Physica A 545 (2020) 123592

Contents lists available at ScienceDirect

Physica A

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

BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer

Mesut Toğaçar [a,*], Kutsal Baran Özkurt [b], Burhan Ergen [c], Zafer Cömert [d]

[a] Department of Computer Technology, Fırat University, Elazığ, Turkey
[b] Buca İnci Özer Tırnaklı Science High School, İzmir, Turkey
[c] Department of Computer Engineering, Faculty of Engineering, Fırat University, Elazığ, Turkey
[d] Department of Software Engineering, Faculty of Engineering, Samsun University, Samsun, Turkey

ARTICLE INFO

Article history:
Received 8 May 2019
Received in revised form 4 September 2019
Available online 22 November 2019

Keywords:
Biomedical image processing
Attention module
Diagnosis system
Breast cancer
Hypercolumn
Deep learning

ABSTRACT

Breast cancer is one of the most commonly diagnosed cancer types in the woman and automatically classifying breast cancer histopathological images is an important task in computer-assisted pathology analysis. Statistics indicate that the breast cancer rate is about 12% in all cancer cases in the world. Also, approximately 25% of women have breast cancer. Therefore, rapid and accurate analysis of breast cancer images is extremely important for diagnosis. Recently, deep learning models have been used in preference for this purpose. In short, the most important reason why we use a deep learning model for the diagnosis of breast cancer is can give faster and more accurate results than existing machine learning based methods. In this study, we come up with a novel deep learning model developed based on a convolutional neural network. The success of the classification was increased by using the proposed model named as BreastNet. The general structure of the BreastNet model is a residual architecture built on attention modules. Each image data is processed by the augmentation techniques before applying it as input to the model. With augmentation techniques, each image is processed one by one and transferred to BreastNet. There is no increase in the number of data. The features of each image are changed using some augmentation techniques, such as flip, shift, brightness change and rotation. Then, each image that comes to the model performs the selection and processing of important key regions of the image via through attention modules. Also, a more stable and accurate classification of the data is performed by using the hypercolumn technique in the model. Other parts of the BreastNet model consist of convolutional, pooling, residual and dense blocks. As a result, 98.80% classification success was achieved with the proposed model. The success rate of the proposed model was better than the success rates of AlexNet, VGG-16 and VGG-19 models performed on the same data set. In addition, the results obtained in this study yielded better results than the other studies that use the current BreakHis dataset.

* Corresponding author.
E-mail addresses: mtogacar@firat.edu.tr (M. Toğaçar), kutsal_baran@hotmail.com (K.B. Özkurt), bergen@firat.edu.tr (B. Ergen), zcomert@samsun.edu.tr (Z. Cömert).

https://doi.org/10.1016/j.physa.2019.123592
0378-4371/

1. Introduction

Cancer is a collection of diseases in which cells in the body come together to form lumps called malignant tumor. These cells grow in an uncontrollable way, spread into surrounding tissues and crowd out the normal cells [1]. From past to present, cancer is one of the most important disease that threaten human health [2]. In the study conducted in 2018, it has been estimated that 18.1 million cancer cases will be added to the existing cancer cases in the world and approximately 9.6 million of these cancer cases will result in death [2,3]. There are more than 100 different types of cancer in medical science [4]. Nowadays, breast cancer is one of the most common type of cancer that is frequently mentioned among women and there are many studies on breast cancer in the scientific sense [5–7]. Breast cancer is the most commonly diagnosed cancer among women in 140 of 184 countries worldwide [8]. It has been estimated that approximately 20% of breast cancers over the world arise from modifiable risk factors, including alcohol use, excess body weight, and physical inactivity [9]. Therefore, early and accurate diagnosis of breast cancer is extremely important.

In the biomedical field, the examination and diagnosis of the breast cancer histopathological images by field experts is a sensitive and labor-intensive process requiring time and high qualification. The diagnosis process can be supported by utilizing existing technological tools and softwares. Thus, the cost and diagnosis effort can be significantly reduced. For this purpose, numerous studies have been conducted based on computational approaches. Support vector machines (SVMs) equipped with a feature selection algorithm has been introduced to detect breast cancer. 99.51% classification success was achieved in the study [1]. An expert system based on the association rules (AR) and neural network (NN) have been suggested for breast cancer diagnosis. AR ensured a reduction in the number of features whereas NN was employed for the intelligent classification. The accuracy of the expert system was 95.6% [10]. The fuzzy systems and evolutionary algorithms have been combined for the same purpose. The model provided a few simple rules for the experts' interpretation [11]. Epithelial (EP) and stromal (ST) that are two types of tissues in histological images have been segmented automatically using the deep convolutional neural network (DCNN) and the model provided satisfactory results. The breast cancer histopathological images have been classified using ImageNet pre-trained DCNN model named as AlexNet. Various experiments were conducted on a public dataset. As a result, the model yielded 90% accuracy with deep fusion rules [12]. A novel breast cancer algorithm, convolutional neural network improvement for breast cancer classification (CNNI-BCC) has been suggested for the diagnosis task. The model was performed on the mammographic breast cancer medical images and reached 90.50% accuracy [13]. CNNI-BCC model has been offered to support medical experts in breast cancer diagnosis in a timely manner. CNNI-BCC model has a capable of classifying the incoming breast cancer medical images according to malignant, benign, and healthy. The model achieved 90.50% classification accuracy [13]. An intelligent diagnosis approach for breast cancer diagnosis has been suggested, which utilize information gain directed simulated annealing genetic algorithm wrapper (IGSAGAW) for feature selection to reveal the top optimal feature performed the cost-sensitive support vector machine (CSSVM) learning algorithm. The model achieved 95.80% classification success [14]. In another study, incremental boosting convolutional networks have been introduced for providing an efficient diagnosis model of breast cancer from histopathological microscopic images. The model yielded an accuracy of 96.4% and 99.5% for four and two classes classification tasks, respectively [15].

It is clearly seen that numerous models have proposed to improve the efficiency of breast cancer diagnosis process. In this regard, one of the common models preferred for the diagnosis of breast cancer is the deep learning approach [16–18]. In this study, a novel deep learning model is proposed for breast cancer diagnosis. In the model that we call BreastNet, attention modules and hypercolumn technique were used together. In addition, more efficient qualifications have been obtained by applying the augmentation technique to the input images. The input data was set to 224×224 pixels. The general architecture of BreastNet consists of convolutional, dense and residual blocks. The CNN models such as AlexNet, VGG-16, VGG-19 were also used to compare the performance of the BreastNet model. The features obtained from the last fully connected layer (FC8) of the CNN models were provided as input to the Softmax activation function. Moreover, the classification performances of the models are examined separately. The experiments are carried out an open-access BreakHis dataset composed of histopathological images [12].

The rest of this study is organized as follows: In Section 2, the used publicly available dataset is described. In Section 3, the proposed novel CNN model, pre-trained deep CNN models, machine-learning algorithms and the steps of the experiments are presented. The experimental results are reported in Section 4. The discussion is given in Section 5. Lastly, concluding remarks are presented in Section 6.

2. Description of the BreakHis dataset

BreakHis is a dataset that includes totally 7909 images and eight sub-classes of breast cancers. The source data comes from 82 anonymous patients of Pathological Anatomy and Cytopathology Lab in Brazil. A normal and cancerous sample recording belongs to the dataset is shown in Fig. 1. The BreakHis dataset is divided into two main groups: benign tumors and malignant tumors. Of these patients, 24 have benign breast cancer and 58 have malignant breast cancer [12]. Histologically benign is a term referring to a lesion that does not match any criteria of malignancy — e.g., marked cellular atypia, mitosis, disruption of basement membranes, metastasize, etc. Normally, benign tumors are relatively "innocents", presents a slow-growing and remains localized. A malignant tumor is a synonym for cancer: lesion can invade and destroy adjacent structures and spread to distant sites to cause death [19,20].

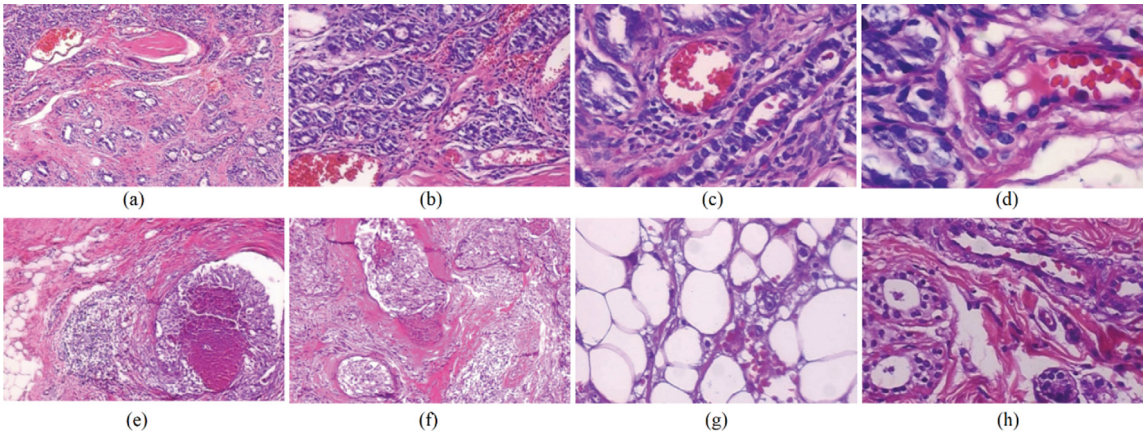

Fig. 1. Histopathological images of benign breast tumor and malignant breast tumor: **(a)** benign tumor with the magnification rate 40X. **(b)** benign tumor with the magnification rate 100X. **(c)** benign tumor with the magnification rate 200X. **(d)** benign tumor with the magnification rate 400X. **(e)** malignant tumor with the magnification rate 40X. **(f)** malignant tumor with the magnification rate 100X. **(g)** malignant tumor with the magnification rate 200X. **(h)** malignant tumor with the magnification rate 400X.

Table 1
The BreakHis dataset provides 7909 histopathological images collected from 82 anonymous patients. The samples were divided into benign and malignant tumors and eight sub-classes tumors that consist of four magnification rates: 40X, 100X, 200X, and 400X.

