JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 40, 919-939 (2024)
DOI: 10.6688/JISE.202407_40(4).0014

Transfer Learning with Fuzzy for Breast Cancer*

PINAR USKANER HEPSAĞ[1,+], SELMA AYSE ÖZEL[2]
AND ADNAN YAZICI[3]
[1]*Department of Computer Engineering*
Adana Alparslan Türkeş Science and Technology University
Adana, 01250 Turkey
[2]*Department of Computer Engineering*
Çukurova University
Adana, 01330 Turkey
[3]*Department of Computer Science*
Nazarbayev University
Nur Sultan, 010000 Kazakhstan
E-mail: puskaner@atu.edu.tr[+]; saozel@cu.edu.tr; adnan.yazici@nu.edu.kz

Deep learning methods have been used to reduce the number of unnecessary breast biopsies. In this study, an accurate hybrid rule-based fuzzy system with transfer learning is developed to classify breast abnormalities as malignant or benign by calculating breast cancer risk from digital mammogram images. Our system consists of three phases: (i) data augmentation methods (*e.g.*, traditional methods, Generative Adversarial Networks (GANs)); (ii) classification of the breast abnormalities on the BCDR-D02 and mini-MIAS databases by fine-tuning transfer learning methods with the deep learning base model Convolutional Neural Network (CNN); and (iii) calculation of breast cancer risk with a rule-based fuzzy system using the results of the second phase to improve the classification of breast abnormality results. Using our CNN baseline model and traditional extension methods, we achieve 64% and 82% accuracy for mini-MIAS and BCDR-D02, respectively. With fine-tuning the transfer learning methods, we obtain 80% and 83% with VGG-16 for mini- MIAS and BCDR-D02, respectively. Using the rule-based fuzzy system, called the risk method, we achieve the highest results for mini-MIAS (93%) and BCDR-D02 (94%). The classification results of our risk method are compared with the other transfer learning and baseline methods, and it is found that the accuracy of breast abnormality classification is improved by using a hybrid rule-based fuzzy system with transfer learning. Our study can serve as a guide that provides useful tips to researchers in the field of breast cancer classification to develop more effective and reliable studies.

Keywords: transfer learning, GANs, breast cancer, mammogram, fuzzy

1. INTRODUCTION

Breast cancer (BC) is the most common cancer in women among the various types of cancer. It occurs when breast cells begin to grow uncontrollably. These cells usually

Received May 10, 2022; revised December 15, 2022 & August 8, 2023; accepted October 19, 2023.
Communicated by Jen-Hui Chuang.
[+] Corresponding author.
[*] This study was funded by Scientific Research Project Unit of Çukurova University (grant number FDK-2016-6931); and Nazarbayev University (Kazakhstan) Faculty-development competitive research (grant number FY2019-FGP-1-STEMM).

form a tumor can be seen on a mammogram or ultrasound, or felt as a lump. BC is divided into two categories: benign and malignant. Benign is known as non-cancerous, while malignant is considered as cancerous [1]. BC has become one of the most important medical issues due to the increasing number of diseases worldwide [2].

To reduce the mortality rate of BC, early detection is very important and requires an accurate diagnostic procedure that allows experts to separate malignants from benigns [3]. Ultrasound and mammography are two of the most common radiological methods to detect BC at an early stage [4]. However, accurate interpretation of mammograms by radiologists is difficult, leading to a high number of false positives and additional investigations. Because of the high sensitivity but low specificity of mammograms, approximately $20-30\%$ breast biopsies are identified as malignant [5, 6]. Improving mammography findings to accurately detect BC is very important. Because a high number of false-positives have several negative consequences, such as anxiety, pain, unnecessary surgery, and health care costs.

The mammography method is commonly used by experts to diagnose BC accurately [7]. Therefore, physicians are increasingly using computer-aided diagnostic methods to interpret mammograms and avoid unnecessary biopsies. Various methods of computer-aided analysis are used including machine learning algorithms [8].

With the increasing success of deep learning (DL) methods in computer vision tasks, CNNs have been applied to many image classification problems, especially in medical [9]. As a result, they have become very popular in the diagnosis of BC [10]. DL methods require a large dataset to learn better during training. However, datasets for medical images are of limited size. Therefore, applying CNN method to solve medical image problems is a challenging task. TL is the most common technique to solve the problem of limited training data. Researchers have used various pre-trained TL models to analyze medical image data [11–14]. The most commonly used TL methods are ResNet [11, 15], AlexNet [12,16], VGG16 [15,17],and Google InceptionV3 [13,15,16] on INbreast, CBIS-DDSM, *etc*. We have used TL with ResNet50 [18] and VGG16 [19] models and obtained higher accuracy compared to other models in classifying mammograms.

The other problem in detecting BC is imbalance data that effects the training quality. Therefore, data augmentation methods have been applied to increase dataset size to get same number of samples in two classes. While traditional data augmentation methods, *i.e*., affine transformations such as flipping or resizing, have been applied to increase the size of the dataset, GANs have become a popular augmentation method due to their success in generating synthetic data. Various researchers in the literature show that GANs have been applied to medical data from different problem domains to synthetically generate new training examples [20–23]. In [24], GANs are used to deal with poor data in liver lesion detection. Similarly, GANs were used to increase the dataset in histopathology [25], retinal fundi [26], and chest radiographs [27]. In our study, we augmented the mini-MIAS and BCDR-D02 datasets for classification of mammograms using GANs and generated 1000 additional images for malignant and benign classes.

Fuzzy systems [28] contain input and output to indicate the degree of values between 1 (yes) and 0 (no). In fuzzy, the output takes values such as possibly yes and possibly no as well as yes or no according to the input. Fuzzy logic is very popular as it gives more realistic results, especially in medicine. Since BC involves some uncertainties, fuzzy logic is an effective solution to get a more realistic term for uncertainties. In [29], a hybrid

neuro-fuzzy decision support system for diagnosing myocardial perfusion from cardiac images was implemented. To our knowledge, there is no study that uses a rule-based TL fuzzy system in combination with a fine-tuning approach to classify mammograms.

Our main objective is to propose a hybrid transfer fuzzy system, a combination of TL and fuzzy rule system for calculating and classifying BC risk from mammograms. There are several works in the literature that use TL for the diagnosis of BC. However, most of these studies have analyzed the performance of one or two TL models. In this study, we first use a CNN as a baseline model and then compare its results with other approaches for classifying mammograms as malignant or benign. Then, we evaluate the performance of TL models with different fine-tuning approaches using preprocessing techniques on the mini- MIAS and BCDR-D02 datasets to make a detailed comparison. Then, we identify the most successful TL model with fine-tuning approach on the datasets used and combine this most successful approach with a rule-based fuzzy system to calculate BC risk scores. We also propose a simple rule-based classifier that uses these risks and performs classification. In this study, we used TL with ResNet50 and VGG16 [30] models because we had limited amount of training data and we observed higher accuracy in classifying mammograms compared to other models.

