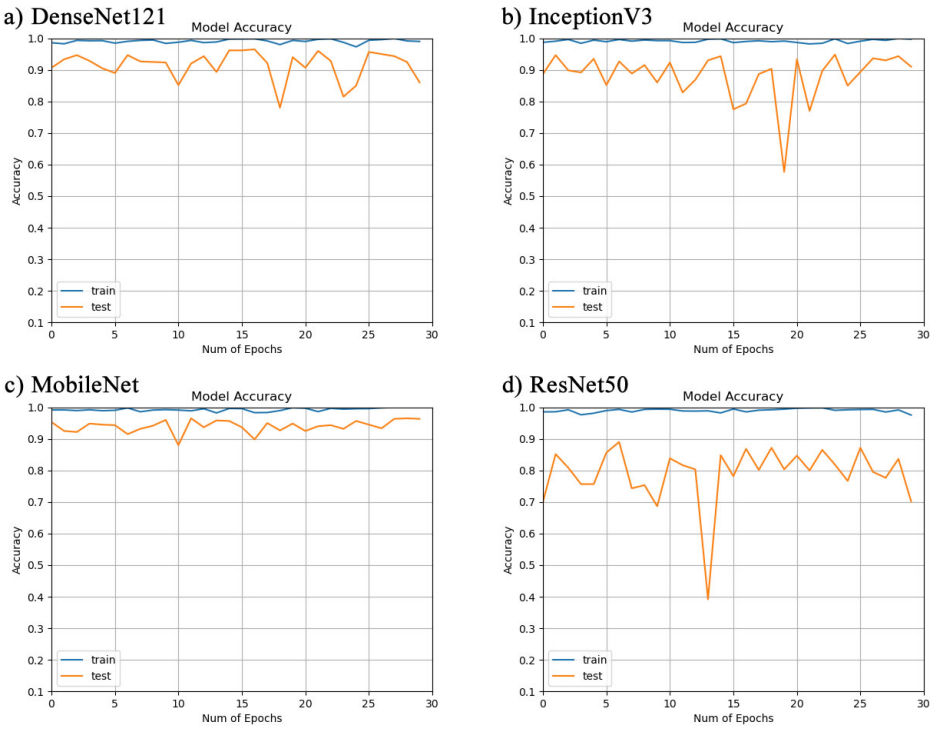

Figure 9. Accuracy graphs of a) DenseNet121-t3, b) InceptionV3-t3, c) MobileNet-t3, d) ResNet50-t3.

proaches 0. Figure 10 shows the loss graphs of t1 training results, Figure 11 shows the loss graphs of t2 training results and Figure 12 shows the loss graphs of t3 training results. When Figures 10–12 are examined, it is seen that there are high fluctuations in the loss graphs of all models. It has been observed that various optimization methods can be effective to correct loss rates.

Confusion matrix was also prepared to see the true and false predictions on the test data of the pretrained models. The confusion matrix for all models and dataset combinations is shown in Figure13, Figure14 and Figure15. When the confusion matrices are examined, it is seen that the pretrained models can often separate the data well.

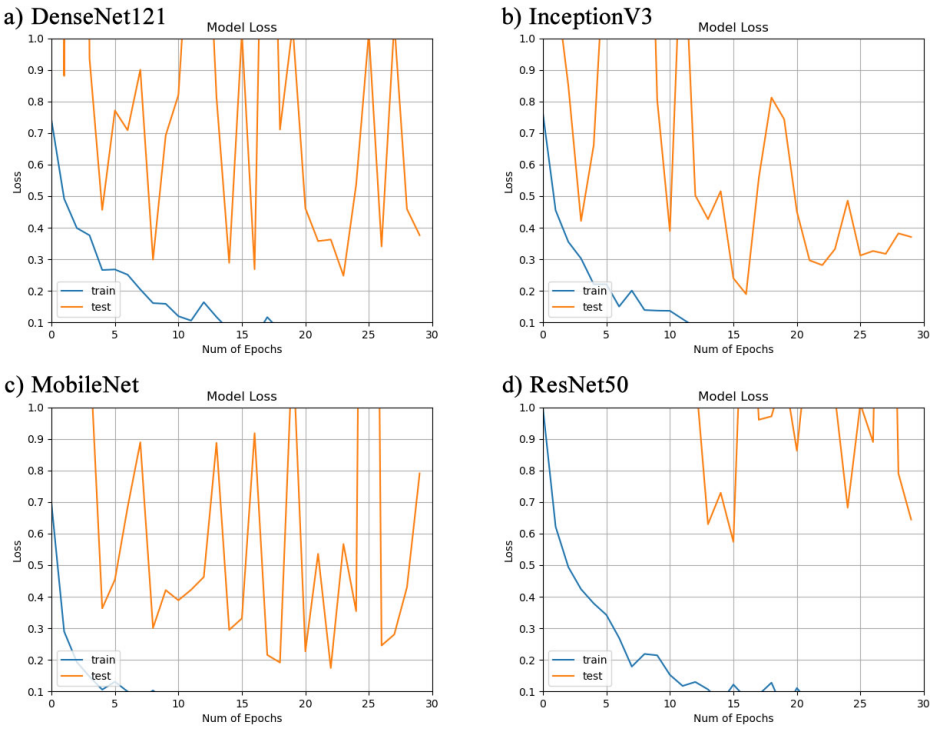

Figure 10. Loss graphs of a) DenseNet121-t1, b) InceptionV3-t1, c) MobileNet-t1, d) ResNet50-t1.