Class	Sub-class	Magnification rate				Total
		40X	100X	200X	400X	
Benign	A	114	113	111	106	444
	F	253	260	264	237	1014
	TA	109	121	108	115	453
	PT	149	150	140	130	569
Malignant	DC	864	903	896	788	3451
	LC	156	170	163	137	626
	MC	205	222	196	169	792
	PC	145	142	135	138	560
Total		1995	2081	2013	1820	7909

BreakHis is separated into benign and malignant tumors that consist of four zooming rates: 40X, 100X, 200X, and 400X. The benign and malignant cells are divided into subclasses according to cell diameter and shape by pathologists. Therefore, the dataset currently contains four histopathological distinct types of benign breast tumors. These, adenosis (A), fibroadenoma (F), phyllodes tumor (PT), and tubular adenoma (TA).

The names of the malignant tumor classes are ductal carcinoma (DC), lobular carcinoma (LC), papillary carcinoma (PC), and mucinous carcinoma (MC) [12]. The statistical information of the disease types is given in Table 1. Also, each image consists of three-channel RGB and each channel occurs at a depth of eight bits. Each image has 700 $\times$ 460 pixel resolution and the image format is PNG. All data given in Table 1 were used in this study.

3. Models and methods

3.1. Deep learning models

In this section, the AlexNet [21], VGG-16, and VGG-19 [22] models are considered and described briefly. These three models have already announced their name in ImageNet competitions. AlexNet is one of the most important CNN architectures. This architecture consists of convolution, pooling and FC layers. The input size of this model is 227 $\times$ 227 pixels. Filters form the output of the related layer by applying convolution to the inputs got from the previous layer. Filters that are hovered over the image in the convolution layer are typically selected as 3 $\times$ 3 or 5 $\times$ 5 pixels. The convolutional layer is based on the process of circulating a particular filter over the input image. This convolution process results in an activation map. The activation map consists of local discriminative features [23]. The pooling layers are used in AlexNet architecture. A pooling layer has a structure that maintains the image features, reduces image size and computation costs [24]. This structure reduces the number of parameters in the model operation and does this by preserving the information obtained from the image [25].

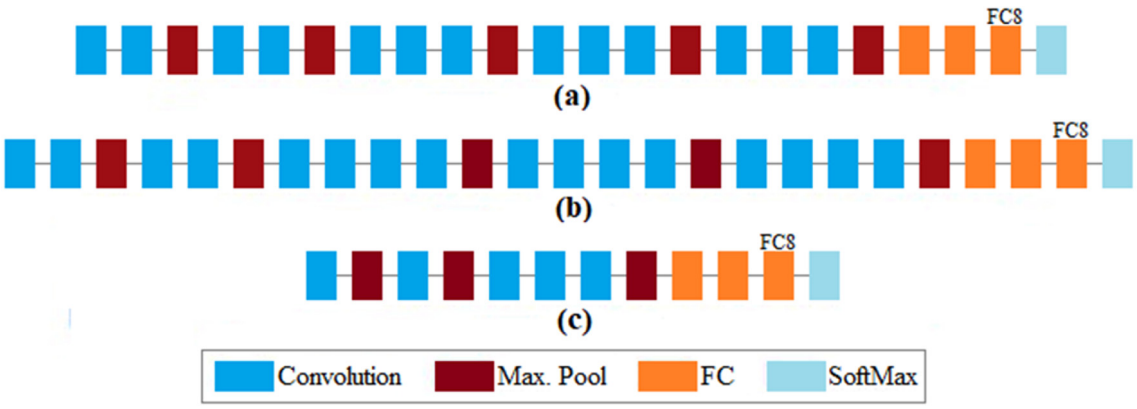

Fig. 2. The schematic diagram of (a) VGG-16, (b) VGG-19 and (c) AlexNet models.

VGG-16 consists of convolutional, pooling and FC layers. It contains a total of 21 layers [22]. The most important feature of this architecture is having an increasing network structure. The input size of the model is 224 × 224 pixels. The filter size in the convolutional layer is 3 × 3. In this architecture, the final layers have FC layers used for feature extraction. In this architecture, Softmax is used as the activation function of the last layer [26,27].

The VGG-19 consists of 24 layers: convolution layers, pool layers, and FC layers. The number of steps is two and the maximum pooling layer is used. The VGG-19 contains about 138 million computational parameters. Compared to AlexNet, the VGG-19 is a deeper CNN architecture with more layers. To reduce the number of parameters in such deep networks, it uses small filters in all convolutional layers. The size of the selected filter in this architecture is 3 × 3 [28].

As mentioned in this study, the last fully connected layer, FC8 activation layer, has been used for deep feature extraction. Softmax function is used as the output function of the last layer. The architectural designs of AlexNet, VGG-16 and VGG-19 models are illustrated in Fig. 2.

3.2. Optimization methods

The main purpose of optimization methods is to update the weights at every batch to find global minima.

In the Stochastic Gradient Descent (SGD) method [16,29], the weights update is performed for each training set. Because of this reason, it tries to achieve the goal as early as possible time. The formulation of SGD optimization is shown in Eq. (1). Here, Θ is the weight vector to be updated, α is the learning coefficient and $\nabla_{\Theta} J(\Theta)$ is the cost function.

$$\Theta_t = \Theta_{t-1} - \alpha \nabla_{\Theta} J\left(\Theta; x^i, y^i\right) \tag{1}$$

Stochastic Gradient Descent with Warm Restarts (SGDR) is a variant of learning rate scheduler, which gradually decreases the learning rate (LR) in defined cycles while model is training. SGDR uses cosine annealing, which decreases the learning rate in the form of half a cosine curve. Herein η_t is the learning rate at time step t (incremented each mini-batch), η^i_{max} and η^i_{min} define the range of desired learning rates, $T_{current}$ represents the number of epochs since the last restart (this value is calculated at every iteration and thus can take on fractional values), and T_i defines the number of epochs in a cycle [30]. The main formula of the SGDR method is shown in Eq. (2). The arguments used in the SGDR method are:

- Minimum LR–Maximum LR: The lower and upper bound of the learning rate range for the experiment.
- Steps per epoch: Number of mini-batches in the dataset. Calculated as (epoch size/batch size).
- LR Decay: Reduce the maximum LR after the completion of each cycle. To reduce the maximum LR by 20% after each cycle, this value was set to 0.8.
- Cycle Length: Initial number of epochs in a cycle.
- Mult. Factor: Scale epochs to restart after each full cycle completion.

$$\eta_t = \eta^i_{min} + \frac{1}{2}\left(\eta^i_{max} - \eta^i_{min}\right)\left(1 + \cos\left(\frac{T_{current}}{T_i}\right)\prod\right) \tag{2}$$

The RMSProp method [29,31] is adapted to the average of the slope weights and maintains learning rates per parameter. This method works well in online and non-stationary situations and performs the parameter update using momentum on the scaled slope.

The ADAM method [32] is one of the methods that update the learning coefficient in each batch. It adopts parameter-learning rates based on the average first moment in RMSProp. It also uses the average of the second moments of the slopes. This method is designed with the advantages of the RMSProp method. In other words, this is one of the optimization methods that update the learning coefficient in each batch. The most important feature of the ADAM method is that it

adjusts the learning rate of the weight parameters by estimating the first and second gradient moments in the model network. The ADAM uses the exponential moving averages that are evaluated on a valid mini-batch event and calculated on the gradient. The past gradient (m_t) and past square gradient (V_t) averages are calculated according to Eqs. (3) and (4). β is the variable used to calculate hyper parameters and usually takes values between 0.9 and 0.999 [33].

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1)\, g_t \tag{3}$$

$$V_t = \beta_2 V_{t-1} + (1 - \beta_2)\, g_t^2 \tag{4}$$

3.3. Machine learning method

The Softmax activation function is often used in the final layer of a neural network-based classifier. Such networks are commonly trained under a log loss or cross-entropy regime, giving a non-linear variant of multinomial logistic regression. The Softmax method is generally used in classification processes where the classification label can take more value [34]. Softmax was used as the activation function in the last layer of the BreastNet model like in the present CNN models used in this study [35,36].

A multi-layer perceptron (MLP) is a forward-class artificial neural network. An MLP consists of at least three-node layers: an input layer, a hidden layer, and an output layer. The MLP employs a supervised learning technique called training for backpropagation. Multiple layers and non-linear activation distinguish MLP from a linear sensor and MLP separates data that cannot be separated linearly [37].

3.4. The proposed CNN model: BreastNet

CNNs are one of the most convenient architectures for the image classification task. CNNs use filters methods to extract the most efficient features from the pixels of an image [38] consist of three bases:

Convolutional layers include convolution filters that are applied to the input image. Mathematical operations are applied to the feature map to obtain a single value in each region where the filter passes. After each convolutional layer typically ReLU activation function applied to prevent vanishing gradient problem [39,40].

Pool layers summarize the presence of features in patches of the feature map. On the other hand, these layers are utilized to reduce the size of the previous layer output. The working principle of these layers is based on the sliding window. Applies a statistical function on values in a certain window-size known as kernel. In general, there are three types of pooling methods. These are max pooling, mean pooling, sum pooling. Maximum pooling is often preferred for CNN-based tasks. It extracts sub-regions of the feature map (e.g., 3×3 sub-region), takes their maximum value and deletes all other values [41].

The dense layer is known as a fully connected layer. Dense layers perform classification on the features extracted from the convolutional and pooling layers. The task of a dense layer, all nodes in the layer is connected to each node in the previous layer [38,42].

CNN is composed of blocks that perform feature extraction. The final dense layer in CNN contains a single artificial neuron node for each target class with a Softmax activation function to generate probabilities between 0–1 for each artificial neuron. The sum of these estimated Softmax probabilities are equal to one [43].