2. MATERIALS AND METHODS

This section composes of two parts. In the first part, datasets, preprocessing steps, data augmentation approaches, and evaluation metrics are presented. In the second part, baseline method (CNN), TL approaches (pre-trained model as a fixed feature extractor, training added layers, layer-wise), and fuzzy system approach are presented.

Datasets

In this study, we use two different BC datasets, namely BC Digital Repository (BC DR-D02) [31] and mini-MIAS [32], to compare our results with those of previous studies.

- **mini-MIAS:** The mini-MIAS [32][1] contains 322 mammograms and each image consists of 1024×1024 pixels. Mammograms are divided into two categories: abnormal (malignant, benign) and normal. Abnormal images, whether malignant or benign, have a type such as calcification, mass, and asymmetry. This database contains 29 calcification images. We use calcifications that contain information about the location of the center and radius, of which 12 are malignant and 10 are benign.
- **BCDR-D02:** The BCDR-D02 consists of Portuguese patients in whom only calcification abnormalities were classified [31][2]. It contains 439 calcification images, of which 42 are malignant and 397 are benign.

Evaluation Metrics

We use the metrics of accuracy, precision, recall, and F-score to evaluate the classification performance of mammograms for the diagnosis of BC [33].

[1] https://www.repository.cam.ac.uk/handle/1810/250394
[2] https://bcdr.eu/

The accuracy represented in Eq. (1) is the ratio of correctly predicted samples to total samples.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

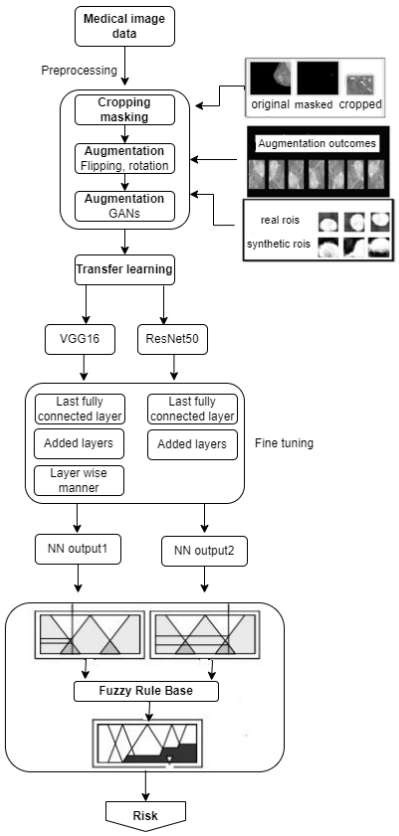


Fig. 1. Our general framework for diagnosing breast cancer.

where TP and TN represent the samples in the positive class (class = YES) and the negative class (class = NO) that were correctly predicted, FP and FN represent the instances in the negative class that were predicted in the positive class, and the instances in the positive class that were predicted in the negative class.

The precision given in Eq. (2) is the ratio of the correctly predicted positive samples to the total predicted positive samples.

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

The recall, represented in Eq. (3), is the ratio of correctly predicted positive samples

to all samples in the actual positive class.

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

F-Score, represented in Eq. (4), is the weighted average of Precision and Recall.

$$Fscore = \frac{2 * Precision * Recall}{Recall + Precision} \tag{4}$$

Preprocessing

The preprocessing steps applied to the dataset are: mini-MIAS and BCDR-D02.

- **Cropping:** Each image is cropped as shown in Fig. 1 to reduce training time. In the mini-MIAS, each image provides valuable information about the type of abnormality as well as the coordinates of the abnormality and the approximate radius of a circle surrounding the abnormality. The abnormal region of interest of each abnormal image is extracted using the x and y coordinates of the center of the abnormal images and the radius. In BCDR-D02, the radius of the circle surrounding the abnormality is unknown. We take advantage of [34] to solve the problem of not knowing the radius of the abnormal region. Therefore, each abnormal image is cropped differently for each dataset. The masses are usually different in size, but according to [34], the masses are rarely larger than 2 cm (length of long axis). Therefore, we reduce the images to the same size to include this 2×2 cm region with a reduction factor of 2.2. We calculate the bounding box of the breast region and crop the mammogram.

- **Balancing Data:** In the mini-MIAS, there are only 10 benign mammograms in 35 calcification images, while in BCDR-D02 there are 42 malignant mammograms in 439 calcification images. Such datasets are referred to as imbalanced, which have a negative impact on minority class classification. To balance the number of instances in the majority class and minority class, m instances are randomly selected from the majority class, where m is the number of samples in the minority class, as in [35]. As a result of undersampling, the total number of data instances is reduced to 20 (10 for malignant and 10 for benign) and 84 (42 for malignant and 42 for benign) instances for mini-MIAS and BCDR-D02, respectively.

Data Augmentation

- **Traditional Methods:** For augmentation, we employ traditional methods to diversify the dataset. This involves combining flipping with rotations at 0, 90, 180, and 270 degrees, creating 8 new instances for each mammogram. Each mammogram undergoes random rotations $\theta \in$ 90,180,270 and subsequent flipping (both vertically and horizontally). Consequently, the total mini-MIAS mammogram data increases from 20 (10 malignant, 10 benign) to 160 (80 malignant, 80 benign). Similarly, the BCDR-D02 dataset expands from 84 (42 malignant, 42 benign) to 672 (336 malignant, 336 benign) instances.

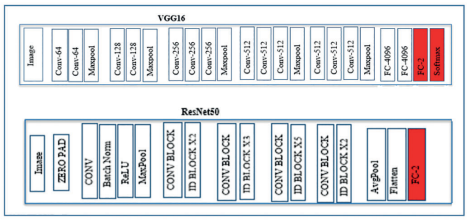


Fig. 2. Updated VGG16 and ResNet50 networks.

- **Generative Adversarial Networks:** GANs, a prominent area in DL research, excel at creating highly realistic synthetic datasets [36]. In our study, we leverage the Deep Convolutional Generative Adversarial Network (DCGAN) [37], known for its success in GAN designs. The generator network comprises 3 convolutional layers ($g_conv1, g_conv2, g_conv3$)$with a kernel size of 5\times5$ and a stride of 2. Likewise, the discriminator network utilizes 3 convolutional layers (d_conv1, d_conv2, d_conv3)$with a kernel size of 4\times4$ and a stride of 2.

METHODS

2.1 Baseline Method

CNNs [38] are image processing models composed of two parts. The initial segment acts as a feature extractor, using convolutional filters to generate feature maps from the input image. These features are then processed through activation functions to form vectors, containing class probability information in the latter part.

