JOURNAL OF SCIENCE

SAKARYA UNIVERSITY

Title: Classification of Breast Cancer Images Using Ensembles of Transfer Learning

Authors: Kadir GUZEL, Gokhan BILGIN

Classification of Breast Cancer Images Using Ensembles of Transfer Learning

Kadir GUZEL[1], Gokhan BILGIN[*2]

Abstract

It is a challenging task to estimate the cancerous cells and tissues via computer-aided diagnosis systems on high-resolution histopathological images. In this study, it is suggested to use transfer learning and ensemble learning methods together in order to reduce the difficulty of this task and better diagnose cancer patients. In the studies, histopathological images with 40× and 100× magnification factors are analyzed. In order to prove the success of the study with experimental studies, firstly, the results provided by pre-modeled deep learning architectures trained by histopathological image dataset, then the results acquired by different transfer learning approaches and the results obtained with the ensembles of deeply learned features using transfer learning methods are presented comparatively. Three different approaches are applied for transfer learning by fine-tuning the pre-trained convolution neural networks. In the experimental section, results of single classifiers (i.e., support vector machines, logistic regression, k-nearest neighbor and bagging) are presented by employing features of CNN models obtained by defined transfer learning approaches. Then, decisions of each classifier model are combined separately by weighted decision fusion (WDF) and stacking decision fusion (SDF) ensemble learning methods that have proven to improve the classification performance of the proposed classification system.

Keywords: Histopathological images, breast cancer, deep learning, transfer learning, ensemble learning.

1. INTRODUCTION

Cancer remains one of the leading causes of human death causes worldwide. Only in 2018, 18.1 million new cancer cases and 9.6 million cancer-related deaths were reported in a total of 185 countries. Recent global cancer statistics show that breast cancer is still the most common form of cancer among women with 24.2% (2.1 million) new cases and 626,679 deaths per year [1]. In all cancer types, early diagnosis and treatment process is very critical in the fight against cancer. In general approach, pathologists look for signs of cancer by analyzing histological descriptive features of tissue sections commonly

[1]Yildiz Technical University, Dpt. of Computer Engineering, 34220 Istanbul, Turkey.
ORCID: https://orcid.org/0000-0002-3664-6810
[*]Corresponding Author: gbilgin@yildiz.edu.tr
[2]Yildiz Technical University, Dpt. of Computer Engineering,
34220 Istanbul, Turkey. ORCID: https://orcid.org/0000-0002-5532-477X

stained with hematoxylin and eosin (H&E), immunohistochemical (IHC) and some specific dyes in a microscopic environment with histopathological whole slides. In recent years, histopathological slides obtained from rapidly developing digital imaging devices are captured as digital high-resolution images and given to expert systems for evaluation. These medical images are an important research topic in the field of biomedical and machine learning to help clinicians and diagnose cancer patients. Computer-aided diagnostic systems gain importance especially in undecided diagnoses and in determining the regions where the doctor's attention should be concentrated.

Computer-aided diagnosis systems (CAD) have been implemented in different medical problems such as mammography and breast image analysis [2], mass detection [3], lung cancer [4], colorectal cancer classification [5], skin lesion classification [6]. Generally, histopathological image processing methods are based on approaches that recognizing patterns which are included local texture and extracting important features from images. In the recent machine learning literature, feature extraction methods can be divided into two parts: *i)* the first one is conventional feature extraction techniques commonly called as handcrafted methods and *ii)* the second one is the deep features extracted from images by deep learning networks [7].

Conventional feature extraction methods have been studied extensively in the biomedical pattern recognition literature; for example, local binary pattern (LBP) that evaluates the value of the center pixel according to neighboring the pixels [8]. A well-known method is the histogram of oriented gradients (HOG) that uses the orientation and magnitude values of the pixels in the image [9]. Gabor filtering is used to capture discriminative features aligned at specific angles on the image [10]. Speeded up robust features (SURF) rely on the determinant of the Hessian matrix for both scale and location [11]. On the other hand, deep features obtained from deep learning models can achieve higher recognition accuracies than handcrafted image features. Especially, the feature extraction from images by using convolution neural networks (CNN) has

made much progress in recent years [12]. Multi-layered forward computing and back-propagation artificial neural networks (ANN) have encountered intense interest by researchers with the use of graphics processors (GPU) and tensor processing unit (TPU). Deep neural models can code the raw pixels of histopathological images into feature vectors that represent high-level concepts for classification and segmentation problems [13].

Various learning methods have been applied to datasets in different medical fields in the literature. Transfer learning and ensemble learning are among the popular learning methods in the field of machine learning. In [14], it has been studied to build ensemble learning models that can achieve more robust results by combining the features or decisions in the classification of histopathological images. In another study, a community-based GoogLeNet architecture was proposed for breast cancer image classification [15]. In their study, the authors proposed a new ensemble scheme to fuse patch probabilities for image-wise classification. In another study, an ensemble of multi-scale CNN algorithm has been proposed to classify H&E stained breast histopathological microscopy images [16]. Transfer learning is a machine learning technique in which a model trained for a task is redesigned in a related second problem. In terms of deep learning, transfer learning involves reusing a pre-trained model on a new problem and fine-tuning stages to adapt to this problem. A deep learning technique based on the ensemble of CNN has been studied to differentiate adenocarcinomas from healthy tissues and benign lesions [17]. A bioimage classification system has been proposed for boosting the performance of trained CNNs by composing multiple CNN models into an ensemble and scores were combined by sum rule [18]. Deep learning based ResNet-50 and DenseNet-161 pre-trained models have been employed to classify histopathology images [19]. An ensemble transfer learning framework has been introduced to classify cervical histopathology images [20].

