

ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE

**SCORING OF CERbB2 TUMORS IN BREAST CANCER BY PATHOLOGICAL
IMAGE ANALYSIS**



M.Sc. THESIS

Gözde Ayşe TATAROĞLU

Department of Computer Science

Computer Science Programme

Thesis Advisor: Assoc. Prof. Behçet Uğur TÖREYİN

JUNE 2018

ISTANBUL TECHNICAL UNIVERSITY ★ INFORMATICS INSTITUTE

**SCORING OF CERbB2 TUMORS IN BREAST CANCER BY PATHOLOGICAL
IMAGE ANALYSIS**



M.Sc. THESIS

**Gözde Ayşe TATAROĞLU
(704151013)**

Department of Computer Science

Computer Science Programme

Thesis Advisor: Assoc. Prof. Behçet Uğur TÖREYİN

JUNE 2018

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ

**PATOLOJİK GÖRÜNTÜ ANALİZİ İLE MEME KANSERİNDE CERbB2
TÜMÖRLERİNİN SKOR DEĞERLERİNİN BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

**Gözde Ayşe TATAROĞLU
(704151013)**

Bilgisayar Bilimleri Anabilim Dalı

Bilgisayar Bilimleri Programı

Tez Danışmanı: Doç. Dr. Behçet Uğur Töreyn

HAZİRAN 2018



Gözde Ayşe Tataroğlu, a M.Sc. student of ITU Informatics Institute student ID 704151013, successfully defended the thesis entitled “ SCORING OF CERbB2 TUMORS IN BREAST CANCER BY PATHOLOGICAL IMAGE ANALYSIS ”, which she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor : **Assoc. Prof. Behçet Uğur TÖREYİN**
İstanbul Technical University

Jury Members : **Assoc. Prof. Hazım Kemal EKENEL**
İstanbul Technical University

Dr. Süha TUNA
Fatih Sultan Mehmet Vakıf University

Date of Submission : 4 May 2018
Date of Defense : 4 June 2018

I notably thank my family and my dear companion, they supported me with patience while I remained working on this subject.

Yearningly commemorating my days at SimitLab where I conducted most of my work, I offer my sincerest favours to each friends of mine I have met there.

I truly thank Assoc. Prof. Hazım Kemal Ekenel who touched our lives by creating this place and gave us the opportunity to work on many different subjects.

Likewise, I thank my advisor Assoc. Prof. Behçet Uğur Töreyn for his guidance. He unsparingly helped me correct my mistakes.

I also offer my thanks to Dr. Abdulkerim Çapar and Kaan Aykut Kabakçı for helping me with my studies. Lastly, I would like to thank my teammates in Arçelik for their support.



FOREWORD

In this thesis, the pathological specimens which are taken from patients with breast cancer were analyzed by whole slides images. The study was carried out by a research group in Istanbul Technical University under the supervision of Pathology Department in Istanbul Medipol University Hospital. Pathological images are obtained from the Istanbul Medipol University Hospital. Detail information of patients are anonymized.

JUNE 2018

Gözde Ayşe TATAROĞLU

TABLE OF CONTENTS

| | <u>Page</u> |
|--|-------------|
| FOREWORD | ix |
| TABLE OF CONTENTS | xi |
| ABBREVIATIONS | xiii |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xi |
| SUMMARY | xix |
| ÖZET | xxiii |
| 1. INTRODUCTION | 1 |
| 1.1 Purpose of Thesis | 1 |
| 1.2 Literature Review | 3 |
| 1.3 Hypothesis..... | 5 |
| 2. METHODS | 7 |
| 2.1 Dataset | 7 |
| 2.2 Proposed Methods for Pathological Image Analysis | 11 |
| 2.2.1 Classification of CerbB2 Scores..... | 11 |
| 2.2.1.1 Cell based biopsy image scoring..... | 12 |
| 2.2.1.1.1 Cell based feature extraction..... | 14 |
| 2.2.1.1.2 Training with Support Vector Machine (SVM)..... | 15 |
| 2.2.1.1.3 Selection of proper hyperparameters..... | 18 |
| 2.2.1.1.4 Rule based tissue scoring..... | 21 |
| 2.2.1.1.5 Experimental results..... | 23 |
| 2.2.1.2 Scoring cell patches with Convolutional Neural Networks (CNNs).. | 25 |
| 2.2.1.2.1 Deep learning Models for cell based classification..... | 26 |
| 2.2.1.2.2 Cell based dataset preparation..... | 30 |
| 2.2.1.2.3 Training with AlexNet and VGG-16..... | 31 |
| 2.2.1.2.4 SVM classification with extracted features from VGG-16 Network..... | 33 |
| 2.2.1.2.5 Accuracy comparison among CNN models..... | 34 |
| 2.2.2 Detection of CerbB2 Scores on Whole Slide Images..... | 37 |
| 2.2.2.1 Detection with CNN models..... | 39 |
| 2.2.2.2 How does Yolo (You Only Look Once) work?..... | 41 |
| 2.2.2.3 Score Detection by Yolo..... | 43 |
| 3. DISCUSSION and CONCLUSION | 51 |
| CONTRIBUTIONS | 55 |
| REFERENCES | 57 |
| CURRICULUM VITAE | 63 |



ABBREVIATIONS

| | |
|----------------|---|
| CNNs | : Convolutional Neural Networks |
| CerbB2 | : Epidermal Growth Factor Receptor |
| IHC | : Immunohistochemistry |
| FISH | : Fluorescent in Situ Hybridization |
| Caffe | : Deep Learning Framework |
| SVM | : Support Vector Machine |
| LIBSVM | : A Library for Support Vector Machines |
| MIH | : Membrane Intensity Histogram |
| AlexNet | : 8-Layered Deep Learning Architecture |
| VGG-16 | : 16-Layered Deep Learning Architecture |
| CAD | : Computer-Aided Diagnosis |
| ROI | : Region of Interest |
| WSI | : Whole Slide Image |
| R-CNN | : Region-based Convolutional Neural Network |
| Yolo | : You Only Look Once Convolutional Network |



LIST OF TABLES

| | <u>Page</u> |
|--|-------------|
| Table 2.1 : Number of tissue patches are used in cell based feature extraction for SVM classification..... | 10 |
| Table 2.2 : Number of dataset that obtained from whole slide pathologic image samples for score detection approach in WSI | 11 |
| Table 2.3 : CerbB2 score scale system based on ASCO/CAP 2013 recommendations [27] | 13 |
| Table 2.4 : Distribution of data to obtain SVM classifier model by using 5 folds cross validation..... | 17 |
| Table 2.5 : Best hyperparameters with kernel functions of SVM | 20 |
| Table 2.6 : Demonstration of cell based confusion matrices. Darker cells represent accurate prediction of every scores as score 0, score 1, score 2, score 3 respectively | 21 |
| Table 2.7 : Comparison of some score estimation of this study and ImmunoMembrane [5] with the pathologist's score remark through the original pathological images | 24 |
| Table 2.8 : Some cell fragments that are prepared for each score type in the dataset | 31 |
| Table 2.9 : Accuracy results of deep learning architectures..... | 35 |
| Table 2.10 : Performance results by weights of Yolov2 | 49 |



LIST OF FIGURES

| | <u>Page</u> |
|--|-------------|
| Figure 1.1 : The hierarchical presentation of the approaches in this study..... | 5 |
| Figure 2.1 : Image acquisition system. Microscope: Zeiss Axio Scope A1, Camera: Argenit Kameram 2 CCD, Software: Argenit Kameram: Computer: Intel i7 processor, 8 GB RAM, 1TB HDD, Win7 Professional OS..... | 8 |
| Figure 2.2 : H and E biopsy tissue sample..... | 8 |
| Figure 2.3 : Sample breast carcinoma tissue images with (a) Score 0, (b) Score 1, (c) Score 2, and (d) Score | 9 |
| Figure 2.4 : Decomposition of H and DAB channels by the color deconvolution method in [14]..... | 10 |
| Figure 2.5 : Workflow of the proposed IHC Stained CerbB2 mutated breast carcinoma automated tissue scoring system. Tissue samples are fed into the system in order to obtain the resulting ASCO/CAP 2013 tissue score [27]..... | 14 |
| Figure 2.6 : Cell based feature extraction to feature vector form for classification.. | 15 |
| Figure 2.7 : Presentation of classification steps..... | 16 |
| Figure 2.8 : Visualization of optimized data that is used in training. This appearance is constituted by PCA (Principal component analysis) [41] with trained data through the 4 different classes..... | 18 |
| Figure 2.9 : Conditional flow of rule based tissue scoring as suggested in ASCO / CAP 2013 [27]. Rule 1: If the percentage of Score 3 found in the breast tissue sample exceeds 10%. Rule 2: If the percentage of Score 2 found in the breast tissue sample exceeds 10%. Rule 3: If the percentage of Score 1 in the tissue sample exceeds the Score 2..... | 22 |
| Figure 2.10 : AlexNet architecture | 27 |
| Figure 2.11 : VGG-16 architecture with tissue sample in our hand-crafted dataset [53] | 27 |
| Figure 2.12 : Softmax activation function with multi-classes [54] | 28 |
| Figure 2.13 : Graph definition of Hinge Loss, red line defines the loss..... | 29 |
| Figure 2.14 : Example of obtaining cell centered fragments | 31 |
| Figure 2.15 : Representation of experienced CNN model's architectures..... | 32 |
| Figure 2.16 : The general outline of the classification with convolutional networks with cell fragment is obtained from the tissue sample | 32 |
| Figure 2.17 : Representation of functions which are used in output layer | 33 |
| Figure 2.18 : Representation of feature extraction by VGG-16 and usage of these features to form 4 classes SVM classifier model | 34 |
| Figure 2.19 : The confusion matrices show accuracy results for each type of score as follows: (a) Alexnet with Softmax loss function, (b) Hinge loss function with AlexNet (c) Softmax lost function with VGG-16 (d) Hinge loss function with VGG-16 | 36 |

Figure 2.20 : Architectures of three main region based convolutional network39

Figure 2.21 : Sample WSI's are : (a) Score 0, (b) Score 1, (c) Score 2, (d) Score 3 respectively 40

Figure 2.22 : Creating training data from sample WSI 41

Figure 2.23 : Yolov2 architecture with its bounding box approach 43

Figure 2.24 : Examples of detection results with (a) Score 0, (b) Score 1, (c) Score 2, and (d) Score 3 samples respectively by the generated model..... 44

Figure 2.25 : Artificial data samples which obtained by bilinear image rotation for each score type as followingly: (a) Score 0, (b) Score 1, (c) Score 2 and (d) Score 3.....46

Figure 2.26 : WSI Score 2 samples which are applied different threshold values by decreasing order from “1” to “0” : (a) test result by threshold 1, (b) test result by threshold 0.5, (c) test result by threshold 0.25 and (d) test result with threshold 0.1..... 47

Figure 2.27 : The score distribution on tissue sample with the original resolution of ROI through the rescaled WSI 48



SCORING OF CERbB2 TUMORS IN BREAST CANCER BY PATHOLOGICAL IMAGE ANALYSIS

SUMMARY

Nowadays, breast cancer is being investigated as a widespread disease and it is frequently a subject for scientific studies. Many types of interdisciplinary studies emerge concerning breast cancer, particularly about the treatment processes. For treatment purpose, targeted therapy/treatment approaches have become widespread in recent years. In the light of targeted therapy, the distribution of CerbB2 which is a protein of the Epidermal Growth Factor Receptor (EGFR) family, is examined. According to the distribution of CerbB2, the stage of the disease is classified. Score types of CerbB2 are proposed to be classified into four main classes. In this regard, this study analyzes CerbB2 score types by detection and classification methods, which are developed to diagnose the related tumor regions on the pathological images by interdisciplinary approaches. For diagnosis, firstly, whole slide tissue images were obtained from the hospital and analyzed under the supervision of pathologists. After the analysis of the images, the implementation of a real-time software tool is intended. Using this targeted tool, it is foreseen that pathological report for a patient can be formed faster and cheaper. To diagnose the phase of breast cancer, the images are analyzed in three different approaches.

In the first approach, specific features of the cells are extracted from tissue specimens to be trained in the Support Vector Machine (SVM). For starters, a hybrid multi-level thresholding and radial line based cell detection method is applied. Following, a Membrane Intensity Histogram (MIH) method is used to feature extraction for each cell on tissue specimens. Then, these cell features are used to fed SVM classifier. These MIH based features are categorized into four separate score classes, namely, Score 0, Score 1, Score 2 and Score 3. These scores are classified based on the rule which is set by American Society of Clinical Oncology and the College of American Pathologists.

In the second approach, a cell basis deep learning methods are used. By using deep learning models, the study proposes a unique approach to classify CerbB2 scores over tissue patches that has breast cancer. For this purpose, Whole Slide Images (WSI) are divided into tissue patches and labeled based on the score types. Then the data set is used for training, validating and testing with deep learning models. CerbB2 tumor scores are generated for cell fragments that are classified with high accuracy by the aid of convolutional neural networks (CNNs). During this classification, unlike traditional methods, CNN models are modified to get better performance. Cell based features are also extracted from these CNN models to be performed on SVM. Following these, the results obtained with deep learning models are compared with handcrafted cell-based feature extraction method.

In the last approach, CerbB2 Scores are detected on WSI directly by using deep learning models. Since there isn't a previously published pathological dataset, a genuine dataset is prepared. The whole slide images are resized because WSI's are so large to be processed. The coordinates of Region of Interests (ROI) on WSI's are kept as optimized labels. Then the dataset is used in training phase of detection with Region Based Convolutional Neural Networks like Yolo using proper libraries. After the training phase, obtained models are used to diagnose score types on WSI's without any division of tissue specimens as in previous works. Also, the existing set of images are retrained using deep learning models with the augmented dataset by bilinear image rotation. Finally, the WSI's with original resolution is split to form the third pathological dataset. Then, by using this dataset, a third detector model is obtained for better analysis of the score type on the resized region of WSI. In this way, multiple labeled regions can be re-examined by this third model.

At the end of the study, it is aimed to develop a software application using these methods in order to shorten the period for diagnosis comparing with traditional methods. Since it is possible to create faster and cheaper methods for diagnosis using computer programs, the relevant tumor regions can be detected rapidly by applying machine learning approaches on pathological images. For a correctly targeted therapy in breast cancer, the results from these methods will be useful. In the light of these objectives, the contributions of this work to pathological image analysis can be listed as follows: a real-time tissue scoring system, the original dataset is shared in public for contributing to similar studies, diagnosis of score types on WSI's. This study can also

be categorized as an innovative study, since it has been devised for the classification of a WSI without fragmentation and direct detection of score types on WSI images. In first two approaches of the study, the images are processed through the texture fragments of pathologic images, whilst in last approach, the scores determined over the WSI are detailed.





PATOLOJİK GÖRÜNTÜ ANALİZİ İLE MEME KANSERİNDE CERbB2 TÜMÖRLERİNİN SKOR DEĞERLERİNİN BELİRLENMESİ

ÖZET

Günümüzde, meme kanseri, yaygınlaşan bir hastalık olarak bilinmekte olup, bilimsel çalışmalara da sıklıkla konu olmaktadır. Yaygınlaşan bu kanser türünde, özellikle tedavi sürecinde başarıyı arttırmaya yönelik olarak disiplinler arası birçok çalışma ortaya çıkmıştır. Bu amaç doğrultusunda da, hedefe yönelik tedavi (targeted therapy/treatment) yaklaşımları son yıllarda yaygınlaşmaktadır. Hedefe yönelik tedavi amacıyla epidermal büyüme reseptörü (EGFR) ailesine ait bir protein olan CerbB2 reseptörünün yoğunluk dağılımı incelenmektedir. Bu dağılıma göre hastalığın aşamaları belirli sınıflara ayrılmaktadır. Sınıflandırmalar dört ana Score türüne göre yapılmaktadır. Bu çalışmada da ilgili tümör bölgelerinin tespiti ve Skor sınıflandırılması disiplinler arası bir yaklaşımla patolojik görüntüler üzerinden yapılmıştır. Tüm slayt doku örneği görüntüleri, hastaneden alınarak patoloğların gözetimi altında incelenmiş ve slaytın doğru etiketlenmesi için, ilgili bölgelerin nasıl tespit edileceği öğrenilmiştir. Görüntülerin analizinden sonra, gerçek zamanlı bir yazılım aracının uygulanması amaçlanmıştır. Hedeflenen bu uygulama sayesinde de hasta için daha hızlı ve daha ucuz bir şekilde patolojik rapor oluşturulması öngörülmektedir. Bu amaç doğrultusunda, görüntüler üç yaklaşımla analiz edilmiştir.

Çalışmanın ilk aşamalarında, doku örnekleri üzerinden Destek Vektör Makineleri (DVM) ile eğitilmek üzere uygun veri kümesi oluşturulmuştur. Öncelikle, hibrid bir çok seviyeli eşikleme ve radyal çizgi tabanlı bir yaklaşımla hücrelerin algılanması yöntemi geliştirilmiştir. İkinci olarak, bir Çeper Yoğunluk Histogramı (ÇYH) metodu ile her bir hücreye özgü öznitelikler oluşturulmuştur. Daha sonra, bu öznitelikler DVM sınıflandırıcısını beslemek için kullanılmıştır. Elde edilen sınıflandırıcı model ile 'Amerikan Klinik Onkoloji Derneği ve Amerikan Patologları Üniversiteleri' tarafından belirlenen kurala göre skor tipleri belirlenmiştir.

Bu çalışmanın ikinci aşamasında, hücre temelli bir derin öğrenme yaklaşımı ele alınmıştır. Çalışma, CerbB2 tümörlerini, meme kanserine ait doku parçaları üzerinden, derin öğrenme modelleri ile sınıflandırmak için özgün bir yaklaşım önermektedir. Bu amaçla, tüm slayt görüntüleri, doku parçalarına bölünüp ve skor türlerine göre etiketlenmiştir. Daha sonra bu veri seti, derin öğrenme modelleri ile eğitim, doğrulama ve test için kullanılmıştır. Hücre parçaları için CerbB2 tümör skorları üretilmiş ve Evrişimsel Sinir Ağları (ESA) yardımıyla yüksek performansla sınıflandırılmıştır. Bu sınıflandırma yaklaşımı sırasında, daha iyi performans elde etmek için modeller değiştirilmiş ve modeller üzerindende hücre gruplarının öznelikleri çıkartılarak DVM sınıflandırıcısı ile eğitilmiştir. Elde edilen sonuçlar manuel öznelik çıkarımı ile elde edilenlerle kıyaslanmıştır.

Bu çalışmanın son aşamasında ise bölgesel tabanlı derin öğrenme modelleri kullanılarak tüm slayt görüntüsü üzerinden CerbB2 reseptörüne ait skor tespiti yapılmıştır. Derin öğrenme yöntemi ile eğitim için patolojik görüntülerden oluşan hazır bir veri seti olmadığından orijinal veri kümesi hazırlanmıştır. Herbir slayt görüntüsü işlenemeyecek boyutta büyük olduğu için de patolojik resimler yeniden boyutlandırılmış ve tümörlerin bulunduğu koordinatlar optimize edilere etiketlenmiş. Daha sonra bu veri kümesi uygun kütüphane kullanımı ile Yolo gibi bölgesel yaklaşımlı evrişimsel sinir ağlarıyla eğitilmiştir. Eğitimler sonunda elde edilen modeller ile önceki çalışmalardan farklı olarak tüm doku örneği üzerinden, skor tespiti yapılmıştır. İkinci bir basamak olarak, var olan görüntü kümesinin rotasyon değişimiyle sayıca arttırılan veri kümesi, derin öğrenme modeli kullanımıyla yeniden eğitim yapılmıştır. Son olarak, tüm slayt görüntüsünden orjinal çözünürlükte yeni bir veri kümesi oluşturulup yine derin öğrenme modelleriyle eğitilip model oluşturuluyor. Daha sonra bu model, birden fazla skor tipiyle etiketlenmiş skor belirsiz bölgeler için kullanılmaktadır. Yeniden boyutlandırılmış tüm slayt görüntüsü üzerinden tespit edilen tümör bölgeleri, tüm slayttan ayrılıp orjinal boyutunda yeniden ölçeklendirilerek, elde edilen son model ile test ediliyor. Böylelikle, yeni ağırlık modelleri ile, ilgili skor bölgeleri detaylı olarak analiz edilip, kurallar dahilinde nihai skor sınıfı ile belirlenebiliyor.

