Multimedia Tools and Applications (2020) 79:15593–15611
https://doi.org/10.1007/s11042-019-07821-9

Transfer learning with pre-trained deep convolutional neural networks for serous cell classification

Elif Baykal[1] · Hulya Dogan[1] · Mustafa Emre Ercin[2] · Safak Ersoz[2] · Murat Ekinci[1]

Received: 1 October 2018 / Revised: 19 February 2019 / Accepted: 21 May 2019 /
Published online: 31 May 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
Serous effusion is a condition of excess accumulation of fluids in serous cavities due to different underlying pathological conditions. The basis of cytopathological assessment of serous effusions is the identification of cells in the fluid based on their morphology and texture. This assessment is a physically and mentally laborious task, and it can also lead to variability among pathologists. In literature, only a small number of feature-based methods are conducted for automated serous cell classification. In this study, a transfer learning with pre-trained deep convolutional neural networks (ConvNets) is proposed to automatically identify 11 different categories of serous cells in effusion cytology. Unlike the methods which rely on the extraction of cellular features such as morphology and texture, this method is an appearance-based machine learning approach. We fine-tuned four pre-trained ConvNet architectures that are AlexNet, GoogleNet, ResNet and DenseNet on the serous cell dataset. To reduce the overfitting effect, we augmented the data by image rotation, translation, and mirroring. The proposed method was evaluated on both original and augmented sets of serous cells derived from a publicly available dataset. Among the four ConvNet models, ResNet and DenseNet obtained the highest accuracies of 93.44% and 92.90%. However, when two models were compared in terms of accuracy and model complexity, ResNet-TL was selected as the best network model. When compared to the results without data augmentation, data augmentation increased the accuracy rate approximately 10%. Results show that higher classification results were achieved than other traditional methods without requiring precise segmentation.

Keywords Serous effusion · Cytopathological assessment · Cell classification · Convolutional neural networks · Transfer learning

1 Introduction

There are serous cavities in the human body, including pleura, peritoneum, and pericardium surrounded by visceral and parietal surfaces. In a healthy person, a small amount of fluid

✉ Elif Baykal
ebaykal@ktu.edu.tr

Extended author information available on the last page of the article.

is secreted by the parietal surface in the serous cavities. It facilitates easy movement of the visceral and parietal surfaces on top of each other. The imbalance in the reabsorption of the fluid by the visceral surface causes serous effusion which is the accumulation of fluid in the serous cavities [15, 40].

Annually, 1.5 million individuals are diagnosed with pleural effusion, which is a type of serous effusion, in the USA [16]. Cancer is one of the frequent causes of an effusion fluids, and it may be the first manifestation of advanced nature of the disease [40]. An accurate identification of the cancer cells in serous effusion cytology gives a chance of early diagnosis and prevents the dissemination of cancer by staging and monitoring these cells [5].

The cytopathological examination carried out by pathologists requires screening of these cell structures in the specimen by hand-eye coordination. In addition to being time-consuming, it also causes inter-observer and intra-observer variability, so an automated system is required [37]. Cell classification is one of the requirements for the identification of malign and benign cells in serous effusion cytology images when considering the cells with high similarity on the appearance.

There is a depth of field in the microscopic imaging systems. In the slide, only the cells falling into this range appear clear, while the objects falling into the rear or front region of this range appear blurred [39, 41]. In real time cytopathological assessment, this problem is commonly solved with the adjustment of the lens focus manually, and the blurred cells are identified. However, this is not possible with images acquired through a digital camera placed on a microscope. Also, the variability in size, shape, and texture of the individual cell and presence of cell clusters lead to diagnostic challenges in the cytopathological examination of serous cytology [3]. It also lays an obstacle for the segmentation and extraction of features such as morphology and texture.

ConvNets are biologically-inspired deep network architectures that can be trained to perform a variety of detection, segmentation and classification tasks. However, because of the limitations on the number of training samples, the large ConvNets may suffer from overfitting in training process. A ConvNet requires a large dataset to ensure that the model is learned to fit the dataset without overfitting since various parameters need to be tuned. To overcome these problems, in this paper, we present transfer learning with pre-trained ConvNets for classification of serous cells containing 11 classes in cytology images. More specifically, instead of training ConvNets from scratch with random initialization [44], the ConvNets trained on ImageNet [17], which has 1.3 million natural well-labeled images in 1000 classes, is fine-tuned to classify serous cells.

The fundamental contributions of this work are: (a) In the literature, there exists only a few feature-based methods conducted for serous cell classification. To the best of our knowledge, none of existing works on serous cell classification has considered deep feature learning from raw data. This is the first study to employ a transfer learning with pre-trained deep convolutional neural networks for the automated classification of serous cells. (b) Unlike other methods that require a set of hand-crafted features (e.g., the ratio of nucleus/cytoplasm area, the position of the nucleus in the cell, etc.) and nucleus/ cytoplasm segmentation, the proposed method is an appearance-based that performs the classification by extracting deep features from raw data. Thus, at the classification stage, there is no loss of accuracy due to inaccurate segmentation. (c) This method also achieves the highest classification accuracy on the 11-class serous cell dataset [28].

This paper is organized as follows: In Section 2, a review of previous related works is presented. Section 3 gives a detailed description of ConvNets and transfer learning with

pre-trained ConvNets. Experiments, as well as the experimental results, are shown in Section 4 and 5 respectively. The discussion is reported in Section 6. Finally, we draw our conclusions.