In this study, fine-tuned models, which ranked first in ImageNet competitions, are used to detect benign and malignant cells in breast cancer

images. The network is trained in three different approaches using pre-trained weights with the transfer learning method [21] and features are extracted using this fine-tuned model. The supervised learning methods used in the study are support vector machines (SVM), logistic regression (LR), k-nearest neighbor (KNN) and bagging (BG). A robust model was proposed to increase the collective success by using two different ensemble learning methods together with the decisions of the four best classifiers among these classifiers.

The proposed method for classifying breast cancer images is described in detail in Section 2. The properties of the experimental dataset are introduced in Section 3. After the experimental studies are discussed comparatively in Section 4, the results are concluded in Section 5.

2. METHODOLOGY

In the classification of breast cancer histopathological images, it is proposed to use transfer learning and ensemble learning methods together. To prove the success of the study with experimental studies; *i)* the results provided by pre-modeled deep learning architectures, *ii)* the results acquired by different transfer learning approaches and, ultimately *iii)* the results obtained with ensembles of deeply learned features using transfer learning methods are presented comparatively.

2.1. Classification via Pre-modeled CNN Models

Deep neural networks are capable of extracting low, medium and high level features in an end-to-end multi-layered manner. In addition, it can be seen that the number of stacked layers can enhance the levels of the features [13, 19]. Deep learning-based CNN architectures, which proved their efficiency in ImageNet competitions in image recognition problems, are used for breast cancer classification. Histopathological images in the dataset are evaluated by two different advanced CNN architectures, namely VGGNet16 [22] and ResNet50 [23].

- VGGNet16 deep learning algorithm is a 16-layer network consisting of 13 convolutions and 3 fully connected dense layers introduced in the ILSVRC-2014 competition. There are approximately 134 million parameters in this network total. The image placed on the input layer should be in the size of 224×224×3. The last layer is the artificial neural network layer. It is a deep learning algorithm that achieves 92.7% accuracy in the ImageNet database which is a dataset of over 14 million images belonging to 1000 classes.

- ResNet50 architecture stands for residual networks and includes five stages each with a convolution and identity block. In this structure, each convolution block has three convolution layers and each identity block also has three convolution layers. The network model has 177 or more layers and there are approximately 23.5 million parameters in total. In addition, this layered structure provides information about how the connections between the layers will be and skip connection is applied before the ReLU (rectified linear unit) activation function. The input layer is in size of 224×224×3. This network learns rich feature representations for a wide variety of images and presents a robust model that is used very frequently in many image processing tasks.

2.2. Classification with Employing Transfer Learning

Transfer learning is a deep learning methodology that allows the model, previously trained on a special problem, to use it in a new task. It is a promising methodology especially in the field of image processing because it makes it easy to create accurate and time-saving models. Sometimes, it may take days, weeks to train deep convolutional neural network models on very large datasets. One way to shorten this process is to reuse model weights from pre-trained convolutional models developed for standard computer vision benchmark datasets such as ImageNet. In this way, the models that are needed for our own problems can be created more easily

without starting from scratch by taking advantage of what the previous models have learned.

The pre-trained models that we employed in our study are architectures trained on a larger benchmark dataset (i.e. ImageNet containing 1.2 million images from 1000 categories) to solve a problem similar to the problem of finding distinctive features in histopathological images. It is common practice to use models that have proven their success in the machine learning literature to achieve more successful results with data sets containing a limited number of samples and classes, taking into account the computational cost of data. In parallel with these explanations, it is also supported by the results in our study that it is useful to use fine-tuned networks over the pre-trained network instead of training the entire network. There are three different approaches for transfer learning with the fine-tuning process of a convolution neural network. These approaches are summarized in Figure 1.
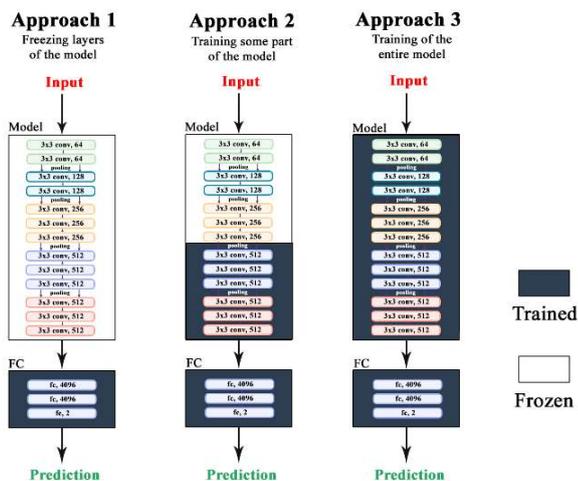

Figure 1. Transfer learning approaches

- **Approach 1 (A1)**: The pre-trained deep model weights are loaded initially. The connections in the fully connected (FC) layers retrained with our own data samples, whereas all other layers are frozen.

- **Approach 2 (A2)**: The pre-trained deep model weights are loaded initially. A certain part of the network is retrained for fine-tuning with our own data samples, whereas other layers remain frozen.

- **Approach 3 (A3)**: The pre-trained deep model weights are loaded initially. Then the entire network is retrained with our own data samples for fine-tuning.