Çalışmanın sonunda pataloglar tarafından, hastalardan alınan doku örneklerine konulan teşhislerin sürelerinin kısaltılması için ele alınan bu yöntemlerin uygun kullanımıyla bir yazılım uygulaması geliştirilmesi hedeflenmektedir. Böylelikle meme

kanserinde doğru tedavi için patolojik görüntüler üzerinden, ilgili tümör bölgelerinin tespitinin, önerilen makine öğrenmesi yöntemleri ile hızlı ve doğru bir şekilde basit bilgisayar programları ile geliştirilebileceği düşünülmektedir. Gerçek hayatta uygulamaya yönelik bu katkı dışında bu çalışmanın diğer katkıları şöyle sıralanabilir: teknik olarak gelecek çalışmalara katkı sağlamak adına özgün dataset üretilerek paylaşılmaktadır, tüm bir slayt görüntüsünün gerek parçalara bölünerek sınıflandırma yöntemi ile gerek parçalara bölünmeden ilgili tümör bölgelerinin tespiti yaklaşımıyla yenilikçi bir çalışma ele alınmıştır. Çalışmanın ilk iki basamağında doku parçaları üzerinden görüntüler işlenirken son basamağında nasıl tüm slayt üzerinden skor tespiti yapıldığı detaylandırılmaktadır.





1. INTRODUCTION

Today, breast cancer is one of the most well-known and the most under-researched cancer type in the world. Since technology has been developing day by day, detection of breast cancer and its treatment has been faster and accurate by co-working of various disciplines. The diagnosis of cancer is made on a biopsy from patient who have been suspected to have cancer. In daily routine, cancer is diagnosed by pathologists, using light microscopes; in this way, the images seen on microscopy can also be used as an optical test sample. As a result, not only physicians are working on detection of cancer on tissue specimen, but also engineers could contribute to this process by technological developments. In this regard, both robotic devices and computer-assisted approaches have started to be used in the cancer diagnosis by microscopic image analysis. Therefore, evaluation of tissue specimens became an engineering problem by utilization of computer and electronical resources.

1.1 Purpose of Thesis

In this work, in order to mitigate the interobserver discrepancy, novel image analysis methods were applied to determine the Score type of the CerbB2 tissue specimens which are taken from patients with breast cancer. For this purpose, whole slide tissue specimens were analyzed both in cellular basis where each cell is identified using membrane based feature extraction and classification methods and tissue patch basis which were handled with deep learning models. The Scores of tissue specimens were analysed under the title of classification and detection problems.

There are four type of CerbB2 Scores namely: Score 0, Score 1, Score 2, Score 3. While Score 0 and Score 1 represents early phase of the breast cancer, Score 2 and Score 3 represents last phase of the cancer. Based on these phases, a targeted therapy is applied to the patients. To detect Score of the tissue specimens, the whole slides tissue specimens were analysed under the three main stages. In first two part of the study Scores were defined under the classification problem and the last part of the study Score were analysed under the detection problem.

In the first two part of the work, tissue images were observed by splitting the WSI into small patches to process them with image processing methods. Since tissue specimens are so big sized images, it is hard to examine them on the simple computer screen. Pathologist screen the images by the microscopes to detect ROI on the tissue sample, so it harder to analyse WSI images patch by patch by microscopes. For this reason to diagnose an whole slide image, the tissue specimens should be examined by various pathologists to avoid from any vital error. In this point this point Computer Aided Diagnosis (CAD) can help by reducing diagnosis phase with machine learning and computer vision techniques [1]. A software application can be applied as an artificial pathologists by experimental technics in this study. If a suitable software application can be applied, expensive medical machines won't be needed or diagnosis period for breast cancer will be reduced. These are the main purposes in this study. For this aim, features of tissue images were extracted by both traditional image processing method and deep learning models to classify Scores accurately. Also ROI were detected over the WSI by suitable Score labels. Therefore one of the challengable problem in pathological image analysis which is Score detection on WSI large images could be solved by proposed deep learning model.

Another purpose in this study to present classifier and detector models with high accuracy to define score type of tissue samples. During creating models original pathological image datasets are produced and shared in public. To analyze score of the tissue sample, WSI's are divided into proper patches to process and each patches labeled with accurate score type. For detection phase of the problem, WSI's are labeled based on the location of ROI on the images and prepared dataset to perform with CNNs. By CNN models, an application is aimed to provide in which the tumorous areas on the pathological tissue sample of breast cancer can be detected quickly and the scoring can be determined. Thus, these regions will be examined by pathologists in more detail, and the cancer stage will be determined in the most accurate way.

To sum up, pathological image analysis is one of the new topics as inter-observer disciplinary study. Those whole slide biopsies with breast cancer consist of healthy, benign and cancerous tissues at various stages and thus, simultaneous detection and classification of diagnostically relevant regions are challenging. In this regard, the study was conducted by a group of computer scientists and medical researchers. Diagnosis period of the breast cancer is long and expensive process, so this study can

prove that this kind of computer vision and machine learning approaches can help to be improved this period. Also this study present different approaches for large size image processing with a reasonable computational power and time.

1.2 Literature Review

Since the cancer is growing rapidly and its treatment is a very difficult disease, it has been the subject of scientific works from almost every disciplines. As the cancer diagnosis process is costly and it is not be allowed error tolerant, technological solutions have been started to develop. Along with the developing technology, computer-aided applications have been developed in order to provide faster and more accurate solutions to cancer diagnosis. In this regard, most of the cancer type is examined over the biopsy images as in breast cancer, prostate cancer. Therefore it is easy to see various studies about pathological image analysis to identify tumor regions over the tissue specimens.

In connection with our work, there are similar works in analysis of biopsy images with breast cancer. These works can differs from each others with difference in dataset type or methods which provide an approach to identify irregular cell division or classification of tumor type. In terms of content, the most similar work with this study has common purposes and one of this work is created in [2]. In that study, the biopsy images were screened with different type of staining whole slide images [3]. Also the cancer regions were classified into five different diagnostic categories of breast cancer, namely as: Non-proliferative changes only (NP), Proliferative changes (P), Atypical ductal hyperplasia (ADH), Ductal carcinoma in situ (DCIS), Invasive cancer (INV)) [2]. That project aimed to analyse WSI by detecting of ROI and classification of that detected region based on the diagnostic categories of breast cancer. They evaluated the WSI images into five different resolution as 20x, 10x, 5x, 2.5x, 1.25x, 0.625x. They zoomed in on the biopsy images as they are given order. They used FCN (Fully Convolutional Network) for detection which is different than in our study [4]. It provide to processing for different size of WSI by using Sliding Windows method by Caffé into Matlab programming tool. That study process the images by small patches and they obtained these patches with their proper labels from Washington University. The dataset were labeled with special tools by 45 pathologists, so they could get more experience with that large and accurate dataset. In our study we formed our dataset by

ourself under the supervision of pathologists in İstanbul Medipol University Hospital. Although, that study contain similarities basically, our study show so different approach by methods, dataset and detection techniques.

Another similiar study in biopsy image analysis in breast cancer was dealt by Helin [5]. In the first stage in our study, we kept our search by comparing tests in free image processing tool which is provided by that study. That stduy provide a tool identify Scores of tissue samples according to H-Blue and D-Brown staining type as in our first approach [5]. However they don't share any information about their classification method in their free tool, the tool gives relatively true predictions for Score type of that pathological images. Even though their dataset type is similiar with our data, our cell-based tissue calssification approach has show better performance than thier proposed free tool. Detail information about that comparison will be given in method section.

The other approaches in digital pathology can be examines into two main titles as image analaysis with CNNs and with traditional image processing methods. In some works the tissue images were evaluated under the segmentation and binary classificaiton problem as in [6]. Firstly, tumor regions were detected as cancerous and non-cancerous by CNNs, then these regions were segmented. On the other hand, CNNs were used for automatical cell detection to classify expressing proteins. The study in [7] applied this approach for classification of Cervical Intraepithelial Neoplasia (CIN) as high grade CIN, or no abnormality (normal cervix). Evaluated cancer type differentiate that study than our work which is related with breast cancer. The CIN was tried to classified into four categories namely; p16 positive, Ki-67 positive, p16 and Ki-67 positive, p16 and Ki-67 negative [7].

Differently from other digital pathology studies, estrogen receptors in breast cancer were evaluated by graphing theory [8]. The research group used spanning trees to extract nuclei of cells on the staining images. They observed CD34 proteins to express classes of biopsy images into three categories. In addition to this study, some research group joined to grand challenges to automatic detection of metatstatic breats cancer in WSI [9]. They used Camelyon 16 dataset, so they can benefited from 400 WSI which are large enough for training with multi-layered CNNs like GoogLeNet. They trained these WSI sentinel lymph node biop-sies by splitting them into small patches with staining H&E (Hematoxylin and Eosin) [9]. The group claim that they decreased errors

from 3% to 1% by 27-layered CNN architecture. Therefore the diagnosis by pathologists could be improved by integration of proposed CAD system.

Although there are various approaches which are increasing day-by-day, we compared the basic studies in digital pathology with our study based on the staining type of the biopsy specimens, cancer type and proposed computer vision and machine learning methods. The Figure 1.1 shows a organization diagram which explains approaches of our study by hierarchical subjects. These analysed studies in this part exemplify every kind of studies in digital pathology through a basic sample.

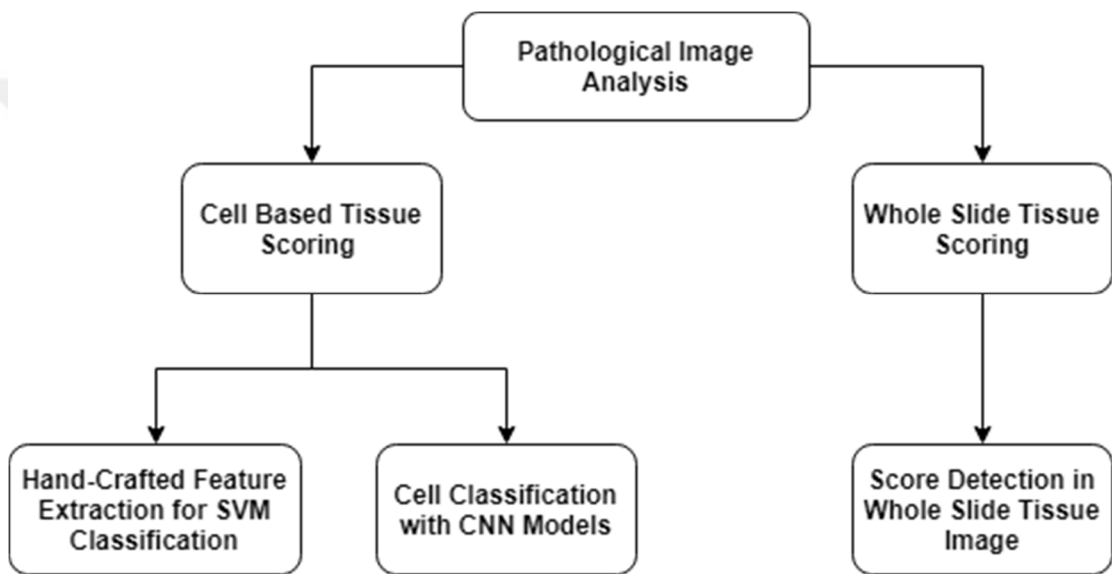


Figure 1.1: The hierarchical presentation of the approaches in this study.

1.3 Hypothesis

This study deals with the suggestion that the error margin can be reduced and the diagnostic period can be shortened in pathological studies performed with CAD and human observation. Computer-aided programs should still be used under the supervision of an extension that people can not tolerate. However, the development of computer vision and machine learning techniques is predicted to be actively used in cancer detection.

In this study, we try to prove that basic computer programs and machine learning techniques can be used to analyze large sized biopsy images. Also by using CNN

models, deeper analysis of that pathological images with their multi-layered and filters can increase accuracy of tumor detection on WSI. Owing to accurate detection of tumors regions, they can be classified with true Scores. For true pathological image analysis, firstly the images should be labeled under the supervision of pathologists. Only in this condition, a proper CAD system can be created and the analysis should be approved by medical doctors. In this regard, some study claims that both computer-aided approaches in digital pathology and pathologists' remark can reduce error rate in diagnosis of cancer by around 85% [9]. Also another approach explains how digital pathology provides an opportunity for better analysis of disease appearance modelling owing to improved prediction of disease phase by deeper observation by computerized imaging tools [10]. In conjunction with this hypothesis, this study aims to prove how the proposed CAD system and pathological remark both decreased error rate of diagnosis of biopsy images.

2. METHODS

2.1 Dataset

In this study raw data obtained from Istanbul Medipol University Hospital was used to prepare data set for analysis of biopsy images. According to the need of the study, the way of data handling changed, so the data was processed and annotated for training phases. Only in first stage of the study was digitized by Argenit Kameram digital microscopy system. In the remaining stages of the study, the raw data were processed by Matlab programme in high computational computers which contain multiple CPUs.

Argenit Kameram consists of Zeiss Axio Scope A1 bright field microscope, 40X objective, 0.63X camera adaptor, Kameram 2 CCD camera (with 1.4 mega pixel sensor resolution), PC and Kameram software (Figure 2.1). The mosaic image acquisition tool of Kameram software is used to capture the panoramic images of tissue regions which are larger than the camera view area. Acquired images are further analyzed using a sequence of image analysis techniques developed specifically for automated tissue scoring. The pathological image on the computer screen exemplify one of the Score 3 tissue image in the dataset.

For analysis of pathological images, staining is needed to make tumor regions colorful, so the images can be screened on the computers. However the staining technique differs based on the receptors or protein type in related tissue specimens. Since different type of proteins which are used to identify tumor Score on the biopsy specimens, show different resistance to staining, different RGB pixel values are obtained. Therefore cancerous and non-cancerous regions were distinguished from each others during the analysis of histopathological images by digitally.

One of the advanced staining technique is called Immunohistochemistry (IHC) which is used for identify specific immunogen in biopsy specimens [11]. IHC was generally used to evaluate spread of estrogen (ER), progesterone (PR), and Human Epidermal

Growth factor 2 (HER2) receptors which is also called as CerbB2 gene [12]. This gene were used to quantify detection of protein increase, so based on the intensity of CerbB2 in the tissue speciemens, the WSI was graded into four classes namely, Score 0,



Figure 2.1 : Image acquisition system. Microscope: Zeiss Axio Scope A1, Camera: Argenit Kameram 2 CCD, Software: Argenit Kameram: Computer: Intel i7 processor, 8 GB RAM, 1TB HDD, Win7 Professional OS.

Score 1, Score 2 and Score 3. Also staining is applied with different type of technique like Hematoxylin (H) and Eosin (E) [13].Hemaloxlylin binds to DNA and change the color of nucleus as blue or purple, while other parts of cells turn to pink. Following Figure 2.2 illustrates H and E tissue sample.

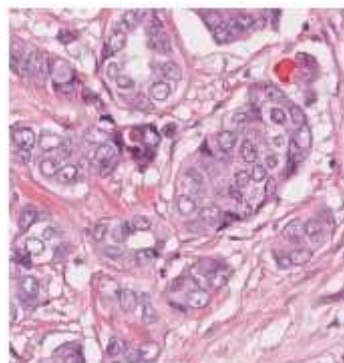


Figure 2.2 : H and E biopsy tissue sample.

The other most common differential staining methods over the color approximations are hematoxylin (H-Blue) and diaminobenzidine (DAB-Brown) which are colored biopsy images are analyzed to detect status of CerbB2 with situ hybridization (FISH) method [14]. Sample tissue patches in Figure 2.3 our dataset ensamples staining with

H and DAB biopsy samples by each Score type. Membranous staining becomes perceptible as scores increase. Tumors with a Score 3 which have intense and uniform circumferential staining are considered as suitable for the targeted CerbB2 protein treatment by pathologists. Also Figure 2.4 presents H-Blue and DAB-Brown tissue image sample by color deconvolution method.

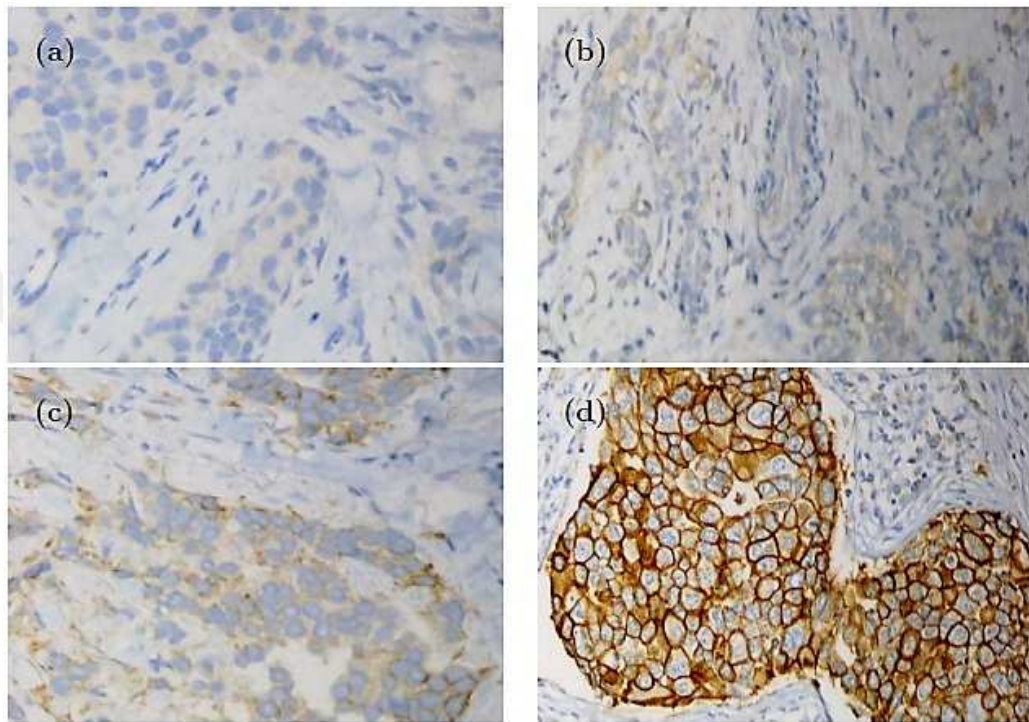


Figure 2.3 : Sample breast carcinoma tissue images with (a) Score 0, (b) Score 1, (c) Score 2, and (d) Score 3.

Furthermore, the number of data was changed according to the computer vision method. In the first stage of this study, “59804” cells were used for cell based Support Vector Machine (SVM) training to create a classifier model. These cells are extracted from total “198” tissue patches which consist of different number of Score type as shown in Table 2.1. Based on the Table 2.1 number of tissue patches in each Score class are close to each other, therefore robust classifier models can be produced. Also these tissue patches are shared as online for other researchers who want to study in this field by link in [40]. In the second stage of the study, these cell patches are used with different CNNs training to obtain a classifier weights. In last stage of the study various WSI images were used for region based detection to form automatic score detection on the tissue specimens. Table 2.2 show number of tissue sample in every proposed

dataset for score detection. Abbreviations in the Table 2.2 are following : “Dataset_1” defines total number of of WSI that obtained from Istanbul Medipol University Hospital, “Dataset_2” defines total number of tissue patches that are obtained by rescaled WSI, “Dataset_3” defines total number of augmented artificial data that are obtain from rescaled WSI, “Dataset_4” defines total number of tissue patches that are obtained from WSI with original resolution. How these datasets are cretated and used to form detector models will be detailed in related section.

