H&E Boyalı Histopatolojik Görüntülerin Sınıflandırılması

Selin Gezer, Sezin Biner
Bilgisayar Mühendisliği Bölümü
Yıldız Teknik Üniversitesi, 34220 Istanbul, Türkiye
l1117029@std.yildiz.edu.tr, l1116068@std.yildiz.edu.tr

***Özetçe* —Hematoksilen - Eozin Boyama (H&E), rutin incelemelerde ve patoloji laboratuvarlarında vazgeçilmez bir histolojik boyama tekniğidir. Temel prensibi, mavi-mor renk veren hematoksilen ile, pembemsi kırmızı renk veren eozin'in hücreleri boyamasıdır.Bu işlemde çekirdek mavi,sitoplazma kırmızı renk alır.Oluşan görüntü sonucunda kanser tipi tespit edilir. Bu tespit süreci yıllardır patologlar tarafından gerçekleştirilmektedir.**

Bu çalışmada; H&E boyanmış normal,iyi huylu (benign), mahalli karsinoma (in situ carcinoma) ve yayılan karsinoma(invasive carcinoma) görüntü tiplerinin sınıflandırılması amaçlanmıştır.Sınıflandırma yapılırken evrişimsel sinir ağı,önceden eğitilmiş sinir ağları ve geleneksel sınıflandırma algoritmaları kullanılmıştır. Başarı oranları karşılaştırmalı olarak özetlenmiştir.

Anahtar Kelimeler—Anahtar Kelimeler— Meme kanseri,histopatoloji,görüntü analizi,evrişimsel sinir ağları, önceden eğitilmiş sinir ağları

***Abstract*—Hematoxylin - Eosin Staining (H&E) is an indispensable histological staining technique in routine experiments and pathology laboratories. Its basic principle is that the hematoxylin, which gives blue-purple color, and eosin, which gives pink-red color, stains the cells.In this process, the nucleus turns blue and the cytoplasm turns red. As a result of the resulting image, the type of cancer is determined.This determination process has been carried out by pathologists for years.**

As well known breast cancer is the most common form of cancer among women.In this project the main purpose is to classify four types of H&E stained cancerous tissues as normal, benign, in situ carcinoma and invasive carcinoma. As deep convolutional neural networks have succeeded in histopathological image analysis Convolutional Neural Networks, Pre-trained Neural Networks and traditional classification algorithms are used for classification. Their performances have been compared and summaraized.

Keywords—Keywords—Breast cancer, Computer-Assisted Analysis, , histopathology, image analysis, Convolutional Neural Networks, Pre-trained Neural Networks

I. INTRODUCTION

Medical imaging is used in the early diagnosis and treatment of many diseases.Early diagnosis is very important for cancer patients.We use methods such as PET, MR, ultrasound, X-ray, mammography, tomography to get the images.Interpretation of these images is executed by specialists in hospitals.However, the rate of diagnosis is low due to human factors. Therefore, computer systems are needed. Recently, the field of digital pathology is developing rapidly.There was great progress in the identification and classification of medical images thanks to the developments in machine learning and deep learning.In cancer research, these approaches important to minimize human intervention, to reinforce relevant second views and to provide traceable clinical information.

In this project; classification of normal, benign, in situ carcinoma and invasive carcinoma image types are aimed. When we classify these images,firstly we use K-nearest Neighbors, Support Vector Machines and Random Forest algorithms. Then we classify with CNN using different models.Finally, we classify with pre-trained DenseNet, InceptionV3, VGG16, VGG19, Resnet50 models.

II. RELATED WORKS

Localization and interpolation of anatomical structures is an important step in the radiological workflow. Radiologists usually accomplish this task by identifying some anatomical signatures, image features that can distinguish one anatomy from others.However, considering the diversity in pathology and human factors, the correct diagnosis rate decreases. At this point, the need for computers is increasing.Medical imaging systems have been an important topic for many researchers.

Researchers presented a method for organ- or bodypart-specific anatomical classification of medical images using deep convolutional networks[1]. Specifically, they trained their deep network by using 4,298 axial 2D CT images to learn 5 anatomical classes, i.e., neck, lungs, liver, pelvis, and legs.They focused on training deep models from scratch[2]. To tackle the problem of data insufficiency in training deep CNNs, they expanded their dataset by scaling, translation, and rotation in random over training samples.