2 Related work

Image classification has received considerable attention in the computer vision field, and many algorithms have been proposed in the literature to solve various classification problems. In the medical images, it is a more challenging task given different appearances of the objects such as cells within the image and increasing complexities of the background textures. Automated screening procedures in the literature involve three main tasks: detection, segmentation, and classification. In this section, we provide a brief overview of the previous studies on serous cell detection, segmentation, and classification.

Lezoray et al. [26] achieved region extraction by mathematical morphology using color information in many color spaces. An extension of the watershed was utilized as an optimal region-growing operator. The segmentation results were 94.5% for the nuclei and 93% for the cytoplasm. In their work, the variability of the success rate is mostly dependent on the cell population. If there is a large number of isolated or touching cells, the method works efficiently. In more complex configurations, the success rate is not a real representative because it is difficult or impossible for a human observer to segment the cells in an absolute way. Also, a color object recognition scheme was proposed that consists of three consecutive steps: segmentation, feature extraction, and classification. In the scheme, more emphasis was placed on segmentation and classification steps. A color watershed with global and local criteria as in [26] and multiple ordinate neural network architecture (MONNA) were used sequentially [28]. There are 46 attributes [27] measured for each cell with a nucleus and cytoplasm segmentation. The total classification accuracy of MONNA was 84%, after feature selection by the sequential floating forward selection (SFFS) method. They also suggested applying two-color pixel classification schemes (Bayesian and k-means) and color watershed together to obtain segmentation. They noted that this approach produces a segmentation closer to what is expected of the experts [25].

The graph of networks (GNN) architecture was proposed to create an automatic cellular sorting system in serous effusion cytology [6]. The cells were pre-segmented, and the ROIs were defined by 46 attributes [27] according to size, shape, color, and texture. The isolated cells were classified by a GNN. Also, a feature selection was applied by using the SFFS method. 83.54% total recognition rate was obtained in this work. Ta et al. [43] proposed a framework of graph-based tools for the segmentation of serous cytology images. This work aims to create a particular scheme that addresses a various type of problems in microscopic image segmentation with graph-based tools. Finally, it is demonstrated that these tools may be easily adaptable to a variety of discrete data which can be represented by a single graph.

A supervised approach to segment microscopic color images was proposed which is described in three steps [29]. In the first step, different segmentation maps were extracted for each color plane by a marginal approach. Secondly, the intersection point of these maps, which is called label concordance map, was simplified. The simplification process is based on two assumptions that the color planes are correlated, and some pixels are unlabeled.

Thirdly, a self-adaptable region growing method was applied to obtain the final segmented areas. It is noted that the segmentation is very close to the expected result. Cheng et al. [9] advocated a discriminative learning approach for cellular image segmentation. Three new features were devised and incorporated to provide a fused result of appearance, shape and context information. Their approach achieved a better result of 98.12% on a serous dataset which contains 10 microscopic images from serous cytology. They also proposed a learning-based method [10] using a bag-of-words model. To sum up, 82.71% pixel-level and 78.99% object-level segmentation accuracy were achieved. Jin et al. [22] proposed a random forest (RF) based superpixel-level segmentation approach. A pixel-level RF is utilized to define the data penalty and superpixel-level RF to specify the discontinuity penalty terms of the graph-cuts algorithm. On a serous dataset which contains 10 microscopic 512×512 images from cytology and ground-truth annotations for cell nuclei, 87.05% pixel-level accuracy was achieved. Also, in our previous work [1], we attempted to detect nuclei positions using Viola-Jones object detection framework and 89.32% detection accuracy was achieved. We [2] also proposed Stacked Sparse Autoencoder (SSAE) approach with 98.30% detection accuracy.

Since the high success rate of ConvNets in the ImageNet Competition, they are the current state of art deep learning method for many computer vision tasks, such as image classification [23, 36]. ConvNets usually require large training datasets, but it is difficult to implement in the medical field depending on the time and labor cost required to label datasets by experts. However, when only a finite set of training data is available, ConvNets may exhibit overfitting [32]. Recent studies have proved that transfer learning [45] can be used to adapt ConvNets to overcome these problems. Transfer learning is the use of knowledge learned in a large well-labeled image dataset to improve learning in another classification task. This method has been successfully applied to various medical studies [7, 8, 33, 46]. However, any application of this method has not been established on the classification of serous cells in serous effusion cytology. So it is still unclear what is the best network depth and width for serous cell classification given limited training data.

3 Methods

3.1 Convolutional neural networks

The architecture of a ConvNet model refers to the structure that contains a sequence of different layers concatenated one next to the other [23, 36]. Each layer is characterized by neurons that perform different differentiable mathematical operations, to make the neurons learnable through backpropagation (BP). BP is a method to calculate the gradient of the loss function for the weights. The weights (W) of the ConvNet are updated via BP to minimize the classification error on the training set [24]. In general, typical layers include the convolutional (CONV) layer, the pooling (POOL) layer, the fully connected (FC) layer, and the activation functions include the rectified linear unit (ReLU) and the softmax.

In the following equations, for a CNN layer, f_i^{in} and f_j^{out} represent its i-th input feature map and j-th output feature map. $bias_j$ represents the bias term of its j-th output map. n_{in} and n_{out} denote the number of input and output feature maps respectively. On the other hand, for an FC layer, n_{in} and n_{out} represent the size of the input and output feature vectors.

3.1.1 Convolutional layer

CONV layer takes a set of input feature maps which are local regions across the input image and convolves them with convolutional filters to generate the output feature maps. The CONV layer can be defined as:

$$f_j^{out} = \sum_{i=1}^{n} fi^{in} \bigotimes g_{i,j} + b_j (1 \leq j \leq n_{out}) \tag{1}$$

where $g_{i,j}$ denotes the convolutional filter applied to i-th input feature map to obtain j-th output feature map.