Table 2.1 : Number of tissue patches are used in cell based feature extraction for SVM classification

| <u>Score 0</u> | <u>Score 1</u> | <u>Score 2</u> | <u>Score 3</u> |
|----------------|----------------|----------------|----------------|
| 41 | 47 | 52 | 58 |

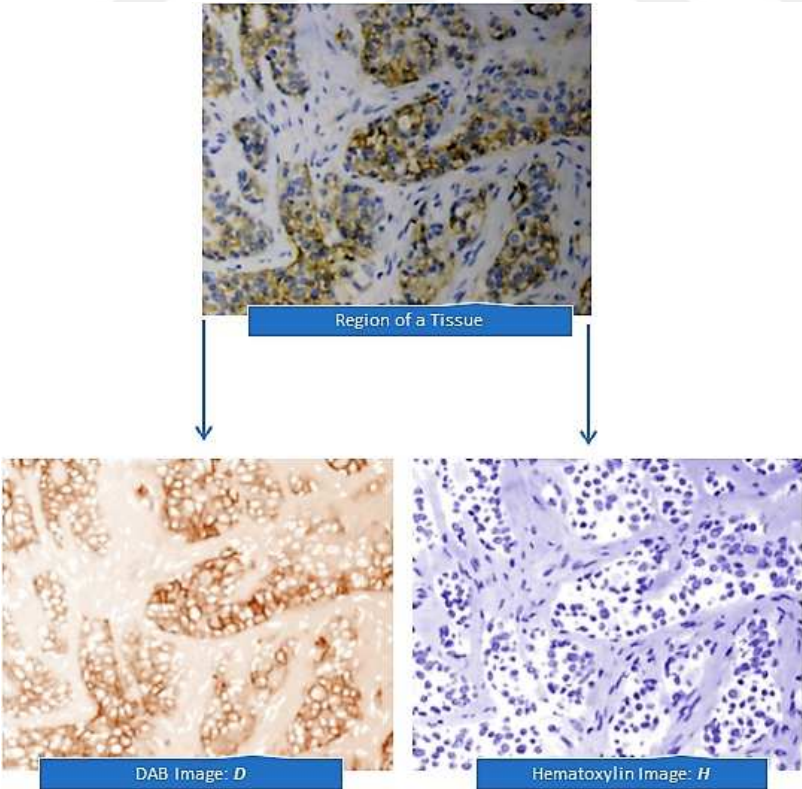


Figure 2.4 : Decomposition of H and DAB channels by the color deconvolution method in [14].

For the preparation of datasets and the training of the data by deep learning models, high-processor machines were used. The type of graphics cards, GPUs, and CPUs in the processors were used during this study are as follows:

- For splitting of Whole Slide Large Images to dataset preparation by Matlab environment: 128 GB Memory with 28 CPUs computers were used.
- For cell based Score classification by Caffe Library : GeForce GTX TitanX 12GB NVIDIA GPU was used.
- For Score detection in WSI by Darknet Library : Virtual Machine of Google Cloud with 8 CPUs and 30 GB Memory, NVIDIA Tesla K80 version was used.

Table 2.2 : Number of dataset that obtained from whole slide pathologic image samples for score detection approach in WSI.

| Dataset_1 | Dataset_2 | Dataset_3 | Dataset_4 |
|------------------|------------------|------------------|------------------|
| 24 | 348 | 608 | 3480 |

2.2 Proposed Methods for Pathological Image Analysis

2.2.1 Classification of CerbB2 Scores

For scoring of CerbB2 receptors in breast cancer tissue specimen images of patients with suspected breast cancer are proposed throughout this analysis, new modalities in the treatment which is called 'targeted therapy' is used for pathological assessment of the receptor status in the tumor with CerbB2 receptors, therefore a therapy is advised for breast cancer. For this purpose, pathological images obtained by using IHC were analyzed to determine CerbB2 score levels at cellular level. Firstly, images are separated into hematoxylin (H-Blue) and diaminobenzidine (DAB-Brown) color channels and cell centers and cell membranes are determined [14]. Subsequently, membrane intensity histograms are extracted by the proposed image analysis method, so features for each cell are obtained. Finally, obtained attributes are trained with the hyper parameters of SVM obtained by experimental methods, therefore a classification model was created [15]. The CerbB2 scoring system consisting of 4 classes with this

model outperformed similar studies with 98.2% accuracy. Each step in the study is explained in detail and experimental results are presented in comparison to other state-of-the-art methods.

2.2.1.1 Cell based biopsy image scoring

Although main aim of the microscopic image analysis which is a topic of digital pathology is to diagnose cancer, until now it is effectively used for cell counting, cell segmentation, and cell scoring [16]. There is interobserver discrepancy in most of the scoring systems used in pathology practice such as percentage of Ki67, ER, PR, Liver Steatosis [17]. In this point, computer-aided scoring systems could help to overcome this discrepancy problem, and speed up the process.

In the treatment of cancer, chemotherapy (CT) is a frequently used choice, but usual CT agents also damage native tissues as well as cancer cells [18]. Recent developments enabled physicians to use methods termed 'targeted therapy' which only destroy the cancer cell, labeled by some cell markers [19]. CerbB2 is a transmembrane protein belonging to epidermal growth factor receptor (EGFR) family and located on the chromosome 17q12 is overexpressed in 15-20% of the breast carcinomas [11,20,21]. CerbB2 mutated breast carcinomas have worse prognosis but are suitable for a targeted therapy agent called trastuzumab [22]. In CerbB2 mutated breast tumors, tumor cells have increased CerbB2 protein, which is located on the cell membrane. Pathologists can detect CerbB2 overexpression with two methods. The gold standard method is In Situ Hybridization (ISH) that shows directly gene amplification on the chromosome [23]. Since this method is expensive, time consuming and not easily applicable, in practice, most laboratories use Immuno Histo Chemistry (IHC) at the first line [24]. By IHC, increased CerbB2 protein on the cell membrane is demonstrated by using antibodies against this protein, after protein-antibody binding, a stain, called chromogen, is used for visualization. With IHC, pathologist evaluates the intensity/darkness, continuity and the extensity of the stain in the tumor cells [25].

For standardization of the evaluation among pathologists and concordance with FISH (Fluorescent in Situ Hybridization) [26], ASCO/CAP 2013 score scale system recommendations are used [27]. According to these recommendations, tumors are categorized into four groups based on their staining observations, namely, Score 0,

Score 1, Score 2 and Score 3 in Table 2.3 and Figure 2.3. Tumors having Score 3, are suitable for the treatment targeted CerbB2 protein. For tumors with Score 0 and 1, this therapy is not effective. Score 2 is the borderline group which means some of the tumors with Score 2 have CerbB2 mutations in DNA and can benefit from targeted therapy and some of them have not. Therefore, for this group of tumors, the gold standard method ISH must be used for higher accuracy [24-27]. Accuracy which is shown as (2.1), measures success of classifiers with proportion of number of correct prediction to total number prediction, is calculated as follows:

$$A = \frac{\sum_{l=0}^n P_l}{\sum_{l=0}^n T_l} \times 100 \quad (2.1)$$

where l express labels, “ P ” symbolize true predictions and “ T ” is used to represent

Table 2.3 : CerbB2 score scale system based on ASCO/CAP 2013 recommendations. [27].

| Score Type | Definition |
|----------------------------|--|
| <i>Score 0 (negative)</i> | <i>No staining is observed or membrane staining that is incomplete and is faint/barely perceptible and within $\leq 10\%$ of tumor cells</i> |
| <i>Score 1 (negative)</i> | <i>Incomplete membrane staining that is faint/barely perceptible and within $> 10\%$ of tumor cells</i> |
| <i>Score 2 (equivocal)</i> | <i>Circumferential membrane staining that is incomplete and/or weak/moderate and within $> 10\%$ of tumor cells or complete and circumferential membrane staining that is intense and within $\leq 10\%$ of tumor cells</i> |
| <i>Score 3 (positive)</i> | <i>Circumferential membrane staining that is complete, intense, and within $> 10\%$ of tumor cells</i> |

total number prediction regardless of correctness. CerbB2 is a prognostic, predictive and therapeutic marker, but the targeted therapy against CerbB2 is expensive and have cardiotoxicity. Hence, it is important to choose correct patients. Also, there is an interobserver discrepancy among pathologists especially for tumors with Score 2, so determination of the tumor CerbB2 status is critical [28].

In this stage of this study, in order to mitigate the interobserver discrepancy, novel image analysis methods are developed to determine the score of the CerbB2 tissue sample. For this purpose, tissue samples are analyzed on a cellular basis where each cell is identified using membrane based feature extraction and classification methods.

2.2.1.1.1 Cell based feature extraction

Stages of the CerbB2 mutated breast carcinoma automated tissue scoring system are presented in Figure 2.5. Sample tissue images, CerbB2 tissue specimen is divided into two sub-regions namely cell nuclei and cell membrane to benefit from pixel values of hematoxylin and diaminobenzidine which are used for background staining. Then multilevel thresholding is applied for detection of cell nuclei and cell membrane by using gray level binary images. In the last part, how the classes are grouped by using IHC, how the SVM classify the scores based on the features is explained in details of subsections in followings. In this study, only related parts with cell based feature extraction and SVM based classification will be detailed.

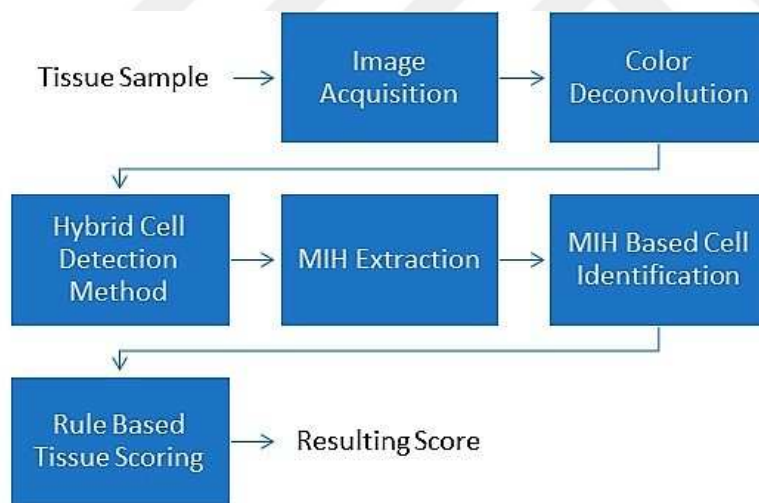


Figure 2.5 : Workflow of the proposed IHC Stained CerbB2 mutated breast carcinoma automated tissue scoring system. Tissue samples are fed into the system in order to obtain the resulting ASCO/CAP 2013 tissue score [27].

After features were extracted based on the intensity analysis of DAB-Brown staining on the membrane. Each cells on the tissue specimens were represented by 16 quantized features in terms of vectors. The vectors were used by SVM multi class train to create

a classifier model. The symbolic representation of feature extraction method and steps of feature usage is shown in Figure 2.6.

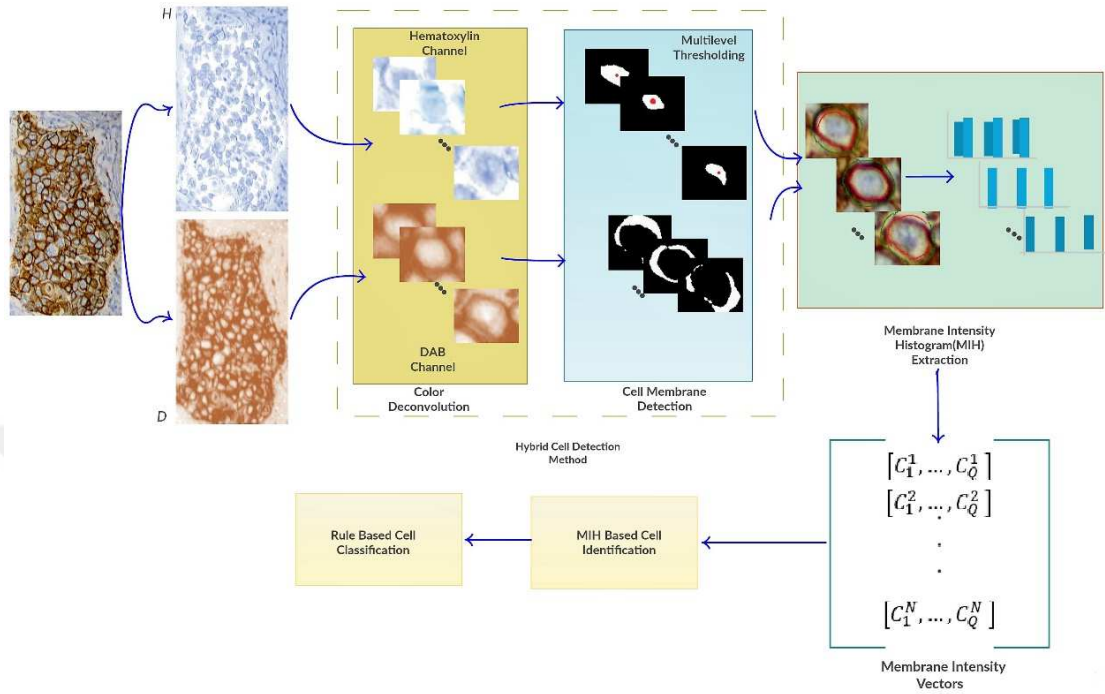


Figure 2.6 : Cell based feature extraction to vector form for classification.

As seen in the Figure 2.6, firstly IHC image is applied color deconvolution, then H-Blue and Dab-Brown channel is divided. After that center of cell is defined by multilevel thresholding and cell membrane is extracted. Based on the intensity of staining in the membrane, Membrane Intensity Histogram (MIH) is extracted which is used as feature vector. Therefore 16 quantized feature vector are obtained for identification each cell. Finally based on the rule that set by ASCO/CAP 2013 [27], these vectors are fed into a cell identification module utilizing Support Vector Machines (SVM) to classify MIH based features into four separate score classes, namely, Score 0, Score 1, Score 2 and Score 3 [29]. For identification purposes at the cellular level, SVM classification method based on LIBSVM (A Library for Support Vector Machines) is used in [30].

2.2.1.1.2 Training with Support Vector Machine (SVM)

After the extraction of features, they are used to fed SVM training to get a classifier model. For this purpose, before the training and test phase, all the data was normalized

that are used in SVM classification to get higher and proper accuracy. On that note, firstly it is found average value of all the data which is called as mean. Secondly, the mean (μ) is used while calculation of normalization which limits the value of train and test data between [-1, 1] as advised in the method [31] and normalization (N) [32] is calculated as follows:

$$N = \frac{(d-\mu)}{\sigma} \quad (2.2)$$

where ' σ ' is used to symbolize standard deviation and ' μ ' symbolize mean which is subtracted from every data value before division to standard deviation. Secondly the normalized data set for the model planned to be created using SVM is appropriately separated for the training and testing phase. Then the 5 fold cross-validation method has been used to more accurately divide the test and training set [33]. Since 10 folds cross validation uses 90% of the data for the training, cross validation can cause wrong prediction of test performance. In this regard, 5 folds cross validation was preferred for training to improve robustness of the classifier [34-36]. In the end, the track with the highest accuracy was used for SVM training set with optimal hyperparameters which are used to determine learning model [37]. The way in which parameters are used is covered in next section. Also, the proposed feasible classification by SVM are lined up step by step in Figure 2.7.

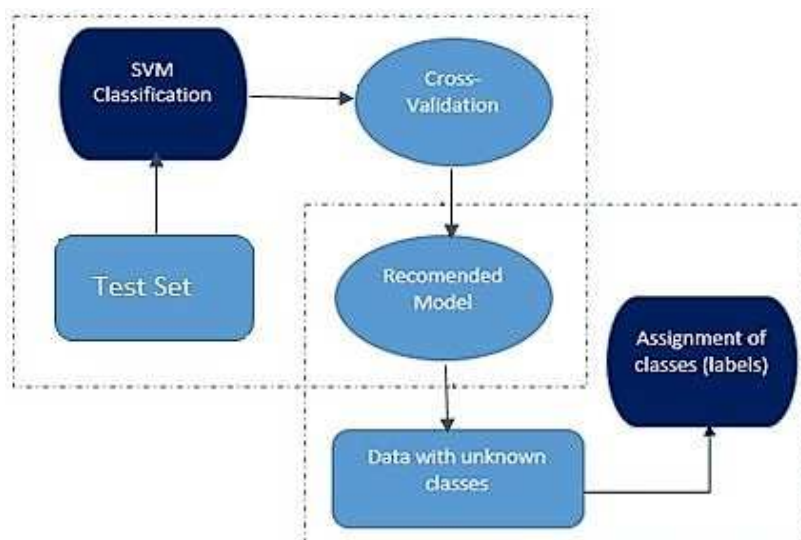


Figure 2.7 : Presentation of classification steps.

In order to evaluate the scoring performance of the proposed technique, following experiments are carried out using tissue specimens from patients with breast cancer at

the department of medical pathology, Istanbul Medipol University Hospital. These specimens were processed over the tissue images consisting of cell patches are obtained to be analyzed for Computer Aided Diagnosis [38,39]. All these tissue samples are diagnosed with cancer and scored by pathologists based on their own expertise and FISH technique which is accepted as the gold standard for breast cancer scoring [27]. Optimization of SVM classification methods parameters used in the proposed MIH-based cell identification technique are presented below. In this way, the SVM model that will make the most accurate classification was tried to be created.

After the features were extracted by membrane intensity histogram using novel methods such as color deconvolution, multilevel thresholding and radial lines a database was formed with 59804 cells from 198 tissue images. 41 of them are diagnosed as 'Score 0' (negative), 47 of them are diagnosed as 'Score 1', 52 of them are diagnosed as 'Score 2' and 58 of them are diagnosed as 'Score 3'. These tissue images are arranged into files which consist of 4 main score classes and are shared accordingly in [40]. Also to the best of authors' knowledge, this is the first publicly available IHC stained tissue image data set for CerbB2 mutated breast carcinoma automated tissue scoring purposes.

Table 2.4 : Distribution of data to obtain SVM model by using 5 folds cross validation.

| Total Cell Num | Train Cell Num | Validation Cell Num | Test Cell Num |
|-----------------------|-----------------------|----------------------------|----------------------|
| 59804 | 37897 | 9474 | 12433 |

Table 2.2 shows how many cells were used during the training and test process by using 5 folds cross validation. Through the total number of cells are shown in Table 2.4, there are 16376 "Score 0", 18030 "Score 1", 13650 "Score 2", 11748 "Score 3" number of cells. Also Figure 2.8. presents the class distribution of training data in optimized form with different colors. The classes distribute based on the number of data in each Score type. According to this figure, while each class type is grouped in a color among themselves, the presence of data from different color classes spanning between these groups shows the similarities and confusions between the data classes

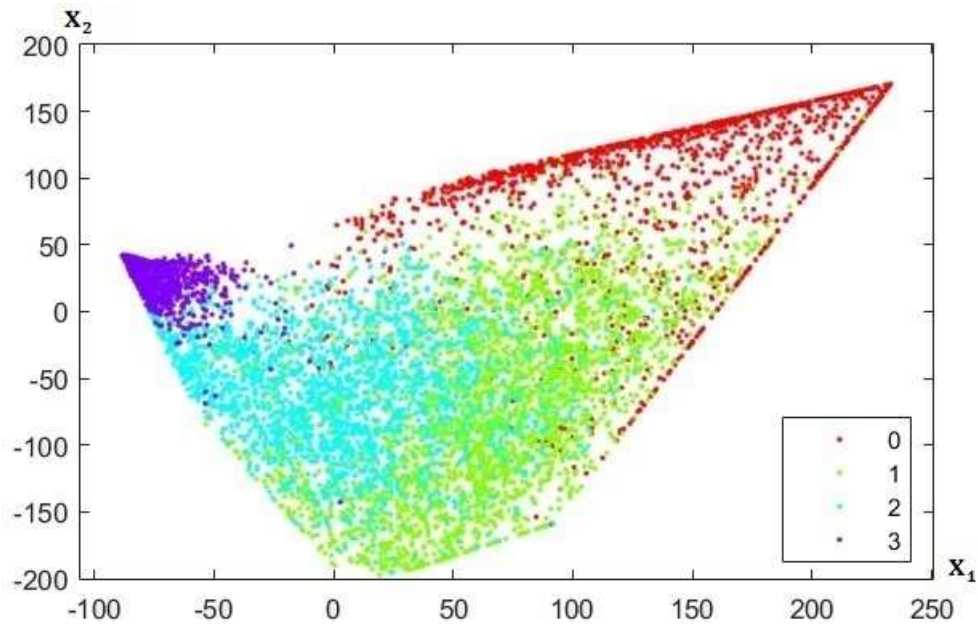


Figure 2.8 : Visualization of optimized data that is used in training. This appearance is constituted by PCA (Principal component analysis) [41] with trained data through the 4 different classes.

used in training to obtain the SVM model. As the Figure 2.8 shows a uniform distribution in general terms, it is observed that in the Score 2 class which is represented in the cyan color, the colors belonging to the other classes are also present in a small amount. The coordinates of x_1 and x_2 principle axes show the range of the largest features obtained with PCA. Each color represents different class based on their real distribution in training set. While purple represents Score 3, cyan represents Score 2 and green performs as Score 1, red stands for Score 0. It is envisaged that a more uniform distribution would improve the classification success with an average of 98.2%.