BreastNet architecture is an end-to-end model. That is, each image data entering the proposed model performs probability distributions between classes without another type of feature encoders or decoders. BreastNet model consists of the following modules in general: convolutional block attention module (CBAM), dense block, residual block and hypercolumn technique. CBAM is an attention layer. It provides identification of important key places in histopathological images. In other words, it selects the important area of the image and allows to model to focus there. Thus gives higher accuracy for the classification process and saves time. The CBAM includes both the channel attention module and the spatial attention module. The Residual block makes the gradients smoother. It was used in the BreastNet model to prevent overfitting and underfitting [44,45]. The Hypercolumn technique examines BreakHis images on various scales and contributes to the understanding of the disease. In other words, it provides more stable results in the classification stage and improves classification performance [46]. The general design and parameter values of the BreastNet model are shown in Fig. 7.

3.4.1. Convolutional Block and Dense Block

These blocks consist of Conv2D/Dense (FC), batch normalization and ReLU activation layers. The main purpose of these blocks is to extract relevant features from input tensors (like the image (2D) or neuron (1D) output). Batch normalization is performed very well when it used just after convolution and dense layers [47]. As shown in Fig. 3, Convolutional Block and Dense Block were integrated into our proposed CNN model. The parameter values used in this study are shown in Fig. 3.

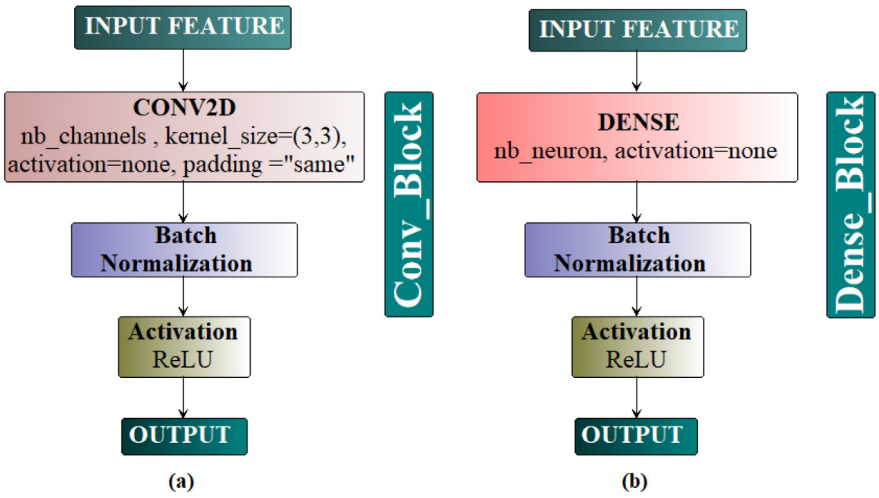

Fig. 3. The schematic diagram of (a) Convolutional Block, (b) Dense block.

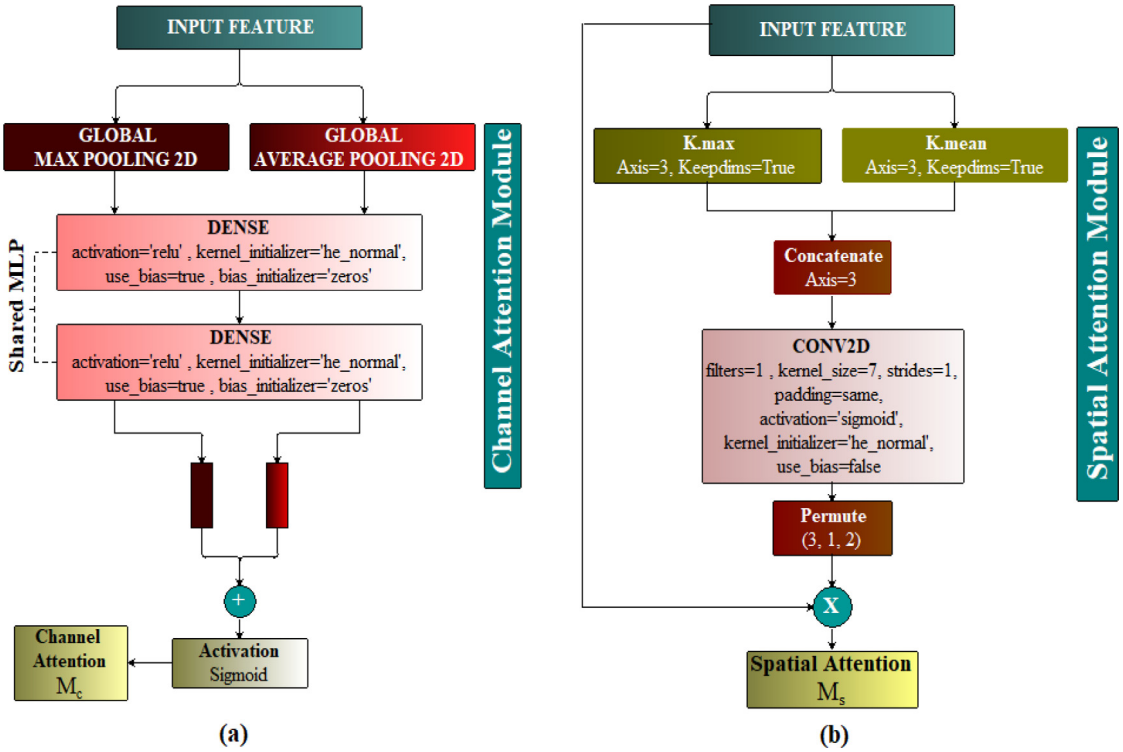

Fig. 4. The schematic diagram of (a) Channel attention module, (b) Spatial attention module.

3.4.2. CBAM

CBAM is an effective module for feed-forward CNN model. CBAM examines the feature map extracted from images at two dimensions: channel and spatial. The aim is to draw attention maps from the channel and spatial dimensions. After, the attention maps are multiplied by the input feature map for adaptive feature refinement. CBAM can be used in any CNN architecture [44]. CBAM includes channel attention module and spatial attention module.

Channel attention module refers to the channels of the feature map. This module shows the regions that would be focused on the features map and contributes to the classification results more efficient. With this module, feature maps obtained from input data are compressed [48].

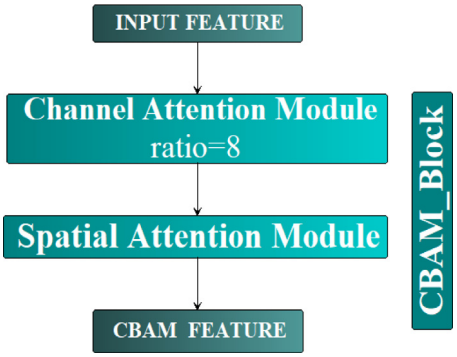

Fig. 5. The schematic diagram of the general design of the CBAM block.

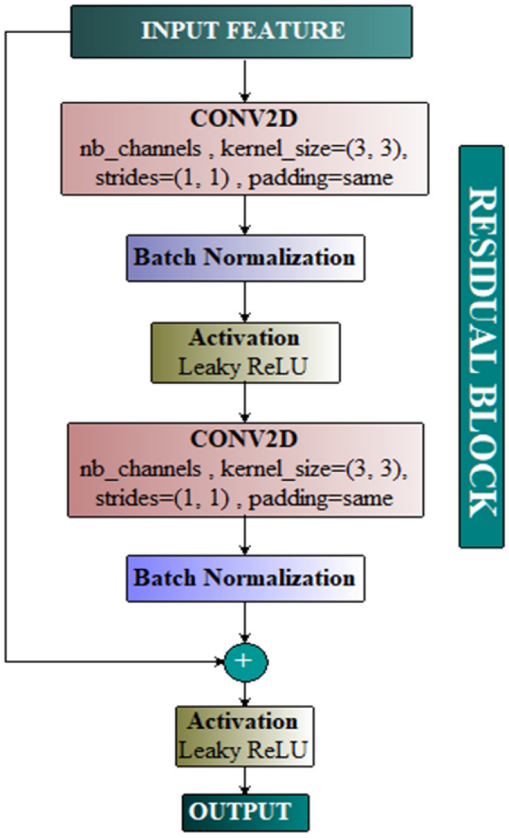

Fig. 6. The schematic diagram of the residual block.

In order to collect spatial information, both the average and maximum pooling information are used together [49]. This information is obtained with both average-pool and max-pool features simultaneously. The design of both modules is shown in Fig. 4. As shown in Fig. 4, the channel module utilizes both max-pooling outputs and average-pooling outputs. The spatial module utilizes similar two outputs that are pooled along the channel axis and forward them to a convolution layer. The values specified in Fig. 4 are the parameter values used in this study.

With channel module, average and maximum pooled features process together. Two features set are collected in a shared network. The shared network is composed of MLP with one hidden layer. It subjects to a classification process with MLP [50].

To compute the spatial attention, we first apply average-pooling and max-pooling operations along the channel axis and concatenate them to generate an efficient feature descriptor. The extracted information is then transferred to the convolutional network.