3.1.2 Pooling layer

POOL layer, which is usually added to the CONV layer, performs down-sampling by taking the maximum or average value of each sub-region. This process decreases the feature map size and calculation time for the subsequent layers and also provides features to be translation invariant.

$$f_{i,j}^{out} = max \begin{pmatrix} f_{m,n}^{in} & \cdots & f_{m,n+p-1}^{in} \\ \vdots & \ddots & \vdots \\ f_{m+p-1,n}^{in} & \cdots & f_{m+p-1,n+p-1}^{in} \end{pmatrix} \tag{2}$$

where p denotes the pooling kernel size.

3.1.3 Fully connected layer

FC layer performs a linear transformation on the input feature vector.

$$f^{out} = W f^{in} + b \tag{3}$$

where W denotes $n_{out} \times n_{in}$ transformation matrix and b denotes the bias term. For the FC layer, the input is a feature vector which is created by the fusion of the feature maps generated by CONV and POOL layers. As a result, in Eq. 3, n_{in} and n_{out} represent the size of the input and output feature vector.

3.1.4 Rectified linear unit

RELU, which implements a nonlinear activation function to each element in the output feature maps with $f(x) = max(0, x)$, is often connected to the CONV layer. It also provides robustness to small changes such as noise in the input data.

3.1.5 Softmax

The softmax layer utilizes a different activation function that may be nonlinear. The softmax activation function is given by

$$h_i^l = \frac{e^{w_i^l h^{l-1} + b_i^l}}{\sum_j e^{w_j^l h^{l-1} + b_j^l}} \tag{4}$$

where w_i^l denotes the i-th row of W^l and b_i^l denotes the i-th bias term of final layer. We can use h_i^l as an estimator of $P(Y = i \mid X)$ where Y is the associated label of input data vector x.

3.2 Transfer learning with pre-trained ConvNets

The procedure of transfer learning with pre-trained ConvNets consists of two main parts: the construction of pre-trained ConvNets and the fine-tuning phase (Fig. 1). More details are given below.

3.3 Construction of the fully pre-trained ConvNets

A variety of ConvNet architectures include AlexNet [23], GoogleNet [42], ResNet [18] and DenseNet [20] have attracted increasing interest in the area of image classification. All these ConvNets models are trained on a subset of the ImageNet [17] that contains over hundreds of thousands of images with 1000 classes.

AlexNet has five CONV layers and three FC layers. Reducing the spatial dimensions of the output is performed by POOL layers, which follow only the first, second, and fifth CONV layers. The local response normalization (NORM), which is helpful for generalization, is also applied after POOL layer of the first and second CONV layers. **GoogleNet** introduces an interception module that concatenates feature maps produced by filters of different sizes. That makes the network wider and deeper. It consists of 22 CONV layers including 9 inception module contains convolutional filters of three different sizes (5×5, 3×3, and 1×1 pixels) and a pooling one of size 3×3. **ResNet** is much deeper than AlexNet and GoogleNet. It builds shortcut connections to jump over some layers to avoid the problem of vanishing gradient. It consists of 49 CONV layers with five CONV blocks, followed by an average POOL layer and one FC layer. **DenseNet** is similar but different from ResNet. It utilizes direct connections from any layer to all subsequent layers that encourages features reuse throughout the network. It consists of four parts: the first CONV layer, multiple dense blocks, transition layers and finally the FC layer. Each dense block is a stack of CONV layers.

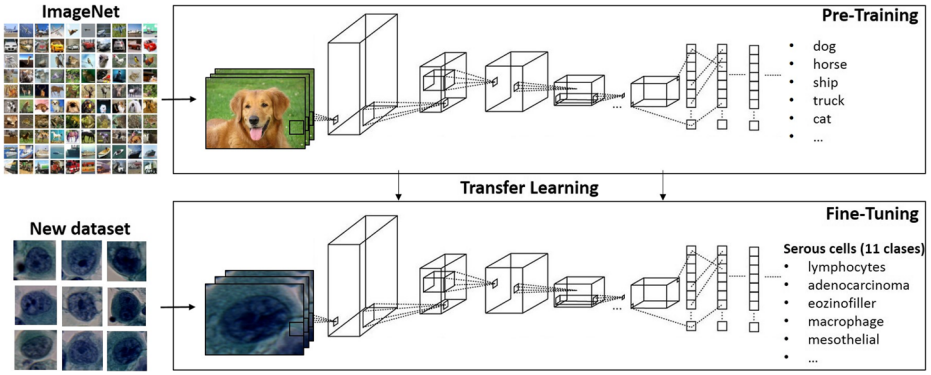


Fig. 1 An illustration of transfer learning architecture. A ConvNet is fine-tuned by transferring the weights from another ConvNet pre-trained on a large image dataset (ImageNet) to serous cell classification

In our experiments, the ConvNet architectures (AlexNet, GoogleNet, ResNet and DenseNet) are utilized as the pre-trained ConvNets. The weights (W) of the ConvNets are updated via BP algorithm [24] to minimize the class error on the training set.

3.4 Fine-tuning the fully pre-trained ConvNets

Fine-tuning is a procedure that aims to train a new ConvNet with transferring the weights of the pre-trained ConvNets (AlexNet, GoogleNet, ResNet, and DenseNet) and they are successfully used in various applications [19, 35, 38]. The number of neurons in the last FC layer of all the pre-trained ConvNets is modified to the number of classes in the new classification task. We have 11-class classification task. To match our dataset with 11 classes, we change the size of output layer of networks from 1000 to 11. The network is fine-tuned using error backpropagation via Stochastic Gradient Descent (SGD).