2.2 Transfer Learning

For knowledge extraction from mammograms, VGG16 [19] and ResNet50 [17] architectures are utilized in this study. Fig. 2 displays the modification of the last fully connected layer (highlighted in red) in these models to adapt them to BC datasets. Leveraging knowledge from the ImageNet dataset [39], mammograms are classified as malignant or benign. Fine-tuning involves adjusting pre-trained CNN weights by training specific layers with the problem dataset. The following fine-tuning methods are employed:

1. ***Pre-trained model as fixed feature extractor:*** Pre-trained models were trained on the ImageNet [39], which consists of 1000 categories. To classify mammograms, we only replaced the last layer of the pre-trained models with our new last layer, as shown in Fig. 2. Thus, the new last layer of the pre-trained network has 2 categories instead of 1000 categories. We use VGG16 and ResNet50 as pre-trained networks by using their weights as initial weight values and training only the classifier with our dataset. We freeze all layers except the last layer, add a new layer FC -2 with a soft-max function, and then re-train the last fully connected layer using the fixed weights of the previous layers.

2. ***Training added layers:***Since common features are learned in the first layers, we freeze some of the first layers in this method. As shown in Fig. 3 , we freeze the

network for the VGG16 up to the fully connected layers and add a stack of fully connected layers with 128, 128, and 2 neurons, respectively. The softmax layer of the network displays the probability for each possible class and classifies the image according to the most likely class. Fig. 3 shows the architecture of the ResNet50, where the newly added layers are shown in red. In ResNet50, the network is frozen up to the Average Pooling layer and we add a global Average Pooling layer to prevent overfitting, followed by Fully Connected and Dropout layers.

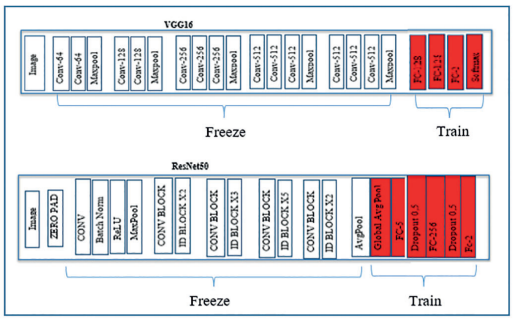


Fig. 3. Training added layers for VGG16 and ResNet50 architectures.

3. ***Layer-wise manner:*** We start by training the output layer of the VGG16 and freeze all layers from block1 to fc-2, as shown in Fig. 4. Next, we freeze the layers from block1 to fc-1 and train the output layer and layer fc-2. Finally, we freeze the layers from block1 to block5 and train the output layers, fc-2 and fc-1. Thus, we continue until we get a lower score. For mini-MIAS we freeze the layers from block1 to block2 and train the layers of block3 for output in a final step. For BCDR-D02, we freeze the layers up to block4 and train the layers of block5 for output in the last step. Since we obtain lower accuracy results by training deeper layers.

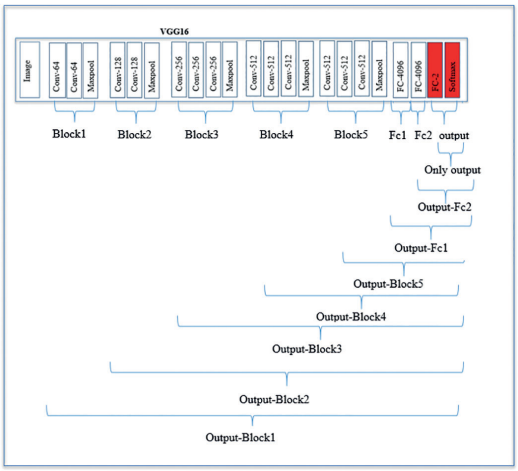


Fig. 4. The layered VGG16 architecture.

2.3 Fusion of the Outputs of the Models with Transfer Learning to Form a Fuzzy Rule-Based System

We have developed a hybrid neuro-fuzzy rule based system to diagnose BC with greater accuracy. This hybrid system consists of two independent parts: Fine-tuning TL methods and a fuzzy expert system that helps us model how an expert determines the risk of BC disease. We trained the most successful fine-tuned TL model and then gave the outputs to the fuzzy system as input. Among all neural network models, we compute risk by using the most successful network structure and present results. The fuzzy expert system matches these inputs with the corresponding outputs to calculate the risk value of BC disease.

We developed a fuzzy system that has two input variables, namely neural network output 1 (NN output 1) and neural network output 2 (NN output 2), and one output variable (BC risk). First, we determined the input of the fuzzy system that are output of the best model (VGG16 with layer-wise fine-tuning).Since the task is a binary classification (benign or malignant), the TL model has two outputs. The first output (NN output 1) and the second output (NN output 2) are the probability that the mammogram belongs to the benign and malignant class, respectively. The outputs are used as input to the fuzzy system to calculate the risk. In Table 1, the computed BC risk scores for the randomly selected 3 cases from both datasets are presented. These scores are computed by using layered outputs for VGG16 without GANs for the mini-MIAS and BCDR-D02 datasets.We give the malignant and benign probabilities, obtained from layered approach for VGG16 without GANs for mini-MIAS and BCDR-D02 datasets, and the calculated risk percentages for these samples. If the probability value of the malignant class is higher than the probability value of the benign class, the risk of that case is higher. For example, in below table, for case 1 of the mini-MIAS dataset, the probability of belonging to benign class is 0.91 and while this probability for malignant class is 0.09, thus the risk is computed as 8%. While actual labels show the real label of the test cases. Label 1 means malign and 0 is benign. To predict labels by using risk scores, we apply a simple rule based classifier that assigns label 0 if risk score is less than or equal to 40%, and label 1 is assigned otherwise. Then, we calculate the accuracy, recall, precision and f-score values for the risk-based classification model and compare with the results of the deep neural network architecture.

Table 1. Risk results of the breast cancer using layered outputs for VGG16 without GANs for mini-MIAS and BCDR-D02 datasets.

Mammogram Images	mini-MIAS			BCDR-D02		
	benign	malign	risk (%)	benign	Malign	risk (%)
case 1	0.91	0.09	8	0.55	0.45	50
case 2	0.25	0.75	78	0.18	0.82	90
case 3	0.76	0.24	15	0.01	0.99	91

Finally, we defined some fuzzy sets and rules to create a fuzzy expert system. The two inputs and the output of the fuzzy system are represented by three fuzzy sets. Fig. 5 shows the fuzzy sets used for the two inputs, NN output 1 and NN output 2, and the fuzzy

Table 2. Actual and predicted class labels for random test cases using risk based classifier for mini-MIAS and BCDR-D02 dataset.

Mammogram Images	mini-MIAS			BCDR-D02		
	benign	malign	risk (%)	benign	Malign	risk (%)
case 1	0	0	8	0	1	50
case 2	1	1	78	1	1	90
case 3	0	0	15	1	1	91

sets for the risk, which is the output of the fuzzy system. We have implemented a fuzzy inference system (FIS) using Matlab. As shown in Fig. 5 we used different membership functions and three linguistic variables (low, medium and high) to define the inputs of the fuzzy system. Triangular and trapezoidal membership functions are commonly used in medical diagnosis support systems due to their simplicity [40]. Therefore, we used trapezoidal membership functions for low and high linguistic terms and triangular membership functions for medium linguistic terms. Each membership function was created based on expert opinion.