2.2.1.1.3 Selection of proper hyperparameters

Since the hyper-parameters provide optimization of data set and usage of kernel owing to its adaptable parameters for both one-class and multi-class SVM, optimal parameters for the highest performance classification were chosen to make the program faster and accurate [42,43]. In this direction, proposed classification approach used by hyper-parameters that have been determined by experiencing the methods in the ongoing steps. Firstly, Support Vector Clustering (SVC) was used to implement the "one-against-one" approach for multi-class classification which construct

classifiers and each classifier trains data with the four classes. Yet, it was not suitable in this work, as SVC make the program slower [44]. To this respect, Shrinking Heuristics of LIBSVM was preferred, thus it speed up the optimization by decreasing number of iteration [45]. Also, kernel was used that enlarge the data as multi-dimensional to get more feasible classification.

Experiments for the determination of the most appropriate parameter functions followed by experiments of combinations of kernel that would give the most probable result during training the model. To this respect, RBF (Radial Basis Function) act as a low-pass filter and makes the model even smoother was preferred [46]. Then, RBF kernel and non-linear parameters C and Gamma are used and large parameters were assigned to C, so the influence of each individual support vectors could be controlled. On the contrary, small parameter values were assigned to Gamma (γ) which is generally obtained by division of 1 to number of features as in following equation:

$$\gamma = \frac{1}{\sum_{i=0}^n F_i} \quad (2.3)$$

where “F” represents number of features. However our classifier model doesn’t show successful performance with lower value of gamma which obtained with this formula, since the model is constrained and can’t capture whole model complexity. For this reason, slightly larger gamma values preferred as shown in Table 2.5. To make the model smooth, low parameters for gamma and bigger parameters for "C" have been tried to be selected. Finally, after all hyper-parameters were changed, thus the training set showed 99% accuracy and test set up to 98.2% accuracy by the learning model that was obtained with the adjusted parameters. The following Table 2.5 shows test accuracy with best optimal hyper parameters were chosed for each experimented kernel function. According to the table, the pathological data set we created from the tissue samples did not show very good test performance with the linear kernel function because it has a very inhomogeneous distribution between the classes. Although the sigmoid function is similar to RBF, it is difficult to adjust its parameters, and its use in classification problems is not recommended [47]. Besides, in the case of confused data sets which don’t consist of definite calsses like in ours dataset, sigmoid function doesn’t show accurate performance. Since the sigmoid does not have positive definite kernel function, in this case, the classification test accuracy is lower than other

functions [48]. The polynomial function performs well when there are large number of features and more hyper parameters are used. It also examines the feature vectors in separable independent spaces such as the linear kernel function [49]. The use of the polynomial function has shown lower performance in our data set than RBF because our data set has only 16 features per cell and does not have a linear distribution. The RBF kernel function generally show more successfull classification performance than other functions by dividing the data into different spaces and classifying it linearly. The complexity of the model you create is also directly proportional to the hyper parameters used, ie parameter selection directly increases performance. Thus, the use of RBF showed the highest classification performance, in accordance with the feature class structure of our non-linear dataset.

Table 2.5 : Best hyperparameters with Kernel Functions of SVM

| Kernel Functions | Values of "c" and "g" | Accuracy |
|------------------------------|------------------------------|-----------------|
| Linear | c = 100, g= 0.01 | 64% |
| Polynomial | c = 100, g = 0.25 | 74% |
| Radial Basis Function | c = 100, g = 0.25 | 98.2% |
| Sigmoid | c = 100, g = 0.1 | 58% |

The following Table 2.6 illustrates how the accuracy is obtained over the confusion matrix. Intersection of columns and rows show accurate predictions in terms of cell number. Column which is called “Actual” shows cell based total test numbers in each Scores. “Recall” and “Precision” shows cell based accuracy for each type of score. While “Precision” shows predicted accuracy rate, “Recall” presents actual accuracy rate. Precision and Recall are used to show false positive predictions. As it is expected, overall test accuracy is between precision and recall with 98.2% percentage. The confusion matrices in Table 2.6 provides test accuracy with each type of score, it also

represents data distribution. Since Score 2 has the most confusion between score types, the least accuracy is obtained in Score 2. On the other hand, wrong prediction in test has caused lower accuracy rate, even though the related score type has high accuracy rates in actually. Overall, this confusion matrix explain that the SVM model which is used in prediction of cell based accuracy for a tissue pacht. Then, “Rule Based Tissue Scoring” is applied which is explained in next section.

Table 2.6 : Demonstration of cell based confusion matrices. Darker cells represent accurate prediction of every scores as score 0, score 1, score 2, score 3 respectively.

| SCORES | 0 | 1 | 2 | 3 | Actual | Recall |
|------------------|-------|-------|-------|-------|--------|---------|
| 0 | 2701 | 0 | 0 | 0 | 2701 | 100.00% |
| 1 | 45 | 3817 | 36 | 23 | 3921 | 97.3% |
| 2 | 5 | 16 | 2694 | 95 | 2810 | 95.87% |
| 3 | 0 | 0 | 0 | 3001 | 3001 | 100.00% |
| Predicted | 2751 | 3833 | 2730 | 3119 | 12433 | 98.3% |
| Precision | 98.1% | 99.5% | 98.6% | 96.2% | 98.1% | 98.2% |

2.2.1.1.4 Rule based tissue scoring

In the previous sections, cell based scoring methods are presented. In this section, cell scores are used to identify the tissue score based on the rules that are summarized in Table 2.3. Based on the rules are set by ASCO / CAP 2013 [27], diagnosis of whole tissue image requires CerbB2 score scale system. Table 2.3 presents what are the rules to determine score type for a tissue. These rules were formed to analyse the whole slide image specimen by patch, so there is no assumption while evaluating score types for whole slide images. In this study, cell-wised tissue scoring was applied in line with rules by ASCO / CAP 2013 [27]. Cells in a tissue pacht are classified by SVM and every score result for every cell were used to general evaluation. Distribution of score type on the tissue were observed and sent a condition step which has specific percentage rules as shown in Figure 2.9. If there are minumum 10% or more Score 3

on the tissue, this specimen accepted as Score 3. In a similiar way, if percentage of Score 3 on the tissue sample are lower and Score 2 has 10% percentage or more, this sample are accepted as Score 2.

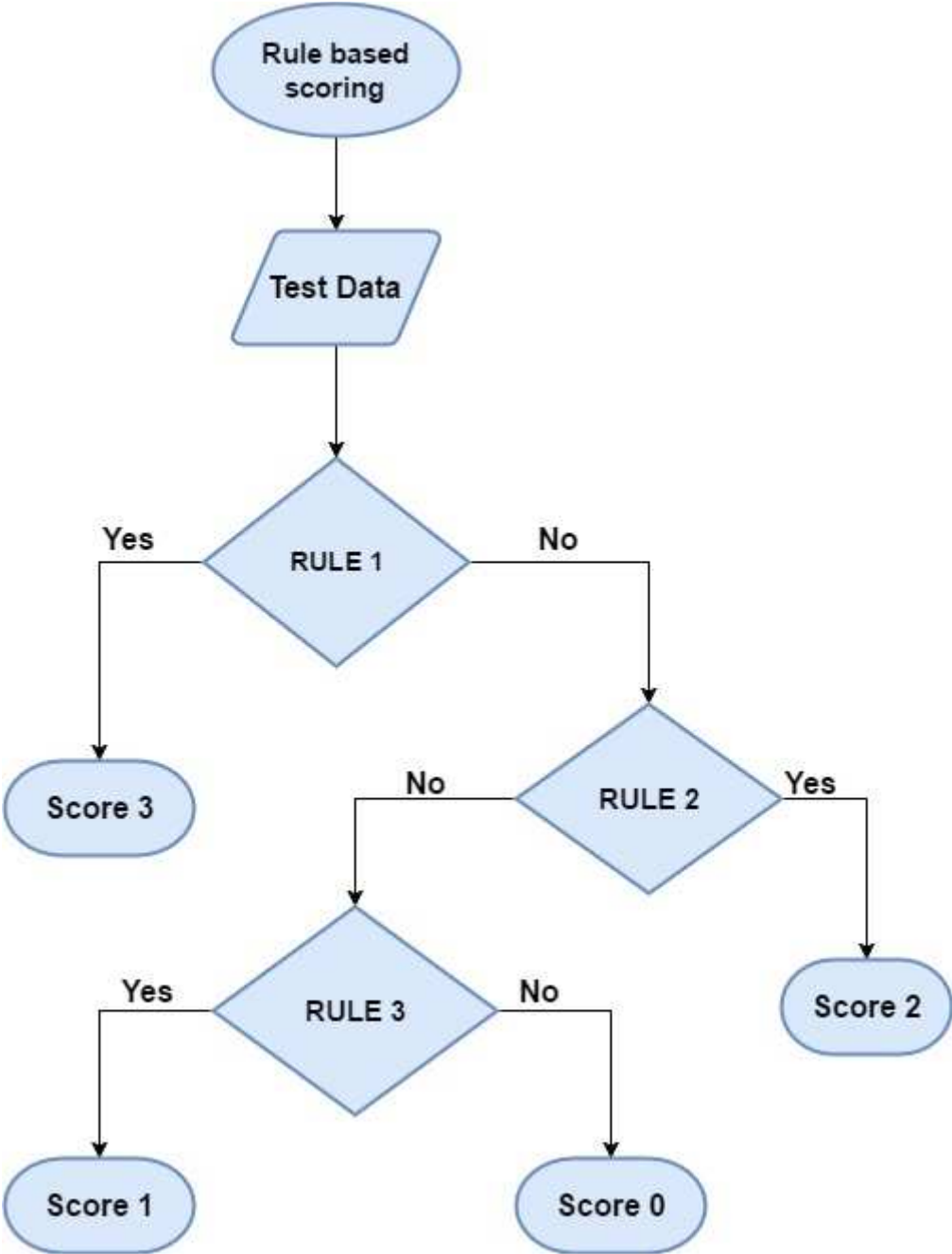


Figure 2.9 : Conditional flow of rule based tissue scoring as suggested in ASCO / CAP 2013 [27]. Rule 1: If the percentage of Score 3 found in the breast tissue sample exceeds 10%. Rule 2: If the percentage of Score 2 found in the breast tissue sample exceeds 10%. Rule 3: If the percentage of Score 1 in the tissue sample exceeds the Score 2.

Score 0 and Score 1 are signed in condition of both Score 2 and Score 3 have lower 10% percentage. Between Score 0 and 1, which type of score was observed highly, provided that the 10% percent limit is exceeded based as shown in Figure 2.9.

An algorithmic representation of the rule based method is given in Figure 2.9. These rules are set by ASCO / CAP 2013 [27] score recommendations which takes into consideration of the whole tissue regions. As opposed to individual cell based approaches [50], the proposed MIH-based Cell Identification technique takes into account the whole tissue sample by following the rule based scoring method summarized in Figure 2.9.

2.2.1.1.5 Experimental results

In recent years, the data used in many image analysis studies conducted with breast cancer data show a significant difference in terms of evaluation. These studies also differ in the use of different chemical dyes in the imaging of tissue samples taken from patients suspected of breast cancer. In this context, different from the type of chemical paint used in other works, IHC biomarker was used for the screening of tissue samples in this study. Then the results of image analysis were compared with the studies using the same chemistry to screen tissue specimens. These comparisons were executed based on the ImmunoMembrane [5] study and higher performance achievements were got. Firstly, features of cells in the tissue images were subtracted by original methods. Subsequently, a SVM classifier model was obtained and scoring success reached up to 98.2%. This success surpasses 80% success of ImmunoMembrane [5]. However the performance of this study varied among the types of scoring especially in the Score 2 prediction. In order to better observation of the differences in success between the types of scoring, the tests were repeated using the open interface that ImmonoMembrane [5] published on the internet and the system proposed by this study. The obtained results were compared with the pathologist's point of view as exemplified. During this comparison, the data set created for this study was utilized. This dataset is also available in [40]. The comparative test results showed that the error rate of ImmunoMembrane's [5] in Score 2 estimation was higher than the system we propose.

Table 2.7 : Comparison of some score estimation of this study and ImmunoMembrane [5] with the pathologist's score remark through the original pathological images.

| Original Tissue Patche Samples | Pathological Remark | Proposed Method's Remark | Immuno-membrane [5] |
|---|---------------------|--------------------------|---------------------|
|  | SCORE 2 | SCORE 2 | SCORE 1 |
|  | SCORE 2 | SCORE 2 | SCORE 3 |
|  | SCORE 2 | SCORE 2 | SCORE 3 |
|  | SCORE 1 | SCORE 1 | SCORE 2 |
|  | SCORE 1 | SCORE 1 | SCORE 2 |

Table 2.7 illustrates this situation and compares pathological images used in the tests in concern with the score estimation for these images. In order to increase the accuracy of the Score 2 estimate, it can be foreseen that data samples of this type need to be increased and optimized to a large extent.

However, this foresight has not yet been realized experimentally since there is no chance of reaching more data during this study. The currently used data in this study are already available online [40] and this contribute the work specificity. According to comparative tests by the interface of ImmunoMembrane [5], there are some challenge between ImmunoMembrane [5] and proposed approach in this study.

Comparisons with tests have shown that the success achieved in the classification technique discussed in this study outperform ImmunoMembrane [5]. While up to 98.2% test success is observed in this study, only 80% success is obtained in ImmunoMembrane [5]. As the reason for this, it can be shown that the correct extraction of the features and the correct optimization of the hyper parameters to obtain the SVM model.

Another distinguishing approach dealt in this study while score classification is the rule based tissue scoring in Figure 2.9. According to this rule, since the score distributions are calculated as a percentage over the partially examined tissue sample, these distributions are compared and scored according to the determined threshold percentages by ASCO / CAP 2013 [27]. These thresholds require conditions that combine SVM-based test results with the rules in Figure 2.9. For this reason, the score estimation made for each cell is used to identify the whole tissue specimen. The score estimation made for each tissue segment are compared with the determined threshold values to learn the final score. In this way, the score type determined by this rule is used instead of the most observed score type.

2.2.1.2 Scoring cell patches with Convolutional Neural Networks (CNNs)

In this study conducted by a team of pathologists and machine learning researchers, unlike the other methods in the literature. The CerbB2 tumor cell scores needed to determine the treatment duration of breast cancer were classified by an approach based on deep learning models. For this purpose, an original data set was obtained by taking patches through pathological tissue samples. Cell patches were created from this dataset, trained with deep learning models and classified as CerbB2 tumors. Since the classical image processing techniques only the superficial meaning of the images have been tried to be extracted, this step of the working was implemented by the idea of achieving a better and faster classification with multi-layered deep learning models has been realized [51].

The Score classification by CerbB2 receptors was tried to be realised by both cell-based and tissue patches biopsy images by using CNNs architecture. In order to train the tissue images with deep learning models, small scale images were needed based on the network models. For this purpose, firstly the WSI tissues were splitted into cell patches and labelled according to their Scores which pre-marked by pathologists. In this way, cell-based dataset was obtained for the first stage of the CNN based classification. The same whole slide images were used to create a dataset compose of tissue patches.

During that stage of the study, firstly the detailed information about the dataset is given by sample biopsy images which were prepared for the study. Secondly, how CNNs work and which CNN models were used for scoring is explained and the section is terminated by discussing the experimental results.

2.2.1.2.1 Deep learning models for cell based classification

In addition to deep learning, the learning of machine learning evolved from attribute engineering to architectural engineering. In previous studies, researchers often rested on the definition of best set of attributes to represent the problem in order to solve the related problem like the extraction of attributes and the selection of those with the highest representation capabilities among these attributes. Due to the multi-layer artificial neural networks, there is no need to find hand-crafted features by traditional image analysis techniques. Also, by its multi-layered architecture, it allows better accuracy rates to be achieved from the results than obtained by known methods. However, the need of large dataset for training phase with deep learning and the need of considerable computer power with high-performance GPUs and multiple CPU structure are shown negative side of the CNN models. After all, along with artificial data production and developing technology, these difficulties have begun to be overcome.

In the light of these information, the following deep learning models that were widely used during the period of this work were used.

AlexNet: It is a pre-trained image classification network that was trained with more than one million images published in 2012, resulting in approximately 1000 categories [52]. Alexnet has 8 layers which consist of 5 convolutional and 2 are fully connected layers 1 softmax layer at the last. The filter size of it is 11×11 and needs “227 \times 227”

sized images. Following Figure 2.10 describe AlexNet architecture which is taken form original paper:

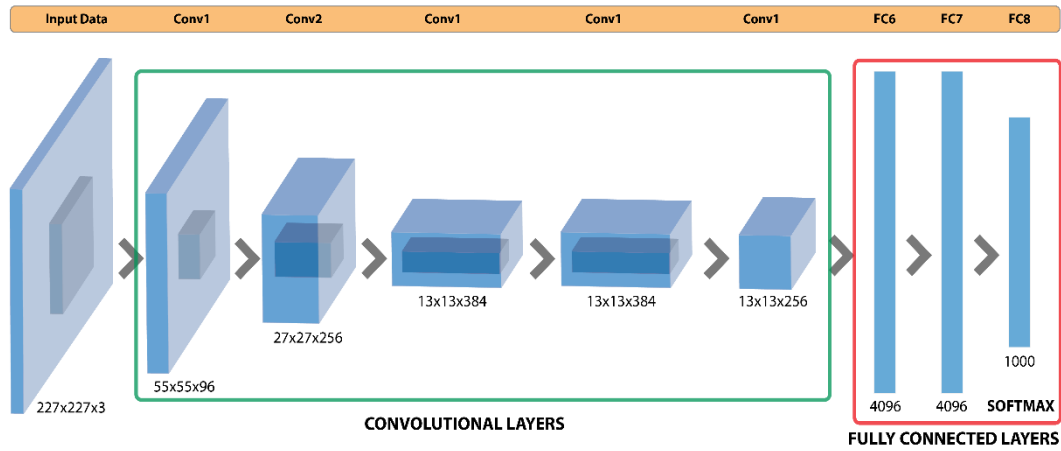


Figure 2.10 : AlexNet architecture.