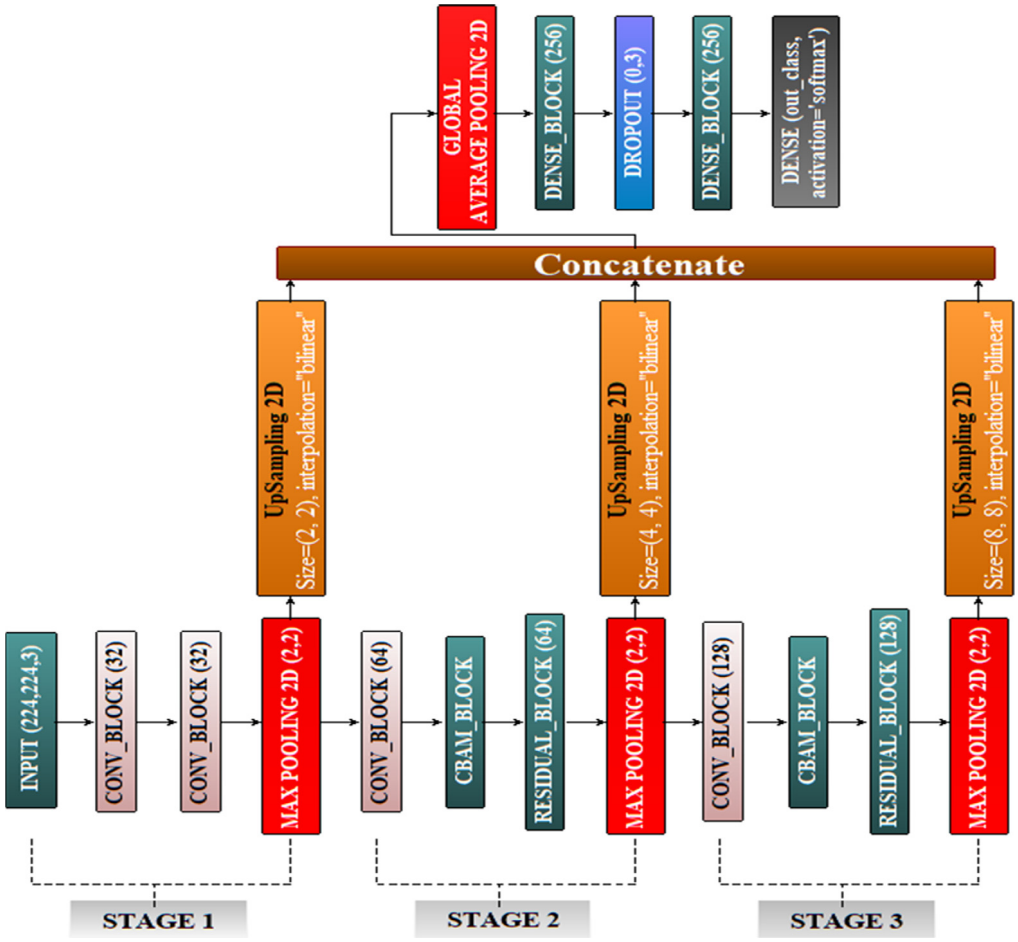

Fig. 7. The general design of the proposed BreastNet model.

As a result, the channel and the spatial module focus on attention. Channel and spatial module focuses on 'what' and 'where' respectively. And they contribute to the classification process by extracting more efficient features [48,49]. The general design of CBAM consisting of the channel and spatial attention modules is shown in Fig. 5.

3.4.3. Residual block

In traditional neural networks, each layer feeds the next layer. In a network with residual blocks, each layer feeds the next layer and initial layer connects the layers about two-three hops away with addition operation [51].

For conventional deep learning networks, they usually have convolution layers then fully connected layers for classification task like AlexNet, ZFNet, and VGGNet, without any skip/shortcut connection, these structures are called sequential networks. When the sequential network is deeper (layers are increased), the problem of vanishing/exploding gradients occurs [52]. To reduce the negative results of this problem we used residual block after CBAM blocks in this study. Used parameters and the design of the residual block are shown in Fig. 6.

3.4.4. Hypercolumn technique

Hypercolumn at a pixel is the vector of activations of all CNN units above that pixel. By this mean, spatial location information can be brought from earlier layers, and have a more accurate prediction result. Since CNN normally uses the output of the last layer as a feature representation, it does not use the features of the preceding layers. Therefore, hypercolumn in a pixel with the hypercolumn technique is the vector of all CNN units over that pixel. In other words, it also includes the features of previous layers. In this way, allows the network to achieve better results. This technique consists of UpSampling2D with bilinear interpolation and "Concatenate" layers [46]. The function of bilinear upsampling2d determines the values of new neighboring pixels by looking at neighboring pixel values. In addition, by enabling the size of the image to be input size; The effect of attention blocks has been observed by the model. Thus, the better estimate of the BreastNet architecture has been provided for this study. The concatenate parameter allows to combine the features of

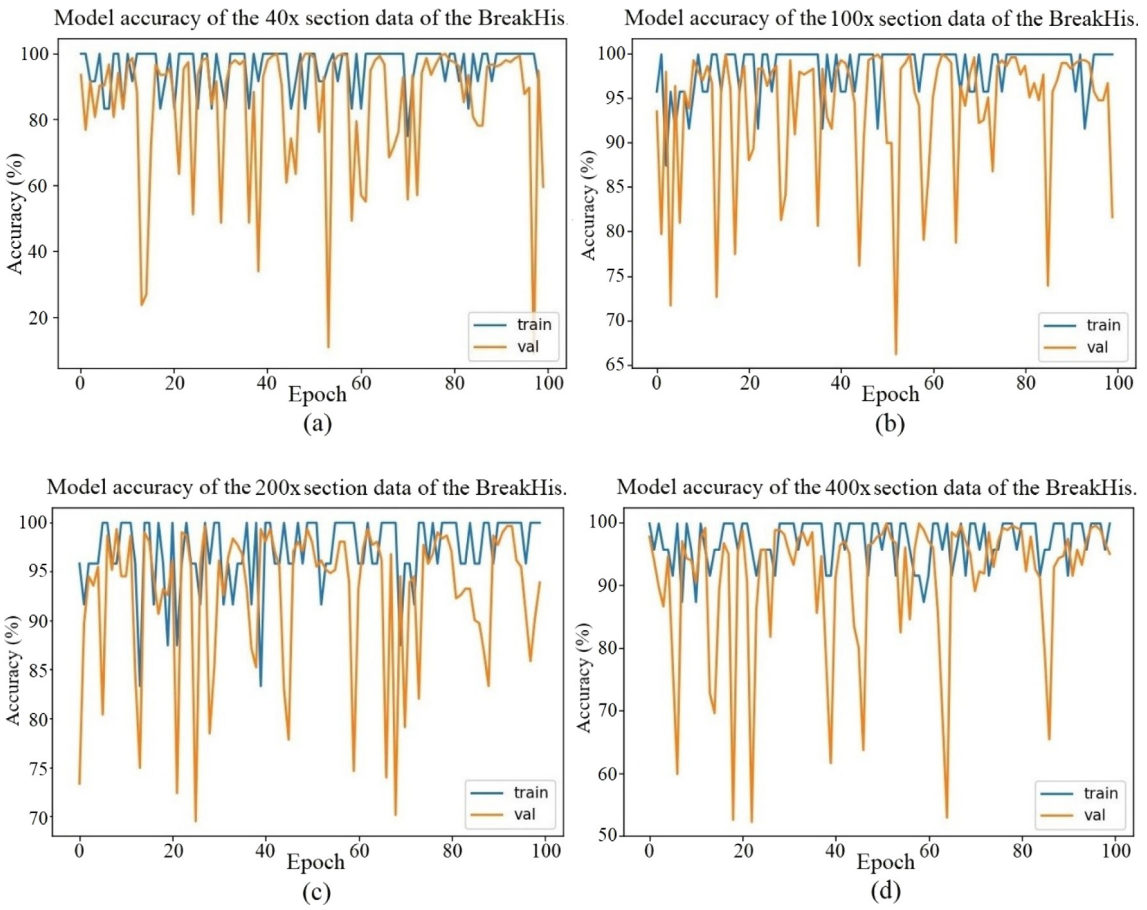

Fig. 8. Visual graph of the BreastNet model training and validation accuracy which has the best score in test dataset among the 5 cross-validation models for all epochs, **(a)** Training and validation accuracy graph for 40X images. **(b)** Training and validation accuracy graph for 100X images. **(c)** Training and validation accuracy graph for 200X images. **(d)** Training and validation accuracy graph for 400X images.

the same size in the third dimension (the channels are combined). In addition, the dropout parameter is used to prevent overfit–underfit problems in this study.

Histopathological images in this study can be captured with different zoom rates. Instead of creating models for every different zoom rates, with hypercolumn, the model examines histopathological images at three different scales (stages) and selects relevant features. Also, hypercolumn is preferred in BreastNet model architecture for providing high accuracy in this study. The general design of the BreastNet model is shown in Fig. 7. The parameters specified in Fig. 7 are the values used in the BreakHis dataset.

4. Results

This study was carried out by using MATLAB (R2018b) and Python software. In all experiments, the present models were compiled using the MATLAB whereas the results of the proposed BreastNet model were obtained using the Python. Also, the BreakHis dataset was divided into two sets as 80% training and 20% testing, respectively. The present CNN models used in this study were used with the transfer learning approach. The parameter values of the present CNN models used in the experiments are given in Table 2. As inferred from Table 2, using these parameters with default values were preferred. Also, mini-batch size was adjusted to 32. Mini-batch is the number of data, you update the weights and biases during backpropagation. Mini-batch value is selected to be less than the total dataset size. Generally, a number that is divisible by the total dataset size is preferred. Mini-batch contributes to learning processes by balancing the convergence rate of the network as well as accurate estimation [53]. However, the size of the mini-batch was not increased more since this is costly in terms of memory usage [54].

The parameter values used in the BreastNet architecture are shown in Table 3. In the proposed model, the K-fold cross-validation value was chosen as five. In the last layer of this architecture, Softmax was selected as the activation function. The epoch value for each cross-validation was set to 100. Data augmentation was performed to extract more