4 Experiments

4.1 Study group

In this study, the dataset consisting of the serous cytology cells suggested in 2003 was utilized [28]. Samples are smeared over slides, fixed and colored by the Papanicolaou [34] international standard of coloration. The dataset consists of 3652 images each containing an individual serous cell. The 11 classes and the number of samples for each class are given in Table 1. 90% of the serous cells in each class were used to train the ConvNet while the remaining were used to test the performance of the method. Table 1 also shows the distribution of the all the serous cells used in this work. Some sample images for each class are shown in Fig. 2. All pre-trained ConvNets are trained on particular input image sizes. The original input size of the AlexNet 227 × 227 pixels and original input size of the GoogleNet, ResNet and DenseNet is 224 × 224. Due to fixed image size required by the pre-trained ConvNets, each image sample was resized to fit the input layer of pre-trained ConvNets. For AlexNet, all images in the serous cell dataset were resized to 227 × 227

Table 1 The type of classes and the number of cells in each class

Class	Serous cell type	# of cells	training set	test set
1	Activated Lymphocytes	356	320	36
2	Adenocarcinoma	87	78	9
3	Altered polymorphonuclear cells	347	312	35
4	Dystrophic mesothelials	38	34	4
5	Eosinophile	282	254	28
6	Macrophages	595	536	59
7	Mesothelial	599	539	60
8	Mesothelioma	27	24	3
9	Non-Activated Lymphocytes	698	628	70
10	Neutrophile	358	322	36
11	Polynuclear lobe	265	239	26
	Total	3652	3286	366

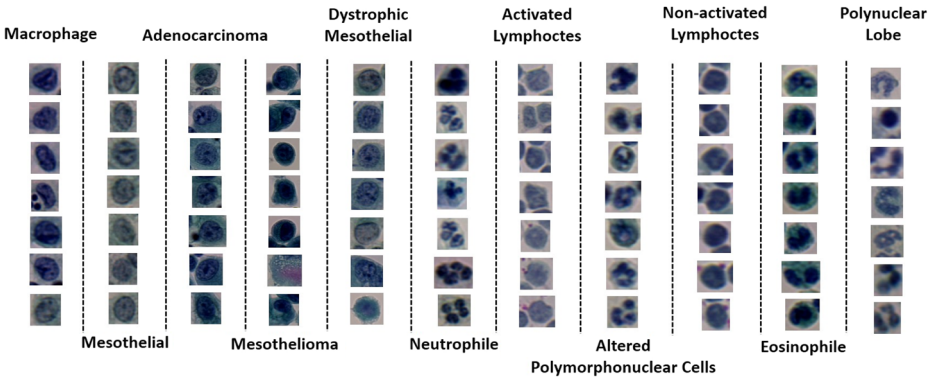


Fig. 2 Typical sample images for each serous cell classes

pixels. For other networks, all images were resized to 224 × 224 pixels. Also, they were processed by applying zero-center normalization and then fed as an input to the network.

4.2 Data augmentation

The easiest and most attracted method to reduce overfitting and improve the accuracy of ConvNets for image classification tasks is to enlarge the dataset by using some transformations [11, 12, 23].

Most common data augmentation techniques are rotating, translating, horizontal-vertical mirroring and scaling the image. In this work, we applied rotation, translation and horizontal-vertical mirroring. The angle of rotation θ was determined by the number of samples for each class in the dataset. We also translated (t pixels) each cell randomly. θ was selected within [-180, 180] and t was selected between [-5, 5]. Scaling was not performed because the size is an essential feature to differentiate abnormal cells from the normal ones.

4.3 Network details

The number of training epochs was varied to 20, 30, **40**, 50, etc., in which one epoch corresponds to the complete training of the entire training set. SGD is utilized to train all models for 40 epochs. To improve the learning speed, the network uses mini-batch that controls how many samples are used to calculate the gradient descent and error for each iteration. The mini-batch sizes of training are 64, 128, 32 and 16 for AlexNet, GoogleNet, ResNet and DenseNet respectively. The base learning rates are 0.001 for all models. Momentum is set to be 0.9 for all networks. All performance evaluations were performed using NVIDIA GeForce GTX 1050 Ti GPU.

4.4 Evaluation metrics

Five metrics have been used in this research including accuracy, sensitivity, specificity, precision, F-measure to evaluate the classification performance. For this reason, the number of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN) was obtained for each classifier. In the 2-class problem; TP is the number of positive samples correctly classified, FP is the number of negative samples incorrectly classified, TN is the

number of negative samples correctly classified and FN is the number of positive samples incorrectly classified.

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5}$$

$$sensitivity \quad (or \quad recall) = \frac{TP}{TP + FN} \tag{6}$$

$$specificity = \frac{TN}{TN + FP} \tag{7}$$

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

$$F - measure = \frac{2 * (precision * recall)}{(precision + recall)} \tag{9}$$

5 Experimental results

The algorithms were implemented in MATLAB R2018b (MathWorks) software using CPU@ 3.4 GHz, 16GB RAM, 64-bit operating system. Transfer Learning with Fully Pre-trained Deep ConvNets was employed to classify 11 serous cell types. We firstly compared the classification results obtained separately on the dataset generated by applying image rotation, translation and mirroring with the dataset without data augmentation. The number of samples for each class in both datasets is given in Table 2. As can be seen in Table 3, before the data augmentation process was applied, 81.42% accuracy was achieved with AlexNet. This accuracy was 91.25% after applying data augmentation. Table 3 shows that data augmentation process improves the accuracy by approximately 10% over all the models.