VGG-16: As its name implies, it has 16 layers which compose of 13 convolutional and remaining 3 ones fully connected layers and softmax at the end. It was published in 2014 and has 1000 class success with the softmax structure of the last layer like AlexNet [52]. It is a simple network model and the most important difference from the preceding models is the use of 2 or 3 of the convolutional layers. It uses “3x3” filters with ”224x224” sized input images. Following Figure 2.11 illustrates the architecture of VGG-16 deep learning model:

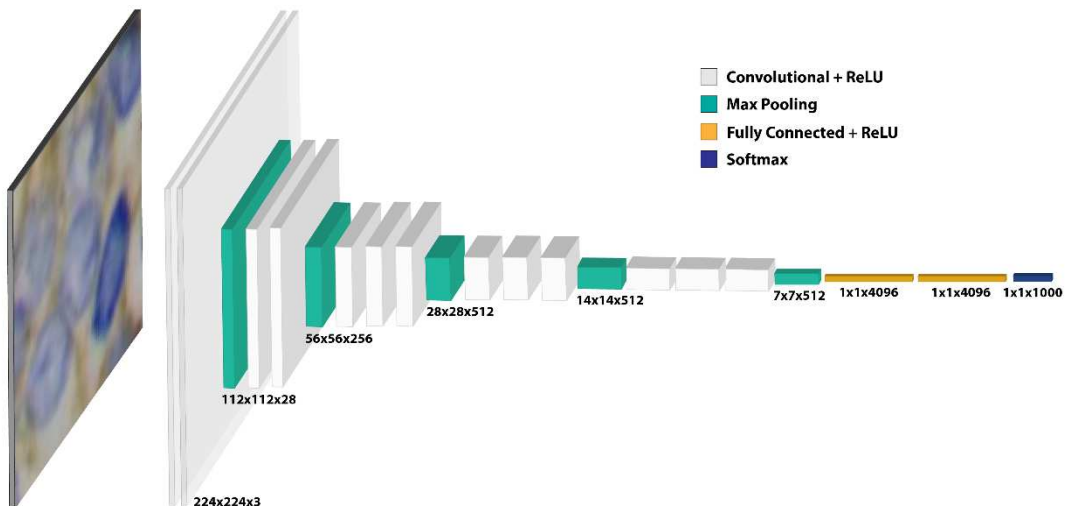


Figure 2.11 : VGG-16 architecture with tissue sample in our hand-crafted dataset [53].

In addition to information about deep learning models that were used in this study, also the information about some activation functions will be shared in this part.

The activation functions are used in the final layer of the convolution architecture, which is usually used in classifier problems. The most common example of this is the Softmax function. Both of the deep learning models that we used in this part of the study have softmax function on the their last layer. In addition to this function, Hinge Loss function is used instead of Softmax function. The results of this change were analyzed together with the data set that we had prepared. In the this section, brief information about Softmax and Hinge Loss functions is shared before experimental results with these functions. How the usage of these functions affected the accuracy of cell based classification will be debated in the following sections.

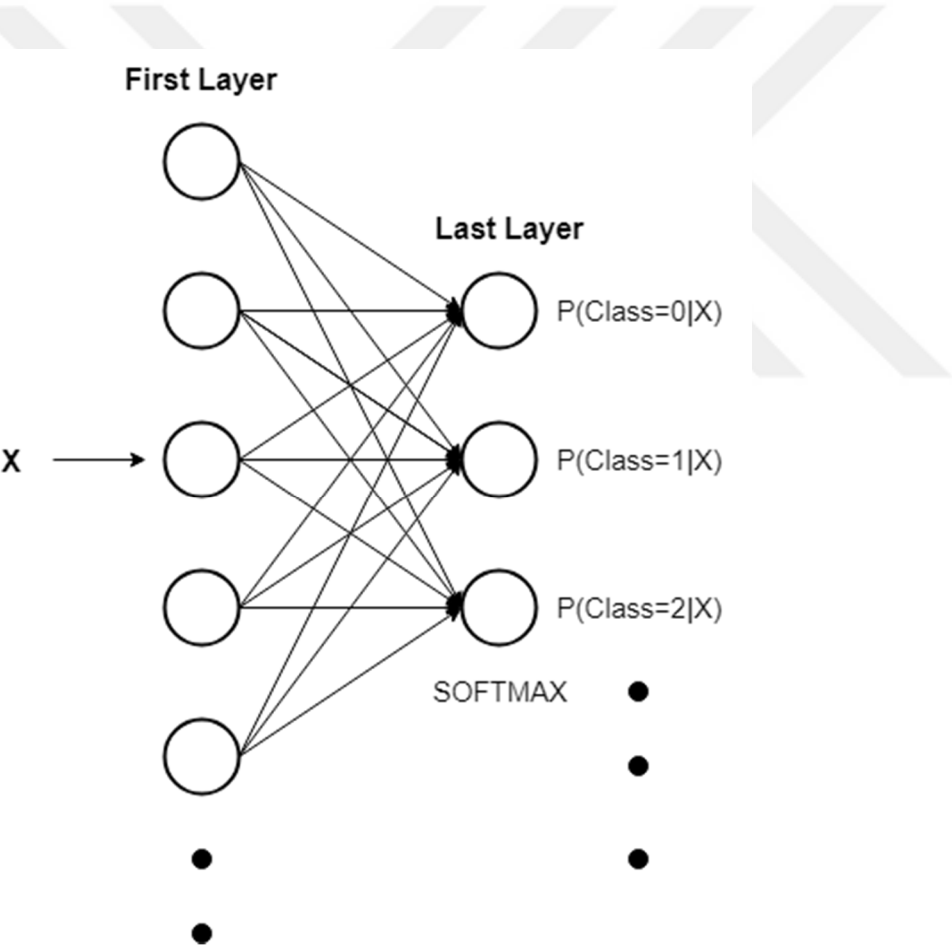


Figure 2.12 : Softmax activation function with multi-classes [54].

Softmax Function: Softmax is an activation function and generates probability-based loss value using score values generated by CNNs [54]. For the result of softmax the probability value for the similarity of each class of test input is generated. The output

values of the CNNs model are the values that are not normalized. Softmax transforms these values into probability values by normalizing them. This transformation is done with the maximum likelihood function. The goal is to maximize the log likelihood value for the correct class of test data. Unlike in the loss function, the goal is to minimize the likelihood; that is, to find the class with the least similarity and the least similarity. Following equation show softmax the function for value of “x” in cluster length of “j”. Also, Figure 2.12 ensamples working principle of Softmax function over the CNNs architecture.

$$\sigma(x_j) = \frac{e^{x_j}}{\sum_i e^{x_i}} \quad (2.4)$$

Hinge Loss Function : The loss layer, which is usually the last layer of CNNs, tries to direct the neural network correctly using the wrong guesses on the data to be tagged. In this context, the effect of the equation used to make the decisions of the loss layer can not be neglected on the whole result. Hinge loss tries to category maximum margin to define a threshold which prevent wrong classifications. It is als known as SVM loss funtion [55].

$$\mathit{hinge_loss}(y) = \mathit{max}(0, x \cdot y) \quad (2.5)$$

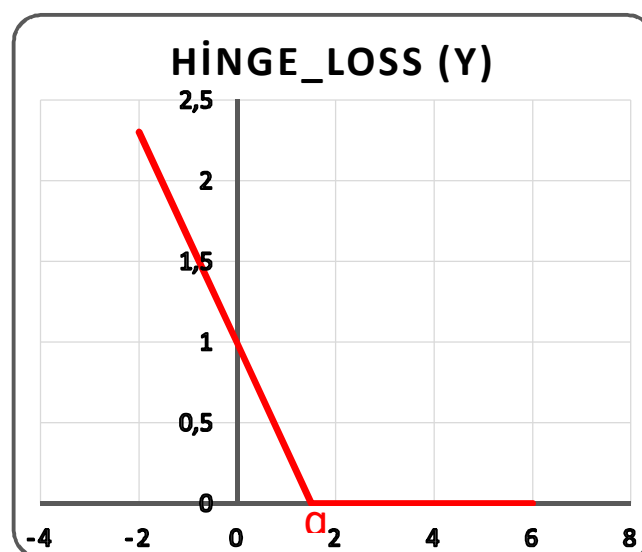


Figure 2.13 : Graph definition of Hinge Loss, red line defines the loss [56].

It defines a region, if a value in this area, it keep the losses, Otherwise, the loss will turn to zero. Following equation gives mathematical definition hinge loss function. While “y” symbolise the predicted class, “x” symbolise expected score for the classification result. Also Figure 2.13 represents graphical definition of Hinge loss.

2.2.1.2.2 Cell based dataset preparation













In this step of the study, the files of the 208 microscopic images were obtain from pathology department of Istanbul Medipol University Hospital were analysed. Then the centers of the cells were determined by basic image processing approaches. The coordinates of the cell centers were kept in files, then cell patches were created by using these center information.

After the obtained tissue slides were digitized, the images were splitted into H and DAB color channels [14]. Due to color deconvolution, the cell nucleus was stained blue, while the tumor cell was stained with brown. While the cell nucleus appears blue on the colored tissue, the cancer cell membranes look brown because the colors are separated from each other to facilitate image processing. After separating the membrane region and cell nucleus based on their color, multilevel thresholding was used to detect cell nuclei. Accordingly, after applying some image enhancement and noise reduction procedures, the Otsu threshold was applied and the variance in the background and front was calculated to solve the multilevel thresholding problem with the gray level histogram [57]. The reason for the application of the multilevel threshold was to separate the cells from each other in contact with each other. For this purpose we have used a cell segmentation technique that allows us to make real distance conversions to perceive the cell [14]. After that, the distance transformation is calculated according to the nearest zero pixel of the objects in the binary image. Lastly, cell objects with the highest pixel values in the view have been determined and the objects have been listed as cell centers by decreasing the threshold value from the highest value to the lowest value.

The cell fragments at a resolution of 84x84 pixels, centered on the detected coordinates, were generated by cutting 208 microscopic tissue images previously determined by pathologists. Thus, a total of 50990 cell fractions were obtained to train, validate and test deep learning models. From these fragments, 12700 were of type Score 0, 14790 of Score 1, 12600 of Score 2, and 10900 of Score 3 type cell pieces.

Then the data set with these cells is divided into test, validation and train files at appropriate rates. The matching score for each cell is also collected in text files with labels ranging from 0-3. Some examples of generated cell fragments are given in Table 2.8. While preparing of these cellular files, not only the cell areas were cut out,

Table 2.8 : Some cell fragments that are prepared for each score type in the dataset [76].

| | | | |
|----------------|---|--|---|
| Score 0 |  |  |  |
| Score 1 |  |  |  |
| Score 2 |  |  |  |
| Score 3 |  |  |  |

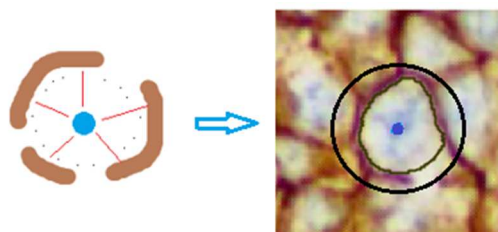


Figure 2.14 Example of obtaining cell centered fragments.

but also the peripheral areas where each cell was adjacent were cut off as shown in Figure 2.14. It illustrates how center of cell is used while creating cell data from tissue patches. The reason for this is that the appearance of the brown origin of the tumor is concentrated in the membrane area of the cell. The determination of the score type is sought in the membrane regions according to this density. Also for medically healthy diagnosis, the cells should be evaluated in the adjacency of the periphery.

2.2.1.2.3 Training with AlexNet and VGG-16

After the whole data set was prepared, 50990 cell patches were obtained and 6373 of the data set were used for testing, 6374 of the data set were used for validation and remaining 38243 number of cell data were used for training set. The trainings were run on TitanX with 12 GB machine. The distribution between total number of classes was applied in both AlexNet and VGG-16. The data set which consist of cell patches, was trained by fine-tuning, then different accuracy results were observed. Since the experienced deep learning models have different architectural structure, they might show different performances. The Figure 2.15 illustrates this difference by comparing convolutional layers of AlexNet and VGG-16. While AlexNet has only 8 convolutional layers, VGGNet has 16 layers and 3 fully connected layers and one Softmax layer for cell based scoring into 4 classes. Also general representation of CNNs with our data is sampled by Figure 2.16.

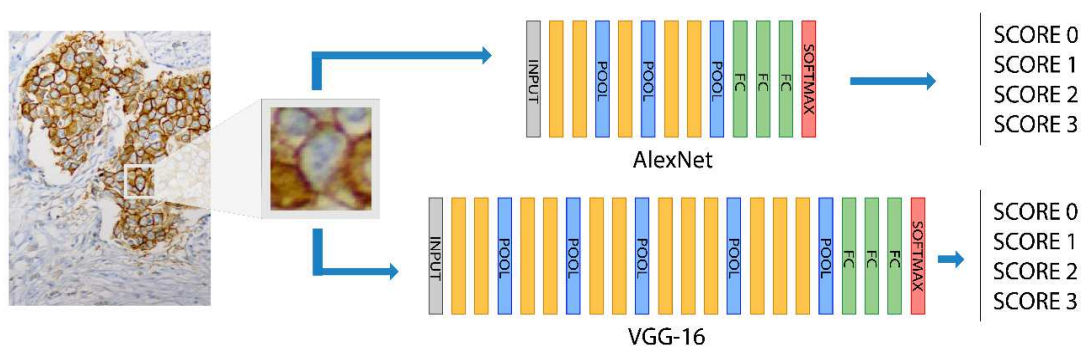


Figure 2.15 : Representation of experienced CNNs model's architectures.

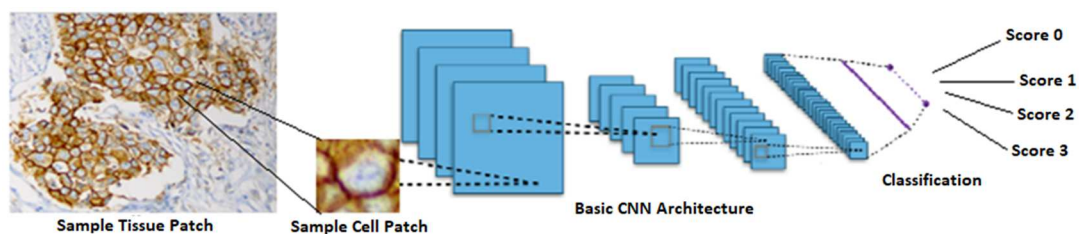


Figure 2.16 : The general outline of the classification with convolutional networks with cell fragment is obtained from the tissue sample [76].

The training, validation and testing of Deep Learning Models has been done using the AlexNet and VGG-16 architectures on the Caffe platform, a development environment for deep learning models [58-60]. Loss function layer is used with convolutional neural networks used in the classification stage. Since the loss functions are trying to direct the neural network correctly, by using incorrect predictions on the data to be labeled [61]. For this reason, the equation used to make the decisions of the missing layer has an effect on the whole result. Since the number of classes in the study is also small, the success of the architects used has been tried to be increased. For this reason, the "Softmax" lost function used in the last layer of CNNs, as well as the "Hinge" lost function, has also been utilized as shown in Figure 2.17 [62,63].

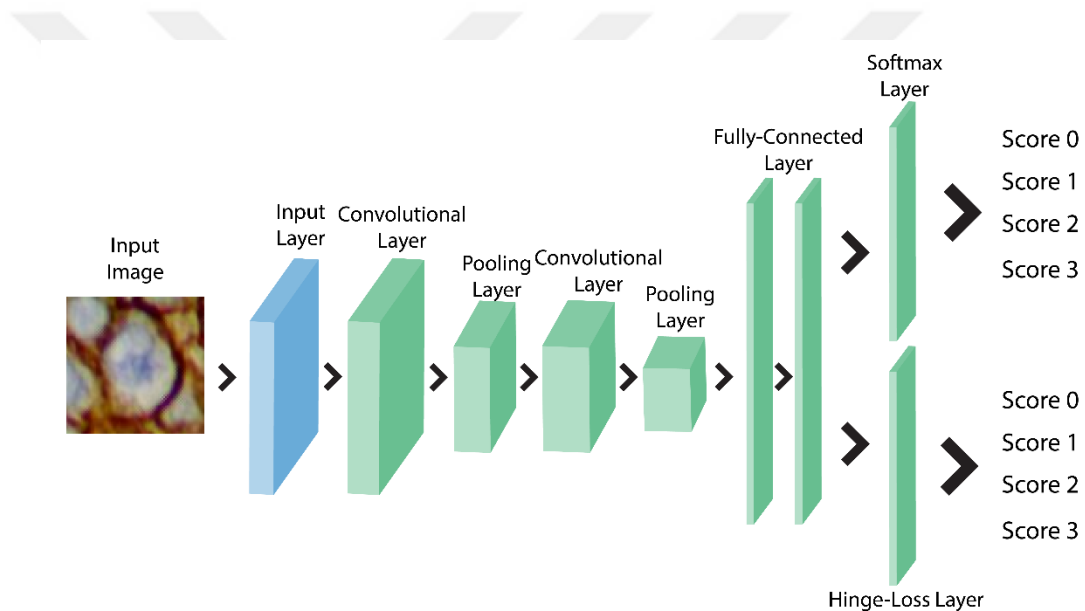


Figure 2.17 : Representation of functions which were used in output layer.

2.2.1.2.4 SVM classification with extracted features from VGG-16 network

In this part of the study not only deep learning classifiers were not used but also SVM based classifier was used. The convolutional network which has highest accuracy at the end of the training with cell based data set, was used for feature extractions. After the 4096 features were obtained, these features were used in SVM training. The tests were performed with two different attributes derived from the VGG-16 architecture in order to observe the effect of loss functions on learning models. One approach that

needs to be noted again in here is that the features were used in SVM were derived from the VGG-16 architecture only because they showed the highest classification success. Since these attributes are high-dimensional, logistic regression is used to simplify this dimension problem while training with the SVM to get best model. The Figure 2.18 demonstrates how feature extraction process works by sample VGG-16 architecture. The logistic regression is classification algorithm which provide to be found coefficients of classes [64]. The regression perform an input sensitivity analysis to determine the 4 most effective rates among these 4096 features obtain by VGG-16 network. It was investigated that the most effective and least effective factor could be obtained from these four factors in order to obtain the most effective SVM model. Following equation ensamples the logistic regression approach in this study. While “ ϵ ” represents error rate, “ α ” represents coefficients of classes (“x”). Logistic regression tries to find a relation between classes by these coefficients.

$$Y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_1 x_2 + \epsilon \tag{2.6}$$

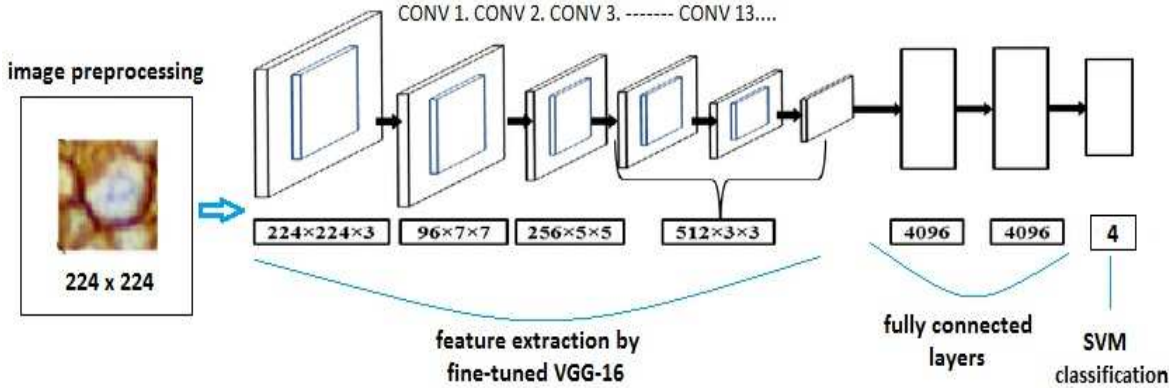


Figure 2.18 : Represents of feature extracted by VGG-16 and usage of these features to form 4 classes SVM classifier model.

2.2.1.2.5 Accuracy comparison among CNN models

In this section, the training of the studies using deep learning architectures, the verification table, the achievement table of the test results and the confusion matrices of each network architecture are shared on the score types. Table 2.9 shows the results of experiments using Softmax, Hinge loss functions and Support Vector Machines (SVM) [55] in AlexNet, VGG-16 architectures. According to these results, it is

observed that the education data is learned about 99% on average, it is seen that the test results are over 90%. The architecture with the highest test performance was VGG-16 with hinge loss use, which showed approximately 96% classification success. The use of hinge loss in AlexNet has reduced test results by 2%. It is also evident that the classic VGG-16 architecture has not experienced such a situation, while the difference between training and verification success has led to over-adaptation of the classical AlexNet architecture over the training data.