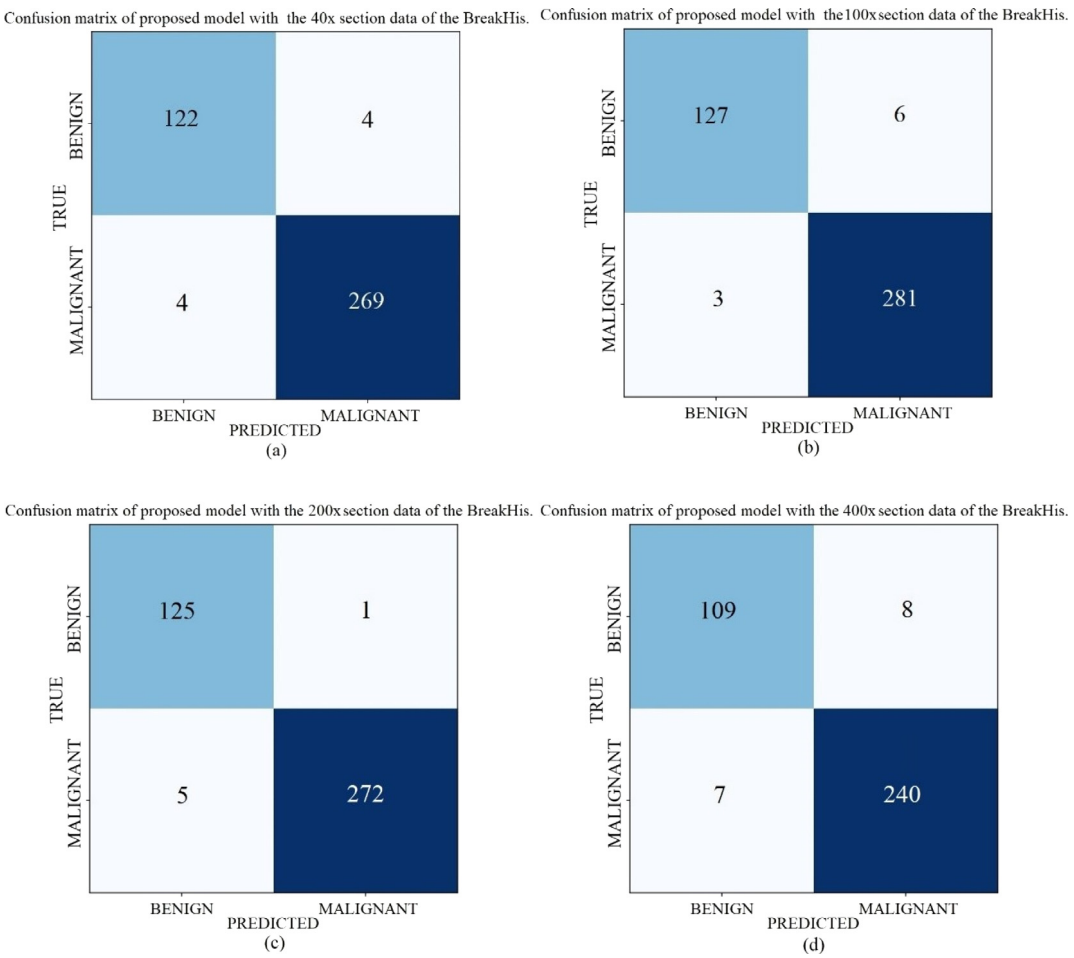

Fig. 9. The confusion matrix of the BreastNet model which has the best score in test dataset among the 5 cross-validation models, **(a)** The confusion matrix of the model trained using 40X images, **(b)** The confusion matrix of the model trained using 100X images, **(c)** The confusion matrix of the model trained using 200X images, **(d)** The confusion matrix of the model trained using 400X images.

Table 2
The values of the parameters used in the CNN architectures.

Used software	CNN architecture	Image size	Optimization	Momentum	Decay	Mini batch	Learning rate
MATLAB	AlexNet VGG-16 VGG-19	227 × 227 224 × 224 224 × 224	SGD	0.9	1e-6	32	0.0001

efficient attributes from each sample. The arguments (flip, shift, brightness change, and rotation, etc.) and hyperparameter values used for this specific purpose are shown in Table 3. ADAM method was chosen for optimization, and SGDR method was used to improve accuracy and learning speed during training. Also, the mini-batch size of the BreastNet model was adjusted to 24. The reason for reducing the mini-batch size arises from the insufficient in graphics card memory.

The models were compiled with GPU support. The simulation environment was run on 64-bit Windows 10 operating system. Other hardware details of the computer are that NVIDIA GeForce GTX 1070 - 8 GB graphics card, Intel®Core™ i7-8700 processor and 16 GB RAM.

The differences in the mini-batch values in Tables 2 and 3 suffered from the following reasons: The mini-batch value in Table 2 was used to compile present CNN models on MATLAB. The mini-batch value in Table 3 was compiled on the BreastNet architecture in Python. These mini-batch values were the largest values that can be used with the existing hardware units in the experiments.

In order to measure the performances of the models, accuracy (Acc), sensitivity (Se), specificity (Sp), precision (Pr) and f1-score metrics derived from confusion matrix were used and the formulations of the metrics are described as

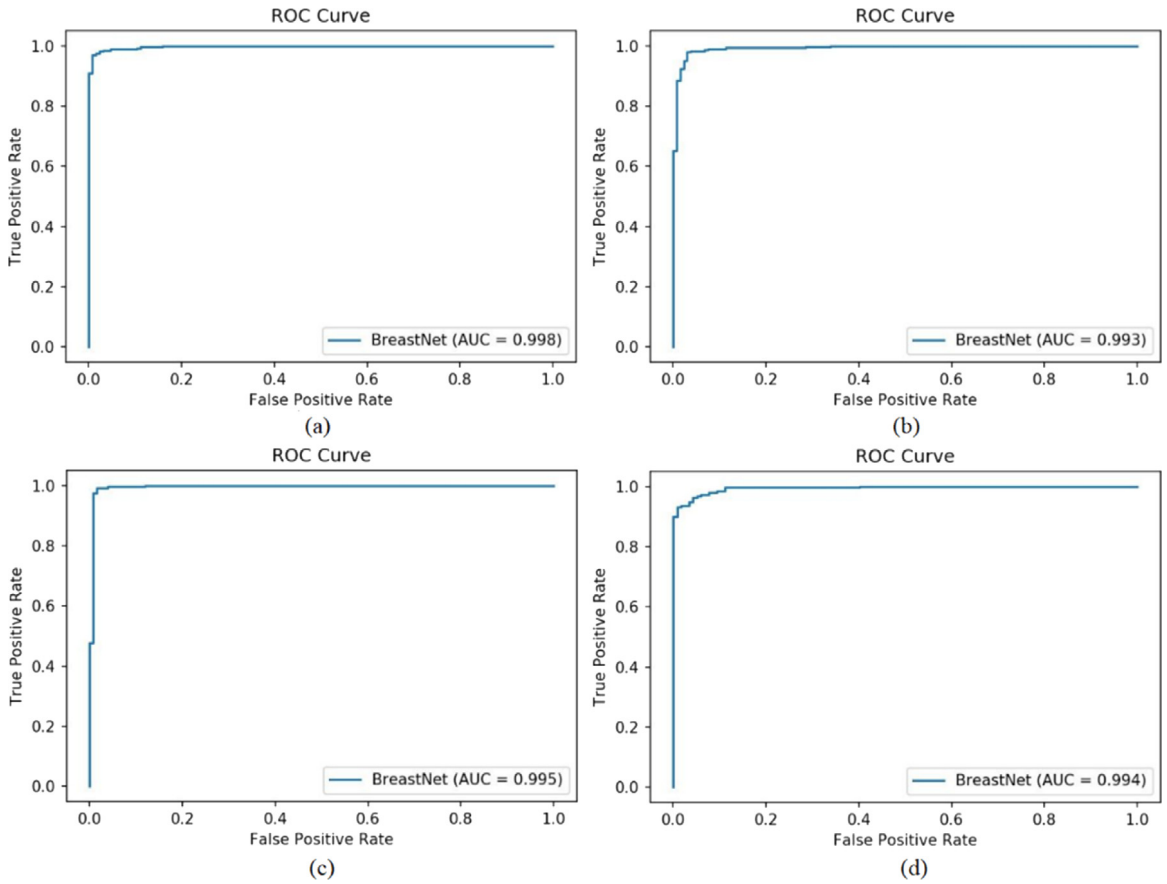

Fig. 10. The ROC curves of the BreastNet model, **(a)** The proposed model was trained using 40X images, **(b)** The proposed model was trained using 100X images, **(c)** The proposed model was trained using 200X images, **(d)** The proposed model was trained using 400X images.

Table 3
The values of the parameters in the proposed BreastNet architecture.

		Optimization & Parameters	Data augmentations parameters	LR scheduler & parameters
Model	BreastNet	**ADAM**	Vertical Flip = 0.5	**SGDR**
Used software	Python	Beta 1 = 0.9	Horizontal Flip = 0.5	Min LR = 1e-6
Image size	224 × 224	Beta 2 = 0.999	Random brightness Contrast = 0.3	Max LR = 1e-3
Mini batch	24	Decay = 0.0	Shift scale rotate = 0.5	Steps Per Epoch = 5
Framework	Keras		Shift limit = 0.2	LR Decay = 0.9
Loss type	Categorical Cross-entropy		Scale Limit = 0.2 Rotate limit = 20°	Cycle length = 10 Mult. Factor = 2.

follows [55]:

$$Acc = \frac{(\mathrm{TP} + \mathrm{TN})}{(\mathrm{TP} + \mathrm{FN}) + (\mathrm{FP} + \mathrm{TN})} \tag{5}$$

$$Se = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FN})} \tag{6}$$

$$Sp = \frac{(\mathrm{TN})}{(\mathrm{TN} + \mathrm{FP})} \tag{7}$$

$$Pr = \frac{(\mathrm{TP})}{(\mathrm{TP} + \mathrm{FP})} \tag{8}$$

Table 4
The comparison results of the existing CNN models and the BreastNet model.

CNN Model	Used BreakHis data type	Acc (%)	Se (%)	Pr (%)	F_1 - Score (%)
AlexNet	40X	70.05	63.25	95.72	76.17
	100X	80.83	77.17	87.56	82.03
	200X	84.22	93.24	73.80	82.39
	400X	84.38	82.35	87.50	84.85
VGG -16	40X	77.81	72.42	89.84	80.19
	100X	72.02	97.75	45.08	61.70
	200X	81.28	96.07	65.24	77.71
	400X	83.24	81.28	86.36	83.74
VGG -19	40X	83.96	91.51	74.87	82.36
	100X	81.87	78.34	88.08	82.93
	200X	81.82	95.42	66.84	78.61
	400X	82.39	80.98	84.66	82.78
BreastNet	40X	**97.99**	97.68	97.68	97.68
	100X	**97.84**	97.21	97.80	97.50
	200X	**98.51**	98.70	97.89	98.28
	400X	**95.88**	95.16	95.36	95.26

Table 5
Results of comparison of existing CNN models with the BreastNet using all of the BreakHis data.