They [3]conducted experiments on datasets of thoraco-abdominal lymph node detection and interstitial lung disease classification to explore how the CNN performance changes according to factors of CNN architectures, dataset characteristics, and transfer learning.

Researchers studied for brain tumor segmentation using CNNs in MR images[4]. In particular, they explored small-sized kernels to have the fewer number of parameters but deeper architectures. They trained different CNN architectures for low and high grade tumors

III. TISSUE PREPARATION AND IMAGING

Before we start to discuss the different algorithms which are used in classification,we give an overview of the preprocessing stage of the histopatological images. The regions of interest in the tissue are nuclei and cytoplasm, which are not visible therefore they are dyed with stains that highlight them. The Hematoxylin, which gives blue-purple color, and eosin, which gives pink-red color are used for staining the cell.In this process, the nucleus turns blue and the cytoplasm turns purple. As a result of the resulting image, the type of cancer is determined. Normally the diagnosis is performed by human pathologists after interpreting the stained tissue using a microscope. After staining process the tissues are shown in Figure 1. Present day with advancements in technology WSI scanners are used for slide digitization. This leads to computer aided diagnosis and aims to reduce the workload in a pathology lab and to improve the quality of the interpretation.

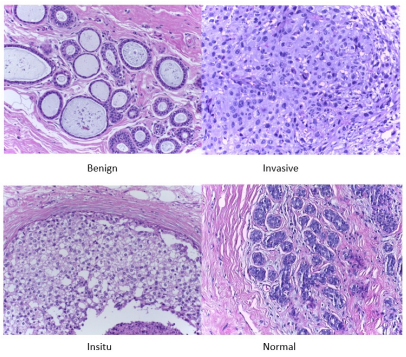

Figure 1 H&E stained tissue after LBP

IV. IMAGE PREPROCESSING

Before applying machine learning algorithms, preprocessing must be performed. When images are classified using deep learning,small sized images are used.Images with large sizes need to be resized. Decrease in the size concludes in decrease in computational power and memory. As our images have a very large image size (1536*2048) mini patches should be extracted in order to preserve the details and features. Otherwise with resizing too much data is lost.Commonly used patch sizes are 256*256,512*512 and each patch is analyzed independently. Also this patch based training reduces model complexity.

Another problem in pathological image analysis is that small number of training data is available. This problem is solved by producing patches. With patch level solution each classes number of samples increase exponentially. Choosing the right patch size is also important. Too small patches will also produce too many look-alike patches and will make it harder to classify.

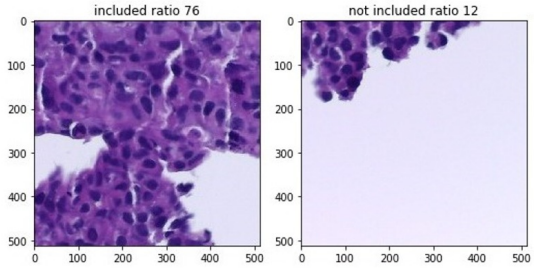

Figure 2 Elimination

An image cancerous tissue can have a part of normal tissue in it, to solve this problem while cropping images and making patches k-means algorithm is also used for detecting regions of interest. After using k-means algorithm for segmentation, we extract patches. After comparing the extracted patches with a treshold value we determine if the patch is useful patch or not.

V. METHODS

Machine learning methods often used in digital pathology image analysis are divided into supervised learning and unsupervised learning. Supervised learning aims to form a function that can map the input images to their labels using the training data. The algorithms for supervised learning include support vector machines, random forest and convolutional neural networks.

A. Traditional Classification Algorithms

We use K-nearest Neighbors, Support Vector Machines and Random Forest algorithms.In addition,the effect on accuracy values were observed by using Local Binary Pattern,before these algorithms.We analyze that applying Local Binary Pattern to the data set has major negative effect on accuracy values because Local Binary Pattern removes too much necessary detail from data set.