We secondly investigated the effect of transfer learning compared to the training from scratch (random initialization) for serous cell classification. As seen in Fig. 3, the AlexNet model with transfer learning was converged in less iteration than the training of the AlexNet from scratch and consequently the computation burden was reduced. In Table 3, we denote the AlexNet with random initialization and transfer learning as AlexNet-RI and AlexNet-TL. Table 3 shows that transfer learning (AlexNet-TL) improves the accuracy by 4.36% over AlexNet-RI. This result indicates that transfer learning provides a much better initialization of ConvNet parameters than random initialization. Due to having more complexity of GoogleNet, ResNet and DenseNet, million of images are required to properly train these models with random initialization [31]. Therefore, random initialization was not performed for these models.

10-fold cross-validation (CV) was applied for the generalization of the classification results throughout the experiments. The three most popular classification algorithms,

Table 2 Number of cells for each class in training dataset without (dataset A) and with data augmentation (dataset B)

	1	2	3	4	5	6	7	8	9	10	11
dataset A	356	87	347	38	282	595	599	27	698	358	265
dataset B	2136	2088	2082	2128	1974	1785	1797	1863	2094	1790	1855

Table 3 Best average results for 11-class serous cell classification using different methods (Best results are bolded)

	Method	Accuracy(%)	Sensitivity(%)	Specificity(%)	Precision(%)	F-measure(%)
without data augmentation	k-NN	52.57	63.88	92.49	47.91	54.75
	SVM	58.26	41.66	95.49	50.00	45.45
	RF	58.80	58.33	95.19	56.75	57.52
without data augmentation	**AlexNet-TL**	81.42	88.89	93.94	61.54	72.73
	GoogleNet-TL	81.97	91.67	96.67	75.00	82.50
	ResNet-TL	82.24	83.33	95.45	66.67	74.07
	DenseNet-TL	85.25	88.89	96.67	74.42	81.01
with data augmentation	**AlexNet-RI**	86.89	69.44	99.70	96.15	80.65
	AlexNet-TL	91.25	88.89	**99.70**	**96.96**	92.74
	GoogleNet-TL	91.53	83.33	**99.70**	96.77	89.55
	ResNet-TL	**93.44**	91.67	99.39	94.29	92.96
	DenseNet-TL	92.90	**97.22**	98.79	89.74	**93.33**

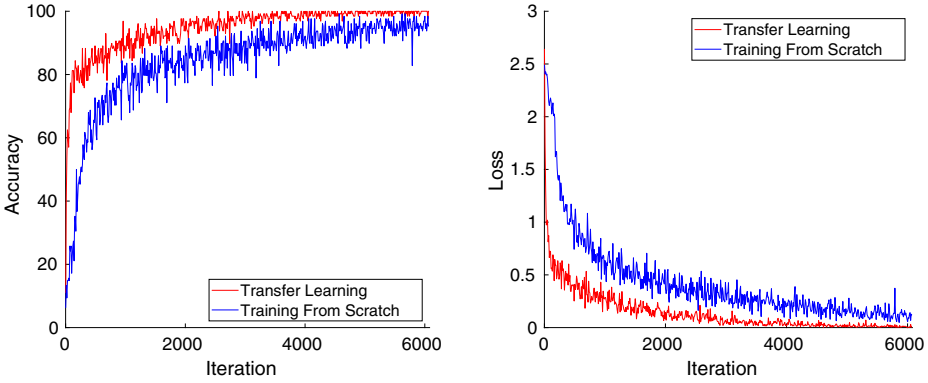


Fig. 3 Comparison of training accuracy and loss when transfer learning is implemented versus training a network from scratch.

namely k-Nearest Neighbors (k-NN) [14], Support Vector Machines (SVM) [13] and Random Forest (RF) [4] were applied for performance evaluation. Its purpose is to demonstrate the superiority of deep learning approaches over classical machine learning approaches. Table 4 shows the optimum configuration parameters of these three classifiers used for the classification regarding the results obtained.

The results of the accuracy, sensitivity, specificity, precision, and F-measure of all proposed and popular classifiers were shown in Table 3. It can be seen that the best accuracy of 93.44% is achieved by using the ResNet with data augmentation (AUG). The specificity and sensitivity, in this case, were 99.39% and 91.67%, respectively. DenseNet had the highest sensitivity of 97.22% and F-measure of 93.33%, but at the expense of a lower precision and specificity than ResNet. Also, its complexity is relatively high and undesirable practical use. AlexNet and GoogleNet had the highest specificity of 99.70%, but a lower accuracy and sensitivity. Sensitivity is clinically important for early diagnosis of cancer. High accuracy can give more correct information about the cancer. ResNet and DenseNet had the highest accuracies of 93.44% and 92.90%. However, when two models were compared in terms of accuracy and model complexity, ResNet-TL was selected as the best network model.

In Fig. 4, the second and third layer convolutional filters of ResNet were visualized. It is seen that some blurry filters have been learned without transfer learning (ResNet-ImageNet). However, in ResNet-TL, the higher level representations of shapes and objects that also allow us to capture cell image details are learned through fine-tuning.