CNN model	Used BreakHis data type	Acc (%)	Se (%)	Pr (%)	F_1- Score (%)
AlexNet	40X & 100X & 200X & 400X	78.49	72.18	92.74	81.18
VGG −16	40X & 100X & 200X & 400X	83.13	78.90	90.46	84.29
VGG −19	40X & 100X & 200X & 400X	83.27	83.13	83.47	83.30
BreastNet	40X & 100X & 200X & 400X	**98.80**	98.35	98.84	98.59

$$F_1 - Score = \frac{(2 \times \mathrm{TP})}{(2 \times \mathrm{TP} + \mathrm{FP} + \mathrm{FN})} \tag{9}$$

where True positive (TP) represents the number of malignant breast images classified as cancerous breast whereas True Negative (TN) represents the number of normal breast images classified as normal breast. Also, False positive (FP) represents the number of normal breast images incorrectly classified as cancerous breast while False Negative (FN) represents the number of cancerous breast images misclassified as normal breast.

The experiment consists of three parts. In the first part, the results of the designed CNN model with the results obtained by using the present CNN models were obtained and compared. The comparison results are shown in Table 4. The BreastNet model compared to the present CNN models, for each data type group (40X, 100X, 200X, 400X) the proposed model achieved superior results. This model achieved the best success rate using 200X data with 98.51%. According to the results of the BreastNet model, the learning accuracy graphs are shown in Fig. 8 and the confusion matrix are given in Fig. 9. ROC curves are also shown in Fig. 10.

In the second part of the experiment, BreakHis data (40X, 100X, 200X, 400X) were combined. The combined data was obtained classification results using the present CNN (AlexNet, VGG-16, VGG-19) models. Then, the combined data was classified by the BreastNet model. It was then compared with the existing CNN models with the BreastNet model. The classification success of the BreastNet increased by 0.29% and obtained as 98.80%. The statistical information on the results is shown in Table 5. The combined data was classified by the proposed model. The learning accuracy, confusion matrix and ROC curve of the classification results are shown in Fig. 11.

Benign and malignant classes of BreakHis data were divided into four sub-classes. Benign subclasses are adenosis, fibroadenoma, phyllodes tumor, tubular adenoma. Malignant subclasses are ductal carcinoma, lobular carcinoma, mucinous carcinoma, papillary carcinoma. In the three-part of the experiment, the performance of the BreastNet model in quadruple sub-classification success was measured. When Table 6 is examined, the average classification success of subclasses of Benign data was 97.78%. Likewise, the average classification success of the subclasses of Malignant data was 96.41%. The training & validation graph and the confusion matrix of the classification results are shown in Fig. 12.

5. Discussion

Breast cancer is on the front bench type among hundreds of cancer diseases. The incidence of this disease is increasing day by day, especially among women. If this disease is not diagnosed in a timely manner, the rate of death is fairly high. The early diagnosis of this disease is associated with rapid and accurate results of the image processing techniques as regarding the computation approaches. In this scope, the CNN models have a great advantage in terms of giving faster and better results compared to the conventional machine learning methods [56]. Especially, these models have a powerful ability to extracting the discriminative local features. Furthermore, the models can also reflect this ability to the classification

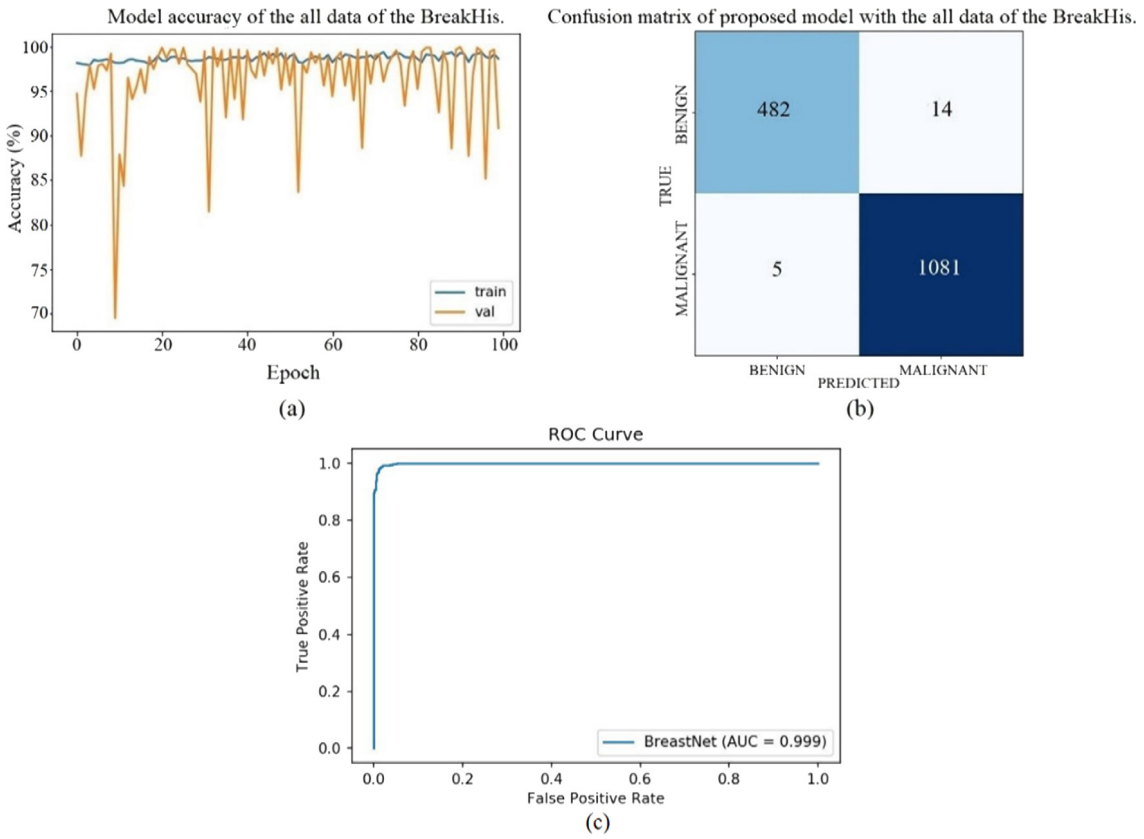

Fig. 11. BreastNet model which has the best score in test dataset among the 5 cross-validation models in whole trained on whole BreakHis data, **(a)** Training and validation accuracy of the BreastNet model, **(b)** The confusion matrix of the BreastNet model, **(c)** The ROC curve of the BreastNet model.

Table 6
The sub-classification success of the BreastNet model using benign and malignant data.

BreakHis data type	Sub classification	Accuracy (%)	Recall (%)	Precision (%)	Average Acc. (%)
Benign	Adenosis	98.59	96.08	97.03	97.78
	Fibroadenoma	97.18	97.54	95.65	
	Phyllodes tumor	96.98	91.25	90.12	
	Tubular adenoma	98.39	94.60	98.13	
Malignant	Ductal	93.46	95.26	94.29	96.41
	Lobular	94.01	72.30	81.68	
	Mucinous	98.90	98.71	93.87	
	Papillary	99.26	97.20	95.41	

process. As a result, numerous studies have been focused on machine learning and deep learning models to early diagnose breast cancer through BreakHis data. With the BreastNet model, we realized a classification analysis with BreakHis data. The results of our study were compared with the other studies using this dataset as shown in Table 7.

Shallu et al. [57] used BreakHis dataset. 40X, 100X, 200X, and 400X have applied the classification by combining the data. They undertook feature extraction using the previously trained VGG-16 architecture. They used LR as a classifier. The success of the classification was 92.60%. As presented in Table 7, The success rate of the proposed model was 98.80%. So, our model produced better results compared to the related study.

Yangqin Feng et al. [58] proposed a new feature extractor, called deep manifold preserving autoencoder, to learn discriminative features from unlabeled data. Then, they integrated the proposed feature extractor with a Softmax activation function to classify breast cancer histopathology images. They were identified the features obtained in the deep model with the sampling patches from the breast histopathology images. They tried to classify the new sampled patches of the data and combine the classification results of these patches to predict the label of the data. They have achieved success by separately applying the model to the benign and malignant data.

Yun Gu et al. [59] are proposed to generate the discriminative binary codes by exploiting the histopathological images with multiple magnification factors from the DCMM model. DCMM model with a mutual guidance learning paradigm

