Biomedical Signal Processing and Control 85 (2023) 104871

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Breast UltraSound Image classification using fuzzy-rank-based ensemble network

Sagar Deep Deb *, Rajib Kumar Jha

Department of Electrical Engineering, Indian Institute of Technology Patna, India

ARTICLE INFO

Keywords:
Breast cancer
CNN
Deep learning

ABSTRACT

Breast Cancer is one of the most commonly occurring cancers in women. Detecting cancer at an earlier stage increases the chance of survival. Thus the development of an automatic CAD system is important for cancer diagnosis. Researchers have used various image modalities like mammograms and Magnetic Resonance Images to detect cancer. This manuscript uses UltraSound Images for detecting breast cancer. We have proposed a fuzzy-rank-based ensemble network for detecting breast cancer. The proposed model uses four different base learners, namely VGG-Net, DenseNet, Xception, and Inception, and it tries to take advantage of the predictions made by the base learners. The weights of the initial layers of the base learners are pre-trained on the ImageNet dataset. In contrast, the final five layers are fine-tuned using a publicly available Breast Ulta Sound Image dataset. The fuzzy rank of the base learners' predictions is used to make the final classification. Conducting five-fold cross-validation using the base learners an accuracy of 77.69 ± 3.22, 83.23 ± 3.14, 78.31 ± 2.27, and 78.62 ± 4.23 were obtained. Furthermore, using the proposed fuzzy-rank-based model, an accuracy of 85.23 ± 2.52 is obtained. We have proved that the proposed fuzzy-rank-based ensemble network increases the classification performance.

1. Introduction

Breast cancer (BC) is among women's most common cancers. Most of the cancer-related deaths occurring in women are because of breast cancer. Early detection of breast cancer is crucial as detecting the disease in its early stages leads to a higher likelihood of successful treatment and recovery. For this reason, developing a computer-aided diagnosis system to help detect breast cancer is important to improve survival rates. The CAD system developed by various researchers would use algorithms and computer technology to analyze medical images like mammograms, and Ultrasound images and potentially detect signs of breast cancer. And the entire algorithm would be automatic without the involvement of human being [1,2]. Due to this development of an automatic Computer-Aided Diagnosis (CAD) system for detecting cancer is essential. Initially, researchers have extensively used mammograms for the detection of Breast cancer. Zhang et al. [3] used a mini-MIAS [4] dataset for detecting BC. They have developed a nine-layer convolutional neural network and compared the model's performance with different activation functions. In their next task, they combined Graph Convolutional Network with CNN for the same task of BC detection [5]. Research has confirmed that mammograms fail to detect breast cancer in young women with dense breasts. Due to this, UltraSound Images (USI) has become an alternative to mammograms [6]. Ultrasound images (USI) have become an alternative to mammograms for several other reasons like, Safety: USI does not use ionizing radiation, which can potentially be harmful to the patient, Comfort: USI is a non-invasive procedure that does not involve any compression of the breast tissue, making it more comfortable for the patient, Effectiveness: USI is effective in detecting breast abnormalities, especially in women with dense breast tissue, which can make it difficult to detect cancer using mammograms, and Cost: USI is often less expensive than mammograms and other imaging methods [7–9].

Initially, researchers used various hand-crafted features for the classification of input USI images. Gomez et al. [10] extracted GLCM features from the input USI and SVM to classify the features into three categories. Local Binary Pattern (LBP) captures the texture description. Huang et al. [11] used LBP features from the input USI images.

Due to its excellent performance in image identification tasks, deep learning algorithms have recently drawn the interest of researchers [12,13]. Convolutional neural networks (CNNs) have been successfully used in Deep Learning to classify and recognize patterns in medical images [14]. DL has been used in the field of breast US for segmentation [15], locate [16], and estimate axillary lymph node size [17].

* Corresponding author.
E-mail addresses: sagar_1921ee20@iitp.ac (S.D. Deb), jharajib@iitp.ac.in (R.K. Jha).

https://doi.org/10.1016/j.bspc.2023.104871
Received 20 December 2022; Received in revised form 7 March 2023; Accepted 18 March 2023
Available online 23 March 2023

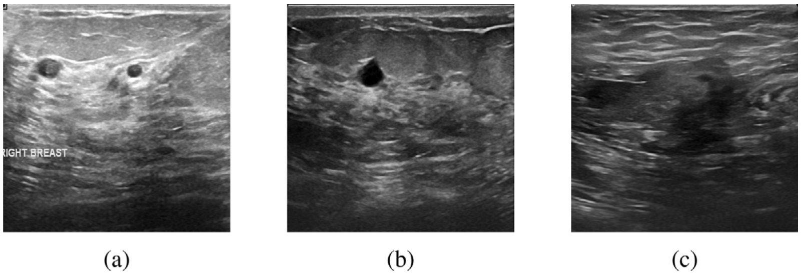


Fig. 1. A sample (a) Benign, (b) Malignant, and (c) Normal image from the Kaggle BUSI dataset [2].

Gu et al. [18] used VGGNet [19] for extracting high-level features from 14,043 Breast USI images that they collected from 32 different sources and classified them into three categories, namely Benign, Malignant, and Normal. [7] employed pre-trained DarkNet-53 for feature extraction. Various feature selection algorithm is applied before the final classification. Few researchers like [20] used segmentation steps ahead of classification to improve performance. The use of deep learning for categorizing breast masses from ultrasound images was explored by Byra et al. [21]. They introduced deep representation scaling layers to enhance the flow of information and utilized transfer learning techniques. The results showed that their method was far superior compared to other techniques. Irfan et al. [22] developed the Dilated Semantic Segmentation Network (Di-CNN) for diagnosing and classifying breast cancer. They used a deep model called DenseNet201 that was pre-trained with transfer learning for feature extraction and combined it with a 24-layered CNN to identify nodules. The results indicated that the fusion process improved recognition accuracy. Jabeen et al. [7] used a pre-trained Darknet model to extract features from ultrasound images for breast cancer detection. Despite the widespread use of transfer learning in deep learning, the authors believe that there is room for improvement. The proposed model is based on four pre-trained networks as base learners, each of which has its own strengths and weaknesses and can complement each other. The proposed fuzzy-rank based model can potentially enhance accuracy and robustness by combining the predictions from multiple base learners, instead of relying on a single model. In machine learning, it is a common practice to compare different models and select the best-performing one for a specific task and dataset. A publicly available dataset is used for training and validating the model.