2.3. Ensemble Learning with Deeply Learned Features

In the convolution layers of CNN networks, features are extracted using various sizes of kernels in the image; because discriminative features in the image are usually local and it is necessary to take into account several neighbor pixels [24]. The pre-trained deep learning models used in the experiments are handled by a method known as inductive learning, which is a kind of transfer learning. The main purpose of inductive learning algorithms is to extract a mapping from previous training samples.

Thus, thanks to transfer learning with the models used, deep features are extracted with three different approach types by using our own histopathology data. For this purpose, after the training is completed, the fully connected layers are removed for creating a model that could extract the discriminative features. In order for the feature vectors to have the same size, the flattened features are first normalized and then PCA is used for dimension reduction to obtain the principal components that contain 95% of the entire variance.

Ensemble learning can be defined as a process in which the decisions of the system consisting of more than one different classifier models are combined. Ensemble learning is primarily used to improve the classification performance of a total system model [20]. Therefore, instead of increasing the accuracy of a single classifier using some of the specified features, it is aimed to make a more accurate decision with the logic of the ensemble by combining different classification results.

In our study, weighted decision fusion (WDF) [25] and stacking decision fusion (SDF) [26] ensemble methods are used when combining decisions. The WDF method is utilized to combine weak learners using the weighting

strategy. Model weights are small positive values and the sum of all weights is equal to one. In the WDF method, weights are adjusted according to the percentage of confidence in each model or the expected performance. In regression problems, weights are multiplied and averaged while decisions for classification problems are multiplied in proportion to the weights of the classifiers. The weights used in the problem we discussed in our article are given according to the success of the learners as a result of the experiments. The method of SDF involves a combiner learning algorithm training to ensemble the predictions of many weak learning algorithms. In this approach, first, single/weak classifiers are trained using our own data. It is then trained to make a final prediction using all predictions of other algorithms through the combiner learning algorithm used. In this study, the logistic regression model, which is generally used as a combiner learning algorithm, is employed.

3. EXPERIMENTAL DATASET

Due to the difficulty of the data collection procedures, time consuming processing and the labeling costs, the high quality datasets in the scientific medical literature consist of limited numbers and limited samples. Therefore, accessing a good large-scale dataset with detailed explanations for researchers is quite laborious. In this study, the Breast Cancer Histopathological Image Classification (BreaKHis) dataset presented to the public by the Prognostic and Diagnostic (P&D) Laboratory in Brazil is used to perform our studies efficiently. The dataset contains microscopic biopsy images of benign and malignant breast tumors which were captured through a clinical study from January 2014 to December 2014 [27]. Breast tissue biopsy slides were stained with hematoxylin and eosin (H&E). The breast cancer dataset images were obtained by an Olympus BX-50 microscope with a 3.3× magnification relay lens combined with a Samsung digital color camera SCC-131AN. Histopathological images were taken in RGB color space true color format using magnification levels of 40×, 100×, 200× and 400×. The dataset totally is composed of 7909 microscopic biopsy breast cancer images collected from 82 patients.

The dataset consists of four different types of histological benign breast tumors (adenosis, fibroadenoma, phyllodes tumor, tubular adenoma) and four types of malignant tumors (ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma). In this study, H&E images captured at 40× and 100× magnification levels from microscopic images at very different magnification levels are used. As can be seen from Table 1, the total number of benign images selected from the dataset is 1269 and the total number of malignant images is 2807. Examples of benign and malignant image samples taken at 40× and 100× magnification levels can be seen in Figure 2.

Table 1. The number of data samples according to magnification level and class label

Magnification	Benign	Malignant	Total
40x	625	1370	1995
100x	644	1437	2081

The magnification factor in digital microscopes plays an important role in the analysis of histology images [28]. In normal procedure, pathologists adjust the magnification factor on the slide so that the objects of interest can be easily seen. Histology images consist of different types of tissues and analysis of these tissues becomes difficult in low magnification levels. Besides that, background changes may occur in digitized images with different magnification levels. Therefore, different magnification factors make it difficult to obtain an accurate diagnosis in automated computer-aided diagnostic (CAD) systems [29].
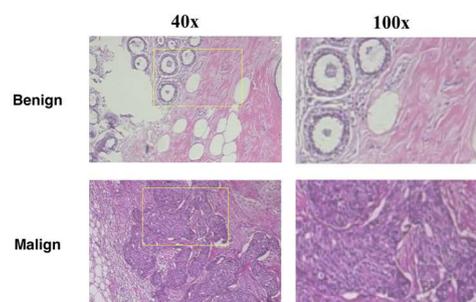

Figure 2. Examples of breast cancer histopathological image samples in different magnification levels [27]

4. EXPERIMENTAL RESULTS

In this study, experimental studies with histopathological images at 40× and 100× magnification levels are analyzed to observe the effects of different magnification factors. The experiments of this research have been developed by using Tensor Processing Units (TPU) with Keras [30], libraries with Scikit-learn [31] and TensorFlow [32] in the background.

In our study, firstly, the results are obtained by performing 10-fold cross-validation with pre-modeled deep learning architectures and presented in Table 2 with standard deviation values in parentheses to make comparative evaluations. The overall accuracy, recall, precision and F1-score metrics of the pre-modeled architectures are presented in Table 2. Note that, when using pre-modeled networks, results are obtained without using pre-trained weights in this table. Original VGGNet16 and ResNet50 network models are adopted and trained with breast cancer histopathological image classification dataset. As can be seen from Table 2, the results of VGGNet16 are better than ResNet50 in at 40× and 100× magnification levels. The results in Table 3 are obtained from experiments by applying 10-fold cross-validation with standard deviation values in parentheses using three different transfer learning approaches explained in previous sections, at 40× and 100× magnification levels of breast cancer histopathological image classification dataset. As can be seen from the Table 3, ResNet50 surpasses VGGNet16 by transfer learning approach 3 (A3) by retraining the entire network with our own data samples. With this fine-tuning approach, ResNet50 architecture reaches 95.09% and 94.62% for at 40× and 100× magnification levels respectively.