The confusion matrix is used to show the classification performance of 11-class task. Table 5 shows the confusion matrix for the ResNet model. CA denotes the classification accuracy (%) that is calculated as the proportion of the number of correctly classified cells in class i against the total number of cells in class i. According to the Table 5, almost all

Table 4 Optimum parameters used for classification

Classifiers	Optimum parameters used for classification
k-Nearest Neighbors (k-NN)	k=5, distance=euclidean, rule=nearest
Support Vector Machine (SVM)	RBF Kernel with kernel width=3
Random Forest (RF)	Number of trees=100

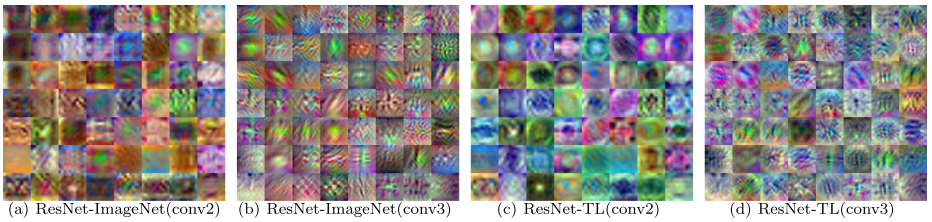


Fig. 4 Visualization of the second (conv2) and the third (conv3) layer convolutional filters of ResNet on the ImageNet and learned on our serous cell dataset using the parameters learned from the ImageNet database as initialization

serous cell types were correctly classified with more than 90% accuracy. The low accuracy of Class 4 (dystrophic mesothelial) is due to the fact that there are only 38 samples of this class in the dataset. Only 4 samples can be used in the test set because data augmentation is applied to the training set only. Therefore, if the number of samples in the dataset is increased, the success is expected to increase.

The proposed method ResNet-TL also yielded the best classification performance as compared to the published works recorded in Table 6. This shows that the proposed method can achieve better results without requiring separate steps such as nucleus/cytoplasm segmentation and feature selection. Only 11 classes of the dataset are publicly shared so we tested the transfer learning approach for 11-class classification using pre-trained deep convolutional networks.

To show the effectiveness of our proposed method, it is also evaluated on the Herlev cervical cell dataset by 5-fold CV. Herlev dataset consists of 917 images each containing one cell. There are seven types of classes diagnosed by two cyto-technicians and a doctor (see [30] for more details). Among the four ConvNets, ResNet-TL and DenseNet-TL obtained the highest classification accuracies of 66.84% and 68.15% for 7-class classification task. Table 7 shows the performance of the method on Herlev dataset, which surpasses the previous deep learning results of 64.5% in [30] and 61.1% in [21].

Table 5 Confusion matrix for 11-class serous cell classification using ResNet-AUG

	1	2	3	4	5	6	7	8	9	10	11	Total
1	33	0	0	0	0	1	0	0	2	0	0	36
2	0	9	0	0	0	0	0	0	0	0	0	9
3	0	0	31	0	0	0	0	0	0	4	0	35
4	0	2	0	2	0	0	0	0	0	0	0	4
5	0	0	0	0	27	0	0	0	0	1	0	28
6	1	0	0	0	0	54	4	0	0	0	0	59
7	0	0	0	0	0	1	59	0	0	0	0	60
8	0	0	0	0	0	0	0	3	0	0	0	3
9	1	0	0	0	0	1	0	0	66	1	1	70
10	0	0	1	0	0	1	0	0	0	34	0	36
11	0	0	0	0	0	0	0	0	1	1	24	26
CA (%)	91.7	100.0	88.6	50.0	96.4	91.5	98.3	100.0	94.3	94.4	92.3	366

Table 6 A comparison of selected studies in the automated classification of serous cells using the serous database (Best result is bolded)

Author	# of Class	Approach	Accuracy (%)
Lezoray et al. [28]	18	Feature selection (SFFS) Feature extraction [27] Classification (MONNA)	84.0
Cardot et al. [6]	18	Feature selection (SFFS) Feature extraction [27] Classification (GNN)	83.54
This work	11	Classification **(ResNet-AUG)**	**93.44**

6 Discussion

The studies in this work show that 11-class serous dataset was correctly classified by using Transfer Learning with Fully Pre-trained Deep ConvNets. One of the advantages of this appearance-based method compared to conventional feature-based methods is that it does not require an accurate nucleus/cytoplasm segmentation.

In general, it is known that deep learning approaches require a large number of training examples. Although the finite number of training samples, the accuracy of the proposed method was relatively high (above 80%) even in the absence of data augmentation. Furthermore, the training samples were augmented by rotation, translation, and reflection, and the accuracy was further improved.

The dataset proposed in [28] was used for the study of serous cell classification. Lezoray et al. [28] extracted 46 features with the use of the regions obtained after applying the segmentation to each cell. When using the SFFS feature selection method together with the MONNA approach, 84% classification accuracy was achieved for 18 classes. Also Cardot et al. [6] used the SFFS feature selection method with the GNN approach, and 83.54% accuracy was achieved. Since 11 classes of the dataset were publicly shared, we tested the transfer learning approach for 11-class classification using pre-trained deep convolutional networks. Among the four ConvNet models, ResNet and DenseNet obtained the highest accuracies of 93.44% and 92.90%. However, when two models were compared in terms of accuracy and model complexity, ResNet-TL was selected as the best network model. We can infer that better classification accuracy can be achieved by using transfer learning with pre-trained deep CovNets without precise cell segmentation. The result can be used efficiently in the automated screening of serous effusion cytology, which may be useful for cytopathological evaluation.