The significant contributions of the manuscript are as follows.

- A fuzzy-rank-based deep ensemble network is proposed. The proposed model is based on four different base learners, namely VGG-Net, DenseNet, Xception, and Inception, and it tries to take advantage of the predictions made by the base learners. The weights of the initial layers of the base learners are pre-trained on the ImageNet dataset. In contrast, the final five layers are fine tunes using a publicly available Breast Ulta Sound Image dataset. As shown in Fig. 2, a fuzzy-rank-based fusion is done using the predictions for making the final classification.
- The proposed fuzzy-rank-based ensemble network is compared with the four base learners using various classification metrics like accuracy, precision, recall, and F1 score.
- Statistical significance tests of the proposed model with the base learners are done to prove that the proposed model is statistically significant.
- Grad CAM visualization of the input images is done to understand the working of the proposed fuzzy-rank-based ensemble model.

2. Dataset description

The training and validation of the fuzzy-rank-based ensemble model was carried out using data obtained from Kaggle, an open-source dataset [2]. The dataset, named the Breast UltraSound Image dataset, consists of 780 images that are separated into three categories: Benign, Malignant, and Normal. This dataset comprises 210 malignant images, 437 benign images, and 133 normal images, all of which were taken from women between 25 and 75 years old and are stored in PNG format with a resolution of 600 × 600. To conduct experiments, all images were resized to 224 × 224 × 3. The Fig. 1 shows samples from each of the three categories, with the first image showing a USI with a benign (non-cancerous) tumor, the second image depicting a USI with a malignant (cancerous) tumor, and the last image representing a USI with no visible tumor.

3. Methodology

This section briefly introduces our proposed fuzzy-rank-based ensemble network. The same schematic diagram is given in Fig. 2. The main idea behind proposing a fuzzy-ensemble-based breast cancer detector is to use multiple customized base learners, each generating a confidence score for the presence of breast cancer. Using multiple base learners provides more robust and accurate results than relying on a single classifier. The fuzzy-ensemble approach then combines these confidence scores using fuzzy set theory, taking into account the uncertainty and variability in the scores. The resulting fuzzy-ensemble score is then used to make a final diagnosis. This approach aims to improve breast cancer detection's accuracy and reliability by utilizing the strengths of multiple classifiers and incorporating uncertainty in the final decision. The weights of the initial layers of four base learners are frozen, whereas the later layers are fine-tuned on the Ultrasound dataset we have used. The brief details about the base learners used are as follows.

- VGG19 [19] The Visual Geometry Group network won the ImageNet classification challenge in 2013. They use smaller kernels 3 × 3 and 2 × 2 to learn the salient features.
- InceptionNet [23] A 22-layer deep learning network, won the 2014 edition of the ImageNet [24] LSVRC challenge. With the introduction of InceptionNet, the models could be more profound, and thus, the idea of Deep Learning was revolutionized. The number of parameters is almost 12 times lesser parameters than AlexNet [25].
- DenseNet One of the most recent advancements in neural networks for visual object detection is DenseNet [26]. ResNet and DenseNet are relatively similar. However, there are several key distinctions. While DenseNet concatenates the previous layer's output with the subsequent layer's output, ResNet utilizes an additive approach that kind of merges the previous layer with the subsequent layer.

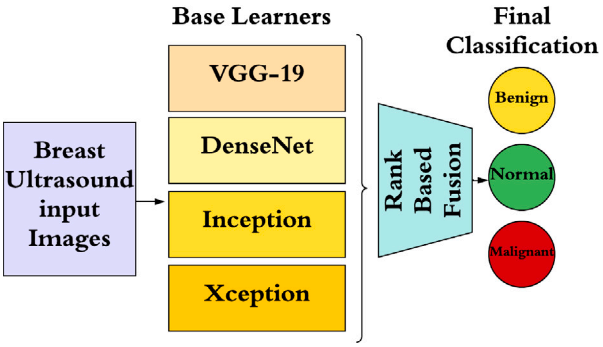


Fig. 2. Schematic diagram of the proposed model.

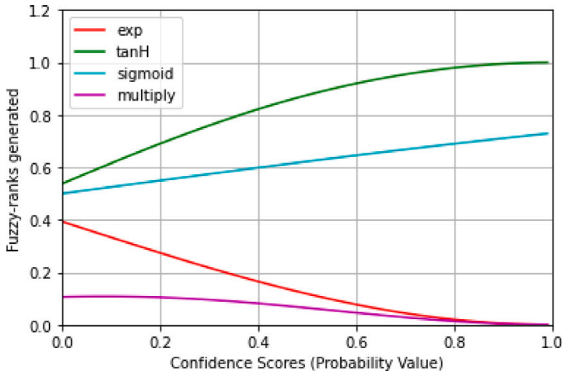


Fig. 3. The non-linear function used.

- Xception uses Depthwise separable convolutional layer instead of conventional convolutional layer. The depthwise separable convolution is a combination of point-wise and depth-wise convolution, which reduces the number of trainable parameters to a great extent [27].

3.0.1. *Cascade of the base learners*

Ensemble learning combines the power of prediction of the base learners [28,29]. The initial layers of the base learners are frozen as the dataset used for classification is not large enough for CNN to be trained from scratch. The final few layers of the base learners are fine-tuned on our dataset. Fine-tuning updates the weights of the final layer according to the dataset used. After the base learners, dense layers containing 1024, 512, 2048, and 102 neurons are attached. A dropout layer with a dropout score of 0.2 is added to avoid overfitting. To show the power of fuzzy rank-based ensemble learning, classification performance using the individual base learners is also done. As the dataset is relatively small so we have used 5-fold cross-validation. After training the base learners, the confidence scores are extracted as shown in Fig. 2. The confidence scores are later passed through a rank-based fusion block which contains three non-linear functions. The first non-linear function is a reward function, which captures how close the confidence score is to 1. Conversely, the second non-linear function calculates the deviation from 1, and finally, the last non-linear function is used for normalization. Details about the non-linear functions and their mathematical justification are given in next subsection.