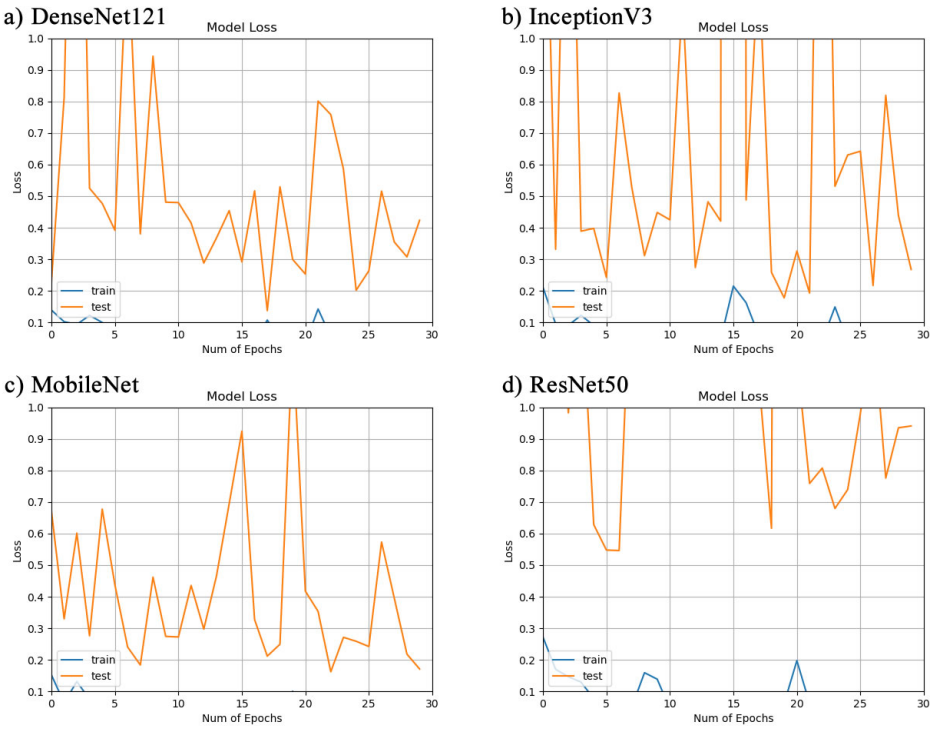

Figure 11. Loss graphs of a) DenseNet121-t2, b) InceptionV3-t2, c) MobileNet-t2, d) ResNet50-t2.