The flowchart of the proposed system is shown in Figure 3. In the third part of the experiments, SVM, LR, KNN and Bagging (BG) classifiers are employed for decision combining. The results of single classifiers and ensemble learning by employing features of CNN models obtained by defined transfer learning approaches are presented in Table 4. Table 4 presents the results

in the subsection structure for the transfer learning approaches used with each pre-modeled network architecture. For the first four rows in each subsection of Table 4, the results of single classifiers (SVM, LR, KNN and BG) are presented by employing features of CNN models obtained by defined transfer learning approaches. Then, in the lower part of each subsection of Table 4, decisions of the each single classifier models are combined by WDF and SDF ensemble learning methods which are used to improve the classification performance of a total system model. The results represented in Table 4 are also obtained by 10-fold cross-validation and standard deviation values are shown in parentheses. According to the Table 4, the results obtained at 40× magnification level are better than 100× magnification level for each deep model and defined transfer learning approach. In general, according to the results obtained with the single classifiers, the classification performances are increased with the use of the WDF and SDF ensemble learning methods. Between these two methods, it can be seen that SDF yields better results than WDF. When all of Table 4 is analyzed, it is seen that the best result of 97.01% is obtained with the ResNet50 deep network with transfer learning approach 3 (A3) via the SDF ensemble method. In addition, the results obtained in the experimental studies in the literature using the same data set are presented for comparison purposes in Table 5. It should be taken into account that different training and test sample numbers are used in these studies.

5. CONCLUSIONS

In this study, it is aimed to develop a framework that can be evaluated as the second reader that can help doctors aiming to automatically classify cancerous tissues in histopathological images. The proposed approach with ensemble learning by employing features of pre-modeled CNN architectures using transfer learning approaches show promising improvement. According to the results in the tables, it is concluded that using traditional machine learning algorithms such as SVM and LR instead of fully connected layers in deep learning architectures yielded better results in the breast cancer dataset. In addition, among

the three different for transfer learning methods, the third approach (A3) where the pre-modeled ResNet50 neural network is fine-tuned with initial weights and the proposal to combine single classifier decisions with the SDF method work solidly than other methods. With this proposed framework structure, overall accuracies of 97.01% and 96.25% have been achieved at 40× and 100× magnification levels respectively.
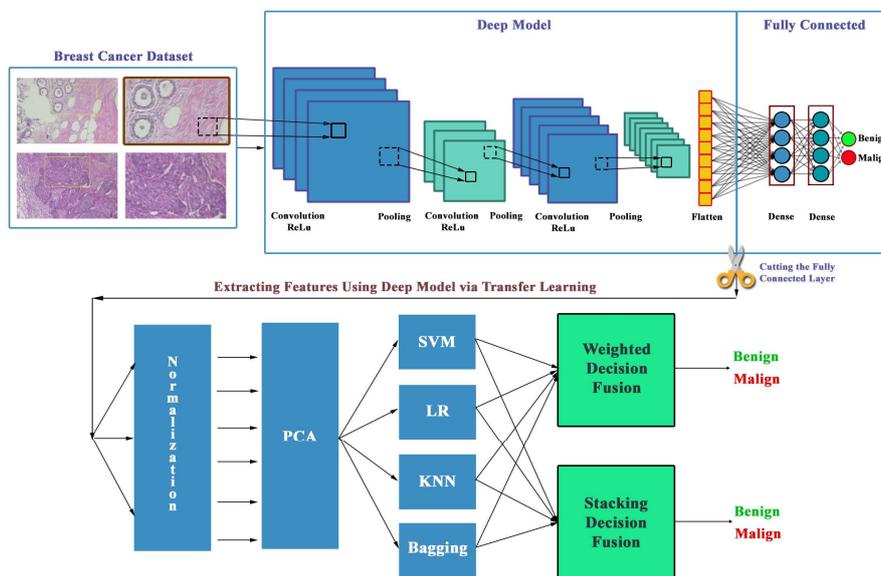

Figure 3. The flowchart of the proposed system

Table 2. The results obtained by pre-modeled networks without using pre-trained weights

Model	40×, 10-fold Cross-Validation				100×, 10-fold Cross-Validation			
	Accuracy	Recall	Precision	F1-Score	Accuracy	Recall	Precision	F1-Score
VGGNet16	86.50 (+/- 1.23)	0.86 (+/- 0.02)	0.81 (+/- 0.02)	0.83 (+/- 0.02)	87.12 (+/- 1.58)	0.85 (+/- 0.02)	0.84 (+/- 0.02)	0.84 (+/- 0.02)
ResNet50	77.50 (+/- 2.00)	0.76 (+/- 0.01)	0.79 (+/- 0.03)	0.76 (+/- 0.01)	64.75 (+/- 0.32)	0.65 (+/- 0.01)	0.74 (+/- 0.01)	0.71 (+/- 0.01)

Table 3. The results of three different transfer learning approaches using pre-trained weights