3.0.2. *Proposed approach*

In the present study, we have designed a rank-based ensemble strategy to combine the four base learners' decision scores. As base learners, we have used four partially pre-trained Convolutional Neural Networks (CNN), namely VGG-Net, DenseNet, Xception, and Inception. These base classifiers have been pre-trained on the ImageNet dataset except for the last five layers. The last five layers of the base learners are fine-tuned on the BUSI dataset. The fusion approach employs a fuzzy ranking-based method where the probability scores from the base classifiers are transformed through three non-linear functions: an exponentially decaying function, the tanH function, and the Sigmoid function. The transformed scores are then used to assign ranks to the class probabilities. The exact process is repeated for each base classifier, and the rank products are added to obtain the final ranks. The aim of using three different non-linear functions with different concavities is to produce complementary results. The final decision is made by combining the multiple ranks associated with an identity and determining a new rank. The class with the lowest sum of rank products is considered the predicted class for the ensemble model. Using two ranks considers the closeness to and deviation from the expected result corresponding to the primary classification result, and the final rank normalizes the entire product. A higher confidence score results in a smaller value of the sum of rank products, indicating a better prediction. For better understanding the points are explained as follows -

- The confidence scores or the class belonging probability for every input breast USI given by the base learner i are $P_i{}^1$, P_i^2, $P_i^3, \dots, P_i{}^C$. Here i = 1,.., 4, as the proposed fuzzy-rank based ensemble model uses four base learners. In our case, we have three different classes (C) as the input USI image needs to be classified into one of the three different classes: Benign, Malignant, and Normal. Thus in our case, i = 4 and C = 3.
- $P_i{}^1$, P_i^2, $P_i^3, \dots, P_i{}^C$ are the confidence scores obtained from the ith base learner. As it represent probabilities, essentially it will follow

$$\sum_{k=1}^{C} p_k^i = 1, \forall i = 1, 2, 3 \tag{1}$$

- Let $(R_1{}^{i_1}, R_2{}^{i_1}, R_3{}^{i_1}, \dots, R_C{}^{i_1})$, $(R_1{}^{i_2}, R_2{}^{i_2}, R_3{}^{i_2}, \dots, R_C{}^{i_2})$ and $(R_1{}^{i_3}, R_2{}^{i_3}, R_3{}^{i_3}, \dots, R_C{}^{i_3})$ are the are fuzzy ranks generated by the three non-linear functions. The fuzzy ranks are generated when the confidence scores collected from the customized base learners are passed through the non-linear functions. The three non linear functions are shown in Fig. 3 and are given as below:

$$R^{i_1}{}_k = 1 - tanH(\frac{(p_k - 1)^2}{2}) \tag{2}$$

$$R^{i_2}{}_k = 1 - exp(-\frac{(p_k - 1)^2}{2}) \tag{3}$$

$$R^{i_3}{}_k = \frac{1}{1 + e^{-p_k}} \tag{4}$$

- The first non-linear function is given in Eq. (2). The tanH function, as shown in Fig. 3 is an increasing function. The function increases with increase in the confidence score or probability of class belonging. It is a rewarding function. More the confidence score approaches 1, more the reward. The range of the function is [0.5378828427399902 0.9999500000000416]
  The second non-linear function, as given in Eq. (3) and shown in Fig. 3 is a decreasing function. It captures the deviation from 1. The range of the function is [4.999875002087428e−05 0.3934693402873666e−01]
  The third non-linear function is a *sigmoid function* and it is used to normalize the over all function.
- Let $RS_1{}^i$, $RS_2{}^i$, $RS_3{}^i, \dots, RS_C{}^i$ be the fused rank scores, where $RS_k{}^i$ is given by following equation,

$$RS^i_k = R_1{}^{i_k} \times R_1{}^{i_k} \times R_1{}^{i_k} \tag{5}$$

  The product of the three non-linear function is also given in Fig. 3. This function is a decreasing function and it decreases with increase in probability.

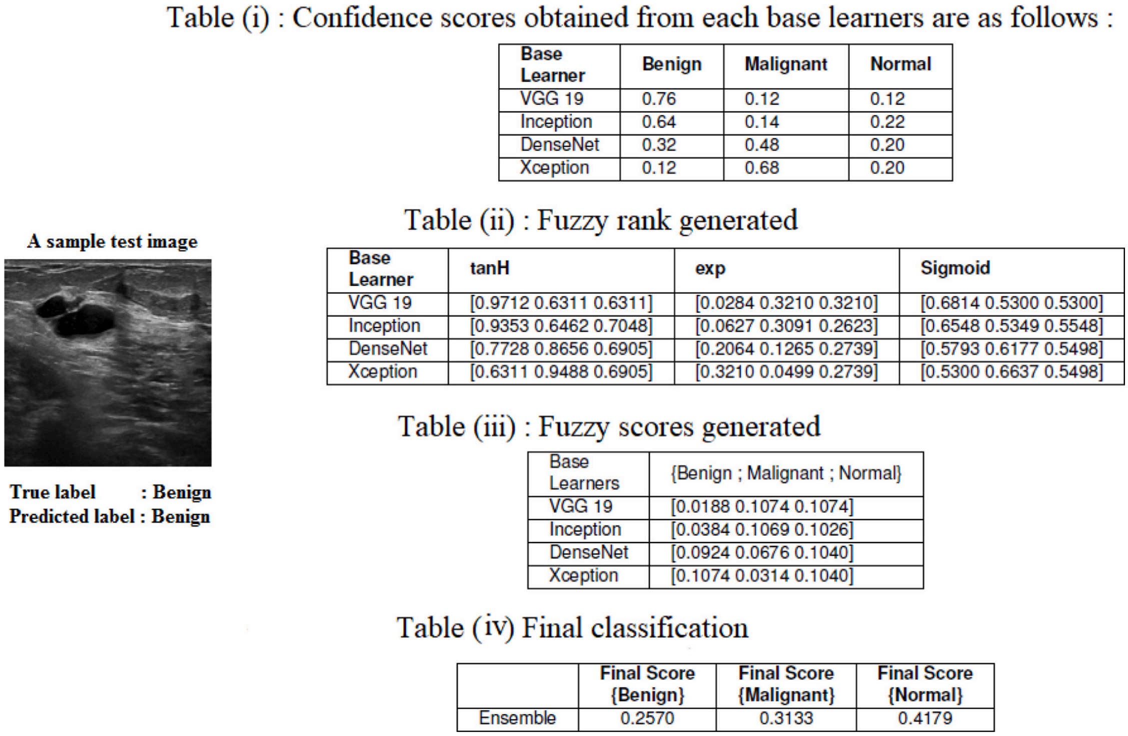


Table (i) : Confidence scores obtained from each base learners are as follows :