We implemented the output of the FIS using trapezoidal and triangular membership functions and very low, low, medium, high and very high linguistic variables. Moreover, the output of the fuzzy system (*i.e.*, the risk of BC disease) is in the range of [0,100] and each FIS input is in the range of [0,1].

Algorithm 1 : Calculation of Risk

procedure RULES(*benign*,*malign*)
    Rules to calculate the risk based on benign and malignant values
    if *benign* is low and, *malign* is low **then**
        Medium Risk
    else if *benign* is low and, *malign* is medium **then**
        High Risk
    else if *benign* is low and, *malign* is high **then**
        Very High Risk
    else if *benign* is medium and, *malign* is low **then**
        Low Risk
    else if *benign* is medium and, *malign* is medium **then**
        Medium Risk
    else if *benign* is medium and, *malign* is high **then**
        High Risk
    else if *benign* is high and, *malign* is low **then**
        Very Low Risk
    else if *benign* is high and, *malign* is medium **then**
        Low Risk
    else
        Medium Risk
    end if
end procedure

For example, if *benign* is 0.12 and *malign* is 0.88, we can say that the class of the mammogram is malignant and the BC risk is very high due to the RULES algorithm.

We determined these rules with the help of a specialist based on the distribution of the data. Based on these rules, the fuzzy system generates a unique number indicating the risk of BC disease using the Center of Gravity (COG) Defuzzification method [41]. The COG is defined by the following formula,

$$x^i = \frac{\int \mu_c(x).xdx}{\int \mu_c(x)dx} \tag{5}$$

where, $\mu_c(x).xdx$ denotes the area of the region bounded by the curve$_{\mu c}$, μ_c is the degree of membership, and x^i is the x-coordinate of center of gravity.

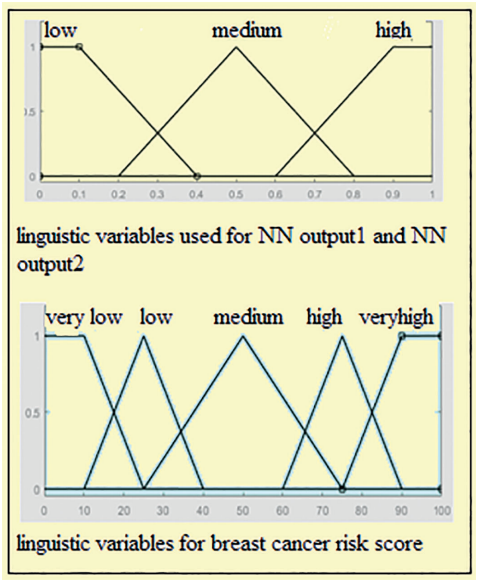


Fig. 5. Linguistic variables.

3. RESULTS

3.1 Implementation Details

K-Fold: Since our dataset is too small to train the DL, we apply K-Fold to train the model with all samples. In *K*-Fold cross-validation, we divide our data into k distinct subsets. We use $k-1$ subsets to train our data and leave the final subset as test data. The average error value indicates the validity of our model.

Hyperparameters

- Data Augmentation (GANs): We use the Adam optimizer [42] for both the generator and discriminator networks with a learning rate of 0.0002 and a beta1 of 0.5.

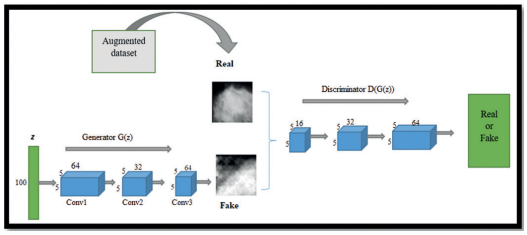


Fig. 6. GAN network for mammogram data.

Using these parameters, as shown in Fig. 6, we generated a total of 2000 additional data samples, 1000 samples for the malignant class and 1000 samples for the benign class. Table 1 presents number of training and test instances with/without data augmentation for both datasets. Before augmentation with GANs, mini- MIAS and BCDR-D02 consist of 160 and 672 instances, respectively.

- Baseline Model (CNN): The CNN takes a 224×224×3 dimensional image as input. It consists of 3 convolutional blocks including 3×3 Convolution – ReLU – Max pooling, with 32, 32 and 64 filters respectively, followed by a pooling layer with global average and a fully connected layer of dimension 2. The last layer is a softmax layer for malignant or benign classification. We also use the Adam Optimizer with a learning rate of 0.001.

Table 3. # of training and testing instances (for each fold) with/out GANs for datasets.

		# of instances before GANs			# of instances after GANs		
		The whole set	Benign	Malign	The whole set	Benign	Malign
mini-MIAS	Training set	128	63	65	2128	1063	1065
	Test set	32	17	15	32	17	15
BCDR-D02	Training set	536	269	267	2536	1269	1267
	Test set	136	67	69	136	67	69

- Transfer Learning: We train the ResNet50 and the VGG16 with the Adam optimizer for all experiments, the base learning rate 10^{-3}, the dropout 0.5, and these values are determined by [38]. We choose the default parameters for the learning rate and dropout such that the regularization effect is maximized when the dropout rate is set to 0.5 [34].

- Layer-wise manner: Consider a CNN with 5 layers L1, L2, L3, L4, and L5, with the last 2 layers fully connected. In the first training round, we freeze layers L1 to L4 and fine tune layer L5. Next, we train layers L4 and L5. Next, we train the layers L3, L4, and L5. We also used the stochastic gradient descent optimizer [43] with a learning rate of 10^{-2} for the last fully linked layer and set the learning rate to 10^{-3} for all other layers. A learning rate of 0.001 resulted in reasonable convergence. Choosing a smaller learning rate slowed convergence, and a larger learning rate resulted in convergence errors.

The all results for presented methods in Section 2 are given in Table 2.

The success of the CNN increases when both datasets are augmented with GANs. This shows that the model learns better when synthetic data is added to the training. The BCDR-D02 achieves higher scores than the mini- MIAS, both with and without GANs. The reason could be the number of samples in the datasets. Also, the mini- MIAS achieves higher values for accuracy and recall than for precision and F1 score. While the recall has the highest value for BCDR-D02 with GANs, the recall is the lowest for BCDR-D02 without GANs.