Table 7 Accuracy comparison of different models for cervical cell classification

Model	Accuracy(%)
AlexNet-TL	62.94
GoogleNet-TL	63.79
ResNet-TL	**66.84**
DenseNet-TL	**68.15**

In this work, the main contribution is the implementation of a transfer learning for the automated classification of serous cells into 11-classes. Also, the advantage of the method is that distinct precise cell segmentation, feature extraction, and feature selection steps are not required.

7 Conclusion

In this work, the transfer learning by adapting four pre-trained deep learning architectures was proposed to achieve an appearance-based cell classification. To improve the classification performance, the number of the training samples was increased by adding more image with making rotation, translation and mirroring processes. 93.44% accuracy with a sensitivity of 91.67% and a specificity of 99.39% were achieved. Based on the works in this paper, the appearance based serous cell classification techniques can be effectively implemented for automated screening of serous effusion cytology. In the future studies, a new serous cell dataset will be gained into literature by collecting more samples from different screening to improve the classification accuracy.

Acknowledgements This work has fully been supported by the TUBITAK Research Project 117E961.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

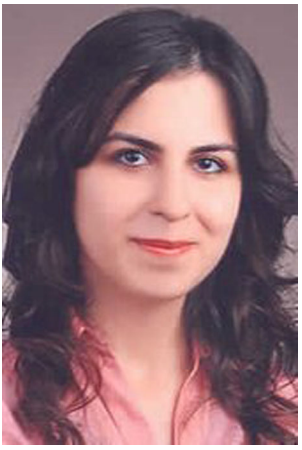
References

1. Baykal E, Dogan H, Ekinci M, Ercin ME, Ersoz S (2017) Automated nuclei detection in serous effusion cytology based on machine learning. In: Signal Processing and Communications Applications Conference (SIU). IEEE, pp 1–4
2. Baykal E, Dogan H, Ercin ME, Ersoz S, Ekinci M (2018) Automated nuclei detection in serous effusion cytology with stacked sparse autoencoders. In: 2018 26Th signal processing and communications applications conference, SIU. IEEE, pp 1-4
3. Bedrossian CW (1998) Diagnostic problems in serous effusions. Diagn Cytopathol 19(2):131–137
4. Breiman L (2001) Random forests. Mach Learn 45(1):5–32
5. Cakir E, Demirag F, Aydin M, Unsal E (2009) Cytopathologic differential diagnosis of malignant mesothelioma, adenocarcinoma and reactive mesothelial cells: a logistic regression analysis. Diagn Cytopathol 37(1):4–10
6. Cardot H, Lezoray O (2002) Graph of neural networks for pattern recognition. In: 2002. Proceedings. 16th international conference on Pattern recognition. IEEE, vol 2, pp 873–876
7. Carneiro G, Nascimento J, Bradley AP (2015) Unregistered multiview mammogram analysis with pre-trained deep learning models. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, pp 652–660
8. Chen H, Ni D, Qin J, Li S, Yang X, Wang T, Heng PA (2015) Standard plane localization in fetal ultrasound via domain transferred deep neural networks. IEEE J Biomed Health Inf 19(5):1627–1636
9. Cheng L, Ye N, Yu W, Cheah A (2011) Discriminative segmentation of microscopic cellular images. Med Image Comput Comput-Assist Inter–MICCAI 2011:637–644
10. Cheng L, Ye N, Yu W, Cheah A (2012) A bag-of-words model for cellular image segmentation. Advances in Bio-Imaging: From Physics to Signal Understanding Issues, pp 209–222
11. Ciregan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: 2012 IEEE Conference on Computer vision and pattern recognition (CVPR). IEEE, pp 3642–3649
12. Cireşan DC, Meier U, Masci J, Gambardella LM, Schmidhuber J (2011) High-performance neural networks for visual object classification. arXiv:11020183
13. Cortes C, Vapnik V (1995) Support-vector networks. Mach Learn 20(3):273–297