Base Learner	Benign	Malignant	Normal
VGG 19	0.76	0.12	0.12
Inception	0.64	0.14	0.22
DenseNet	0.32	0.48	0.20
Xception	0.12	0.68	0.20

A sample test image

True label : Benign
Predicted label : Benign

Table (ii) : Fuzzy rank generated

Base Learner	tanH	exp	Sigmoid
VGG 19	[0.9712 0.6311 0.6311]	[0.0284 0.3210 0.3210]	[0.6814 0.5300 0.5300]
Inception	[0.9353 0.6462 0.7048]	[0.0627 0.3091 0.2623]	[0.6548 0.5349 0.5548]
DenseNet	[0.7728 0.8656 0.6905]	[0.2064 0.1265 0.2739]	[0.5793 0.6177 0.5498]
Xception	[0.6311 0.9488 0.6905]	[0.3210 0.0499 0.2739]	[0.5300 0.6637 0.5498]

Table (iii) : Fuzzy scores generated

Base Learners	{Benign ; Malignant ; Normal}
VGG 19	[0.0188 0.1074 0.1074]
Inception	[0.0384 0.1069 0.1026]
DenseNet	[0.0924 0.0676 0.1040]
Xception	[0.1074 0.0314 0.1040]

Table (iv) Final classification

	Final Score {Benign}	Final Score {Malignant}	Final Score {Normal}
Ensemble	0.2570	0.3133	0.4179

Fig. 4. The figure shows all the necessary calculations that are to be carried out for the proposed Rank Based Ensemble Classifier.

- Now the fused score tuple is $(FS_1, FS_2, FS_3, \dots, FS_C)$, where FS_k is given by following equation.

$$FS_k = \sum_{i=1}^{L} RS_k{}^i, \forall k = 1, 2, \dots, C \tag{6}$$

- This fused score can be viewed as the final score for each class. The class with the smallest fused score is determined to be the winning class.

$$class(I) = \min_{\forall k} FS_k \tag{7}$$

Example of how the algorithm works is given in Fig. 4.

4. Implementation details

The experiments were conducted on Google Collab, equipped with a Tesla K80 graphics card. As explained above, the proposed fuzzy rank-based ensemble learner is based on four base learners: VGG-Net, DenseNet, Inception Net, and Xception Network. All the base learners are trained on the ImageNet dataset. The final five layers are fine-tuned using the publically available Breast UltraSound Image dataset for our task of classifying the input USI images into three classes, namely Benign, Malignant, and Normal.

The customized classification segment is attached after all the base learners. The customized classification segment is three layers deep. The number of neurons in the input layer of the classification segment depends upon the base learner, after which the classification segment is attached.

For VGG-Net and DenseNet, the classification segment attached has 1024 neurons in the input layer, followed by a dropout layer with a dropout rate of 0.5. The second layer of the customized classification segment has 256 neurons, followed by a ReLU activation function. The final layer has three neurons corresponding to the three classes used in our task. The final layer of the classification segment has a sigmoid activation function.

For Xception, the number of neurons in the customized classification segment is 256, and for Inception, the number is 1028. The first base learner contains a total of 28,545,603 parameters, and the model weights are updated from the layer named *block5_conv1*. For the second base learner, the total number of parameters is about 40,308,467, and the model is fine-tuned from the layer named *Conv5_block32_1_conv*. Finally Inception and Xception networks contain 30,291,239 and 29,258,667 parameters respectively. Inception network is trained from the layer named *conv2d_485*, and finally, Xception is trained from the layer named *block14_SepConv1*. We have used the RMSprop optimizer for conducting all our experiments. RMSprop is almost similar to the gradient descent algorithm with momentum [30]. Batch size and learning rate of 32 and 10^{-2} are used.

As we are performing multi-class classification, so the Categorical Cross Entropy loss function is used. Early stopping is used to stop training when the monitored metric (validation loss) has stopped improving. The patience of early stopping is considered equal to 3 for our case. So basically, if the validation loss does not decrease further for three consecutive epochs, the model would stop training.

As the BUSI dataset does not have a train, test, or validation set, we have performed a 5-fold cross-validation. The results section contains further details.

5. Results and discussions

This section reports the results obtained by the base learners and the proposed fuzzy-rank-based ensemble model. The fuzzy-rank-based ensemble model takes an UltraSound Breast Image as input and classifies it into the Benign, Malignant, or Normal class. As explained above, the proposed model has four base learners, which produces confidence scores. We have used three different non-linear functions to generate the fuzzy ranks from those confidence scores. The first non-linear

Table 1
Performance of the proposed model on the Breast Ulta Sound Image dataset.