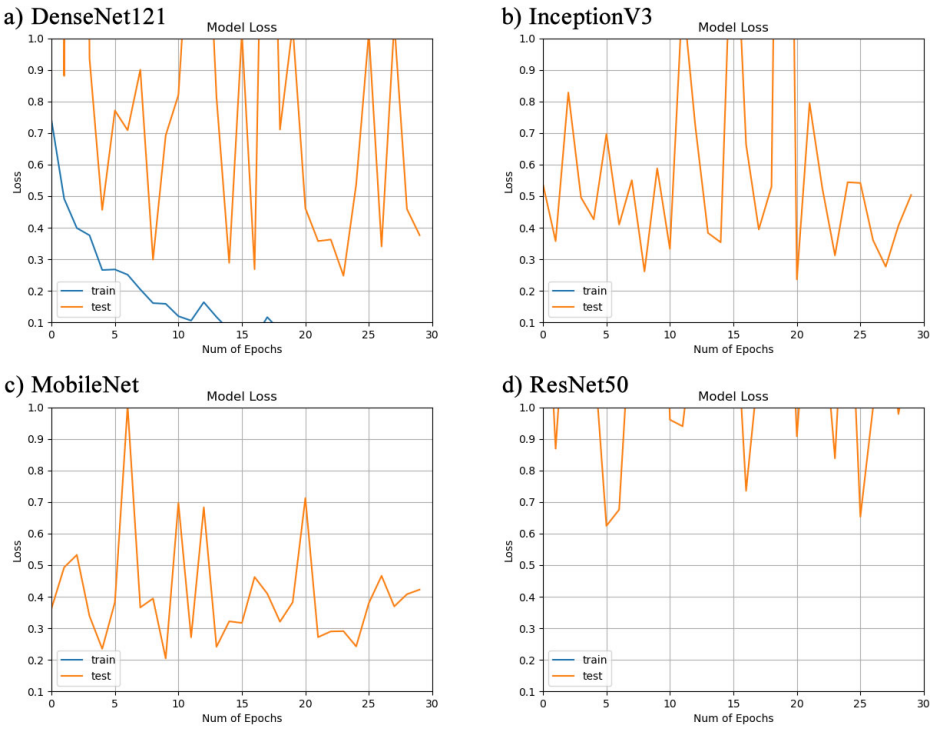

Figure 12. Loss graphs of a) DenseNet121-t3, b) InceptionV3-t3, c) MobileNet-t3, d) ResNet50-t3.

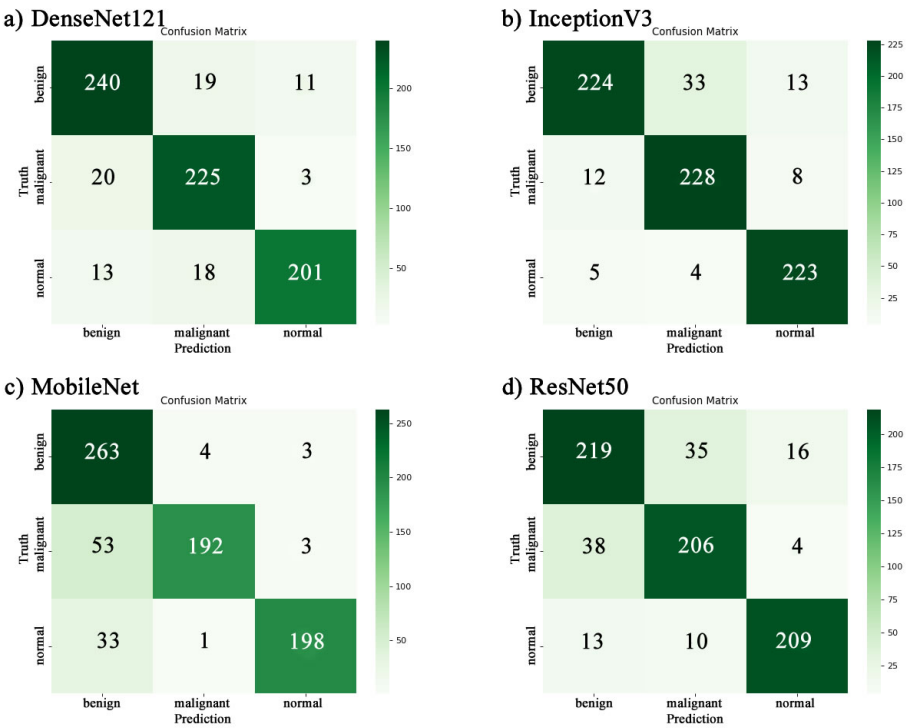

Figure 13. Confusion matrices a) DenseNet121-t1, b) InceptionV3-t1, c) MobileNet-t1, d) ResNet50-t1.