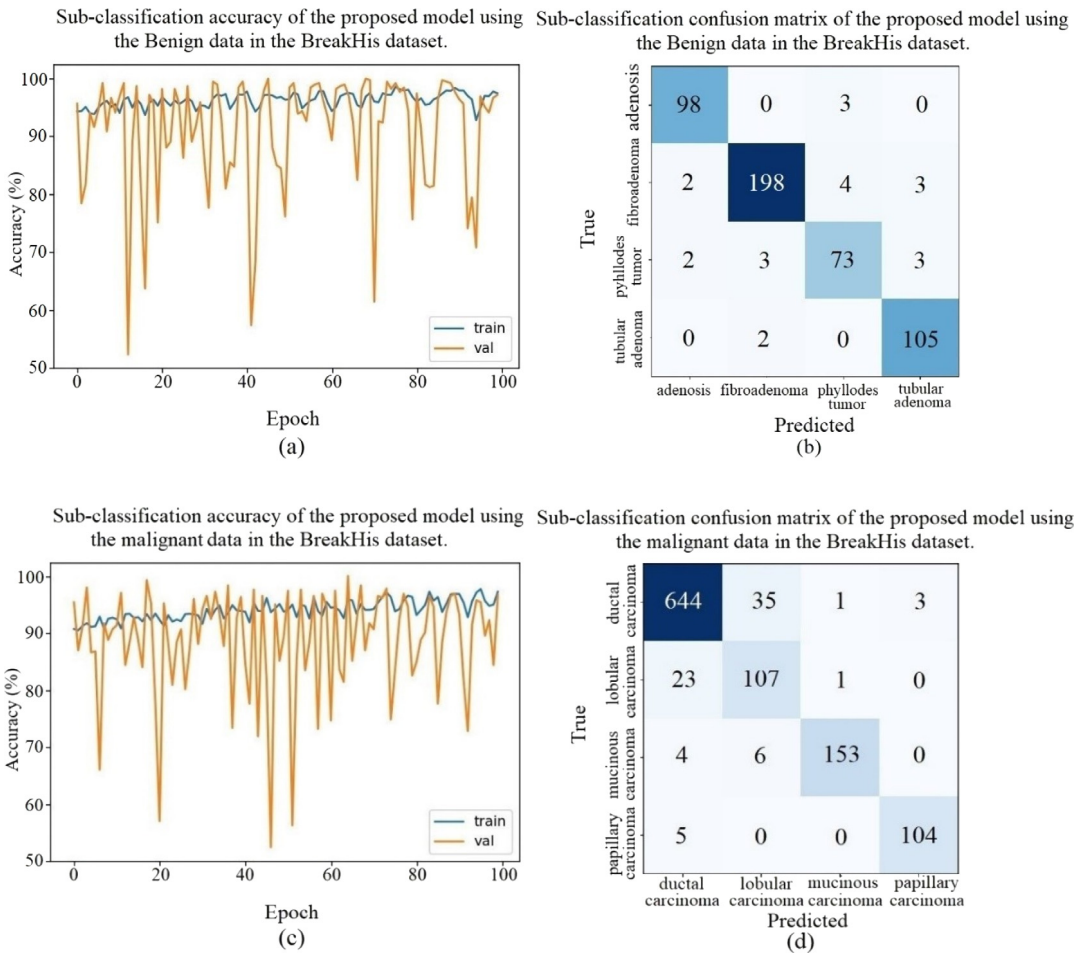

Fig. 12. Sub-classification using benign and malignant data with the BreastNet model, (a) training and validation accuracy graph of the BreastNet model in benign data, (b) the confusion matrix of the BreastNet model in benign data, (c) training and validation accuracy graph of the BreastNet model in malignant data, (d) the confusion matrix of the BreastNet model in malignant data.

is conducted based on high-low magnification pairs of image data and a densely-connected architecture is applied to fully utilize the cross-magnification information. They used multiple magnification levels to learn the binary codes for histopathological BreakHis dataset. The best performance in the classification was achieved by using the 400X dataset at a rate of 96.31%.

Hamed Erfankhah et al. [60] used 3 different datasets. They used the LBP method in BreakHis and other datasets to extract features and extracted the texture features from a circularly symmetric pixel neighborhood. Histogram values were extracted to differentiate between homogeneous and heterogeneous regions of image tissue. The homogeneous regions have low values, while the non-homogeneous regions have high values in the heterogeneity images. In their study, in the LBP method, feature extraction was based on the heterogeneous and homogeneous status of tissue regions. They used the SVM method as the classifier and achieved the best success at 88.30% in the BreakHis dataset.

Daniel Lichtblau et al. [61] proposed a combination of six machine learning techniques employed for the histopathological image classification task. In their work, they used the AlexNet model and the pre-trained weights and without fine-tuning. In the classification stage, five classifiers (random forests, nearest neighbors, LR, naive Bayes, and SVM) exhibiting different features were used. Each classifier was independently trained and applied to the test set. In addition, Fourier trig transform (FTT) with principal component analysis (PCA) was used in the model. The success rate was determined as 86.67% by classifying the selected features with various classifiers.

Zhongyi Han et al. [62] proposed CSDCNN model in their study. This model consists of an input, convolutional, pooling layer that similar the CNN architecture. In addition, the CSDCNN model has used its own data augmentation technique. Also, they choose Softmax as the activation function in the last layer of the architecture in their study. The success of the model was 97.1%.

Table 7
Comparison of the success of studies using the BreakHis data set with the success of the BreastNet model.

Study	Year	Used BreakHis data type	# of Classes	Model	Classifier/ Function	Acc (%)
Shallu et al. [57]	2018	All data	Benign & Malignant	VGG-16	Linear Regression (LR)	92.60
Yangqin Feng et al. [58]	2018	40X	Benign Malignant	Deep Manifold Persevering Autoencoder (DMPA)	Softmax	95.03 84.56
		100X	Benign Malignant			93.15 82.98
		200X	Benign Malignant			99.36 83.57
		400X	Benign Malignant			95.27 83.45
Yun Gu et al. [59]	2018	40X 100X 200X 400X	Benign & Malignant	The Densely-Connected Multi-Magnification (DCMM)	Linear layer	95.62 95.03 97.04 **96.31**
Hamed Erfankhah et al. [60]	2019	40X 100X 200X 400X		Local Binary Patterns (LBP)	Support Vector Machine (SVM)	88.30 88.30 87.10 83.40
Daniel Lichtblau et al. [61]	2019	40X 100X 200X 400X		AlexNet	Various classifiers	81.61 84.47 86.67 83.15
Zhongyi Han et al. [62]	2017	40X 100X 200X 400X		Class Structure-based Deep Convolutional Neural Network (CSDCNN)	Softmax	97.10 95.70 96.50 95.70
Proposed model	2019	40X 100X 200X 400X		Designed CNN Model	Softmax	**97.99** **97.84** **98.51** 95.88

We achieved the best results using the proposed model compared to all the studies given in Table 7. The most efficient results were obtained from 40X, 100X and 200X histopathological images. Especially, the best result in Table 7 was obtained with the proposed CNN model by using the 200X images with 98.51% classification success was achieved.

6. Conclusion

The current study focused on improving the classification accuracy on the BreakHis data. It was comprised of the histopathological images separated into two different classes as benign or malignant. The classification accuracies of the BreastNet model is superior or approximate to the previously attempted techniques on the same dataset. This model can be used in all microscopic images at different magnification rates. The classification was carried out without using any preprocessing procedure for the histopathological images. As a result, the BreastNet model in this study and four different magnification factors (40X, 100X, 200X, 400X) were classified as binary classification. The best classification result was 98.51%. The binary classification was repeated by combining all data of the magnification factors. The best classification result was obtained with an increase of 98.80%. The same model was applied to the subclasses of BreakHis data. The classification results of the subclasses consisting of four categories can be evaluated as promising.

In the future, we will examine the proposed model on a different dataset. The proposed method can be generalized to the design of high-performance computer-aided diagnosis systems for other medical imaging tasks in the future.

Data availability and open source code

The source codes of the proposed model can be downloaded on https://github.com/Goodsea/BreastNet.

Funding

There is no funding source for this article.

Ethical approval

This article does not contain any data, or other information from studies or experimentation, with the involvement of human or animal subjects.

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.

References

[1] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Expert Syst. Appl. 36 (2009) 3240–3247, http://dx.doi.org/10.1016/j.eswa.2008.01.009.
[2] P. Fribert, L. Paulová, P. Patáková, M. Rychtera, K. Melzoch, Alternativní metody separace kapalných biopaliv z média při fermentaci, Chem. List 107 (2013) 843–847, http://dx.doi.org/10.3322/caac.21492.
[3] World Health Statistics, World Health Organization, 2018.
[4] Y. Guo, X. Shang, Z. Li, Identification of cancer subtypes by integrating multiple types of transcriptomics data with deep learning in breast cancer, Neurocomputing 324 (2019) 20–30, http://dx.doi.org/10.1016/j.neucom.2018.03.072.
[5] American Cancer Society, Breast Cancer Facts &amp Figures 2012–2014, Breast Cancer Facts Fig, 2013, pp. 1–44.
[6] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, A dataset for breast cancer histopathological image classification, IEEE Trans. Biomed. Eng. 63 (2016) 1455–1462, http://dx.doi.org/10.1109/TBME.2015.2496264.
[7] G.A. Tataroglu, A. Genc, K.A. Kabakci, A. Capar, B.U. Toreyin, H.K. Ekenel, I. Turkmen, A. Cakir, A deep learning based approach for classification of CerbB2 tumor cells in breast cancer, in: 2017 25th Signal Process. Commun. Appl. Conf, SIU, 2017, http://dx.doi.org/10.1109/SIU.2017.7960587.
[8] L.A. Torre, F. Islami, R.L. Siegel, E.M. Ward, A. Jemal, Global cancer in women: Burden and trends, Cancer Epidemiol. Biomark. Prev. 26 (2017) 444–457, http://dx.doi.org/10.1158/1055-9965.EPI-16-0858.
[9] G. Danaei, S. Vander Hoorn, A.D. Lopez, C.J.L. Murray, M. Ezzati, Causes of cancer in the world: Comparative risk assessment of nine behavioural and environmental risk factors, Lancet 366 (2005) 1784–1793, http://dx.doi.org/10.1016/S0140-6736(05)67725-2.
[10] M. Karabatak, M.C. Ince, An expert system for detection of breast cancer based on association rules and neural network, Expert Syst. Appl. 36 (2009) 3465–3469, http://dx.doi.org/10.1016/j.eswa.2008.02.064.
[11] C.A. Peña Reyes, M. Sipper, A fuzzy-genetic approach to breast cancer diagnosis, Artif. Intell. Med. 17 (1999) 131–155, http://dx.doi.org/10.1016/S0933-3657(99)00019-6.
[12] F.A. Spanhol, L.S. Oliveira, C. Petitjean, L. Heutte, Breast cancer histopathological image classification using convolutional neural networks, in: Proc. Int. Jt. Conf. Neural Networks, 2016, pp. 2560–2567, http://dx.doi.org/10.1109/IJCNN.2016.7727519, Octob (2016).
[13] F.F. Ting, Y.J. Tan, K.S. Sim, Convolutional neural network improvement for breast cancer classification, Expert Syst. Appl. 120 (2019) 103–115, http://dx.doi.org/10.1016/j.eswa.2018.11.008.
[14] N. Liu, E.-S. Qi, M. Xu, B. Gao, G.-Q. Liu, A novel intelligent classification model for breast cancer diagnosis, Inf. Process. Manag. 56 (2019) 609–623, http://dx.doi.org/10.1016/j.ipm.2018.10.014.
[15] D.M. Vo, N.-Q. Nguyen, S.-W. Lee, Classification of breast cancer histology images using incremental boosting convolution networks, Inf. Sci. (Ny) 482 (2019) 123–138, http://dx.doi.org/10.1016/j.ins.2018.12.089.
[16] A.C. Wilson, R. Roelofs, M. Stern, N. Srebro, B. Recht, The marginal value of adaptive gradient methods in machine learning, 2017, arXiv:1705.08292v2 [stat.ML] (22 May 2018).
[17] Q. Liao, Y. Ding, Z.L. Jiang, X. Wang, C. Zhang, Q. Zhang, Multi-task deep convolutional neural network for cancer diagnosis, Neurocomputing (2018) http://dx.doi.org/10.1016/j.neucom.2018.06.084.
[18] S. Khan, N. Islam, Z. Jan, I.U. Din, J.J.P.C. Rodrigues, A novel deep learning based framework for the detection and classification of breast cancer using transfer learning, Pattern Recognit. Lett. 125 (2019) 1–6, http://dx.doi.org/10.1016/j.patrec.2019.03.022.
[19] K.A. Chun, Case reports on the differentiation of malignant and benign intratracheal lesions by 18 F-FDG PET/CT, Med. (United States) 94 (2015) e1704, http://dx.doi.org/10.1097/MD.0000000000001704.
[20] H. Kjellin, H. Johansson, A. Höög, J. Lehtiö, P.J. Jakobsson, M. Kjellman, Differentially expressed proteins in malignant and benign adrenocortical tumors, PLoS One 9 (2014) http://dx.doi.org/10.1371/journal.pone.0087951.
[21] A. Krizhevsky, I. Sutskever, G.E. Hinton, Machine learning and computer vision group deep learning with tensorflow, 2012, http://cvml.ist.ac.at/courses/DLWT_W17/.
[22] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, 2014, ArXiv Prepr. ArXiv:1409.1556.
[23] J. Koushik, Understanding convolutional neural networks, 2016, http://dx.doi.org/10.1016/j.jvcir.2016.11.003.
[24] K. O'Shea, R. Nash, An introduction to convolutional neural networks, 2015, http://dx.doi.org/10.1007/978-3-642-28661-2-5.
[25] N. Passalis, A. Tefas, Learning bag-of-features pooling for deep convolutional neural networks, in: Proc. IEEE Int. Conf. Comput. Vis. 2017-Octo, 2017, pp. 5766–5774. http://dx.doi.org/10.1109/iccv.2017.614.
[26] M.E. Sertkaya, B. Ergen, M. Togacar, Diagnosis of eye retinal diseases based on convolutional neural networks using optical coherence images, in: 2019 23rd Int. Conf. Electron, 2019, pp. 1–5. http://dx.doi.org/10.1109/electronics.2019.8765579.
[27] M. Toğaçar, B. Ergen, M.E. Sertkaya, Zatürre Hastalığının Derin Öğrenme modeli ile tespiti detection of pneumonia with deep learning model, 31, 2019, pp. 223–230.
[28] Z. Huang, Nasrullah J. Wen, S. Song, M. Mateen, Fundus image classification using VGG-19 architecture with PCA and SVD, Symmetry (Basel) 11 (2018) 1, http://dx.doi.org/10.3390/sym11010001.
[29] S. Ruder, An overview of gradient descent optimization, 2016, pp. 1–14.
[30] I. Loshchilov, F. Hutter, SGDR: Stochastic gradient descent with warm restarts, 2017, pp. 1–16.
[31] M.D. Zeiler, ADADELTA: An adaptive learning rate method, 2012, http://doi.acm.org.ezproxy.lib.ucf.edu/10.1145/1830483.1830503.
[32] R. Shindjalova, K. Prodanova, V. Svechtarov, Modeling data for tilted implants in grafted with bio-oss maxillary sinuses using logistic regression, AIP Conf. Proc. 1631 (2014) 58–62, http://dx.doi.org/10.1063/1.4902458.