Model	App.	40×, 10-fold Cross-Validation				100×, 10-fold Cross-Validation			
		Accuracy	Recall	Precision	F1-Score	Accuracy	Recall	Precision	F1-Score
VGGNet16	A1	89.57 (± 2.54)	0.89 (± 0.01)	0.86 (± 0.04)	0.87 (± 0.04)	89.91 (± 1.51)	0.89 (± 0.02)	0.87 (± 0.03)	0.88 (± 0.02)
	A2	92.33 (± 1.66)	0.92 (± 0.02)	0.90 (± 0.02)	0.91 (± 0.02)	89.96 (± 1.77)	0.88 (± 0.03)	0.85 (± 0.01)	0.87 (± 0.02)
	A3	**93.34** (± 2.38)	0.94 (± 0.04)	0.89 (± 0.05)	**0.92** (± 0.05)	**92.56** (± 3.85)	0.93 (± 0.02)	0.90 (± 0.07)	**0.91** (± 0.06)
ResNet50	A1	70.12 (± 2.10)	0.65 (± 0.05)	0.58 (± 0.04)	0.59 (± 0.04)	71.35 (± 3.22)	0.66 (± 0.03)	0.60 (± 0.03)	0.61 (± 0.03)
	A2	64.75 (± 8.25)	0.68 (± 0.07)	0.63 (± 0.01)	0.61 (± 0.05)	70.23 (± 4.42)	0.60 (± 0.01)	0.64 (± 0.08)	0.60 (± 010)
	A3	**95.09** (± 2.53)	0.96 (± 0.02)	0.93 (± 3.62)	**0.94** (± 0.03)	**94.62** (± 2.14)	0.96 (± 0.01)	0.92 (± 0.03)	**0.93** (± 0.03)

Table 4. The results of single classifiers and ensemble learning by employing features of CNN models obtained by defined transfer learning approaches