Table 4. Classification results for mini-MIAS and BCDR-D02.

		mini-MIAS		BCDR-D02	
		without GANs	with GANs	without GANs	with GANs
CNN	Accuracy	0.64	0.68	0.82	0.89
	Precision	0.50	**0.52**	0.88	**0.87**
	Recall	0.73	**0.77**	0.75	**0.93**
	F1-Score	0.59	**0.61**	0.80	**0.90**
VGG16♣	Accuracy	0.75	0.91	0.91	**0.92**
	Precision	0.80	**0.90**	0.90	0.93
	Recall	0.69	0.92	0.91	**0.93**
	F1-Score	0.73	0.90	0.91	**0.92**
Resnet50♣	Accuracy	0.76	0.81	0.69	0.89
	Precision	0.66	**0.90**	0.66	**0.94**
	Recall	0.65	0.72	0.87	0.83
	F1-Score	0.65	0.79	0.74	0.88
VGG16♢	Accuracy	0.80	**0.91**	0.83	0.84
	Precision	0.86	**0.90**	0.83	0.83
	Recall	0.82	**0.91**	0.83	0.87
	F1-Score	0.80	**0.90**	0.83	0.84
Resnet50♢	Accuracy	0.68	0.80	0.71	**0.95**
	Precision	0.68	0.85	0.65	**0.95**
	Recall	0.70	0.73	0.88	**0.90**
	F1-Score	0.71	0.78	0.78	**0.90**
Layer-wise	Accuracy	**0.89**	-	**0.93**	-
	Precision	**0.89**	-	**0.93**	-
	Recall	**0.91**	-	**0.95**	-
	F1-Score	**0.88**	-	**0.94**	-
Risk model	Accuracy	**0.93**	-	**0.94**	-
	Precision	**0.93**	-	**0.94**	-
	Recall	**0.94**	-	**0.95**	-
	F1-Score	**0.93**	-	**0.94**	-

The first approach to fine-tuning is to use a pre-trained model as a fixed feature extractor, where all layers of the model are frozen and only the last fully linked layer is trained. In Table 4 , where ♣ shows the results of the pre-trained model as a fixed feature extractor, we can see that the accuracy of mini- MIAS data for VGG16 with GANs has improved from 75% to 91%, while the accuracy of mini- MIAS data for Resnet50 has

increased from 76% to 81%. Thus, VGG16 generalizes more by using the larger dataset. We obtain better results with VGG16 than with ResNet50 for both datasets. We also see that the results with GANs are better than the results without GANs. However, the mini-MIAS show worse performance than BCDR-D02 both with and without GANs. However, the mini- MIAS without GANs has higher accuracy than BCDR-D02 without GANs.

The second approach for fine-tuning is to replace the last fully connected layer with a small mini-network. Table 4 , where ♢ shows the results of training the added layers, demonstrates the results for mini- MIAS and BCDR-D02 with ResNet50 and VGG16. Overall, the results show that we obtain better results with GANs, both for mini-MIAS and BCDR-D02. Comparing the VGG16 results, the best results are obtained by using mini-MIAS with GANs. Also, the ResNet50 results show the highest score for BCDR-D02 with GANs. On the other hand, the VGG16 results are better than ResNet50 results both with and without GANs except for BCDR-D02 with GANs. This shows that the deeper network ResNet50 learns better with GANs by adding new layers and training these layers with pre-trained weights.

When comparing the results of ♣ and ♢ for BCDR-D02, we see that the results of ResNet50♣ are lower than those of ResNet50♢ and the results of VGG16♣ are higher than those of VGG16♢. Also, when comparing the CNN results with the results of ♣ and ♢, we get better results with fine-tuning than with CNN. Moreover, the results of ♢ for BCDR-D02 with GANs using VGG16 are lower than those of CNN and Resnet50. While using BCDR-D02 with/without GANs on ResNet50♢ leads to higher values than using mini-MIAS, using mini-MIAS with GANs on VGG16♢ leads to higher values than BCDR-D02.

The last approach is the layered method, where we perform the training in several layers. For this method, we have the experimental results with mini- MIAS and BCDR-D02 without GANs only for VGG16. The reason is that using data with GANs or with Resnet50 leads to memory errors.

Table 5. Training Layer-wise Manner using VGG16.

		Fine-tuned VGG16 without GANs						
		Only Output Layer	Output fc2	Output fc1	Output flatten	Output block5	Output bloack4	Output block3
mini-MIAS	Accuracy	0.64	**0.89**	0.88	0.88	0.89	0.78	0.53
	Precision	0.72	**0.89**	0.86	0.88	0.88	0.81	0.15
	Recall	0.75	**0.91**	0.88	0.89	0.90	0.74	0.25
	F1-Score	0.64	**0.88**	0.88	0.88	0.88	0.75	0.14
BCDR-D02	Accuracy	0.90	**0.93**	0.92	0.92	0.92		
	Precision	0.91	**0.93**	0.92	0.93	0.93	-	-
	Recall	0.94	**0.95**	0.94	0.93	0.94	-	-
	F1-Score	0.90	**0.94**	0.93	0.93	0.93	-	-

Layer-wise Findings

As seen in Table 5, updating only the output layer has the second worst performance for mini-MIAS. Meanwhile, fine-tuning the last two layers (output-fc2) significantly outperforms fine-tuning the deeper layers. These results show that fine-tuning the last few layers is sufficient to achieve better results than deep fine-tuning.

For BCDR-D02, we obtain similar results for almost all fine-tuning layers. The

lowest accuracy (90%) is obtained by fine-tuning only the output layer, and the highest accuracy (93%) is obtained by fine-tuning the output-fc2 layers. The lowest recall rate (74%) is obtained by fine-tuning the output-block4.

When we compare overall results, we can say that fine-tuning of the output-fc2 layer of VGG16 is superior to the other fine-tuning layers for both datasets.

When we look at the overall results, we can say that the use of GANs improves the classification results. Moreover, if we compare the results of CNN with the results of TL, we can say that TL is more successful than the baseline method. Also, when we analyze all the experimental results of our study, we find that the layered method outperforms all other fine-tuning approaches.

Transfer learning-Fuzzy Rule Based System Findings

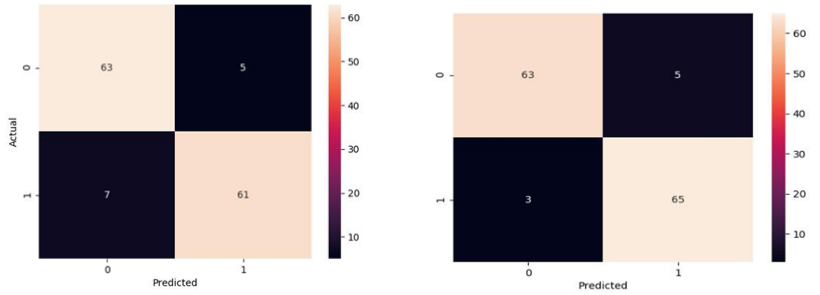
Among all fine-tuned TL models, we compute the risk with the most successful of them. We calculated BC risk scores with layered outputs for VGG16 without GANs with mini-MIAS and BCDR-D02. We give the malignant and benign probabilities, obtained from layered approach for VGG16 without GANs for mini-MIAS and BCDR-D02 datasets, and the calculated risk percentages for these samples. Based on the training results of the neural network models, we obtained two different outputs. One is the probability that the image belongs to the malignant class (NN output 2); the other is the probability that the image belongs to the benign class (NN output 1). Membership degree of each class is computed according to linguistic variables. And then, we apply fuzzy rule algorithms using the linguistic variables of the NN output 1 and NN output 2 to find the liguistic variable of the risk. Finally, we apply center of gravity method for the defuzzification of the risk score.