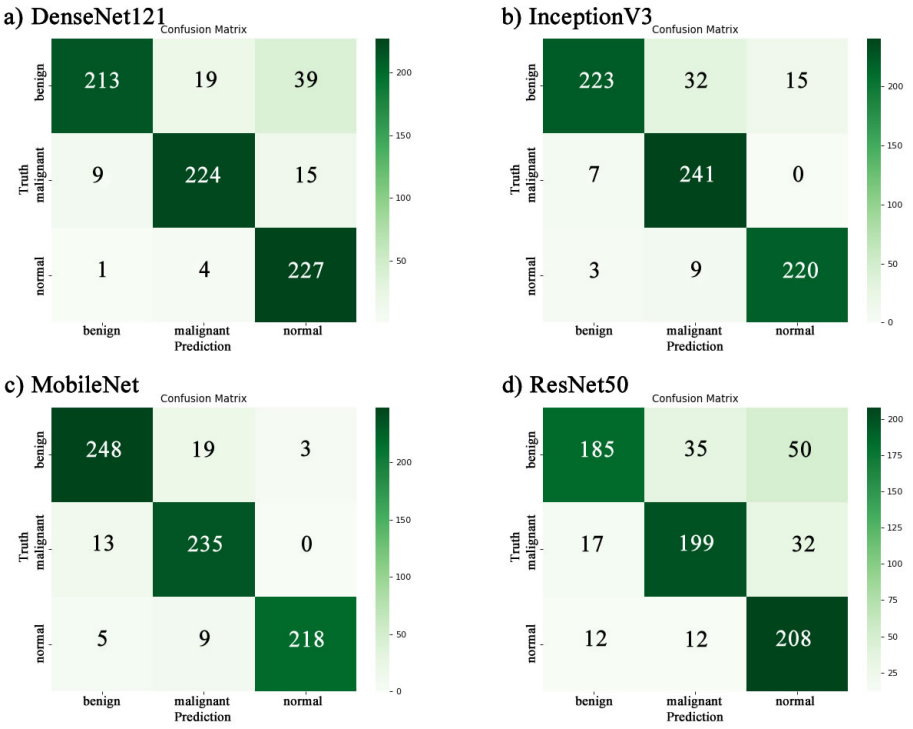

Figure 14. Confusion matrices a) DenseNet121-t2, b) InceptionV3-t2, c) MobileNet-t2, d) ResNet50-t2.

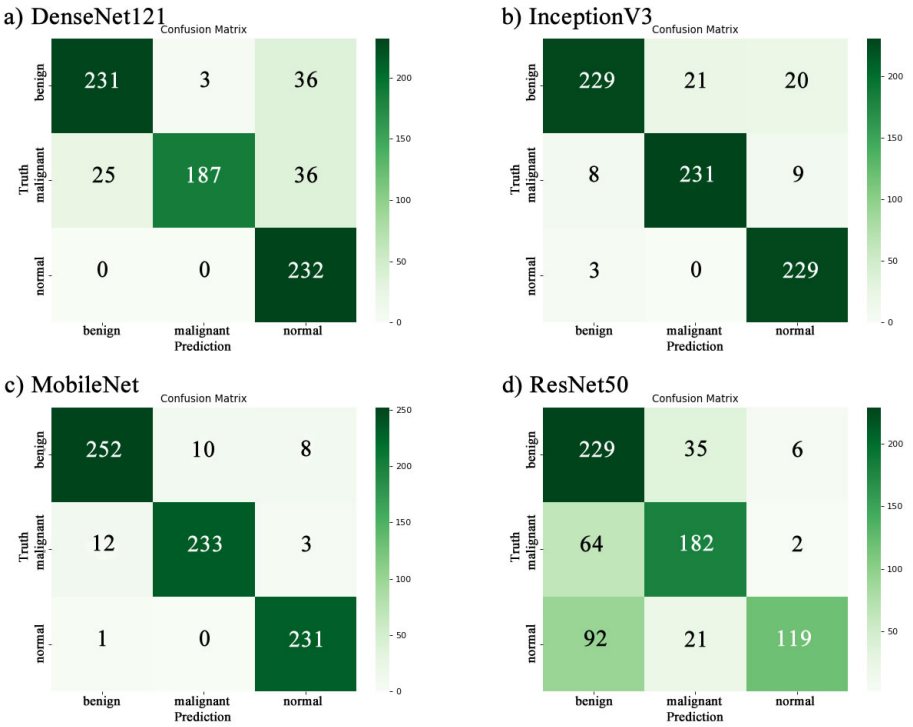

Figure 15. Confusion matrices a) DenseNet121-t3, b) InceptionV3-t3, c) MobileNet-t3, d) ResNet50-t3.

Performance evaluation scores of transfer learning models are given in Table 2.

Table 2. Performance evaluation scores of pretrained models.

Metric	Pretrained model	Train number	Benign	Malignant	Normal
Precision	DenseNet121	t1	0.88	0.86	0.93
		t2	0.96	0.91	0.81
		t3	0.90	0.98	0.76
	InceptionV3	t1	0.93	0.86	0.91
		t2	0.96	0.85	0.94
		t3	0.95	0.92	0.89
	MobileNet	t1	0.75	0.97	0.97
		t2	0.93	0.89	0.99
		t3	0.95	0.96	0.95
	ResNet50	t1	0.81	0.82	0.91
		t2	0.86	0.81	0.72
		t3	0.59	0.76	0.94
Recall	DenseNet121	t1	0.89	0.91	0.87
		t2	0.79	0.90	0.98
		t3	0.86	0.75	1.00
	InceptionV3	t1	0.83	0.92	0.96
		t2	0.83	0.97	0.95
		t3	0.85	0.93	0.99
	MobileNet	t1	0.97	0.77	0.85
		t2	0.92	0.95	0.94
		t3	0.93	0.94	1.00
	ResNet50	t1	0.81	0.83	0.90
		t2	0.69	0.80	0.90
		t3	0.85	0.73	0.51
F1 score	DenseNet121	t1	0.88	0.88	0.90
		t2	0.86	0.91	0.88
		t3	0.88	0.85	0.87
	InceptionV3	t1	0.88	0.89	0.94
		t2	0.89	0.91	0.94
		t3	0.90	0.92	0.93
	MobileNet	t1	0.85	0.86	0.91
		t2	0.93	0.92	0.96
		t3	0.94	0.95	0.97
	ResNet50	t1	0.81	0.83	0.91
		t2	0.76	0.81	0.80
		t3	0.70	0.75	0.66
Support			270	248	232

As seen in Table 2, the most successful results were obtained as a result of the third training with the MobileNet model. The MobileNet model obtained 0.94, 0.95, 0.97 F1 score, 0.93, 0.94, 1.00 recall, and 0.95, 0.96, 0.95 precision values from benign, malignant, and normal classes, respectively.