Table 2.9 : Performance results of deep learning architectures were used.

| CNNs | Training | Validation | Testing |
|--------------------------|-----------------|-------------------|----------------|
| AlexNet (Softmax) | 0.9825 | 0.9154 | 0.9320 |
| AlexNet (Hinge) | 0.9954 | 0.9089 | 0.9107 |
| VGG-16 (Softmax) | 0.9973 | 0.9605 | 0.9499 |
| VGG-16 (Hinge) | 0.9980 | 0.9670 | 0.9602 |
| SVM (Softmax) | 1 | 0.9597 | 0.9457 |
| SVM (Hinge) | 1 | 0.9565 | 0.9451 |

This has increased the performance of the VGG-16 architecture by about 1% compared to Softmax and also indicate that the loss functions (Softmax, Hinge) have effect on the deep learning models. Apart from deep learning architectures, the results of SVM trainings are also observed in Table 2. The tests were performed with two different features derived from the VGG-16 architecture in order to observe the effect of loss functions on learning models.

In order to show the test performances given in Table 2.9 in more detail for each class, the confusion matrices obtained from 4 different architectures are presented in Figure 2.19. In the confusion matrices, the labels of the classes are the same as the previously described cell scores.

While dark areas on matrices show the correct prediction rates for the type of score in the corresponding row and column, areas with light darkness show low rates of confusion with other types of scoring. Another point that is noteworthy when looking at the confusion matrices is that more neighbor cell scores are mixed with each other.

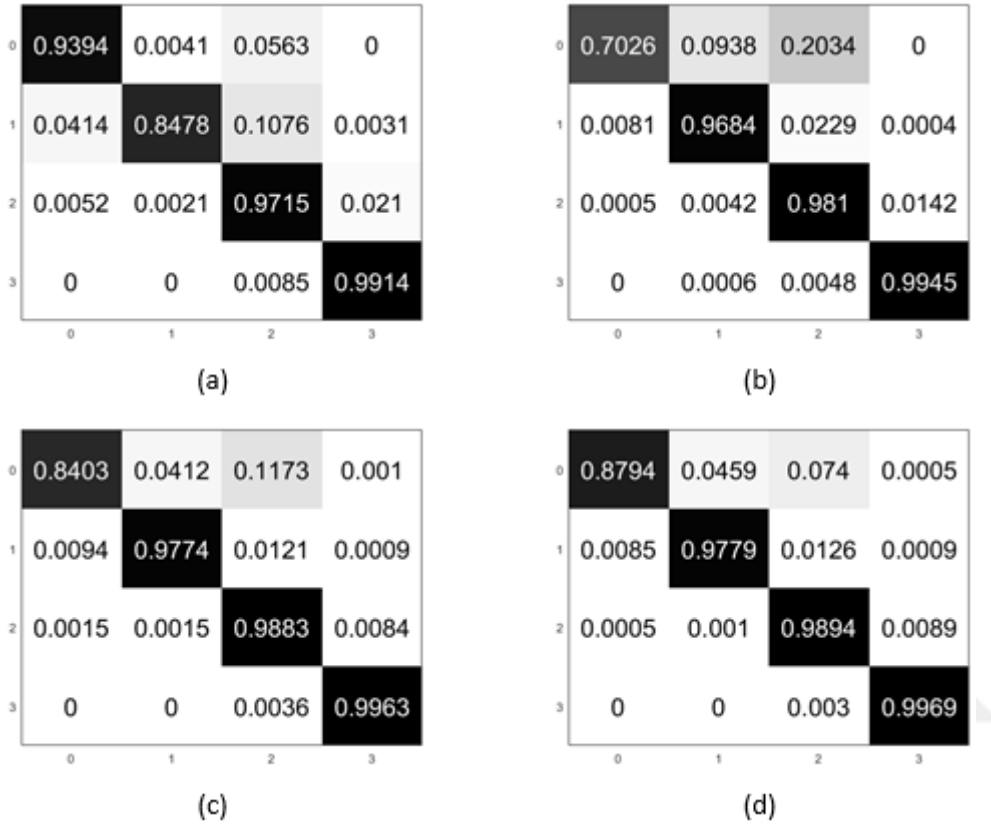


Figure 2.19 : The confusion matrices show accuracy results for each type of score as follow: (a) Alexnet with Softmax loss function, (b) Hinge loss function with AlexNet (c) Softmax lost function with VGG-16 (d) Hinge loss function with VGG-16 [76].

This verifies that as the difference in cell scores increases, the probability of confusion between the two classes decreases with the change in distribution of tumor regions. In this study, the classification of the lesions of breast tumors was performed with deep learning models, and the VGG-16 architecture showed good test and training success. The use of the hinge loss function has also diversified the work and has increased this success. Thus, it has been observed that the effect of loss functions is largely dependent not only on the data set but also on the deep learning architecture used. When the resultant confusion matrices were examined, Score 0 and Score 2 were confused higher than other score type. When the data set generated is examined, it is observed that the reason for this confusion is the presence of tumor parts similar to the Score 2 type in

the Score 0 regions. But this confusion has a low percentage in the way that final score pattern of the cell fragment will not change. It is aimed to obtain the achievement of the classification study obtained with existing deep learning models through an entire pathological slide image. For this purpose, it is envisaged that cellular and regional studies will be carried out in order to automatically detect the tumor regions of interest on the slide images with the help of CNNs in next section of this study.

2.2.2 Detection of CerbB2 Tumors on Whole Slide Images

In this part of the study, unlike the previous study steps, the ROI was tried to analyze on the WSI with breast cancer by detection methods with region based learning models. The innovations in this part of the study are: analysis of the tumor areas on the whole slide and this analysis is dealt with detection method instead of classification approach. The difference from the classification of detection is that it makes it possible to mark the tumor regions on the pathological image directly by the appropriate score type. Thus, these areas, which are only marked on the whole picture, can be examined by pathologists and time can be saved in determining the score of tissue image. The pathologist can review the relevant tumor regions to be examined by creating a convenience in the direction of the classes suggested by the detection method. Unfortunately, WSI's are very large, tumor regions are labeled by scaling the images instead of directly processing all images. Although so far, in the methods dealt with detection, all slide images have to be separated and scaled, but these methods can be applied by scaling the images more than once. Differently, in this study, the images are scaled only one time and divided into segments while creating a data set for region-based detection. For this purpose ROIs on the WSI were marked based on the tumor coordinates to prepare a dataset with biopsy images. We propose a complete computer-aided diagnosis (CAD) system for efficient localization of ROI in WSI by employing state-of-the-art deep learning techniques. Therefore, it is expected that the proposed system will detect tumor regions on WSI without spending time and power. Then these automatically detected regions could be scored in a shorter time. However, these studies focused on well-annotated and cropped patches, whereas a fully automated CAD system requires whole slide histopathology image processing which is, in fact, enormous in size and, therefore, difficult to process with a reasonable computational power and time.

In order to detect tumor regions, region based deep learning models were used. Firstly the data set was prepared for the training phase then region based models were analysed. After the analysis, Yolo (You Only Look Once) was preferred to be used, because it was claimed that its success was better as a result of the investigations [65]. Firstly, a brief information about the recently used the most common region based CNNs like R-CNN [66].

R-CNN architecture consists mainly of 4 sections. Firstly the input images are sent and followingly, the related region selection is made with Selective Search algorithm [67]. The algorithm produces 2000 different regions to find high probability of an object existence. In the third chapter, each region proposal is given to the similarly designed with CNN architecture on AlexNet. Finally, if an object is observed at the end of the CNN output, an arrangement is made on the region specified by selective search algorithm to produce the final result. R-CNN has some disadvantages. Firstly, each region in each image requires forwarding 2000 times over the CNNs for per image, so this test leads to time loss. Secondly, the architecture has to train three different models separately. These consists of CNN to create image features, regression model to compress constraining frames and the classifier estimating the class of the object. This process is very difficult to train the pipeline and takes about 50 seconds per average picture.

To overcome these problems Fast R-CNN [68] model and Faster R-CNN [69] have been developed. Fast R-CNN runs CNN once and share this calculation among the 2000 proposals instead of running CNN about 2000 times for each image, thus training and testing time are shortened. Faster R-CNN has added a proposal network (Region Proposal Network (RPN)) for the region after the last convolution layer. The RPN quickly and efficiently scans each location to assess whether further processing is required to detect the desired obesity in the relevant sites [69]. It does this by producing a bounding box proposition labeled with a binary classification at each locus that represents the probability of object or not. The following Figure 2.20 displays the architecture of three region based networks with their main differences. While R-CNN applies SVM based classification after the convolutional layers, Fast-RCNN use bbox regression with softmax function and Faster-RCNN uses region proposal network before the bbox regressor.

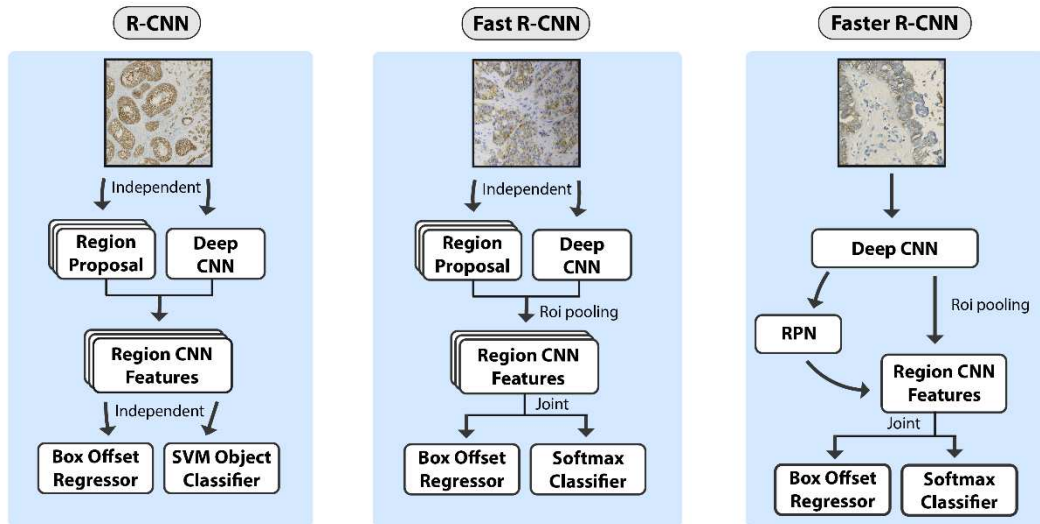


Figure 2.20 : Architectures of three main region based convolutional network.

2.2.2.1 Detection with CNN models

Since the region based CNN models for object detection have less success than the Yolo model, a pathological data set has been prepared to train with the Yolo model in this study. Firstly, the reported WSI's by pathologists were obtained from pathological department of Istanbul Medipol University. Every WSI was examined by pathologists and a labeled with proper Score type. Unfortunately these WSI's are enormous size so it is hard to use them for training directly. The minimum sized of image was "21275x9788" pixels, whilts the maximum size of image into the data set that is obtained from hospital was up to "112762x80059" pixels resolution. The compressed form of some WSI's with resolution info tag below images are ensamples for each score type in Figure 2.21. The sample WSI images were reduced by 10% for visualization. There are "24 WSI in total, 5 of them are Score 0, 7 of them are Score1, 7 of them Score 2 and remaining 5 of the Score 3. Unfortunately score types in the data set show irregular distribution according to class types and also the resolution of the WSI differs from each others. For this reason, the WSI's were preprocessed and their size and number were balanced between classes before the training phase and original pathological dataset was prepared.

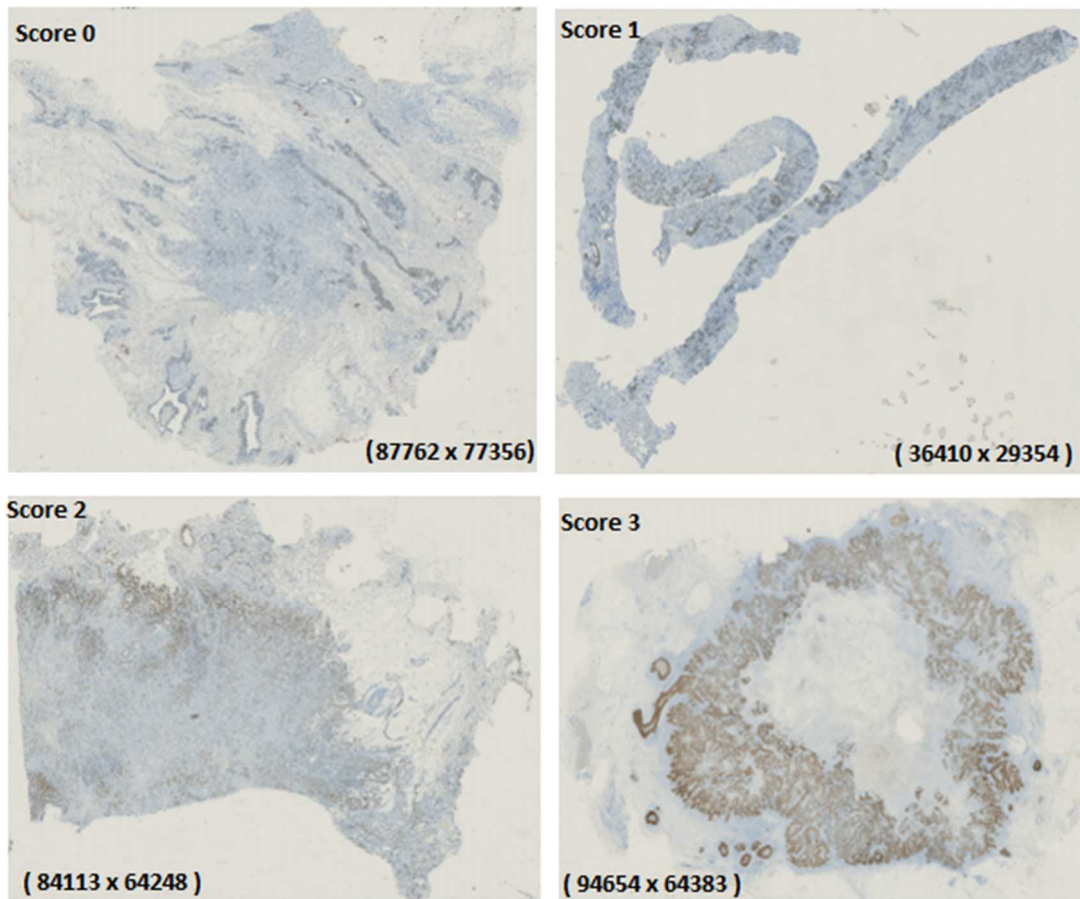


Figure 2.21 : Sample WSI's are : (a) Score 0, (b) Score 1, (c) Score 2, (d) Score 3 respectively.

Firstly, images were divided into smaller patches in Matlab programme which was set up into high computational computer (32 CPU, 4 GPU capacity). Every WSI divided into "10000x10000" pixels resolution and obtained 348 total tissue patches and 90 of them are Score 0, 100 of them are Score 1, 66 of them are Score 2, 92 of them are Score 3. The dimensions of the obtained tissue patches were reduced by 10% again before training so that they could be suitable for region based CNN training. Then ROIs were labeled on these images, thus coordinate information of each relevant object is created. These labels were optimized between 0-1 before the training phase. Every ROI on the tissue patches were marked under the supervision of pathologists. The following Figure 2.22 illustrates how the training data set was obtained with sample tissue images.

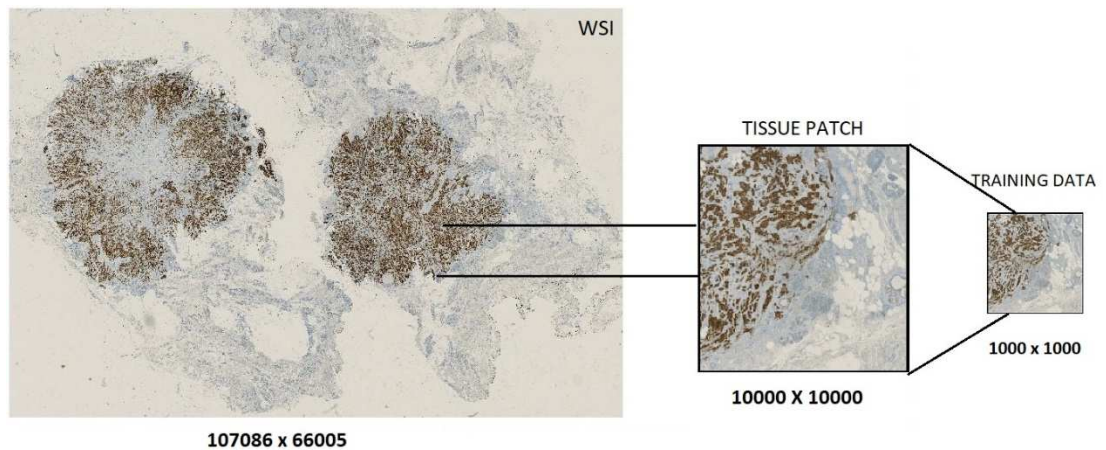


Figure 2.22 : Creating training data from sample WSI

2.2.2.2 How does Yolo (You Only Look Once) work?

YoLo is a region based object detection approach which is also called single shot detection [65]. In this part of study, how tumor regions on the WSI's were detected and labeled with proper score type will be explained.

Unlike sliding window and region based object detection CNNs, Yolo evaluate the entire image during training and test time so it indirectly encodes contextual information about classes with their appearance. In this way, It also uses contextual information for descriptive features of classes. In addition to this, Yolo has 24 convolutional layers and 2 fully layered in the last layer of its architecture, unlike the other region based CNN structures.

The Yolo network actually learns how the views of objects can be processed in a generalizable format by its fully connected CNN architecture. The architecture of Yolo divides input image into " $M \times M$ " grid feature maps. Each grid structure contains 5 bounding boxes and the probabilities of which object classes in these boxes can belong to real class label of object. Thus, the probability of detection is shared in condition of the center point of the searched object is in the one of bounding boxes on entire image. It uses features of entire image to learn class probabilities of each bounding boxes with grid cells on the image. Every grid cell has confidence score where show how the prediction is true. The confidence score is defines by the Intersection Over Union (IOU) which represents the interception between the predicted box and the ground truth box [70,71]. Each bounding box contain coordinate information of center point

of detected objects, size information of prediction box and confidence score which is show with class probabilities " $P(ClassID/Searched_Object)$ ". The following probabilistic calculation show both the class probability which is shown on the box and how this class prediction matches with object's real class. The equation (2.7) gives conditional class probabilities and box confidence score [71].

$$(P(Object) * IOU) * P(ClassID | Object) = P(ClassID) * IOU \quad (2.7)$$

As stated, Yolo searches through the entire image for the object, without using the merely classification, which makes it different from other regional-based object detection methods. However Yolo has 3 different versions within it, and the second version of Yolo is used in this work. Although Yolov3 has just been released and Yolov2 does not have remarkable distinct characteristic, its performance was not guaranteed in the period in which this study was addressed [72]. Yolov2 has a widespread use, with nearly twice the accuracy and speed performance of version 1 of Yolo. Yolov2 propose boundary box based class prediction instead of cell based Following Figure 2.23 demonstrates the architecture of Yolov2 by integrating with bounding box approach [72]. As shown in the Figure 2.23, there is no fully connected layers which are aborted and anchor boxes are started to be used for prediction bounding boxes. In addition to this "416x416" input images are preferred to be used to catch large sclae images' center points in bounding boxes, thus there is no need to search every adjacency of the boxes. The ROIs are represented by 5 probable bounding boxes as shown in Figure 2.23 and the object searched is found by finding the bounding box with the most intersecting region with the ground truth.

In this part of the study, the detection model was created with the dataset that I prepared by using the shared working principle of Yolov2 network and the results of the tests performed with this generated model were elaborated.