Fold	Learners	Class	Precision	Recall	F1 score	Support	Accuracy
1	Base learner 1 (VGG-19)	Benign	0.83	0.78	0.80	73	74.62
		Normal	0.75	0.55	0.63	22	
		Malignant	0.62	0.80	0.70	35	
	Base learner 2 DenseNet	Benign	0.92	0.90	0.91	73	86.92
		Normal	0.75	0.95	0.84	22	
		Malignant	0.87	0.74	0.80	35	
	Base learner 3 (InceptionNet)	Benign	0.81	0.92	0.86	73	80.00
		Normal	0.73	0.73	0.73	22	
		Malignant	0.84	0.60	0.70	35	
	Base learner 4 (Xception)	Benign	0.86	0.88	0.87	73	78.46
		Normal	0.59	0.73	0.65	22	
		Malignant	0.76	0.63	0.69	35	
	Ensemble	Benign	0.93	0.93	0.93	73	88.46
		Normal	0.73	0.86	0.79	22	
		Malignant	0.90	0.80	0.85	35	
2	Base learner 1 (VGG-19)	Benign	0.77	0.88	0.82	73	76.15
		Normal	0.81	0.59	0.68	22	
		Malignant	0.71	0.63	0.67	35	
	Base learner 2 DenseNet	Benign	0.80	0.95	0.87	73	82.31
		Normal	0.89	0.73	0.80	22	
		Malignant	0.85	0.63	0.72	35	
	Base learner 3 (InceptionNet)	Benign	0.76	0.93	0.84	73	80.00
		Normal	0.90	0.82	0.86	22	
		Malignant	0.86	0.51	0.64	35	
	Base learner 4 (Xception)	Benign	0.81	0.81	0.81	73	75.38
		Normal	0.68	0.59	0.63	22	
		Malignant	0.68	0.74	0.71	35	
	Ensemble	Benign	0.81	0.96	0.88	73	84.61
		Normal	1.00	0.73	0.84	22	
		Malignant	0.86	0.69	0.76	35	
3	Base learner 1 (VGG-19)	Benign	0.80	0.90	0.85	73	80.77
		Normal	0.84	0.73	0.78	22	
		Malignant	0.82	0.66	0.73	35	
	Base learner 2 DenseNet	Benign	0.83	0.92	0.87	73	84.61
		Normal	0.94	0.73	0.82	22	
		Malignant	0.84	0.77	0.81	35	
	Base learner 3 (InceptionNet)	Benign	0.75	0.93	0.83	73	77.69
		Normal	1.00	0.55	0.71	22	
		Malignant	0.78	0.60	0.68	35	
	Base learner 4 (Xception)	Benign	0.81	0.89	0.85	73	80.77
		Normal	0.76	0.73	0.74	22	
		Malignant	0.83	0.69	0.75	35	
	Ensemble	Benign	0.88	0.88	0.88	73	86.15
		Normal	0.90	0.86	0.88	22	
		Malignant	0.81	0.83	0.82	35	
4	Base learner 1 (VGG-19)	Benign	0.84	0.86	0.85	73	81.54
		Normal	0.82	0.64	0.72	22	
		Malignant	0.76	0.83	0.79	35	
	Base learner 2 DenseNet	Benign	0.86	0.90	0.88	73	83.85
		Normal	0.73	0.86	0.79	22	
		Malignant	0.89	0.69	0.77	35	
	Base learner 3 (InceptionNet)	Benign	0.77	0.93	0.84	73	79.23
		Normal	0.88	0.64	0.74	22	
		Malignant	0.81	0.60	0.69	35	
	Base learner 4 (Xception)	Benign	0.85	0.95	0.90	73	84.62
		Normal	0.85	0.77	0.81	22	
		Malignant	0.83	0.69	0.75	35	
	Ensemble	Benign	0.85	0.93	0.89	73	85.38
		Normal	0.78	0.82	0.80	22	
		Malignant	0.93	0.71	0.81	35	

(*continued on next page*)

function, as given in Eq. (2), is rewarding. It takes a confidence score as input and rewards values closer to 1. The second non-linear function calculates the deviation from 1, and the final sigmoid function, as given in Eq. (4), is for normalization. Fig. 3 shows the three non-linear functions used to generate the fuzzy scores.

The table (i) of Fig. 4 shows the confidence scores generated by the four base learners for a test USI. According to the first base learner (VGG-Net), the test USI belongs to benign, malignant, and normal classes with a probability of 0.76, 0.12, and 0.12, respectively. Similarly, according to the second base learner (DenseNet), the test USI

Table 1 (*continued*).

Fold	Learners	Class	Precision	Recall	F1 score	Support	Accuracy
5	Base learner 1 (VGG-19)	Benign	0.82	0.75	0.78	72	75.38
		Normal	0.62	0.78	0.69	23	
		Malignant	0.74	0.74	0.74	35	
	Base learner 2 DenseNet	Benign	0.81	0.85	0.83	72	78.46
		Normal	0.70	0.83	0.76	23	
		Malignant	0.79	0.63	0.70	35	
	Base learner 3 (InceptionNet)	Benign	0.71	0.94	0.81	72	74.62
		Normal	0.78	0.61	0.68	23	
		Malignant	0.94	0.43	0.59	35	
	Base learner 4 (Xception)	Benign	0.76	0.81	0.78	72	73.85
		Normal	0.67	0.61	0.64	23	
		Malignant	0.73	0.69	0.71	35	
	Ensemble	Benign	0.80	0.92	0.85	72	81.53
		Normal	0.77	0.74	0.76	23	
		Malignant	0.92	0.66	0.77	35	

belongs to benign, malignant, and normal classes with a probability of 0.64, 0.14, and 0.22, respectively. The fuzzy ranks generated by the three non-linear functions are given in the second table of Fig. 4.

The fuzzy score as shown in the table (iii) of the same figure, is given by the following equation.

$$RS_k^i = {R_c}^{i_1} \times {R_c}^{i_2} \times {R_c}^{i_3} \tag{8}$$

where, ${R_c}^{i_1}$,is the fuzzy rank generated for a particular class C using a particular base learner i. In our case, i = 1,2,3,4 as we have used four base learners and C = 1,2,3 as the dataset used has three different class.

The fused tuple score, which is nothing but the summation of all the fuzzy scores over all the base learners, is given in the table (iv) of Fig. 4. The test sample image belongs to the class having the lowest fused tuple score as the nature of the curve, which multiplies the three non-linear functions, is decreasing, as shown in Fig. 3.

It is worth mentioning that the individual base learners classified the sample test image into different classes. As shown in the table (i) of Fig. 4, the first base learner classified the sample test image into a benign category as the confidence score is highest for the benign class. Similarly, the second, third and final base learners classified the sample image into benign, malignant, and malignant categories, respectively. Moreover, finally, according to the fuzzy-rank-based ensemble classifier, the test sample belongs to the benign category. The true level of the sample test image is Benign.

5.1. Comparison with the individual models

To prove the superiority of the proposed fuzzy-rank-based ensemble model, it is essential to compare it with the performance of the individual base learners. As we conduct three class classifications, the performance matrices used for comparison are accuracy and class-wise Precision, Recall, and F1 score. The equations for the same are given below. Provided the number of true positives (T.P), false positives (F.P), true negatives (T.N), and false negatives (F.N), these measures are mathematically expressed as follows:

$$Precision = \frac{T.P}{T.P + F.P} \tag{9}$$

$$Recall = \frac{T.P}{T.P + F.N} \tag{10}$$

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{11}$$

$$Accuracy = \frac{T.P + T.N}{T.P + T.N + F.P + F.N} \tag{12}$$

As mentioned in the section Dataset description, the number of images available is less. Again, the dataset does not specify any separate train, test, or validation set, so we have conducted a 5-fold cross-validation. The performance comparison with the base learners and the class-wise precision, recall, and F1 score over different folds are given in the Table 1.