The outputs obtained from each training cycle of the 4 different transfer learning models used for the proposed 1st approach were fitted with the DB-DETL model. In Table 3, the validation results performed with the transfer learning models and the 3 different validation results performed with the proposed 1st approach are given. When the results were examined, the DB-DETL model, in which the results of the second training were used, achieved the highest success rate with 97.67%.

Table 3. DB-DETL-1 approach scores.

Pretrained models	Accuracy score	DB-DETL-1 approach Accuracy score
DenseNet121_t1	90.67%	96.00%
InceptionV3_t1	89.00%	
MobileNet_t1	84.67%	
ResNet50_t1	83.83%	
DenseNet121_t2	90.83%	97.67%
InceptionV3_t2	92.33%	
MobileNet_t2	94.17%	
ResNet50_t2	79.33%	
DenseNet121_t3	86.00%	96.67%
InceptionV3_t3	91.00%	
MobileNet_t3	96.33%	
ResNet50_t3	70.17%	

The outputs obtained from the three training results of same transfer learning model were used for the proposed 2nd approach were fitted with the DB-DETL model. In Table 4, validation results with transfer learning models and 4 different validation results with the proposed 2nd approach are given. When the results are examined, the DB-DETL model, in which the three training results of the MobileNet model are used, reached the highest success rate with 97.33%.

Table 4. DB-DETL-2 approach scores.

Pretrained models	Accuracy score	DB-DETL-2 approach Accuracy score
DenseNet121_t1	90.67%	95.00%
DenseNet121_t2	90.83%	
DenseNet121_t3	86.00%	
InceptionV3_t1	89.00%	95.17%
InceptionV3_t2	92.33%	
InceptionV3_t3	91.00%	
MobileNet_t1	84.67%	97.33%
MobileNet_t2	94.17%	
MobileNet_t3	96.33%	
ResNet50_t1	83.83%	87.00%
ResNet50_t2	79.33%	
ResNet50_t3	70.17%	

All the outputs of the 4 different transfer learning models used for the proposed 3rd approach were fitted with the DB-DETL model. In Table 5, the validation results performed with the transfer learning models and the validation result performed with the proposed 3rd approach are given. When the results are examined, the DB-DETL model, in which all training results are used, has reached a success rate of 98.17%.

Table 5. DB-DETL-3 approach scores.

Pretrained models	Accuracy score	DB-DETL-3 approach Accuracy score
DenseNet121_t1	90.67%	98.17%
DenseNet121_t2	90.83%	
DenseNet121_t3	86.00%	
InceptionV3_t1	89.00%	
InceptionV3_t2	92.33%	
InceptionV3_t3	91.00%	
MobileNet_t1	84.67%	
MobileNet_t2	94.17%	
MobileNet_t3	96.33%	
ResNet50_t1	83.83%	
ResNet50_t2	79.33%	
ResNet50_t3	70.17%	

5. Discussion

In this study, 3 different Dirichlet distribution community learning structures based on transfer learning models are proposed. It has been observed that retesting with transfer learning models improves performance.

In the study, 3 repetitive tests were conducted with each of the DenseNet121, InceptionV3, MobileNet, and ResNet50 models using previous training weights. In the study, train and validation accuracy rates were calculated for each model during training, and then accuracy rates with test data were calculated. The best accuracy rates of the DenseNet121 model were obtained as 96.96% training, 90.83% validation and 88.53% testing at the end of the second training. The best accuracy rates of the InceptionV3 model were obtained as 99.17% training, 91.00% validation, and 91.87% testing at the end of the third training. The best accuracy rates of the MobileNet model were obtained as 100.00% training, 96.33% validation, and 95.47% testing at the end of the third training. The best accuracy rates of the ResNet50 model were obtained as 98.58% training, 84.83% validation, and 84.53% testing at the end of the first training. Performance metrics belonging to 3 different classes in the dataset were examined. In the benign class; it is seen that the precision value is high in InceptionV3 and MobileNet models, the recall value is high in the MobileNet model, and the F1 score value is also high in the MobileNet model. In the malignant class; it is seen that precision value is high in DenseNet121 and MobileNet models, recall value is high in InceptionV3 and MobileNet models, and F1 score value is high in InceptionV3 and MobileNet models. In the normal class; it is seen that precision value is high in MobileNet model, recall value is high in InceptionV3 and MobileNet models, and F1 score value is high in InceptionV3 and MobileNet models. When we evaluate in general, it is seen that the MobileNet model is the most successful model among the transfer learning models used in the study. The second successful model is the InceptionV3 model.