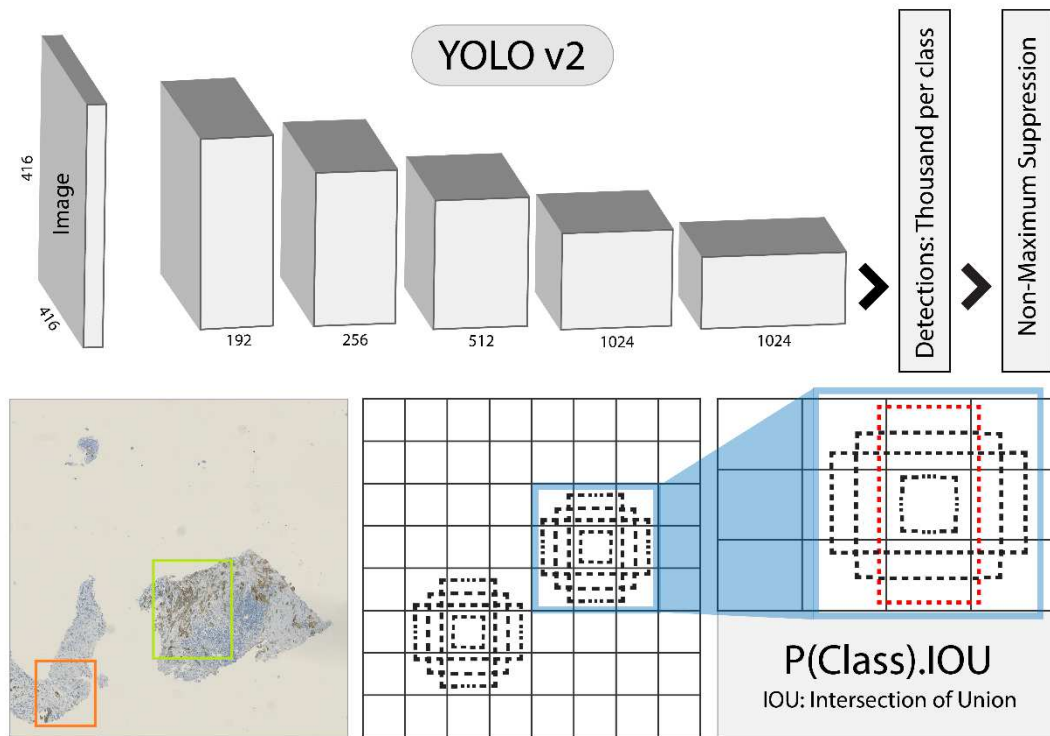


Figure 2.23 : Yolov2 architecture with its bounding box approach.

2.2.2.3 Score detection by Yolo

In the study, firstly, the prepared dataset was fine-tuned and trained with Darknet library in Google VM Instance with Nvidia Tesla K80 GPU. Darknet is the compiler library written for YoLo CNN in C language [73]. In the first step of this training, a detection model of 28000 thousand iterations made with the scaled pathological images was created. This model allows the ROI to be determined in the appropriate score classes on WSI. The purpose of this study is to detect the Score over the entire scaled image, not just the detection of the tumor regions, over the WSI tissue image. For this purpose, the OpenCv which is one of the computer vision libraries that can work with the Darknet library, has been used. Thus, the zoomed-in tissue sample image of the 4 major score classes tested with the model obtained after the training is presented visually in Figure 2.24.

These images in the Figure 2.24 also show that our model successfully determines the score classes according to their locations on the pathological images. This performance is monitored in more detail as the threshold is decreased. Threshold specifies that objects can be marked with appropriate labels when estimating the minimum performance ratio.

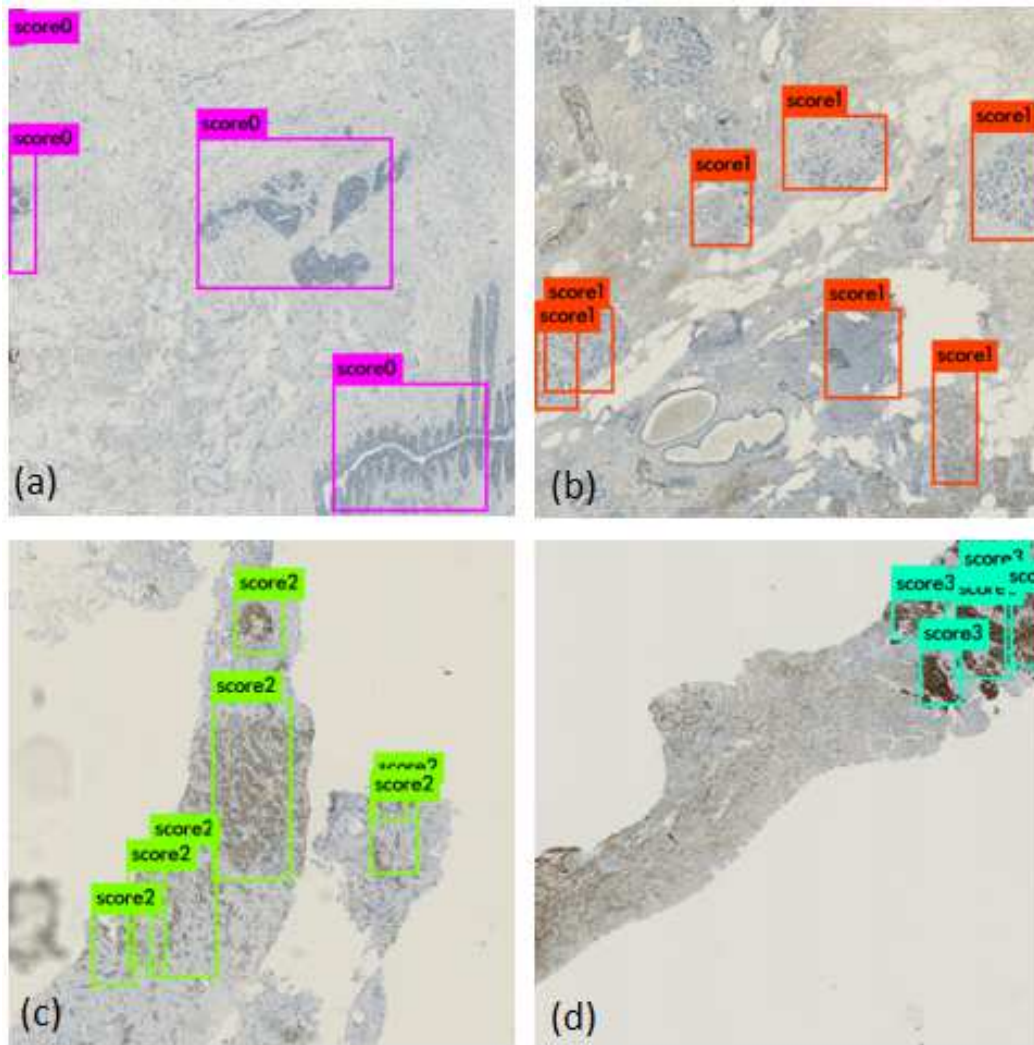


Figure 2.24: Examples of detection results with (a) Score 0, (b) Score 1, (c) Score 2, and (d) Score 3 samples respectively by the generated model.

For example, we see that the Figure 2.26 ensamples different threshold approaches for Score 2 class. As it is seen Figure 2.26, even the regions with lowest prediction ratio are labeled, since the proposed detector model is not robust enough, the model needs more data with further iteration. As can be seen in the figures, more than one label can be created for an ROI region because these images are scaled images so that close classes' information can be squeezed with 10% scale in one region. In this case, the marked region can be cut and re-grown and scaled. Then these resized images were augmented by rotational changes approaches. Therefore a new larger dataset was used for new training to obtain new detector model. A simply bilinear was applied to the existing images and the rotation of images were changed [74]. Therefore a new dataset was constructed by increasing of existing data nearly two times, so the "608" tissue

patches were obtained. However, this increased number of data set did not provide spectacular increase in detection performance, since we could only benefited from bilinear image rotation which didn't still provide data set with enough number. Because the analysis of pathological data is performed by staining of tumor regions, so no data enhancement methods (gaussian blurring, pixel and channel change etc.) can not be applied which causes color change which destroy original image texture. Therefore only bilinear image rotation could be used to increase the existing data set and artificial image samples were constructed as shown in Figure 2.25. The Figure 2.25 ensamples of each score type that were obtained by rotational approach by comparing them with their original images in each row. Every score sample in Figure 2 was obtained by different angle transformation approach.

The second detector model which was obtained by the artificial data production, is used for second level labeling as shown in Figure 2.27. Since the ROI on a WSI can be labeled with different type of score, detail score estimations on a ROI can be observed on a re-scaled image sample to decide a final score type for that region by the rules of ASCO / CAP 2013 [27]. Figure 2.27 ensamples this situation with the red bounding box that is on the top of WSI. This region is labeled with both Score 1, Score 2 and Score 3, then this region is re-scaled with higher resolution and observed score estimation in that region. Since the re-scaled region has over 10% Score 3 label, this region is labeled as Score 3 at final step. This situation can be shown as a result of resampling the compressed regions.

All the whole slide images of the tissue sample were used for the formation of a second model with retraining by dividing them into dimensions of "1000x1000" in their original appearance. Thus, the score estimation over the whole slayt tissue specimens compressed by 10 to 1 are compared with the proposed score predicitions on the original dimensions of these regions. For example, the region estimated as Score 3 on WSI is extracted with its original size and re-examined. In thissituation it can be said that this region is correctly analyzed if the class result is 3 according to ASCO / CAP 2013 [27] description. ASCO / CAP 2013 [27] is described in the first section of the work, section 2.2.1.1.4 under the heading "Rule Based Tissue Scoring", and the estimation of the score of the entire tissue sample contains the necessary rules.

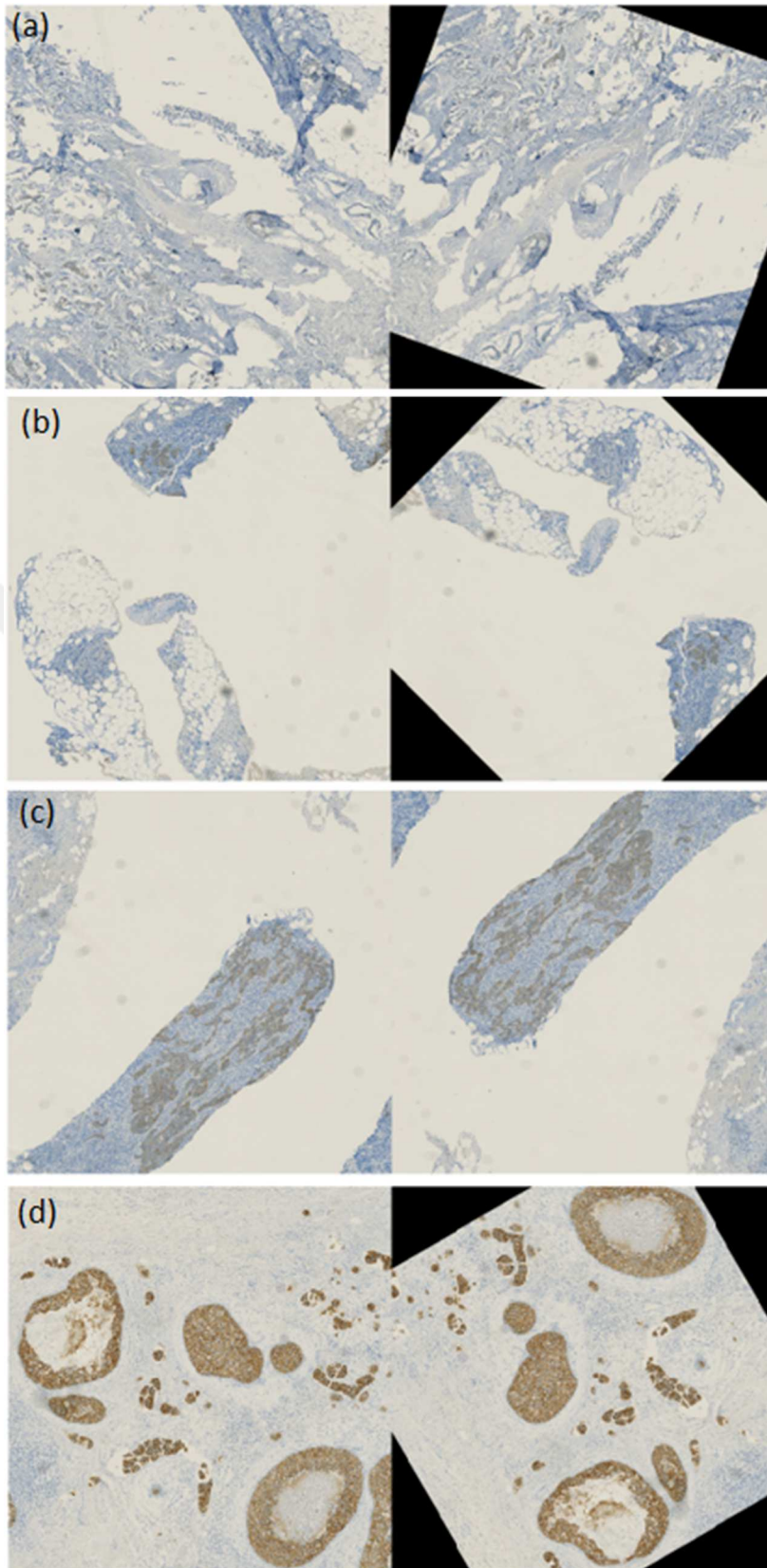


Figure 2.25 : Artificial data samples which obtained by bilinear image rotation for each score type as followingly: (a) Score 0, (b) Score 1, (c) Score 2 and (d) Score 3.

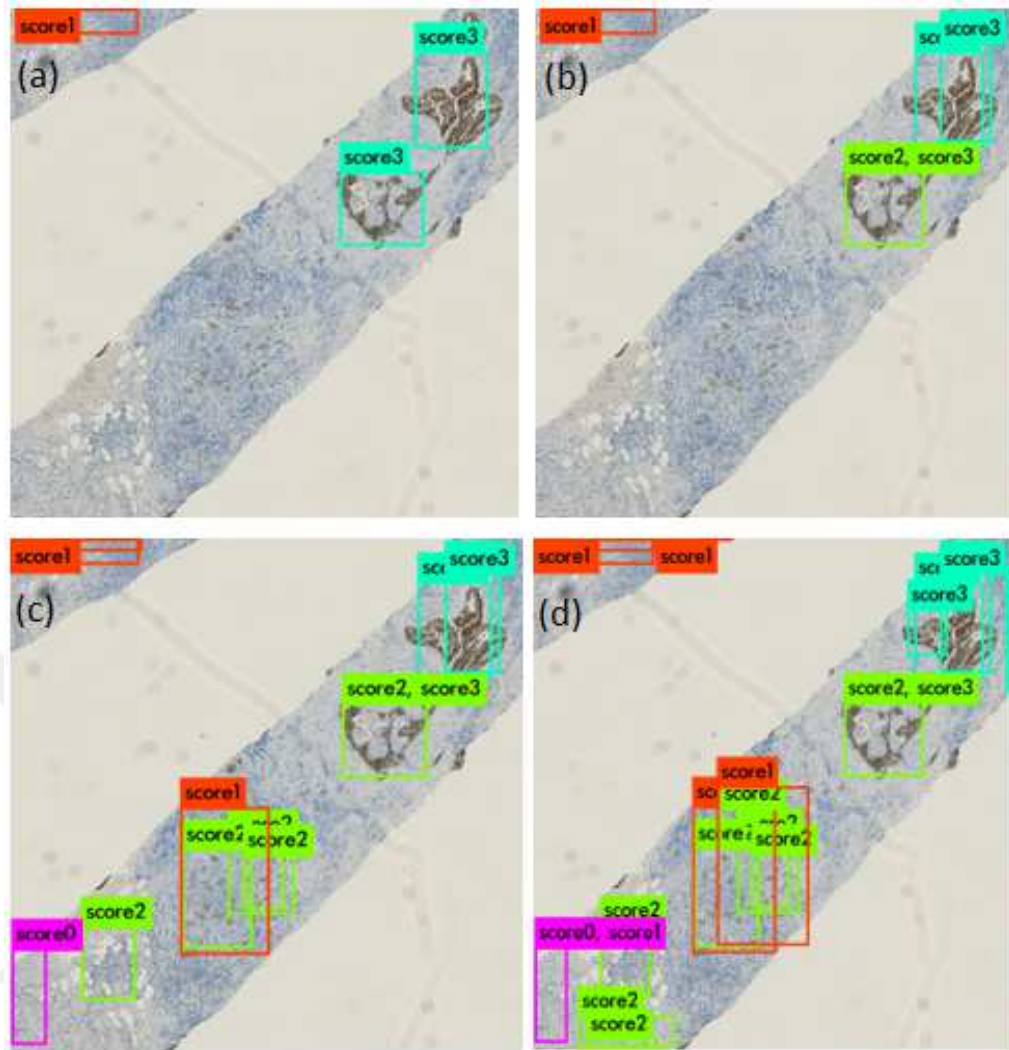


Figure 2.26 : Score 2 WSI samples which are applied different threshold values by decreasing order from “1” to “0” : (a) test result by threshold 1, (b) test result by threshold 0.5, (c) test result by threshold 0.25 and (d) test result with threshold 0.1.

According to this proposal, if the score is Score 2 or Score 3, which is 10% or more, this tissue sample is labeled as 2 or 3 according to the majority score type in the population. On the other hand, Score 0 and Score 1, in a condition of the other score types proportion is less than 10%, the tissue sample is labeled 0 or 1 according to the percentage distribution of the score types. The Table 2.10 compares accuracy of models by mAP (Mean Average Precision) and IOU [75]. While the mAP defines average maximum precisions of object detections, IOU means how much ground truth and predictor bounding boxes intersect with one another. [75].

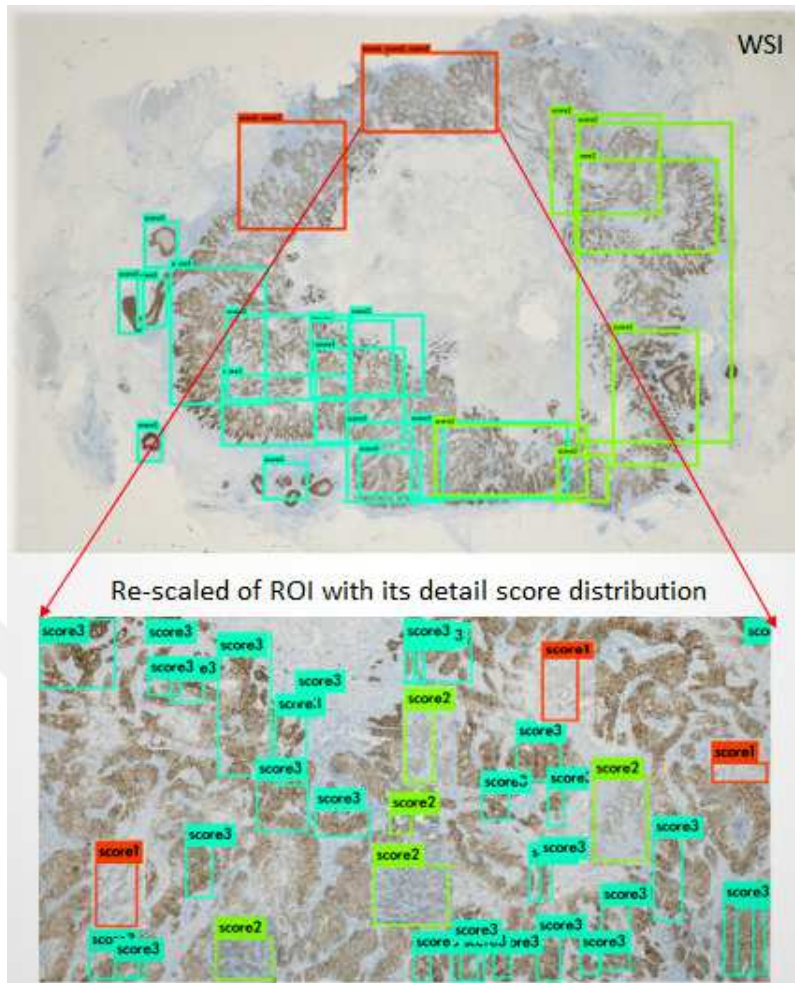


Figure 2.27 : The score distribution on tissue sample with the original resolution of ROI through the rescaled WSI.