Model	Alg.	40×, 10-fold Cross-Validation				100×, 10-fold Cross-Validation			
		Accuracy	Recall	Precision	F1-Score	Accuracy	Recall	Precision	F1-Score
VGGNet16+A1	SVM	89.82 (± 1.39)	0.90 (± 0.02)	0.86 (± 0.02)	0.88 (± 0.02)	89.48 (± 1.94)	0.89 (± 0.04)	0.86 (± 0.02)	0.87 (± 0.03)
	LR	**90.27** (± 1.90)	0.89 (± 0.02)	0.87 (± 0.03)	0.88 (± 0.02)	89.62 (± 1.60)	0.89 (± 0.03)	0.87 (± 0.02)	**0.88** (± 0.02)
	KNN	87.67 (± 1.68)	0.90 (± 0.02)	0.81 (± 0.02)	0.84 (± 0.02)	88.95 (± 1.87)	0.90 (± 0.02)	0.84 (± 0.02)	0.86 (± 0.02)
	BG	84.16 (± 1.63)	0.83 (± 0.02)	0.79 (± 0.02)	0.80 (± 0.02)	85.39 (± 1.79)	0.84 (± 0.03)	0.81 (± 0.02)	0.82 (± 0.02)
	WDF	89.92 (± 1.67)	0.90 (± 0.02)	0.86 (± 0.02)	0.88 (± 0.02)	**89.96** (± 1.70)	0.90 (± 0.03)	0.86 (± 0.02)	**0.88** (± 0.02)
	SDF	**90.27** (± 1.90)	0.89 (± 0.02)	0.88 (± 0.03)	**0.89** (± 0.02)	89.62 (± 1.60)	0.88 (± 0.03)	0.87 (± 0.02)	0.87 (± 0.02)
VGGNet16+A2	SVM	90.53 (± 1.10)	0.91 (± 0.02)	0.87 (± 0.02)	0.88 (± 0.01)	86.57 (± 2.96)	0.91 (± 0.02)	0.87 (± 0.02)	0.89 (± 0.02)
	LR	93.59 (± 1.47)	0.93 (± 0.02)	0.92 (± 0.02)	**0.92** (± 0.02)	**91.49** (± 1.14)	0.90 (± 0.01)	0.89 (± 0.02)	**0.90** (± 0.01)
	KNN	89.72 (± 1.52)	0.91 (± 0.02)	0.85 (± 0.02)	0.87 (± 0.02)	90.54 (± 1.23)	0.91 (± 0.02)	0.86 (± 0.02)	0.88 (± 0.02)
	BG	86.21 (± 1.28)	0.85 (± 0.02)	0.81 (± 0.02)	0.83 (± 0.02)	87.36 (± 1.77)	0.87 (± 0.02)	0.83 (± 0.02)	0.84 (± 0.02)
	WDF	91.58 (± 0.89)	0.92 (± 0.02)	0.88 (± 0.01)	0.90 (± 0.01)	91.25 (± 2.03)	0.91 (± 0.03)	0.88 (± 0.03)	0.89 (± 0.03)
	SDF	**93.62** (± 1.17)	0.92 (± 0.02)	0.92 (± 0.01)	**0.92** (± 0.02)	**91.49** (± 1.14)	0.90 (± 0.01)	0.89 (± 0.02)	**0.90** (± 0.01)
VGGNet16+A3	SVM	94.29 (± 1.19)	0.95 (± 0.01)	0.92 (± 0.01)	0.93 (± 0.01)	93.51 (± 2.73)	0.94 (± 0.03)	0.91 (± 0.04)	0.92 (± 0.03)
	LR	95.54 (± 1.39)	0.95 (± 0.02)	0.94 (± 0.02)	**0.95** (± 0.02)	**94.14** (± 2.40)	0.94 (± 0.03)	0.93 (± 0.03)	**0.93** (± 0.03)
	KNN	90.68 (± 2.12)	0.93 (± 0.02)	0.86 (± 0.03)	0.88 (± 0.02)	92.65 (± 2.39)	0.93 (± 0.02)	0.89 (± 0.03)	0.91 (± 0.03)
	BG	90.28 (± 1.58)	0.90 (± 0.02)	0.88 (± 0.02)	0.88 (± 0.02)	89.86 (± 1.81)	0.89 (± 0.02)	0.87 (± 0.02)	0.88 (± 0.02)
	WDF	94.59 (± 1.24)	0.95 (± 0.01)	0.92 (± 0.01)	0.94 (± 0.01)	93.56 (± 2.73)	0.94 (± 0.03)	0.91 (± 0.04)	0.92 (± 0.03)
	SDF	**96.04** (± 1.42)	0.96 (± 0.02)	0.94 (± 0.02)	**0.95** (± 0.02)	**94.14** (± 2.40)	0.94 (± 0.02)	0.93 (± 0.03)	**0.93** (± 0.03)
ResNet50+A1	SVM	92.67 (± 1.03)	0.93 (± 0.01)	0.91 (± 0.02)	0.91 (± 0.01)	91.91 (± 1.99)	0.91 (± 0.01)	0.90 (± 0.01)	0.90 (± 0.02)
	LR	89.83 (± 1.03)	0.88 (± 0.03)	0.89 (± 0.01)	0.88 (± 0.02)	90.08 (± 1.98)	0.89 (± 0.01)	0.88 (± 0.03)	0.89 (± 0.02)
	KNN	90.67 (± 0.24)	0.90 (± 0.02)	0.87 (± 0.01)	0.89 (± 0.01)	88.96 (± 2.42)	0.87 (± 0.02)	0.85 (± 0.02)	0.86 (± 0.02)
	BG	84.17 (± 1.43)	0.83 (± 0.01)	0.78 (± 0.01)	0.80 (± 0.01)	84.32 (± 2.73)	0.85 (± 0.01)	0.77 (± 0.03)	0.79 (± 0.04)
	WDF	**93.83** (± 0.85)	0.94 (± 0.01)	0.92 (± 0.01)	**0.92** (± 0.01)	**92.32** (± 0.69)	0.92 (± 0.01)	0.90 (± 0.03)	**0.91** (± 0.02)
	SDF	**93.83** (± 0.85)	0.94 (± 0.01)	0.92 (± 0.01)	**0.92** (± 0.01)	92.16 (± 1.98)	0.91 (± 0.02)	0.90 (± 0.03)	**0.91** (± 0.02)
ResNet50+A2	SVM	89.25 (± 2.25)	0.92 (± 0.02)	0.85 (± 0.02)	0.87 (± 0.02)	87.52 (± 2.15)	0.88 (± 0.04)	0.81 (± 0.02)	0.84 (± 0.03)
	LR	94.25 (± 0.75)	0.95 (± 0.01)	0.93 (± 0.01)	**0.93** (± 0.01)	**91.00** (± 1.37)	0.89 (± 0.03)	0.90 (± 0.01)	**0.90** (± 0.02)
	KNN	89.75 (± 0.75)	0.93 (± 0.01)	0.85 (± 0.02)	0.87 (± 0.01)	86.80 (± 2.91)	0.86 (± 0.01)	0.86 (± 0.01)	0.84 (± 0.03)
	BG	90.00 (± 3.50)	0.90 (± 0.03)	0.88 (± 0.04)	0.89 (± 0.04)	86.92 (± 3.48)	0.85 (± 0.03)	0.84 (± 0.05)	0.84 (± 0.04)
	WDF	92.75 (± 1.25)	0.94 (± 0.01)	0.90 (± 0.01)	0.91 (± 0.01)	90.76 (± 1.39)	0.91 (± 0.02)	0.87 (± 0.02)	0.88 (± 0.01)
	SDF	92.75 (± 1.25)	0.94 (± 0.01)	0.90 (± 0.01)	0.91 (± 0.01)	90.76 (± 1.39)	0.91 (± 0.02)	0.87 (± 0.02)	0.88 (± 0.01)
ResNet50+A3	SVM	96.84 (± 1.50)	0.97 (± 0.01)	0.96 (± 0.02)	**0.96** (± 0.02)	**96.25** (± 1.62)	0.96 (± 0.02)	0.95 (± 0.02)	**0.95** (± 0.02)
	LR	96.04 (± 1.61)	0.96 (± 0.01)	0.95 (± 0.03)	0.95 (± 0.02)	95.39 (± 2.05)	0.95 (± 0.02)	0.94 (± 0.03)	0.94 (± 0.03)
	KNN	96.89 (± 1.77)	0.97 (± 0.02)	0.96 (± 0.03)	**0.96** (± 0.02)	**96.25** (± 2.05)	0.96 (± 0.02)	0.95 (± 0.03)	**0.95** (± 0.03)
	BG	92.53 (± 1.75)	0.91 (± 0.02)	0.91 (± 0.02)	0.91 (± 0.02)	92.03 (± 1.68)	0.90 (± 0.02)	0.91 (± 0.02)	0.90 (± 0.02)
	WDF	96.94 (± 1.60)	0.97 (± 0.01)	0.96 (± 0.02)	**0.96** (± 0.02)	**96.25** (± 1.62)	0.96 (± 0.02)	0.95 (± 0.02)	**0.95** (± 0.02)
	SDF	**97.01** (± 1.56)	0.97 (± 0.01)	0.96 (± 0.02)	**0.96** (± 0.02)	**96.25** (± 1.62)	0.96 (± 0.02)	0.95 (± 0.02)	**0.95** (± 0.02)