In the second part of the study, the weights obtained as a result of training with transfer learning models were reweighted with different ensemble approaches. Among the suggested approaches, the DB-DETL-1 approach obtained 97.67% accuracy from the validation images, the DB-DETL-2 approach 97.33% and the DB-DETL-3 approach 98.17% accuracy. The DB-DETL-3 model, in which the weights obtained from all trainings were used, showed the highest performance. When the results are compared, it is observed that the ensemble learning architecture gives better results than the transfer learning architectures.

Detection of BC from ultrasound images has been the subject of study by many researchers. Since we conducted experiments on the BUSI dataset within the scope of our study, the researchers, working methods

and accuracy rates of the study results are given in Table 6, which has worked on the BUSI dataset before.

Table 6. Comparison of classification results against state-of-the-art.

Author	Accuracy	Precision	Recall	F1 score	Specificity
Singh et al. (2016) [6]	95.86%	-	95.13%	-	96.57%
Zhuang et al. (2021) [25]	95.00%	98.11%	93.92%	95.71%	98.33%
Wu al. (2012) [28]	95.24%	91.67%	97.78%	-	93.33%
Lu et al. (2022) [30]	94.10%	98.14%	94.93%	96.50%	-
Mishra et al. (2022) [34]	96.24%	97.14%	97.37%	97.25%	93.95%
Joshi et al. (2022) [35]	96.31%	-	92.63%	92.99%	96.71%
Eroğlu et al. (2021) [37]	95.60%	-	94.89%	94.74%	97.30%
Mishra et al. (2021) [38]	97.40%	95.80%	96.00%	95.90%	98.00%
Our model (2023)	98.17%	Benign: 94.00% Malignant: 95.00% Normal: 97.00%	Benign: 93.00% Malignant: 94.00% Normal: 100.00%	Benign: 95.00% Malignant: 96.00% Normal: 95.00%	-

The studies and accuracy results given in Table 6 show that many researchers have achieved an accuracy rate of 95%–96%. The best performance among the studies performed is 97.40% [38]. In the trainings made with transfer learning models, it is seen that the MobileNet model has a performance equivalent to other studies. When the values we obtained with our ensemble learning approaches, which constitute the structure of our study, are examined, it is seen that the DB-DETL-3 approach is more successful than other models with an accuracy rate of 98.17%.

When other performance metrics (precision, recall, F1 score and specifity) are analyzed, the highest precision value is 98.14% [30], the highest recall value is 97.78% [28], the highest F1 score is 97.25%[34] and the highest specificity value It is seen that it is 98.33% [25]. In the proposed study, it is stated that the performance of the MobileNet model is higher than other models. When the results obtained after the third training with the MobileNet model are examined in the classifications, the precision value is 94.00%, 95.00%, 97.00% for the benign, malignant, and normal classes, respectively. Recall value for benign, malignant, and normal class is 93.00%, 94.00%, 100.00%, respectively. The F1 score value was calculated as 95.00%, 96.00%, and 96.00% for the benign, malignant, and normal classes, respectively. When the results are compared, it is seen that the precision, recall and F1 score values of our model are approximately close to the results obtained in the studies in the literature. If all performance metrics are evaluated in general, we can say that our proposed study is positively comparable with the studies in the literature and shows a successful result for the detection of breast cancer from ultrasound images.

6. Conclusion

In this study, a Dirichlet distribution based ensemble learning approach using transfer learning models for BC detection from ultrasound images is proposed. BUSI dataset was used in the study. Since the BUSI dataset has an unbalanced structure, the dataset has been balanced with the data augmentation technique in order to increase the performance. By conducting repetitive trainings with transfer learning models, the results were reevaluated with ensemble learning. Our proposed DB-DETL-3 approach achieved 98.17% accuracy on the BUSI dataset. Considering the success rate and high performance of our model, we can say that our proposed approach will be a useful aid for radiologists and clinicians in detecting BC from ultrasound images.

One of the key findings of this study is that ensemble learning can outperform ensemble-generating models. In addition, another finding is that transfer learning models give good results in retrainings by giving

the weights obtained in the first training as input to the model again. Another important finding is that addressing the class imbalance problem and balancing classes can improve performance.

Deep learning methods produce consistent predictions for an ultrasound image, reducing misdiagnoses by radiologists, improving the rate of early detection, and speeding up image interpretation time in clinics [3,4]. When deep learning methods were compared with radiologists' diagnoses, it was stated that the proposed method outperformed the expert's analysis in terms of accuracy [20].