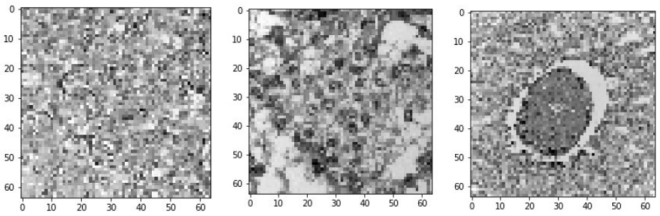

Figure 3 Image examples after LBP

We use GridSearchCV to determine which parameter values give higher results in classification algorithms.As a result of GridSearchCV, we determine the parameters that give the best accuracy values.The highest accuracy values obtain using these parameters.

1) K-nearest Neighbors Algorithm: K-nearest neighbors algorithm [5] is a supervised machine learning algorithm that can be used to solve both classification and regression problem in given dataset. In the k-nearest neighbors algorithm algorithm; Among the n_neighbors':[5, 10, 15, 20, 25], 'weights': ['uniform','distance'], 'algorithm': ['auto','ball_tree','kd_tree','brute'] parameters; algorithm='auto', n_neighbors=10, weights='uniform' parameters provided the highest accuracy values.

2) Support Vector Machines Algorithm: Support Vector Machines Algorithm is a supervised machine learning algorithm which can be used for classification or regression problems. It uses a technique called the kernel trick to transform your data and then based on these transformations it finds an optimal boundary between the possible outputs. In the Support Vector Machines Algorithm ; Among the C: [1, 10, 100], 'gamma': ['auto','scale'], 'kernel': ['rbf','poly','sigmoid'] parameters; C=1,gamma='auto',kernel='rbf' parameters provided the highest accuracy values in given dataset.

3) Random Forest Algorithm: Random Forest Algorithm is one of the many machine learning algorithms used for supervised learning, this means for learning from labelled data and making predictions based on the learned patterns. Support Vector Machines Algorithm can be used for both classification and regression tasks. In the Support Vector Machines Algorithm; Among the 'n_estimators':[100,200],'min_samples_leaf':[2,3] parameters; n_estimators=200 , min_samples_leaf=2 parameters provided the highest accuracy values in given dataset.

B. Deep Learning Algorithms for Classification

Deep learning algorithms such as convolutional neural network starts from feature extraction. Features and classifiers are optimized at the same time. This way deep learning outperforms other traditional ways in histopathological image analysis.

1) Convolutional Neural Networks (CNN) : In the study, the two different CNN models were trained.First model with a simple structure which has 3-4 convolutional layers was only be able to classify images in their original size with low accuracy and high loss values.As that model didn't succeed in classification patches so a more complex model were designed.

Complex model was trained with two different patch sizes (256 x 256 and 512 x 512 pixels) extracted from images and then classified by combining the patches. In the complexed model blocks of layers were used inspired by pre-trained models structures. These blocks include a set of convolutional layers with various kernel sizes. To increase the high-level feature extraction in the models last layers more filters were used. To avoid overfitting more dropout layers were added. To increase accuracy values the models were trained for 200 epochs. Also used checkpoints to stop epochs when the model's validation loss value stopped improving.

We also observed that with imbalanced datasets the models resulted in predicting the class with most samples. To avoid this we equalized all classes number of samples to the class with minimum samples.

2) Pre-trained Convolutional Models: To increase the accuracy value and drop the loss value we used pretrained models. The models used are DenseNet201, InceptionV3, VGG16, VGG19, Resnet50. These networks were trained on more than a million images from the ImageNet database. The pretrained network can classify images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. As a result, the network has learned rich feature representations for a wide range of images. Fine tuning was implemented on models as in this study we only have four classes. We trained with patches of 512 x 512 pixels in all models.

VI. Experimental Results

In the study we classify four types of cancerous tissues labeled as Benign,Invasive,Insitu and Normal. Original Bach Iciar Dataset included 100 images of each class with producing patches this number grow to 5456 images in patch size 512 and 7515 in patch size 256. With patch implementation the classes become imbalanced. The models which are trained without balancing the number of samples in classes were not successful during prediction. They generally predict the class with more samples. To solve this problem Random Under Sampling was used. Before training the number of samples in all classes equalized by eliminating some images randomly. Training with balanced datasets provided better accuracy values. We measured accuracy, precision, recall, and F1 score on the patch-based classification. The experimental results of the experiments are summarized in Table II.