If the probability of the malignant class is higher than that of the benign class, the risk for this case is higher. For example, in Table 7, for a patient in BCDR-D02, the probability for the malignant class is 0.82 and the probability for the benign class is 0.18, so the risk value is 90%.

Table 6. Actual and predicted class labels for sample test cases using risk based classifier for mini-MIAS and BCDR-D02 dataset.

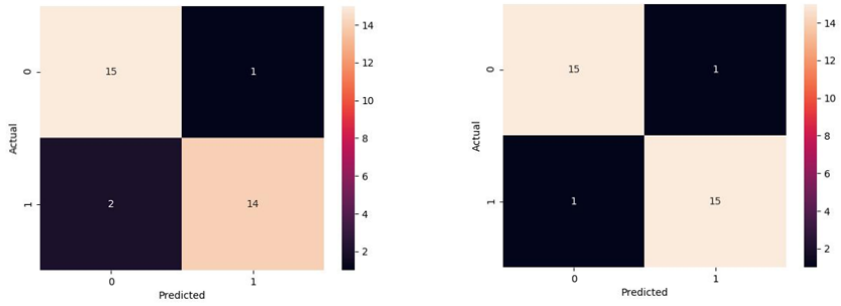
Patient test cases	BCDR-D02			Mini-MIAS		
	Actual Label	Predicted Label	Risk %	Actual Label	Predicted Label	Risk %
case 1	0	1	60	0	1	52
case 2	1	1	91	1	0	15
case 3	1	1	50	1	1	91

Table 6 presents the predicted class labels using risk scores for BCDR-D02 and mini-MIAS datasets for some test cases. While actual labels show the real label of the test cases. The label 1 means malignant and 0 is benign. To predict labels based on risk scores, we apply a simple rule-based classifier that assigns label 0 if the risk score is less than or equal to 40% and assigns label 1 otherwise. The risk-based model achieves more accurate results than layered approach for both datasets. The risk model for BCDR-D02 achieves the highest values for accuracy (94%), f-score (94%), and recall (95%). When we examine all results, we can say that the risk model is superior to the layered approach and improves the accuracy of diagnosis of BC. Figs. 7 and 8 show the confusion



(a) layered approach (b) risk model

Fig. 7. Confusion matrices for BCDR-D02.



(a) layered approach (b) risk model

Fig. 8. Confusion matrices for mini-MIAS.

matrix heatmaps of layered approach and risk model for both dataset BCDR-D02 and mini-MIAS. Each confusion matrix shows the number of mammogram images correctly and incorrectly classified.

While the rows of confusion matrices illustrate the actual classes, the columns show the predicted class. In confusion matrix heatmaps the top left presents the number of TNs, top right: FPs, bottom left: FNs, bottom right: TPs. There are 63 TNs for layered approach using BCDR-D02 and 15TNs for layered approach using mini-MIAS. For layered approach using BCDR-D02 The number of TPs is 61 and for layered approach using mini-MIAS is 14. As the heat maps show, the TP and TN values are high while the FP and FN values are low, which indicates that the risk-based model has high f-score.

Comparison With Previous Studies

In the literature, there are many studies in classifying medical image data using deep learning methods. In the related work section of this study, the most similar studies to our study are included, and we compare these studies with our study in terms of: (i) the dataset used, (ii) method applied, (iii) fine-tuning methods used, (iv)whether segmentation is used or not (v) augmentation method employed, and (vi) the obtained results. A summary of these comparisons is presented in Table 8. If we compare our work with the studies in [44] and [45], we can say that our results are better than the results of these two studies. In [44],

Table 7. Risk results of breast cancer using layered outputs for VGG16 without GANs for mini-MIAS and BCDR-D02 dataset.

Mammogram images	mini-MIAS			BCDR-D02		
	Without GANs			Without GANs		
	benign	malign	Risk %	benign	malign	Risk %
case1	0,91	0,09	8	0,99	0,01	8
case2	0,25	0,75	78	0,18	0,82	90
case3	0,76	0,24	15	0,01	0,99	91

the researchers have tested DCE-MRI dataset using only the ResNet model and tried only one fine-tuning method. In [45], the authors have used only the AlexNet model and fine-tuned by modifying only the last few layers. In addition, in both studies, researchers only have benefited from the classical augmentation method. In our study using GANs and layered fine-tuning helped us to get accuracy higher than 80%. And also, we can say that using VGG16 model is better than the AlexNet model for those two datasets, mini-MIAS and BCDR-D02. As for the studies at [46, 47], the authors have used only Inception V3 model and one fine-tuning method. In [46], they have obtained an accuracy rate, which is 85% by using the classical augmentation method and obtained an accuracy score about 87% in [47]. However, in our work, we have reached better accuracy results. According to the study in [15], the researchers have used only the VGG16 model and have applied different training method. First of all, feature extraction from the VGG16 model was made by using mini-MIAS dataset and then these features were used for neural network training. Even if they have used the VGG16 model and mini-MIAS data which also have been used in our work, we have achieved higher accuracies by using mini-MIAS and the VGG16 model with GANs. In our study, contrary to manual segmentation in [48], we produced a mask to find a region of interest in the BCDR-D02 dataset. And also, compared to [15, 48], we take advantage of GANs to augment BCDR-D02 and mini-MIAS datasets as well as traditional augmentation methods. However, compared to [48], we analyzed various fine-tuning approaches such as layer-wise manner, appending mini-network using the VGG and ResNet architectures different from the Inception model. On the other hand, when we compare [15] with our work, we tried to test different fine-tuning methods as well as updating the weights in the final layers of the VGG16 model on mini-MIAS dataset for only calcification data. To sum up, in this study, we compared results of several deep learning methods by using BCDR-D02 and mini-MIAS datasets and showed their effectiveness. In our study, we have used BCDR-D02 which includes digital mammograms and mini-MIAS dataset to classify only calcification abnormalities while other researches use different datasets. In [48], the authors have used BCDR-F03 which consists of film mammograms. In [15], researchers have used mini-MIAS and DDSM dataset. In our work, we have used CNN as a baseline method, and compare with ResNet50, and VGG16 to evaluate performances of deep learning methods. However, to our knowledge, none of the previous studies have made such a comparison. Also, we have compared four different fine-tuning methods, while other studies use one or two fine-tuning approaches in their research. Nevertheless, in the BCDR-D02 dataset, the radius of a circle surrounding the abnormality is unknown. So, we produce a mask using

Table 8. Comparison of studies involving deep learning methods to classify medical images.