Table 5. Comparative analysis table of deep learning methods in the literature for the BreakHis dataset

Author [citation]	Methodology	Features / Application	Results (%)	
			40×	100×
Bayramoglu et al. [33]	Single Task CNN	• Better accuracy than handcrafted methods • Combined image data from many more resolution	80.97	80.92
Kassani et al. [34]	VGG19, MobilNetV2, DenseNet201	• A three-path ensemble architecture is used by transfer learning and fine-tuning with different number of training and test samples from our study • Deep feature extraction and fused features	98.13	
Vo et al. [35]	ResNetV1 + GBT, InceptionV3 + GBT	• Fused model via Gradient Boosting Trees • Fine-tuning deep learning models	93.50	95.30
Alom et al. [36]	IRRCNN + Aug.	• Better accuracy than similar studies • Applied different data augmentation techniques with different number of training and test samples from our study	97.95	97.57
Spanhol et al. [37]	AlexNet CNN	• Used a deep learning approach to avoid hand-crafted features • Proposed several strategies for training	89.60	85.00
Han et al. [38]	CSDCNN + Aug.	• A new deep learning model has been proposed	92.80	93.90
Gandomkar et al. [39]	ResNet	• A framework called ResNet based MuDeRN has been proposed with transfer learning • Prediction with more class	95.60	94.89
Mehra [40]	VGG16 + LR, VGG19 + LR, ResNet50 + LR	• Used a fine-tuned model via logistic regression classifier • Fine-tuned pre-trained network is more robust	92.60	
Zhu et al. [41]	Hybrid CNN	• Multiple hybrid models with the same architecture are fused. • Used different subset with the voting	85.7	84.2
Kumar et al. [42]	VGG16 + SVM, VGG16 + RF	• Proposed a variant of VGGNet-16, a FC layer was removed and experimented with various classifiers	94.11	95.12

The Declaration of Conflict of Interest / Common Interest

No conflict of interest or common interest has been declared by the authors.

Authors' Contribution

All authors have contributed in experimental study and writing of the manuscript equally.

The Declaration of Ethics Committee Approval

The authors declare that this document does not require an ethics committee approval or any special permission.

The Declaration of Research and Publication Ethics

The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1]   F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A cancer Journal for Clinicians, vol. 68, no. 6, pp. 394–424, 2018.

[2]   A. Hamidinekoo, E. Denton, A. Rampun, K. Honnor, and R. Zwiggelaar, "Deep learning in mammography and breast histology, An overview and future trends," Medical Image Analysis, vol. 47, pp. 45-67, 2018.

[3]   M. N. Gurcan, L. E. Boucheron, A. Can, A. Madabhushi, N. M. Rajpoot, and B. Yener, "Histopathological image analysis: A review," IEEE Reviews in Biomedical Engineering, vol. 2, pp. 147-171, 2009.

[4]   N. Coudray, P. S. Ocampo, T. Sakellaropoulos, N. Narula, M. Snuderl, D. Fenyö, and A. Tsirigos, "Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning," Nature Medicine, vol. 24, no. 10, pp. 1559-1567, 2018.

[5]   K. Sirinukunwattana, S. E. A. Raza, Y. W. Tsang, D. R. Snead, I. A. Cree, and N. M. Rajpoot, "Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images," IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1196-1206, 2016.

[6]   T. Majtner, S. Yildirim-Yayilgan, and J. Y. Hardeberg, "Combining deep learning and hand-crafted features for skin lesion classification," IEEE 6th International Conference on Image Processing Theory, Tools and Applications, IPTA'16, pp. 1-6, 2016.

[7]   N. Hatipoglu and G. Bilgin, "Cell segmentation in histopathological images with deep learning algorithms by utilizing spatial relationships," Medical & Biological Engineering & Computing, vol. 55, no. 10, pp. 1829-1848, 2017.

[8]   K. Guzel and G. Bilgin, "Textural feature extraction and ensemble of extreme learning machines for hyperspectral image classification," IEEE 26th Signal Processing and Communications Applications Conference, SIU'18, pp. 1-4, 2018.

[9]   N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," International Conference on Computer Vision and Pattern Recognition, CVPR'15, vol. 2, pp. 886–893, 2005.

[10]   D. A. Clausi and M. E. Jernigan, "Designing Gabor filters for optimal texture separability," Pattern Recognation, vol. 33, no. 11, pp. 1835–1849, 2000.

[11]   A. Abbas, T. Zehra, and F. Li, "Object detection in a cluttered scene using SURF for computer assisted histopathology," International Conference on Electrical, Mechanical and Industrial Engineering, Atlantis Press, 2016.

[12] A. Albayrak and G. Bilgin, "Automatic cell segmentation in histopathological images via two-staged superpixel-based algorithms," Medical & Biological Engineering & Computing, vol. 57, no. 3, pp. 653–665, 2019.

[13] A. F. Spanhol et al., "Deep features for breast cancer histopathological image classification," IEEE International Conference on Systems, Man, and Cybernetics, SMC'17, pp. 1868-1873, 2017.