A. Traditional Classification Algorithms

We use methods such as K-nearest Neighbors, Support Vector Machines and Random Forest algorithms for classification.We analyze that support vector machine classification algorithm has the best accuracy and K-nearest neighbors classification algorithm has lowest accuracy in our test images.In addition to these algorithms, we use Local Binary Pattern method in pre-processing state for comparing the results. We analyze that after applying Local Binary Pattern to our training and test images, results decrease significantly.

Table 1 Traditional Classification Algorithms

	KNN	**SVM**	**RF**
accuracy without preprocessing	0.367	0.542	0.538
accuracy after LBP	0.247	0.286	0.254

B. Convolutional Neural Networks (CNN)

Resizing the original images to size 256 resulted in too much data and feature loss. To avoid this we continued with patch-based classification. Extracting too small patches resulted in producing similar samples in every class. This hardened the classification and decrease the accuracy value. In order to solve this problem bigger patches were extracted.

Patch size 512 produced the best accuracy value compared to smaller patches. With size 512 we kept the distintive features in the datasets and minimized the feature loss. Optimizer RMSprop is used. Instead of computing gradient for each patch with RMSprop the model used similar gradients for adjacent patches.

	CNN Model1 (original size)	CNN Model2 (256*256 patch)	CNN Model2 (512*512 patch)
Accuracy	0.562500	0.557751	0.675824
Loss	0.972106230	1.1010276	0.87643847
Precision	0.562500	0.557751	0.675824
Recall	0.562500	0.557751	0.675824
F1 score	0.539816	0.506360	0.673355
Cohens kappa	0.421965	0.418536	0.564789

Figure 4 LayersConvolutional Neural Networks

C. Pre-trained Convolutional Layers

We used pre-trained models to perform better classification. Patches sized 512x512 are used in all models and resized to 224x244 for every model. RMSprop is used as optimizer.

	DenseNet201	InceptionV3	VGG16	VGG19	Resnet
Accuracy	0.948718	0.904762	0.769231	0.793040	0.866300
Loss	0.30607450	0.7017	0.71966958	0.6121	0.8663
Precision	0.948718	0.904762	0.769231	0.793040	0.866300
Recall	0.948718	0.904762	0.769231	0.793040	0.866300
F1 score	0.948834	0.905483	0.771427	0.793640	0.867769
Cohens kappa	0.93367	0.871782	0.690764	0.722688	0.867769

Figure 5 Pre-trained Convolutional

VII. Conclusion

In this study,classification of normal, benign, in situ carcinoma and invasive carcinoma image types are aimed. Firstly we use K-nearest Neighbors, Support Vector Machines and Random Forest algorithms.Among these algorithms, the highest success has been achieved with Support Vector Machines. Then CNN models were designed to increase the accuracy value. When using original sized images, 0.562500 accuracy was achieved with a simple CNN structure. It is aimed to increase feature extraction rate by using two convolution layers one after another.When this structure is tested with patches, the accuracy rate has decreased to zero.In the second model, a complex structure was established. The convolution layers used were supported with max pooling layers and higher success was achieved.Classification with CNN has given better results than traditional classification methods. The expected success in the test data has not been achieved with CNN and traditional classification methods.Therefore, training has been made with pre-trained DenseNet201, InceptionV3, VGG16, VGG19, Resnet50 models.The highest success in these models was achieved in DenseNet201 model.Very close accuracy values to DenseNet201 were observed in the InceptionV3 model.A lower accuracy value was obtained in VGG16 and VGG19 models compared to DenseNet and InceptionV3.Nevertheless, higher accuracy values were obtained than CNN and traditional classification methods.

References

[1] H. R. Roth, C. T. Lee, H. Shin, A. Seff, L. Kim, J. Yao, L. Lu, and R. M. Summers, "Anatomy-specific classification of medical images using deep convolutional nets," *CoRR*, vol. abs/1504.04003, 2015. [Online]. Available: http://arxiv.org/abs/1504.04003

[2] H. R. Roth, L. Lu, J. Liu, J. Yao, A. Seff, K. Cherry, L. Kim, and R. M. Summers, "Improving computer-aided detection using convolutional neural networks and random view aggregation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1170–1181, 2016.

[3] H. Shin, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, and R. M. Summers, "Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[4] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in mri images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[5] G. Guo, H. Wang, D. A. Bell, Y. Bi, and K. Greer, "Knn model-based approach in classification," 2003.