Study	Dataset	Method	Fine-tuning	Segmentation	Augmentation	Result
[24]	Liver Lesion	CNN	-	-	Classic GANs	Sensitivity 0.857 Specifity 0.924
[25]	DRIVE	U-net	-	-	GANs	accuracy 0.88
[26]	Histopathology image	CNN	-	-	GANs	-
[27]	Chest X-rays	DCNN			GANs	Accuracy 0.921
[44]	DCE-MRI	ResNet	Only final layer adapted to of classes	-	Classic	Accuracy 0.80
[45]	DDSM	AlexNet	Fine-tuned modifying the last few layers	-	Classic	Accuracy 0.80
[46]	BreaKHis	Google's Inception v3	Only final layer adapted to of classes	-	Classic	Accuracy 0.85
[47]	DDSM	Inception V3	Fine-tuned modifying the last few layers	-	-	Accuracy 0.87
[49]	BreaKHis	ResNet Inception V3		-	-	Accuracy ResNet 0.98 Accuracy Inception 0.94
[48]	BCDR-F03	Inception V3	Fine-tune the layers with a per-layer exponentially decaying learning rate	Manual	Classic	Accuracy 0.975
[15]	mini-MIAS DDSM	VGG16	Pre-trained VGG-16 to extract features from mini-MIAS and then used to train NN from DDSM	-	-	Best Accuracy 0.90
Our study	mini-MIAS BCDR-D02	ResNet VGG CNN	Training the classifier alone Training added layers Layer-Wise Manner	Cropping roi with/out Radius	Classic Classic+GANs	Best values: Accuracy 0.96 Precision 0.96 Recall 0.95 F1-Score 0.95

the provided outlines of lesion to separate the lesions. We have cropped the region of interest, according to x and y coordinates with/without radius information provided by the dataset. But, other studies don't use such a cropping method using a mask. Furthermore, while other studies use only GANs or only classical augmentation methods, we have used the combination of GANs and classical augmentation methods in our study. As Table 8 shows, we have made a detailed comparison of the deep learning methods for image classification and the best results obtained are 96% precision,96% precision, 95% recall and 95% f1-cores, which are compatible or higher than the values previously observed.

4. CONCLUDING REMARKS

In our study, we used CNN as a baseline method and compared it with ResNet50 and VGG16 to evaluate the performance of fine-tuning TL methods. Our study shows how cautious preprocessing, GANs, and TL can deal with the problem of a limited amount of training data in the medical domain. TL and fine-tuning allow us to use DL models when we have a small amount of training data, even when the source and target domains are different. Moreover, the layer-wise method outperforms all other fine-tuning approaches.This method provides a practical solution to achieve the best performance when we work with small medical data. Using the fuzzy rule-based system, we also evaluate the risk of BC disease based on the results of the layer-wise approach. These risk results help to diagnose BC disease more accurately. In addition, as far as we know, we have developed the first hybrid system of TL with fine-tuning and fuzzy classification for classifying BC from mammogram images and our system has the best classification accuracy values compared to the other methods.

In our future study, we plan to test other recently developed architectural models. In addition, the use of larger datasets with more powerful devices can help physicians to obtain more reliable and accurate results when making biopsy decisions.

REFERENCES

1. G. Zorluoglu and M. Agaoglu, "Diagnosis of breast cancer using ensemble of data mining classification methods," *International Journal of Oncology and Cancer Ther-*

apy, Vol. 2, 2017, pp. 24-27.

2. H. K. K. Zand, "A comparative survey on data mining techniques for breast cancer diagnosis and prediction," *Indian Journal of Fundamental and Applied Life Sciences*, Vol. 5, 2015, pp. 4330-4339.
3. K. Sivakami, "Mining big data: Breast cancer prediction using DT-SVM hybrid model," *International Journal of Scientific Engineering and Applied Science*, Vol. 1, 2015.
4. S. V. Sree, E. Y.-K. Ng, R. U. Acharya, and O. Faust, "Breast imaging: a survey," *World Journal of Clinical Oncology*, Vol. 2, 2011, p. 171.
5. P. Stephan, "What is a breast biopsy?" https://www.verywellhealth.com/open-surgical-breast-biopsy-429949, 2021.
6. J. M. E. C. for Clinical, "Having a breast biopsy: A review of the research for women and their families," *Comparative Effectiveness Review Summary Guides for Consumers*, 2005.
7. P. Ruiz, "Understanding and visualizing resnets," https://towardsdatascience.com/understanding-and-visualizing-resnets-442284831be8, 2018.
8. S. Sahran, A. Qasem, K. Omar, D. Albashih, A. Adam, S. N. H. S. Abdullah, A. Abdullah, R. I. Hussain, F. Ismail, N. Abdullah, S. Hayati Md Pauzi, and N. Abd Shukor, "Machine learning methods for breast cancer diagnostic," *Breast Cancer and Surgery*, Vol. 5, 2018, pp. 57-76.
9. S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *Journal of Big Data*, Vol. 6, 2019, pp. 1-18.
10. S. A. Alanazi, M. Kamruzzaman, M. N. Islam Sarker, M. Alruwaili, Y. Alhwaiti, N. Alshammari, and M. H. Siddiqi, "Boosting breast cancer detection using convolutional neural network," *Journal of Healthcare Engineering*, Vol. 2021, 2021.
11. E. Wu, K. Wu, D. D. Cox, and W. Lotter, "Conditional infilling gans for data augmentation in mammogram classification," *CoRR*, Vol. abs/1807.08093, 2018, http://arxiv.org/abs/1807.08093
12. C. Haarburger, P. Langenberg, D. Truhn, H. Schneider, J. Thüring, S. Schrading, C. K. Kuhl, and D. Merhof, "Transfer learning for breast cancer malignancy classification based on dynamic contrast-enhanced MR images," in *Bildverarbeitung für die Medizin*, A. Maier, T. M. Deserno, H. Handels, K. H. Maier-Hein, C. Palm, and T. Tolxdorff, (eds.), Springer, Berlin Heidelberg, 2018, pp. 216-221.
13. D. R. Shettar, "Transfer learning on pre-trained deep convolutional neural network for classification of masses in mammograms," *IOSR Journal of Computer Engineering*, Vol. 19, 2017, pp. 50-55.
14. J. Chang, J. Yu, T. Han, H.-J. Chang, and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer," in *Proceedings of IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*, 2017, pp. 1-4.
15. H. Chougrad, H. Zouaki, and O. Alheyane, "Convolutional neural networks for breast cancer screening: Transfer learning with exponential decay," *arXiv Preprint*, 2017, arXiv:1711.10752.