For conducting a 5-fold cross-validation, the entire dataset is divided into five parts. During the first fold, the first four-fold of the dataset was used for training, and the final fold was used for testing all the base learners and the proposed fuzzy-rank-based ensemble model. A total of 130 images (benign (73), normal (22), and malignant (35)) were used for testing. Precision finds out what fraction of predicted positives is actually positive. The precision score of 0.83, 0.92, 0.81, and 0.86 for the benign class was obtained using the four base learners as shown in the Table 1. The fuzzy-rank-based ensemble classifier's precision score for the benign class of 0.93 is obtained. As shown in the Table 1, in the first fold, an accuracy of 74.62%, 86.92%, 80.00%, and 78.46% is obtained using the four base learners. In contrast, the proposed fuzzy-rank-based ensemble model achieved an accuracy of 88.46%. The performance of the proposed model in terms of accuracy increased by about 18.54%, 1.77%, 10.57%, and 12.74%, respectively. The training progress is shown in Fig. 5. Early stopping is used for training the base models. As shown in Fig. 5(a), the VGG-Net stops at the 40th epoch. The second base learner (DenseNet) saturates at the 32nd epoch. The third (Inception) and the final (Xception) saturate at the 38th and 42nd epochs.

The confusion matrix using the four base learners and the proposed fuzzy-rank-based ensemble model is shown in Fig. 6.

Finally, the accuracy obtained using the four base learners in terms of mean and standard deviation is given as 77.69 ± 3.22, 83.23 ± 3.14, 78.31 ± 2.27, and 78.62 ± 4.23. Furthermore, using the proposed fuzzy-rank-based model, an accuracy of 85.23 ± 2.52 is obtained.

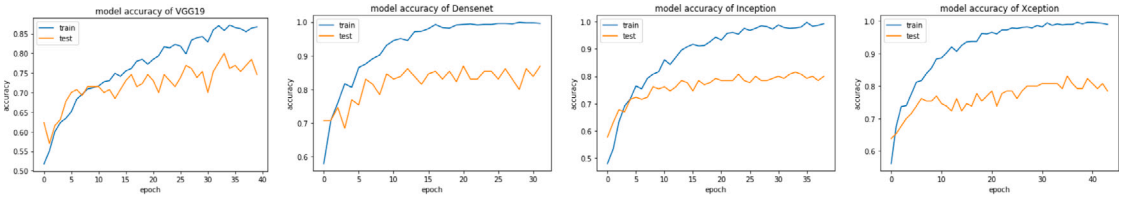
The grad-CAM [35] visualization is given in Fig. 7. The first column of the figure shows a sample benign and a malignant image, the second column shows the heat map generated, and the third column shows the superimposed image.

5.2. Comparison with other ensemble techniques

This section presents the comparison of our proposed fuzzy-rank-based ensemble model with other ensemble techniques present in the literature. The accuracy of 83.66 ± 5.18 is obtained by conducting five-fold cross-validation using the majority voting ensemble technique. Weighted majority voting is another ensemble technique, and its performance is believed to be better than majority voting [36]. Conducting 5-fold cross-validation using weighted average majority voting, an accuracy of 84.87 ± 3.21 is reported. Furthermore, simply concatenating features extracted from the base learners and classifying them into three categories: Benign, Malignant, and normal class accuracy of 82.61±5.23.

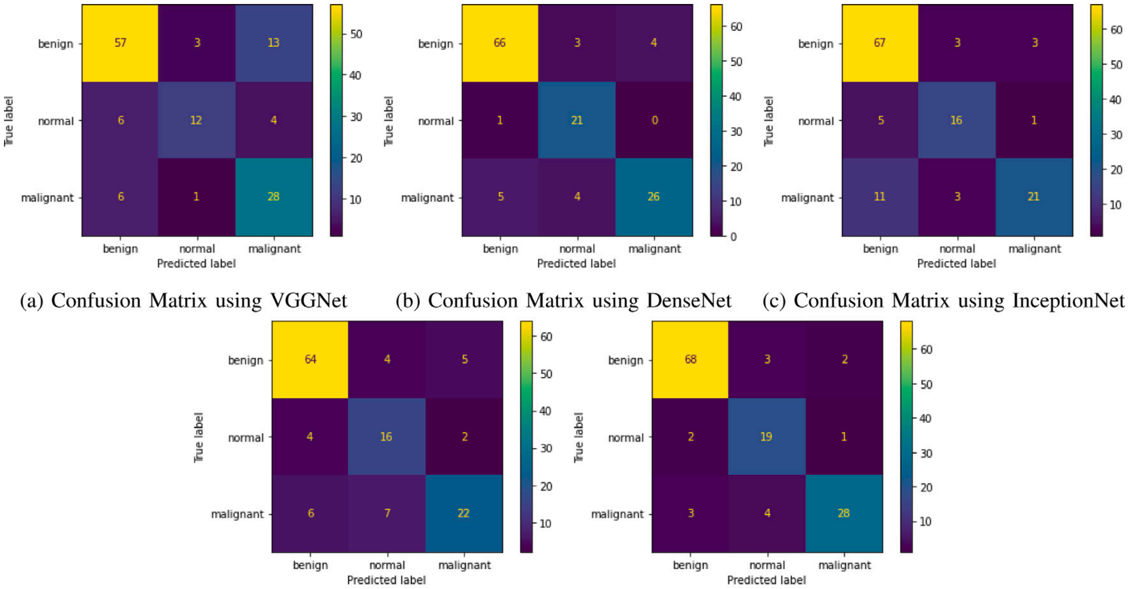
Table 2
Comparison with the state-of-the-art.