14. Cover T, Hart P (1967) Nearest neighbor pattern classification. IEEE Trans Inf Theory 13(1):21–27
15. Davidson B, Firat P, Michael CW (2011) Serous effusions: Etiology, Diagnosis, Prognosis and Therapy. Springer Science & Business Media
16. DeBiasi EM, Pisani MA, Murphy TE, Araujo K, Kookoolis A, Argento AC, Puchalski J (2015) Mortality among patients with pleural effusion undergoing thoracentesis. Eur Respir J 46(2):495–502
17. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: 2009. CVPR 2009. IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp 248–255
18. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 770–778
19. Hoo-Chang S, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM (2016) Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35(5):1285
20. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: CVPR, vol 1, pp 3
21. Jantzen J, Norup J, Dounias G, Bjerregaard B (2005) Pap-smear benchmark data for pattern classification. Nature inspired Smart Information Systems (NiSIS 2005), pp 1–9
22. Jin M, Govindarajan LN, Cheng L (2014) A random-forest random field approach for cellular image segmentation. In: 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI). IEEE, pp 1251–1254
23. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp 1097–1105
24. LeCun Y, Bottou L, Orr G (2012) Efficient backprop in neural networks: Tricks of the trade (orr, g. and müller, k., eds.) Lecture Notes in Computer Science
25. Lezoray O, Cardot H (2002) Cooperation of color pixel classification schemes and color watershed: a study for microscopic images. IEEE Trans Image Process 11(7):783–789
26. Lezoray O, Elmoataz A, Cardot H, Gougeon G, Lecluse M, Elie H, Revenu M (1998) Segmentation of cytological image using color and mathematical morphology. In: European conference on Stereology, pp 10–pp
27. Lezoray O, Elmoataz A, Cardot H, Revenu M (1999) Arctic: an automatic system for cellular sorting by image analysis. In: Proceedings of Vision Interface, vol 99, pp 312–319
28. Lezoray O, Elmoataz A, Cardot H (2003) A color object recognition scheme: application to cellular sorting. Mach Vis Appl 14(3):166–171
29. Lezoray O (2011) Supervised automatic histogram clustering and watershed segmentation. application to microscopic medical color images. Image Anal Stereol 22(2):113–120
30. Lin H, Hu Y, Chen S, Yao J, Zhang L (2018) Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks. arXiv:181006058
31. Lu L, Zheng Y, Carneiro G, Yang L (2017) Deep Learning and Convolutional Neural Networks for Medical Image Computing. Springer, Springer
32. Lyndon D, Kumar A, Kim J, Leong PHW, Feng D (2015) Convolutional neural networks for medical clustering. In: CLEF (Working Notes)
33. Margeta J, Criminisi A, Cabrera Lozoya R, Lee DC, Ayache N (2017) Fine-tuned convolutional neural nets for cardiac mri acquisition plane recognition. Comput Methods Biomech Biomed Eng: Imaging Vis 5(5):339–349
34. Papanicolaou GN (1942) A new procedure for staining vaginal smears. Science 95(2469):438–439
35. Penatti OA, Nogueira K, dos Santos JA (2015) Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp 44–51
36. Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M et al (2015) Imagenet large scale visual recognition challenge. Int J Comput Vis 115(3):211–252
37. Schneider TE, Bell AA, Meyer-Ebrecht D, Böcking A, Aach T (2007) Computer-aided cytological cancer diagnosis: cell type classification as a step towards fully automatic cancer diagnostics on cytopathological specimens of serous effusions. In: Medical Imaging, International Society for Optics and Photonics, pp 65,140G–65,140G
38. Sharif Razavian A, Azizpour H, Sullivan J, Carlsson S (2014) Cnn features off-the-shelf: an astounding baseline for recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 806–813
39. Sheppard C, Wilson T (1978) Depth of field in the scanning microscope. Opt Lett 3(3):115–117
40. Shidham VB, Atkinson BF (2007) Cytopathologic Diagnosis of Serous Fluids E-Book. Elsevier Health Sciences

41. Shotton DM (1989) Confocal scanning optical microscopy and its applications for biological specimens. J Cell Sci 94(2):175–206
42. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1–9
43. Ta VT, Lézoray O, Elmoataz A, Schüpp S (2009) Graph-based tools for microscopic cellular image segmentation. Pattern Recogn 42(6):1113–1125
44. Tajbakhsh N, Shin JY, Gurudu SR, Hurst RT, Kendall CB, Gotway MB, Liang J (2016) Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE Trans Medical Imaging 35(5):1299–1312
45. Yosinski J, Clune J, Bengio Y, Lipson H (2014) How transferable are features in deep neural networks? In: Advances in neural information processing systems, pp 3320–3328
46. Zhang L, Lu L, Nogues I, Summers R, Liu S, Yao J (2017) Deeppap: Deep convolutional networks for cervical cell classification. IEEE Journal of Biomedical and Health Informatics

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Elif Baykal was born in Trabzon, Turkey. She graduated from Karadeniz Technical University and received her BScE in Computer Engineering from Karadeniz Technical University (KTU), Turkey, in 2012 and 2015, respectively. She is still a PhD student at KTU. She is currently a full time Research Assistant in Computer Engineering Department at KTU. She works on image processing, computer vision and machine learning.

Hulya Dogan was born in Bursa, Turkey. She graduated from Karadeniz Technical University and received her BScE in Computer Engineering from Karadeniz Technical University (KTU), Turkey, in 2011 and 2014, respectively. She is still a PhD student at KTU. She is currently a full time Research Assistant in Computer Engineering Department at KTU. She works on image processing, computer vision and machine learning.

Mustafa Emre Ercin graduated from Karadeniz Technical University, Faculty of Medicine in 2009. He completed his pathology residency in 2014 from Kirikkale University, Faculty of Medicine. He is still a Molecular Pathology PhD student at Yildirim Beyaz't University. He is currently Ass. Prof. in Pathology Department in Karadeniz Technical University (KTU), Turkey. He works on bioinformatic, molecular pathology, and digital pathology.

Safak Ersoz was born in Trabzon, Turkey. She graduated from Karadeniz Technical University Faculty of Medicine in 1993. She had succesfully accomplished pathology residency programme between 1997-2002 in the same instution. Currently, she is working as a proffesor of pathology in the Faculty of medicine.

Murat Ekinci was received his BScE in Electronical Engineering and MScE in Computer Engineering from Karadeniz Technical University (KTU), Turkey, in 1990 and 1993, respectively. He obtained his PhD degree from the University of Bristol, England, in 1997. He is currently a full time Professor in Computer Engineering Department at KTU. He works on computer vision, pattern recognition and image and video processing.

Affiliations

Elif Baykal[1] · Hulya Dogan[1] · Mustafa Emre Ercin[2] · Safak Ersoz[2] · Murat Ekinci[1]

Hulya Dogan
hulya@ktu.edu.tr

Mustafa Emre Ercin
drmustafaemreercin@ktu.edu.tr

Safak Ersoz
sersoz@ktu.edu.tr

Murat Ekinci
ekinci@ktu.edu.tr

[1] Department of Computer Engineering, Karadeniz Technical University, 61080, Trabzon, Turkey

[2] Faculty of Medicine, Department of Pathology, Karadeniz Technical University, 61080, Trabzon, Turkey