[33] A. Wibowo, P.W. Wiryawan, N.I. Nuqoyati, Optimization of neural network for cancer microRNA biomarkers classification, J. Phys. Conf. Ser. 1217 (2019) 012124, http://dx.doi.org/10.1088/1742-6596/1217/1/012124.
[34] P. Vamplew, R. Dazeley, C. Foale, Neurocomputing Softmax exploration strategies for multiobjective reinforcement learning, Neurocomputing 263 (2017) 74–86, http://dx.doi.org/10.1016/j.neucom.2016.09.141.
[35] M. Toğaçar, B. Ergen, Deep learning approach for classification of breast cancer, in: 2018 Int. Conf. Artif. Intell. Data Process, 2018, pp. 1–5. http://dx.doi.org/10.1109/idap.2018.8620802.
[36] Z. Cömert, A.F. Kocamaz, Fetal hypoxia detection based on deep convolutional neural network with transfer learning approach, in: R. Silhavy (Ed.), Softw. Eng. Algorithms Intell. Syst., Springer International Publishing, Cham, 2019, pp. 239–248.
[37] B. Choubin, S. Khalighi-Sigaroodi, A. Malekian, Ö. Kişi, Multiple linear regression multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals, Hydrol. Sci. J. 61 (2016) 1001–1009, http://dx.doi.org/10.1080/02626667.2014.966721.
[38] T. Wang, C. Wen, H. Wang, F. Gao, T. Jiang, S. Jin, Deep learning for wireless physical layer: Opportunities and challenges, 2017, http://dx.doi.org/10.1109/CC.2017.8233654.
[39] Y. Altuntaş, Z. Cömert, A.F. Kocamaz, Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach, Comput. Electron. Agric. 163 (2019) 104874, http://dx.doi.org/10.1016/j.compag.2019.104874.
[40] M.G. Alaslani, L.A. Elrefaei, Convolutional neural network based feature extraction for IRIS recognition, Int. J. Comput. Sci. Inf. Technol. 10 (2018) 65–78, http://dx.doi.org/10.5121/ijcsit.2018.10206.
[41] Z. Cömert, A. Şengür, Ü. Budak, A.F. Kocamaz, Prediction of intrapartum fetal hypoxia considering feature selection algorithms and machine learning models, Health Inf. Sci. Syst. 7 (2019) 17, http://dx.doi.org/10.1007/s13755-019-0079-z.
[42] V. Suárez-paniagua, I. Segura-bedmar, Evaluation of pooling operations in convolutional architectures for drug-drug interaction extraction, 19, 2018, http://dx.doi.org/10.1186/s12859-018-2195-1.
[43] M. Geist, Soft-max boosting, Mach. Learn. 2958 (2015) 305–332, http://dx.doi.org/10.1007/s10994-015-5491-2.
[44] S. Woo, J. Park, J.-Y. Lee, I.S. Kweon, CBAM: Convolutional block attention module, 2018, Eccv.
[45] S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, 2016, http://dx.doi.org/10.1109/cvpr.2017.634.
[46] B. Hariharan, P. Arbeláez, R. Girshick, J. Malik, Object instance segmentation and fine-grained localization using hypercolumns, IEEE Trans. Pattern Anal. Mach. Intell. 39 (2017) 627–639, http://dx.doi.org/10.1109/TPAMI.2016.2578328.
[47] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, 2015, http://arxiv.org/abs/1502.03167.
[48] J. Park, S. Woo, J.-Y. Lee, I.S. Kweon, BAM: Bottleneck attention module, 2018, http://arxiv.org/abs/1807.06514.
[49] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, T.-S. Chua, SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning, 2016.
[50] N. Schilling, M. Wistuba, L. Drumond, L. Schmidt-Thieme, Hyperparameter Optimization with Factorized Multilayer Perceptrons, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9285, 2015, pp. 87–103, http://dx.doi.org/10.1007/978-3-319-23525-7_6.
[51] H. Wen, J. Shi, W. Chen, Z. Liu, Deep residual network predicts cortical representation and organization of visual features for rapid categorization, Sci. Rep. 8 (2018) 3752, http://dx.doi.org/10.1038/s41598-018-22160-9.
[52] A. Nugaliyadde, K.W. Wong, F. Sohel, H. Xie, Language modeling through long term memory network, 2019, http://arxiv.org/abs/1904.08936.
[53] P. Reeskamp, Is comparative advertising a trade mark issue ? Eur. Intellect. Prop. Rev. 30 (2008) 130–137, http://dx.doi.org/10.1145/2623330.2623612.
[54] Z. Yang, C. Wang, Z. Zhang, J. Li, Mini-batch algorithms with online step size, Knowl.-Based Syst. 165 (2019) 228–240, http://dx.doi.org/10.1016/j.knosys.2018.11.031.
[55] M. Toğaçar, B. Ergen, Z. Cömert, Application of breast cancer diagnosis based on a combination of convolutional neural networks, ridge regression and linear discriminant analysis using invasive breast cancer images processed with autoencoders, Med. Hypotheses (2019) 109503, http://dx.doi.org/10.1016/j.mehy.2019.109503.
[56] Z. Cömert, A.F. Kocamaz, V. Subha, Prognostic model based on image-based time-frequency features and genetic algorithm for fetal hypoxia assessment, Comput. Biol. Med. (2018) http://dx.doi.org/10.1016/j.compbiomed.2018.06.003.
[57] Shallu R. Mehra, Breast cancer histology images classification: Training from scratch or transfer learning? ICT Express 4 (2018) 247–254, http://dx.doi.org/10.1016/j.icte.2018.10.007.
[58] Y. Feng, L. Zhang, J. Mo, Deep manifold preserving autoencoder for classifying breast cancer histopathological images, IEEE/ACM Trans. Comput. Biol. Bioinform. (2018) 1, http://dx.doi.org/10.1109/TCBB.2018.2858763.
[59] G. Yun, Y. Jie, Densely-connected multi-magnification hashing for histopathological image retrieval, IEEE J. Biomed. Heal. Inform. XX (2018) 1–10, http://dx.doi.org/10.1109/JBHI.2018.2882647.
[60] H. Erfankhah, M. Yazdi, M. Babaie, H.R. Tizhoosh, Heterogeneity-aware local binary patterns for retrieval of histopathology images, IEEE Access 7 (2019) 18354–18367, http://dx.doi.org/10.1109/ACCESS.2019.2897281.
[61] D. Lichtblau, C. Stoean, Cancer diagnosis through a tandem of classifiers for digitized histopathological slides, PLoS One 14 (2019) 1–20, http://dx.doi.org/10.1371/journal.pone.0209274.
[62] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, S. Li, Breast cancer multi-classification from histopathological images with structured deep learning model, Sci. Rep. 7 (2017) 1–10, http://dx.doi.org/10.1038/s41598-017-04075-z.