Studies	Algorithm used	Dataset used	Performance
Gu et al. [18]	Pre-trained VGG	Private	Accuracy = 86.40
Dhabyani et al. [2]	Pre-trained ResNet	BUSI	Accuracy = 82
Tanaka et al. [1]	Ensemble of VGG-Net and Res-Net	BUSI	AUC = 0.951
Ghefalti et al. [31]	Vision transformer	BUSI	Accuracy = 85.3
Hanet et al. [32]	Google-Net based model	Private	Accuracy = 90
Qi et al. [33]	Mt-Net	Private	Accuracy = 87.39
Xiao et al. [34]	Inception V3	BUSI	Accuracy = 85.13%
Proposed	Fuzzy ensemble based model	BUSI	Accuracy = 85.23 ± 2.52



(a) (b) (c) (d)

Fig. 5. Training progress of the four base learners used for developing the fuzzy-rank based ensemble learner (First Fold) (a) VGG-Net, (b) DenseNet, (c) Inception, and (d) Xception.



(a) Confusion Matrix using VGGNet (b) Confusion Matrix using DenseNet (c) Confusion Matrix using InceptionNet

(d) Confusion Matrix using XceptionNet (e) Confusion Matrix using Ensemble

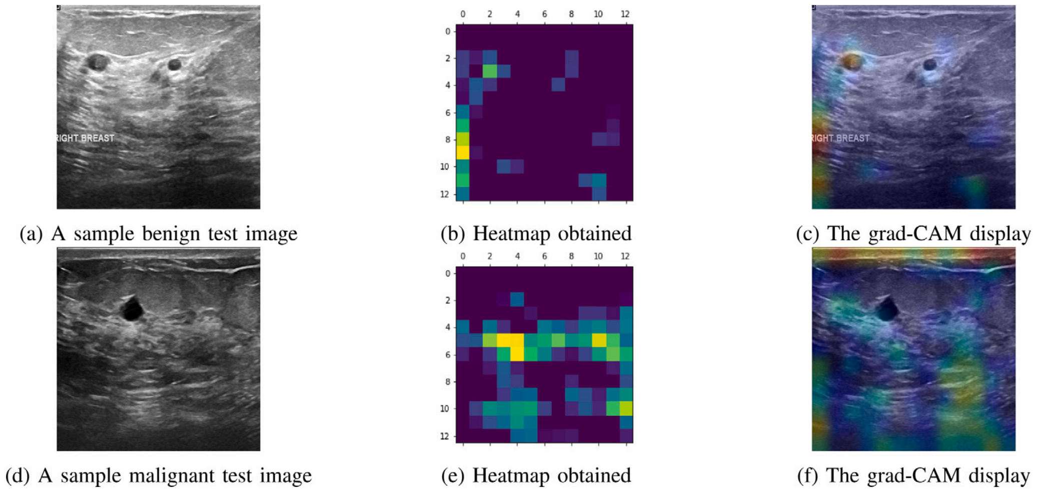
Fig. 6. Confusion matrix over various base learners and the ensemble model (First Fold)

5.3. Comparison with other models in the literature

Gu et al. [18] collected 14,043 ultrasound images from almost 32 different hospitals located in different parts of China and trained a VGGNet to classify the images into three categories of benign, malignant, and normal classes. A total of 5012 women took part in the study. Despite such a large dataset, their model could achieve an accuracy of 86.40%. [37] used a publicly available BUSI dataset for training a Deep Neural Network to classify the images into three categories as the dataset is not that large, so they have used pre-trained ResNet architecture for extracting features from the input images. They have achieved an accuracy of 82%.

Similarly, Tanaka et al. [1] also conducted the experiments using the BUSI image dataset. They have extracted high-level features using an ensemble of VGG-net and ResNet from the input USI images. Conducting a binary classification, an AUC of 0.951 is reported. More recently, using the BUSI dataset, Gheflati et al. [31] used Vision Transformer for Breast UltraSound Images. They have reported an accuracy of 85.3%.

Han et al. [32] collected ultrasound images from 5151 patients. A GoogLeNet-based model was trained from scratch using 7408 breast images. They have achieved an accuracy of almost 90% for binary classification. Similarly, Qi et al. [33] collected breast ultrasound images from 2047 patients. They have collected almost 8200 images and developed an automatic breast image classifier. Their proposed Mt-Net (BASIC) achieved a two-class accuracy of 93.52%. The proposed fuzzy-rank-based model, which combines the prediction made by the base learners, achieved an accuracy of 85.23 ± 2.52. Further details are provided in Table 2.



(a) A sample benign test image (b) Heatmap obtained (c) The grad-CAM display

(d) A sample malignant test image (e) Heatmap obtained (f) The grad-CAM display

Fig. 7. Grad-CAM [35] visualization on two sample images.

5.4. Statistical significance test

To prove the robustness of the proposed fuzzy-ensemble-based ensemble network, we have performed a t-test. The t-test is selected as it is applicable in the case of small datasets like ours, and it follows students' t-distribution. We have used the performance metric accuracy of the four different base learners and the proposed ensemble model on the BUSI image dataset. A p-test and t-test value of −4.21 and 0.00 is obtained when the first base learner is compared with the proposed model. Similarly, a p-test and t-test value of −1.26 and 0.279 is obtained compared to the second base learner. Finally, t-test values of −4.62 and −2.96 and p-test value of 0.00 and 0.01 is obtained when compared with the third and fourth base learner. The differences in performance between the proposed method and existing approaches are statistically significant, which implies that the results are statistically significant.

6. Conclusions

A fuzzy-rank-based ensemble network is proposed for breast cancer detection from UltraSound Images. The input USI images are classified into three classes: benign, malignant, and normal. The proposed network contains four base learners. The weights of the initial layers of the base learners are pre-trained on the ImageNet dataset, whereas the final five layers are fine tunes using our target dataset. Conducting five-fold cross-validation an accuracy of 77.69 ± 3.22, 83.23 ± 3.14, 78.31 ± 2.27, and 78.62 ± 4.23 were obtained. Furthermore, using the proposed fuzzy-rank-based model, an accuracy of 85.23 ± 2.52 is obtained. Grad-Cam visualization is shown to understand the working of the proposed model. A statistical significance test is also conducted to prove the robustness of the model. In the future, we aim to incorporate segmentation information of the USI before the final classification.