Although the proposed approaches show high performance, it is thought that it would be beneficial to apply different optimization methods to reduce the fluctuations in the accuracy and loss graphs obtained as a result of the trainings and to transfer the learning models for a more robust architecture. It is common practice to use a pretrained model as a feature extractor and fine-tune it in transfer learning studies [36]. In future work, we plan to use more transfer learning models and different equilibrium techniques with fine-tuning and optimization. In addition, we are thinking of developing a structure that includes first segmentation and then classification studies on the image. In this way, we aim to refine our proposed approach and develop a reliable and robust CAD system to assist radiologists and clinicians in their work.

Acknowledgment

The author declares no conflict of interest and did not receive support from any organization for the submitted work.

References

[1] World Health Organization. WHO report on cancer: setting priorities, investing wisely and providing care for all. 2020.

[2] Cancer Today. https://gco.iarc.fr/today/home acces date: 08.10.2022

[3] Zhu YC, AlZoubi A, Jassim S, Jiang Q et al. A generic deep learning framework to classify thyroid and breast lesions in ultrasound images. Ultrasonics 2021; 110: 106300.

[4] Balaha HM, Saif M, Tamer A, Abdelhay EH. Hybrid deep learning and genetic algorithms approach (HMB-DLGAHA) for the early ultrasound diagnoses of breast cancer. Neural Computing and Applications 2022; 34 (11): 8671-8695.

[5] Liu B, Cheng HD, Huang J, Tian J et al. Fully automatic and segmentation-robust classification of breast tumors based on local texture analysis of ultrasound images. Pattern Recognition 2010; 43 (1): 280-298.

[6] Singh BK, Verma K, Thoke AS. Fuzzy cluster based neural network classifier for classifying breast tumors in ultrasound images. Expert Systems with Applications 2016; 66: 114-123.

[7] Misra S, Jeon S, Managuli R, Lee S et al. Bi-modal transfer learning for classifying breast cancers via combined B-mode and ultrasound strain imaging. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control 2021; 69 (1): 222-232.

[8] Yap MH, Pons G, Marti J, Ganau S et al. Automated breast ultrasound lesions detection using convolutional neural networks. IEEE journal of biomedical and health informatics 2017; 22 (4): 1218-1226.

[9] Dabass M, Vashisth S, Vig R. A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images. Computers in biology and medicine 2022; 147: 105680.

[10] Gupta K, Chawla N. Analysis of histopathological images for prediction of breast cancer using traditional classifiers with pre-trained CNN. Procedia Computer Science 2020; 167: 878-889.

[11] Calp MH. Use of Deep Learning Approaches in Cancer Diagnosis. In Kose U, Alzubi J. Alzubi (editor). Deep Learning for Cancer Diagnosis 2021, pp. 249–267. Springer, Singapore.

[12] Savaş S. Detecting the stages of Alzheimer's disease with pre-trained deep learning architectures. Arabian Journal for Science and Engineering 2022; 47 (2): 2201-2218.

[13] Buyrukoğlu S. Early Detection of Alzheimer's Disease Using Data Mining: Comparison of Ensemble Feature Selection Approaches. Konya Journal of Engineering Sciences 2021; 9 (1); 50–61.

[14] Asiri N, Hussain M, Al Adel F, Alzaidi N. Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey. Artificial intelligence in medicine 2019; 99: 101701.

[15] Buyrukoğlu S, Akbaş A. Machine Learning based Early Prediction of Type 2 Diabetes: A New Hybrid Feature Selection Approach using Correlation Matrix with Heatmap and SFS. Balkan Journal of Electrical and Computer Engineering 2022; 10 (2): 110–117. https://doi.org/10.17694/bajece.973129

[16] Alafif T, Tehame AM, Bajaba S, Barnawi A, Zia S. Machine and deep learning towards COVID-19 diagnosis and treatment: survey, challenges, and future directions. International journal of environmental research and public health 2021; 18 (3): 1117.

[17] Bütüner R, Calp MH. Diagnosis and Detection of COVID-19 from Lung Tomography Images Using Deep Learning and Machine Learning Methods. International Journal of Intelligent Systems and Applications in Engineering 2022; 10 (2): 190–200. https://ijisae.org/index.php/IJISAE/article/view/1843

[18] Savaş S, Topaloğlu N, Kazcı Ö, Koşar PN. Classification of carotid artery intima media thickness ultrasound images with deep learning. Journal of medical systems 2019; 43 (8): 1-12.

[19] Al-Saedi DKA, Savaş S. Classification of Skin Cancer with Deep Transfer Learning Method. Computer Science, IDAP-2022 : International Artificial Intelligence and Data Processing Symposium 2022; 202-210.

[20] Khan S, Islam N, Jan Z, Din IU, Rodrigues JJC. A novel deep learning based framework for the detection and classification of breast cancer using transfer learning. Pattern Recognition Letters 2019; 125: 1-6.