16. A. H. Shayma'a, M. S. Sayed, M. I. Abdalla, and M. A. Rashwan, "Breast cancer masses classification using deep convolutional neural networks and transfer learning," *Multimedia Tools and Applications*, Vol. 79, 2020, pp. 30 735-30 768.
17. L. G. Falconi, M. Perez, W. G. Aguila, and A. Conci, "Transfer learning and fine tuning in breast mammogram abnormalities classification on CBIS-DDSM database," *Advances in Science, Technology and Engineering Systems*, Vol. 5, 2020, pp. 154-165.
18. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770-778.
19. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Preprint*, 2014, arXiv:1409.1556.
20. X. Peng, Z. Tang, F. Yang, R. S. Feris, and D. Metaxas, "Jointly optimize data augmentation and network training: Adversarial data augmentation in human pose estimation," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2226-2234.
21. M. M. Jadoon, Q. Zhang, I. U. Haq, S. Butt, and A. Jadoon, "Three-class mammogram classification based on descriptive cnn features," *BioMed Research International*, Vol. 2017, 2017.
22. A. Yu and K. Grauman, "Semantic jitter: Dense supervision for visual comparisons via synthetic images," in *Proceedings of IEEE International Conference on Computer Vision*, 2017, pp. 5570-5579.
23. X. Wang, A. Shrivastava, and A. Gupta, "A-fast-rcnn: Hard positive generation via adversary for object detection," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2606-2615.
24. Y.-X. Wang, R. Girshick, M. Hebert, and B. Hariharan, "Low-shot learning from imaginary data," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7278-7286.
25. J. T. Guibas, T. S. Virdi, and P. S. Li, "Synthetic medical images from dual generative adversarial networks," *arXiv Preprint*, 2017, arXiv:1709.01872.
26. M. Frid-Adar, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Synthetic data augmentation using gan for improved liver lesion classification," in *Proceedings of IEEE 15th International Symposium on Biomedical Imaging*, 2018, pp. 289-293.
27. L. Hou, A. Agarwal, D. Samaras, T. M. Kurc, R. R. Gupta, and J. H. Saltz, "Unsupervised histopathology image synthesis," *arXiv Preprint*, 2017, arXiv:1712.05021.
28. S. Valarmathi, A. Sulthana, R. Rathan, K. Latha, S. Balasubramanian, and R. Sridhar, "Prediction of risk in breast cancer using fuzzy logic tool box in matlab environment," *International Journal of Current Research*, Vol. 4, 2012, pp. 072-079.
29. M. Negnevitsky, "Design of a hybrid neuro-fuzzy decision-support system with a heterogeneous structure," in *Proceedings of IEEE International Conference on Fuzzy Systems*, Vol. 2, 2004, pp. 1049-1052.
30. A. Rosebrock, "Imagenet: Vggnet, resnet, inception, and xception with keras," https://www.pyimagesearch.com/, 2017.
31. M. Lopez, N. Posada, D. C. Moura, R. R. Pollán, J. M. F. Valiente, C. S. Ortega, M. Solar, G. Diaz-Herrero, I. Ramos, J. Loureiro, T. Fernandese, and B. Araujo,

"BCDR: a breast cancer digital repository," in *Proceedings of the 15th International Conference on Experimental Mechanics*, Vol. 1215, 2012, pp. 113-120.
32. P. SUCKLING J, "The mammographic image analysis society digital mammogram database," *Digital Mammo*, 1994, pp. 375-386.
33. J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, MA, 2011.
34. B. Sahiner, H.-P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images," *IEEE Transactions on Medical Imaging*, Vol. 15, 1996, pp. 598-610.
35. T. Hasanin, T. M. Khoshgoftaar, J. L. Leevy, and R. A. Bauder, "Severely imbalanced big data challenges: investigating data sampling approaches," *Journal of Big Data*, Vol. 6, 2019, pp. 1-25.
36. N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE Transactions on Medical Imaging*, Vol. 35, 2016, pp. 1299-1312.
37. D. Aadil Hayat, "Building a simple generative adversarial network (GAN) using tensorflow," https://blog.paperspace.com/implementing-gans-in-tensorflow/, 2018.
38. D. Lévy and A. Jain, "Breast mass classification from mammograms using deep convolutional neural networks," *arXiv Preprint*, 2016, arXiv:1612.00542.
39. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248-255.
40. V. Kreinovich, O. Kosheleva, and S. N. Shahbazova, "Why triangular and trapezoid membership functions: A simple explanation," *Recent Developments in Fuzzy Logic and Fuzzy Sets*, 2020, pp. 25-31.
41. "Chapter 4 fuzzy rules and inferences," https://cse.iitkgp.ac.in/ dsamanta/courses/archive/sca/Archives/Chapter%204%20Fuzzy%20Rules%20and%20Inferences.pdf.
42. D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv Preprint*, 2014, arXiv:1412.6980.
43. S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv Preprint*, 2016, arXiv:1609.04747.
44. E. Wu, K. Wu, D. Cox, and W. Lotter, "Conditional infilling gans for data augmentation in mammogram classification," in *Proceedings of the 3rd International Workshop on Image Analysis for Moving Organ, Breast, and Thoracic Images*, LNCS Vol. 11040, 2018, pp. 98-106.
45. C. Haarburger, P. Langenberg, D. Truhn, H. Schneider, J. Thüring, S. Schrading, C. K. Kuhl, and D. Merhof, "Transfer learning for breast cancer malignancy classification based on dynamic contrast-enhanced MR images," in *Proceedings of Workshops on Image Processing for Medicine*, Vol. 11, 2018, pp. 216-221.
46. M. Basanth and R. Shettar, "Transfer learning on pre-trained deep convolutional neural network for classification of masses in mammograms," *IOSR Journal of Computer Engineering*, Vol. 19, 2017, p. e5.
47. J. Chang, J. Yu, T. Han, H.-J. Chang, and E. Park, "A method for classifying medical images using transfer learning: A pilot study on histopathology of breast cancer," in

Proceedings of IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 2017, pp. 1-4.
48. M. H. Motlagh, M. Jannesari, H. Aboulkheyr, P. Khosravi, O. Elemento, M. Totonchi, and I. Hajirasouliha, "Breast cancer histopathological image classification: A deep learning approach," *BioRxiv*, 2018, p. 242818.
49. P. Baldi and P. J. Sadowski, "Understanding dropout," *Advances in Neural Information Processing Systems*, Vol. 26, 2013, pp. 2814-2822.

Pinar Uskaner Hepsağ received the BS degree from the Department of Mathematics, Çukurova University and MS degree from the Computer Science Department, Case Western Reserve University, and the Ph.D. degree in Computer Engineering from Cukurova University. Her research areas include machine learning and deep learning in medicine.

Selma Ayse Özel is a Professor of Computer Engineering in the Department of Computer Engineering, at Çukurova University. Her research interests include text mining, data mining, machine learning, natural language processing,database systems, information retrieval systems, nature inspired computing and their applications to problem solving in several disciplines. She has published more than 100 scientific papers in international and national journals/conferences in the fields of her research interests.

Adnan Yazici is a Professor of Computer Science in the Department of Computer Science, at Nazarbayev University. He has published over 200 international technical papers and co-authored/edited three books entitled Fuzzy Database Modeling (Springer), Fuzzy Logic in its 50th Year: New Developments, Directions and Challenges (Springer), and Uncertainty Approaches for Spatial Data Modeling and Processing: A Decision Support Perspective (Springer). His research interests include data science, multimedia databases and information retrieval, wireless multimedia sensor networks, and fuzzy database modeling. He is currently an Associate Editor of IEEE Transactions on Fuzzy Systems and a member of ACM, IEEE Computational Intelligence Society and the Fuzzy Systems Technical Committee.