CRediT authorship contribution statement

Sagar Deep Deb: Conceptualization, Methodology, Investigation, Writing – original draft. **Rajib Kumar Jha:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request

References

[1] H. Tanaka, S.-W. Chiu, T. Watanabe, S. Kaoku, T. Yamaguchi, Computer-aided diagnosis system for breast ultrasound images using deep learning, Phys. Med. Biol. 64 (23) (2019) 235013.

[2] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images. Data brief 28: 104863, 2019.

[3] Y.-D. Zhang, C. Pan, X. Chen, F. Wang, Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling, J. Comput. Sci. 27 (2018) 57–68.

[4] J. Suckling, The mammographic images analysis society digital mammogram database, in: Exerpta Medica. International Congress Series, 1994, Vol. 1069, 1994, pp. 375–378.

[5] Y.-D. Zhang, S.C. Satapathy, D.S. Guttery, J.M. Górriz, S.-H. Wang, Improved breast cancer classification through combining graph convolutional network and convolutional neural network, Inf. Process. Manage. 58 (2) (2021) 102439.

[6] H.-D. Cheng, X. Shi, R. Min, L. Hu, X. Cai, H. Du, Approaches for automated detection and classification of masses in mammograms, Pattern Recognit. 39 (4) (2006) 646–668.

[7] K. Jabeen, M.A. Khan, M. Alhaisoni, U. Tariq, Y.-D. Zhang, A. Hamza, A. Mickus, R. Damaševičius, Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion, Sensors 22 (3) (2022) 807.

[8] S. Pavithra, R. Vanithamani, J. Justin, Computer aided breast cancer detection using ultrasound images, Mater. Today: Proc. 33 (2020) 4802–4807.

[9] R. Guo, G. Lu, B. Qin, B. Fei, Ultrasound imaging technologies for breast cancer detection and management: a review, Ultrasound Med. Biol. 44 (1) (2018) 37–70.

[10] W. Gómez, W.C.A. Pereira, A.F.C. Infantosi, Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound, IEEE Trans. Med. Imaging 31 (10) (2012) 1889–1899.

[11] Q. Huang, F. Zhang, X. Li, Machine learning in ultrasound computer-aided diagnostic systems: a survey, BioMed. Res. Int. 2018 (2018).

[12] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (7553) (2015) 436–444.

[13] S.D. Deb, R.K. Jha, Modified double U-net architecture for medical image segmentation, IEEE Trans. Radiat. Plasma Med. Sci. (2022).

[14] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. Van Der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.

[15] Y. Hu, Y. Guo, Y. Wang, J. Yu, J. Li, S. Zhou, C. Chang, Automatic tumor segmentation in breast ultrasound images using a dilated fully convolutional network combined with an active contour model, Med. Phys. 46 (1) (2019) 215–228.

[16] Z. Cao, L. Duan, G. Yang, T. Yue, Q. Chen, An experimental study on breast lesion detection and classification from ultrasound images using deep learning architectures, BMC Med. Imaging 19 (1) (2019) 1–9.

[17] M.H. Yap, G. Pons, J. Marti, S. Ganau, M. Sentis, R. Zwiggelaar, A.K. Davison, R. Marti, Automated breast ultrasound lesions detection using convolutional neural networks, IEEE J. Biomed. Health Inf. 22 (4) (2017) 1218–1226.

[18] Y. Gu, W. Xu, B. Lin, X. An, J. Tian, H. Ran, W. Ren, C. Chang, J. Yuan, C. Kang, et al., Deep learning based on ultrasound images assists breast lesion diagnosis in China: a multicenter diagnostic study, Insights Imaging 13 (1) (2022) 1–14.

[19] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, arXiv preprint arXiv:1409.1556.

[20] M. Muhammad, D. Zeebaree, A.M.A. Brifcani, J. Saeed, D.A. Zebari, Region of interest segmentation based on clustering techniques for breast cancer ultrasound images: A review, J. Appl. Sci. Technol. Trends 1 (3) (2020) 78–91.

[21] M. Byra, Breast mass classification with transfer learning based on scaling of deep representations, Biomed. Signal Process. Control 69 (2021) 102828.

[22] R. Irfan, A.A. Almazroi, H.T. Rauf, R. Damaševičius, E.A. Nasr, A.E. Abdelgawad, Dilated semantic segmentation for breast ultrasonic lesion detection using parallel feature fusion, Diagnostics 11 (7) (2021) 1212.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[24] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 248–255.

[25] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.

[26] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.

[27] F. Chollet, Xception: Deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251–1258.

[28] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, R. Sarkar, A fuzzy rank-based ensemble of CNN models for classification of cervical cytology, Sci. Rep. 11 (1) (2021) 14538.

[29] S.D. Deb, R.K. Jha, K. Jha, P.S. Tripathi, A multi model ensemble based deep convolution neural network structure for detection of COVID19, Biomed. Signal Process. Control 71 (2022) 103126.

[30] F. Zou, L. Shen, Z. Jie, W. Zhang, W. Liu, A sufficient condition for convergences of adam and rmsprop, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 11127–11135.

[31] B. Gheflati, H. Rivaz, Vision transformers for classification of breast ultrasound images, in: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society, EMBC, IEEE, 2022, pp. 480–483.

[32] S. Han, H.-K. Kang, J.-Y. Jeong, M.-H. Park, W. Kim, W.-C. Bang, Y.-K. Seong, A deep learning framework for supporting the classification of breast lesions in ultrasound images, Phys. Med. Biol. 62 (19) (2017) 7714.

[33] X. Qi, L. Zhang, Y. Chen, Y. Pi, Y. Chen, Q. Lv, Z. Yi, Automated diagnosis of breast ultrasonography images using deep neural networks, Med. Image Anal. 52 (2019) 185–198.

[34] T. Xiao, L. Liu, K. Li, W. Qin, S. Yu, Z. Li, Comparison of transferred deep neural networks in ultrasonic breast masses discrimination, BioMed. Res. Int. 2018 (2018).

[35] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[36] A. Dogan, D. Birant, A weighted majority voting ensemble approach for classification, in: 2019 4th International Conference on Computer Science and Engineering, UBMK, IEEE, 2019, pp. 1–6.

[37] W. Al-Dhabyani, M. Gomaa, H. Khaled, F. Aly, Deep learning approaches for data augmentation and classification of breast masses using ultrasound images, Int. J. Adv. Comput. Sci. Appl 10 (5) (2019) 1–11.