[14] K. Guzel and G. Bilgin, "Classification of nuclei in colon cancer images using ensemble of deep learned features," IEEE Medical Technologies Congress, TıpTekno'19, pp. 1–4, 2019.

[15] Y. S. Vang, Z. Chen, and X. Xie, "Deep learning framework for multi-class breast cancer histology image classification," International Conference Image Analysis and Recognition, ICIAR'18, pp. 914-922, 2018.

[16] Z. Yang, L. Ran, S. Zhang, Y. Xia, and Y. Zhang, "EMS-Net: Ensemble of multiscale convolutional neural networks for classification of breast cancer histology images," Neurocomputing, pp. 46-53, 2019.

[17] F. Ponzio, E. Macii, E. Ficarra and S. Di Cataldo, "Going deeper into colorectal cancer histopathology," International Joint Conference on Biomedical Engineering Systems and Technologies, BIOSTEC'18, pp. 114-131, 2018.

[18] L. Nanni, S. Ghidoni, and S. Brahnam, "Ensemble of convolutional neural network for bioimage classification," Applied Computing and Informatics, 2018.

[19] M. Talo, "Automated classification of histopathology images using transfer learning," Artificial Intelligence in Medicine, vol. 101, no. 101743, 2019.

[20] C. Li et al., "Cervical histopathology image classification using ensembled transfer learning," International Conference on Information Technologies in Biomedicine, ITIB'19, pp. 26-37, Springer, Cham, 2019.

[21] B. Harangi, "Skin lesion classification with ensembles of deep convolutional neural networks," Journal of Biomedical Informatics, vol. 86, pp. 25–32, 2018.

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

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," IEEE Conference on Computer Vision and Pattern Recognition, CVPR'16, pp. 770–778, 2016.

[24] L. Liu, C. Shen, and A. van den Hengel, "The treasure beneath convolutional layers: Cross-convolutional-layer pooling for image classification," IEEE Conference on Computer Vision and Pattern Recognition, CVPR'15, pp. 4749–4757, 2015.

[25] L. I. Kuncheva and J. J. Rodríguez, "A weighted voting framework for classifiers ensembles," Knowledge and Information Systems, vol. 38, no. 2, pp. 259–275, 2014.

[26] J. Sill, G. Takács, L. Mackey, and D. Lin, "Feature-weighted linear stacking," arXiv preprint arXiv:0911.0460, 2009.

[27] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," IEEE Transactions on Biomedical Engineering, vol. 63, no. 7, pp. 1455–1462, 2015.

[28] T. L. Sellaro, R. Filkins, C. Hoffman, J. L. Fine, J. Ho, A. V. Parwani, and M. Montalto, "Relationship between magnification and resolution in digital pathology systems," Journal of Pathology Informatics, vol. 4, 2013.

[29] M. L. Giger and K. Suzuki, "Computer-aided diagnosis," Biomedical Information Technology, Academic Press, pp. 359-XXII, 2008.

[30] F. Chollet, "Keras: The python deep learning library," Astrophysics Source Code Library, 2018.

[31] F. Pedregosa, et al. "Scikit-learn: Machine learning in Python," Journal of machine learning research, vol.12, pp. 2825-2830, 2011.

[32] M. Abadi, et al. "Tensorflow: A system for large-scale machine learning," 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI'16, pp. 265-283, 2016.

[33] N. Bayramoglu, J. Kannala, and J. Heikkilä, "Deep learning for magnification independent breast cancer histopathology image classification," IEEE 23rd International Conference on Pattern Recognition, ICPR'16, pp. 2440–2445, 2016.

[34] S. H. Kassani, P. H. Wesolowski, M. J. Schneider, and K. A. Deters, "Classification of histopathological biopsy images using ensemble of deep learning networks," arXiv preprint arXiv:1909.11870, 2019.

[35] D. M. Vo, N. Q. Nguyen, and S. W. Lee, "Classification of breast cancer histology images using incremental boosting convolution networks," Information Sciences, vol. 482, pp. 123-138, 2019.

[36] M. Z. Alom, C. Yakopcic, M. S. Nasrin, T. M. Taha, and V. K. Asari, "Breast cancer classification from histopathological images with inception recurrent residual convolutional neural network," Journal of Digital Imaging, vol. 32, no. 4, pp. 605-617, 2019.

[37] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," IEEE International Joint Conference on Neural Networks, IJCNN'16, pp. 2560-2567, 2016.

[38] Z. Han, B. Wei, Y. Zheng, Y. Yin, K. Li, and S. Li, "Breast cancer multi-classification from histopathological images with structured deep learning model," Scientific Reports, vol. 7, no. 1, pp. 1-10, 2017.

[39] Z. Gandomkar, P. C. Brennan, and C. Mello-Thoms, "MuDeRN: Multi-category classification of breast histopathological image using deep residual networks," Artificial Intelligence in Medicine, vol. 88, pp. 14-24, 2018.

[40] R. Mehra, "Breast cancer histology images classification: Training from scratch or transfer learning?," ICT Express, vol. 4, no. 4, pp. 247-254, 2018.

[41] C. Zhu, F. Song, Y. Wang, H. Dong, Y. Guo, and J. Liu, "Breast cancer histopathology image classification through assembling multiple compact CNNs," BMC Medical Informatics and Decision Making, vol. 19, no. 1, pp. 198, 2019.

[42] A. Kumar, S. K. Singh, S. Saxena, K. Lakshmanan, A. K. Sangaiah, H. Chauhan, and R. K. Singh, "Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer," Information Sciences, vol. 508, pp. 405-421, 2020.