[21] Lu J, Behbood V, Hao P, Zuo H et al. Transfer learning using computational intelligence: A survey. Knowledge-Based Systems 2015; 80: 14-23.

[22] Salehi AW, Khan S, Gupta G, Alabduallah BI et al. A Study of CNN and Transfer Learning in Medical Imaging: Advantages, Challenges, Future Scope. Sustainability 2023; 15 (7): 5930.

[23] Boldbaatar EA, Lin LY, Lin CM. Breast tumor classification using fast convergence recurrent wavelet Elman neural networks. Neural Processing Letters 2019; 50 (3): 2037-2052.

[24] Zeimarani B, Costa MGF, Nurani NZ, Bianco SR et al. Breast lesion classification in ultrasound images using deep convolutional neural network. IEEE Access 2020; 8: 133349-133359.

[25] Zhuang Z, Yang Z, Raj ANJ, Wei C et al. Breast ultrasound tumor image classification using image decomposition and fusion based on adaptive multi-model spatial feature fusion. Computer Methods and Programs in Biomedicine 2021; 208: 106221.

[26] Pavithra S, Vanithamani R, Justin J. Computer aided breast cancer detection using ultrasound images. Materials Today: Proceedings 2020; 33: 4802-4807.

[27] Moon WK, Lee YW, Ke HH, Lee SH et al. Computer-aided diagnosis of breast ultrasound images using ensemble learning from convolutional neural networks. Computer methods and programs in biomedicine 2020; 190: 105361.

[28] Wu WJ, Lin SW, Moon WK.Combining support vector machine with genetic algorithm to classify ultrasound breast tumor images. Computerized Medical Imaging and Graphics 2012; 36 (8): 627-633.

[29] Karthik R, Menaka R, Kathiresan GS, Anirudh M, Nagharjun M. Gaussian Dropout Based Stacked Ensemble CNN for Classification of Breast Tumor in Ultrasound Images. IRBM 2022; 43 (6): 715-733.

[30] Lu SY, Wang SH, Zhang YD. SAFNet: A deep spatial attention network with classifier fusion for breast cancer detection. Computers in Biology and Medicine 2022; 148: 105812.

[31] Qu X, Lu H, Tang W, Wang S et al. A VGG attention vision transformer network for benign and malignant classification of breast ultrasound images. Medical Physics 2022; 49 (9): 5787-5798

[32] Acharya UR, Meiburger KM, Koh JEW, Ciaccio EJ et al. A novel algorithm for breast lesion detection using textons and local configuration pattern features with ultrasound imagery. IEEE Access 2019; 7: 22829-22842.

[33] Yang Z, Gong X, Guo Y, Liu W. A temporal sequence dual-branch network for classifying hybrid ultrasound data of breast cancer. IEEE Access 2020; 8: 82688-82699.

[34] Mishra AK, Roy P, Bandyopadhyay S, Das SK. Achieving highly efficient breast ultrasound tumor classification with deep convolutional neural networks. International Journal of Information Technology 2022; 14 : 3311–3320.

[35] Joshi RC, Singh D, Tiwari V, Dutta MK. An efficient deep neural network based abnormality detection and multi-class breast tumor classification. Multimedia Tools and Applications 2022; 81 (10): 13691-13711.

[36] Masud M, Eldin Rashed AE, Hossain MS. Convolutional neural network-based models for diagnosis of breast cancer. Neural Computing and Applications 2020; 34:11383–11394.

[37] Eroğlu Y, Yildirim M, Çinar A. Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign, malignant, and normal using mRMR. Computers in biology and medicine 2021; 133: 104407.

[38] Mishra AK, Roy P, Bandyopadhyay S, Das SK. Breast ultrasound tumour classification: A Machine Learning—Radiomics based approach. Expert Systems 2021; 38 (7): e12713.

[39] Podda AS, Balia R, Barra S, Carta S et al. Fully-automated deep learning pipeline for segmentation and classification of breast ultrasound images. Journal of Computational Science 2022; 63: 101816.

[40] Zhuang Z, Kang Y, Joseph Raj AN, Yuan Y et al. Breast ultrasound lesion classification based on image decomposition and transfer learning. Medical Physics 2020; 47 (12): 6257-6269.

[41] Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A. Dataset of breast ultrasound images. Data in brief 2020; 28: 104863.

[42] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017: pp. 4700-4708.

[43] Szegedy C, Liu W, Jia Y, Sermanet P et al. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2015: pp. 1-9.

[44] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016: pp. 2818-2826.

[45] Howard AG, Zhu M, Chen B, Kalenichenko D et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint 2017; arXiv:1704.04861

[46] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016;pp. 770-778.

[47] Madsen RE, Kauchak D, Elkan C. Modeling word burstiness using the Dirichlet distribution. In Proceedings of the 22nd international conference on Machine learning 2005; pp. 545-552.