According to result Table 2.10, mAP is lower than IOU, because some false positive predictions occurred with these weights. Abbreviations in the Table 2.10 are following: “mAP” is Mean Average Precision, “IOU” is Intersection of Union, “Weight1” is weight that obtained with compressed WSI, “Weight2” is weight that obtained with augmented rescaled WSI, “Weight3” is weight that obtained with tissue patches that are original resolution. The lowest accuracy is observed by weight1 which is obtained with smaller dataset that consist of resized WSI and the highest accuracy is obtained by weight3 which is produced by training with 10 times data of weight1 that are prepared from WSI with original resolution. The slight increase in accuracy of weight2 can be explained by data augmentation owing to bilinear image rotation. Even though compressed WSI’s are used as in weight 1, the data accuracy is increased in weight 2 because the data set is almost doubled with data augmentation. In fact, the number of data and image resolution can be shown for certain reasons of the difference in

accuracy between weights, as the number of iterations of all training is very close to each other and varies from 25000 to 28000. In general, this level of accuracy is envisaged to be increased by pathologic WSI, with more and better annotated datasets, although the accuracy is sufficient for the initial study. By reassessing of images that can not be accurately scored by pathologists, the pathologists will improve the data set that we created for the trainings. This situation is also considered to support of increase in accuracy as another factor.

Table 2.10 : Performance results of weights with Yolov2.

| <u>Detector Model</u> | <u>mAP</u> | <u>IOU</u> |
|------------------------------|-------------------|-------------------|
| Weight1 | 45% | 48.2% |
| Weight2 | 62% | 65% |
| Weight3 | 78.2% | 82% |

As a result, with this application, score estimation is made on the whole slide, and automatically detected regions are extracted to be analysed at their original resolution by re-scaling. Thus, the wrong score estimation due to the compression of the image is prevented. Also detail pathological examination will be provided for the labeled re-scaled regions. It is believed that this will shorten the diagnosis time of the breast cancer stage.



3. DISCUSSION and CONCLUSION

A novel image analysis based automated tissue scoring system is developed to determine the score of CerbB2 receptors in breast cancer tissue specimen images of patients with suspected breast cancer. Images of pathological tissue specimens of breast cancer is investigated in computer environment and analysed under the headings of classification and detection to find score types which define the cancerous phase. The diagnosis of breast cancer is determined in the four main classes Score 0, Score 1, Score 2, Score 3 according to the presence of CerbB2 receptors. These score types have been determined in the computer environment through staining techniques like IHC. According to the presence of staining, tissue images are analyzed by three approaches: original handcrafted feature extraction method to cell based scoring by SVM classifier, score classification for cell patches by CNN architectures, and score detection through WSI.

In first approach, tumor distribution on the membranes of cells are observed, and multiple features are extracted to each cell and these features are used in SVM training to form a classifier model with high accuracy. By this classifier, scores of tissue patches are determined based on score distribution of cells in the guidance with ASCO / CAP 2013 [27] rules. The accuracy is also compared with other approaches and higher accuracy is observed in cell based score classification. However accuracy of each score type has changed, some score class consists of multiple score information and similar to close classes. Therefore, some wrong score predictions occur for tissue samples, this situation is observed especially in tissue samples with Score 2. It is seen that samples belonging to Score 2 class can be confused with Score 3 and Score 1 class. This situation shows similar results in other approaches in this field.

In the second approach, WSI are divided into tissue patches at the cellular level and proposed an original dataset [76]. The result by CNN based feature extraction method is not get over the SVM classifier success which is achieved by handcrafted methods in first approach. Because the handcrafted feature extraction method works only for

the region of cell size, while deep learning models make sense out of the whole image. This reduces the success of SVM supported classifiers against with CNN based approach, even though higher accuracy is obtained by SVM based approach. However, CNN based classifier performance can be increased up to 99% by more accurate annotation of the pathological dataset for deep learning training. One last thing to note is that we do not need complex CNN models like GoogleNet in this classification approach [77], since these 2 dimensional pathological images don't consist of complex features. There is no need any in-depth analysis for observation of tumor distribution on the staining regions. These pathological images painted with only two color channels with H-Blue and Dark-Brown and only these two color channel are taken into consideration during feature extraction. Also score classes are visually similar to each other. For these reasons, there is no need for multilayered feature extraction. If this study is dealt with by multilayered CNN architectures, it could show the same success as either lesser layered architectures, or it could cause confusion between score classes.

In the third approach, diagnosis of score type for all pathological tissues is dealt by CNN based detection approach. The ROI on the WSI can be directly labeled by the deep learning networks with the relevant scores. For this purpose, three different datasets are prepared, and three different detector models are created, by resulting in trainings using Yolov2 CNN architecture. The first model is created by scaling WSI's to smaller dimensions. The second model is derived from a set of artificial data that is augmented by rotating of existing WSI instances. The third model is obtained from the WSI's with original resolution and about 10 times as many data sets as the other data sets. The third model is used for rescaling the detected regions for more detailed analysis of the pathological image. Therefore, the ROI that are marked with more than one label on the entire rescaled tissue image can be reanalyzed by pathologists. The accuracy of existing models can be increased by improved annotation to larger datasets.

To increase the accuracy of the models, pathological images with uncertain score values may need to be re-examined by pathologists and image annotation may need to be re-established within the decisions taken. Because some tissue samples may show similarity with more than one score classes, and to determine the exact score type of the samples with uncertain score value, more pathologists should examine the samples

in detail than the other tissue samples. This approach can be targeted as a next study, in which false positives in tests can be reduced and more robust detector models can be created.

To sum up, this study is implemented in the guidance with pathologists. The pathological images that could not be directly processed were examined and diagnosed with computer aided approaches. The aim of this study is to improve the pathological tissue diagnosis process of breast cancer and to provide a computer-assisted and self-renewing program owing to machine learning approaches. These approaches can reduce the false positive problems that may arise during the score diagnosis. Therefore, it is anticipated that CAD will reduce diagnosis of cancer phase, so it is encouraged to be used more actively in the medical world. The cost of medical devices are too high to be handled for undeveloped countries, so computer aided applications can meet the needs in the guidance with pathologists instead of expensive medical tools. In this context, we believe that the work we are dealing with can be an example when it is arranged to transform into practice.



CONTRIBUTIONS

This study examines pathologic images of breast cancer in three main approaches by computer support. Following details contain distinct sides of these approaches.

An original handcrafted cell based features extraction method is implemented. By using these features, a SVM classifier model with high accuracy is created. The score of each cell in the WSI is determined by this classifier model. Then, the score class of the related tissue sample is labeled according to the cell based score distribution in the whole image within the rules defined by ASCO/CAP [27]. The accuracy of this classification method outperforms to similar approaches. Also a dataset created during this study is shared online and is available for all researchers. Since preliminary annotated pathologic dataset is difficult to find, this study contributes to similar by sharing its dataset.

Another dataset is created which consists of pathological tissue patches by the cellular regions but also by the peripheral regions of the cells. Then, the dataset is trained with multi layered CNN architectures. The cell patches are tested by generated CNN classifier models rather than the whole tissue sample. Thus, a more detailed scoring system has been realised. Since the cells are addressed with their environmental neighborhoods, the tumor distribution can be examined in more detail. This is regarded as a background study for pathological image scoring by WSI. Also the accuracy of the models has been increased by changing the classifier loss functions of CNN architectures and features of CNN's extracted to form second SVM classifier model.

Score detection is performed on WSI using three different models with datasets. By this approach the ROI detection method provides direct scoring of WSI. Whole tissue sample is analyzed and scored without fragmenting. The reason for preparing multiple datasets is to examine the distribution of scores at original resolution of WSI while the other reason is to increase the success rate achieved with small datasets. Thus, unlike other studies related with pathological image analysis of breast cancer, the scoring type of the relevant regions is labeled on the WSI directly. More than one detector model

have been created so the relevant regions are examined in detail within the rules defined by ASCO/CAP [27].

All scoring approaches and annotation of tissue samples with breast cancer are supported by the methods in the way pathologists teach.



REFERENCES

- [1] **Doi, K.** (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5), 198-211.
- [2] **Geçer, B.** (2016). Detection and classification of breast cancer in whole slide histopathology images using deep convolutional networks (Doctoral dissertation, Bilkent University).
- [3] **Hecke, D. V.** (2002). Routine immunohistochemical staining today: choices to make, challenges to take. *Journal of histotechnology*, 25(1), 45-54.
- [4] **Simonyan, K., & Zisserman, A.** (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [5] **Helin, H. O., Tuominen, V. J., Ylinen, O., Helin, H. J., & Isola, J.** (2016). Free digital image analysis software helps to resolve equivocal scores in HER2 immunohistochemistry. *Virchows Archiv*, 468(2), 191-198.
- [6] **Janowczyk, A., & Madabhushi, A.** (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. *Journal of pathology informatics*, 7.
- [7] **Sheikhzadeh, F., Carraro, A., Korbelik, J., MacAulay, C., Guillaud, M., & Ward, R. K.** (2016, March). Automatic labeling of molecular biomarkers on a cell-by-cell basis in immunohistochemistry images using convolutional neural networks. In *Medical Imaging 2016: Digital Pathology* (Vol. 9791, p. 97910R). International Society for Optics and Photonics.
- [8] **Basavanahally, A., Feldman, M., Shih, N., Mies, C., Tomaszewski, J., Ganesan, S., & Madabhushi, A.** (2011). Multi-field-of-view strategy for image-based outcome prediction of multi-parametric estrogen receptor-positive breast cancer histopathology: Comparison to Oncotype DX. *Journal of pathology informatics*, 2.
- [9] **Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H.** (2016). Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.
- [10] **Madabhushi, A., & Lee, G.** (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33, 170-175.

- [11] **Atkins, D., Reiffen, K. A., Tegtmeier, C. L., Winther, H., Bonato, M. S., & Störkel, S.** (2004). Immunohistochemical detection of EGFR in paraffin-embedded tumor tissues: variation in staining intensity due to choice of fixative and storage time of tissue sections. *Journal of Histochemistry & Cytochemistry*, 52(7), 893-901.
- [12] **Layfield, L. J., Frazier, S., Esebua, M., & Schmidt, R. L.** (2016). Interobserver reproducibility for HER2/neu immunohistochemistry: a comparison of reproducibility for the HercepTest™ and the 4B5 antibody clone. *Pathology-Research and Practice*, 212(3), 190-195.
- [13] **Fischer, A. H., Jacobson, K. A., Rose, J., & Zeller, R.** (2008). Hematoxylin and eosin staining of tissue and cell sections. *Cold Spring Harbor Protocols*, 2008(5), pdb-prot4986.
- [14] **Ruifrok, A. C., & Johnston, D. A.** (2001). Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology*, 23(4), 291-299.
- [15] **Hearst, M. A., Dumais, S. T., Osuna, E., Platt, J., & Scholkopf, B.** (1998). Support vector machines. *IEEE Intelligent Systems and their applications*, 13(4), 18-28.
- [16] **Wang, Y., Qiu, Y., Thai, T., Moore, K., Liu, H., & Zheng, B.** (2017). A two-step convolutional neural network based computer-aided detection scheme for automatically segmenting adipose tissue volume depicting on CT images. *Computer methods and programs in biomedicine*, 144, 97-104.
- [17] **Loukas, C. G., & Linney, A.** (2004). A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature. *Computer Methods and Programs in Biomedicine*, 74(3), 183-199.
- [18] **Wu, W. W., Gu, M., & Lu, D.** (2014). Cytopathologic, histopathologic, and immunohistochemical features of intrahepatic clear cell bile duct adenoma: a case report and review of the literature. *Case reports in pathology*, 2014.
- [19] **Kalvapalle, P. B.** (2016). *FUNCTIONAL ANALYSIS OF TYROSINE KINASE MUTANTS IN CANCER* (Doctoral dissertation, Johns Hopkins University).
- [20] **Johnson, K. L., Zhen, D. K., & Bianchi, D. W.** (2000). The use of fluorescence in situ hybridization (FISH) on paraffin-embedded tissue sections for the study of microchimerism. *Biotechniques*, 29(6), 1220-1224.
- [21] **Garrison, L. P., & Veenstra, D. L.** (2009). The economic value of innovative treatments over the product life cycle: the case of targeted trastuzumab therapy for breast cancer. *Value in health*, 12(8), 1118-1123.
- [22] **Tchrakian, N., Flanagan, L., Harford, J., Gannon, J. M., & Quinn, C. M.** (2016). New ASCO/CAP guideline recommendations for HER2 testing increase the proportion of reflex in situ hybridization tests and of HER2 positive breast cancers. *Virchows Archiv*, 468(2), 207-211.

- [23] **Mass, R. D., Press, M. F., Anderson, S., Cobleigh, M. A., Vogel, C. L., Dybdal, N., ... & Slamon, D. J.** (2005). Evaluation of clinical outcomes according to HER2 detection by fluorescence in situ hybridization in women with metastatic breast cancer treated with trastuzumab. *Clinical breast cancer*, 6(3), 240-246.
- [24] **Hammond, M. E. H., & Hicks, D. G.** (2015). American Society of Clinical Oncology/College of American Pathologists human epidermal growth factor receptor 2 testing clinical practice guideline upcoming modifications: proof that clinical practice guidelines are living documents. *Archives of pathology & laboratory medicine*, 139(8), 970-971.
- [25] **Di Cataldo, S., Ficarra, E., Acquaviva, A., & Macii, E.** (2010). Automated segmentation of tissue images for computerized IHC analysis. *Computer methods and programs in biomedicine*, 100(1), 1-15.
- [26] **Rakha, E. A., Starczynski, J., Lee, A. H., & Ellis, I. O.** (2014). The updated ASCO/CAP guideline recommendations for HER2 testing in the management of invasive breast cancer: a critical review of their implications for routine practice. *Histopathology*, 64(5), 609-615.
- [27] **Wolff, A. C., Hammond, M. E. H., Hicks, D. G., Dowsett, M., McShane, L. M., Allison, K. H., ... & Hanna, W.** (2013). Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *Archives of Pathology and Laboratory Medicine*, 138(2), 241-256.
- [28] **Layfield, L. J., Frazier, S., Esebua, M., & Schmidt, R. L.** (2016). Interobserver reproducibility for HER2/neu immunohistochemistry: a comparison of reproducibility for the HercepTest™ and the 4B5 antibody clone. *Pathology-Research and Practice*, 212(3), 190-195.
- [29] **Lo, C. S., & Wang, C. M.** (2012). Support vector machine for breast MR image classification. *Computers & Mathematics with Applications*, 64(5), 1153-1162.
- [30] **Chang, C. C., & Lin, C. J.** (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3), 27.
- [31] **Caruana, R., & Niculescu-Mizil, A.** (2006, June). An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd international conference on Machine learning* (pp. 161-168). ACM.
- [32] **Ali, S., & Smith-Miles, K. A.** (2006, December). Improved support vector machine generalization using normalized input space. In *Australasian Joint Conference on Artificial Intelligence* (pp. 362-371). Springer, Berlin, Heidelberg.
- [33] **Arlot, S., & Celisse, A.** (2010). A survey of cross-validation procedures for model selection. *Statistics surveys*, 4, 40-79.

- [34] **An, S., Liu, W., & Venkatesh, S.** (2007). Fast cross-validation algorithms for least squares support vector machine and kernel ridge regression. *Pattern Recognition*, 40(8), 2154-2162.
- [35] **Dietterich, T.** (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, 27(3), 326-327.
- [36] **Reunanen, J.** (2003). Overfitting in making comparisons between variable selection methods. *Journal of Machine Learning Research*, 3(Mar), 1371-1382.
- [37] **Elisseeff, A., & Weston, J.** (2002). A kernel method for multi-labelled classification. In *Advances in neural information processing systems* (pp. 681-687).
- [38] **Doi, K.** (2007). Computer-aided diagnosis in medical imaging: historical review, current status and future potential. *Computerized medical imaging and graphics*, 31(4-5), 198-211.
- [39] **Šajin, L., & Kukar, M.** (2011). Image processing and machine learning for fully automated probabilistic evaluation of medical images. *Computer methods and programs in biomedicine*, 104(3), e75-e86.
- [40] **Tataroglu, G. A., Kabakçı, K.A., Çapar, A., Töreyn, B. U.,** (2017). "Regional Tissue Samples of Immunohistochemistry Stained Breast Cancer" <https://kovan.itu.edu.tr/index.php/s/Ey3dCu14SCK18of>.
- [41] **Saini, S., Sharma, S., & Singh, R.** (2015, December). Better utilization of correlation between metrics using Principal Component Analysis (PCA). In *India Conference (INDICON), 2015 Annual IEEE* (pp. 1-6). IEEE.
- [42] **Mayoraz, E., & Alpaydin, E.** (1999, June). Support vector machines for multi-class classification. In *International Work-Conference on Artificial Neural Networks* (pp. 833-842). Springer, Berlin, Heidelberg.
- [43] **Duan, K., Keerthi, S. S., & Poo, A. N.** (2003). Evaluation of simple performance measures for tuning SVM hyperparameters. *Neurocomputing*, 51, 41-59.
- [44] **Angulo, C., Parra, X., & Catala, A.** (2003). K-SVCR. A support vector machine for multi-class classification. *Neurocomputing*, 55(1-2), 57-77.
- [45] **Zhu, K., Wang, H., Bai, H., Li, J., Qiu, Z., Cui, H., & Chang, E. Y.** (2008). Parallelizing support vector machines on distributed computers. In *Advances in Neural Information Processing Systems* (pp. 257-264).
- [46] **Musavi, M. T., Ahmed, W., Chan, K. H., Faris, K. B., & Hummels, D. M.** (1992). On the training of radial basis function classifiers. *Neural networks*, 5(4), 595-603.
- [47] **Lin, S. L., & Liu, Z.** (2007). Parameter selection in SVM with RBF kernel function. *Journal-Zhejiang University of Technology*, 35(2), 163.
- [48] **Lin, H. T., & Lin, C. J.** (2003). A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. submitted to *Neural Computation*, 3, 1-32.

- [49] **Novakovic, J., & Veljovic, A.** (2011, September). C-support vector classification: Selection of kernel and parameters in medical diagnosis. In *Intelligent Systems and Informatics (SISY), 2011 IEEE 9th International Symposium on* (pp. 465-470). IEEE.
- [50] **Rajpoot, K., & Rajpoot, N.** (2004, September). SVM optimization for hyperspectral colon tissue cell classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 829-837). Springer, Berlin, Heidelberg.
- [51] **Schmidhuber, J.** (2015). Deep learning in neural networks: An overview. *Neural networks*, 61, 85-117.
- [52] **Krizhevsky, A., Sutskever, I., & Hinton, G. E.** (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105).
- [53] **Simonyan, K., & Zisserman, A.** (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- [54] **Zeiler, M. D., & Fergus, R.** (2014, September). Visualizing and understanding convolutional networks. In *European conference on computer vision* (pp. 818-833). Springer, Cham.
- [55] **Tang, Y.** (2013). Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*.
- [56] **Mao, X., Li, Q., Xie, H., & Rao, Y.** (2014, April). Popularity tendency analysis of ranking-oriented collaborative filtering from the perspective of loss function. In *International Conference on Database Systems for Advanced Applications* (pp. 451-465). Springer, Cham.
- [57] **Otsu, N.** (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1), 62-66.
- [58] **Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., ... & Darrell, T.** (2014, November). Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 675-678). ACM.
- [59] **Ng, J. Y. H., Hausknecht, M., Vijayanarasimhan, S., Vinyals, O., Monga, R., & Toderici, G.** (2015, June). Beyond short snippets: Deep networks for video classification. In *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on* (pp. 4694-4702). IEEE.
- [60] **Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L.** (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4), 834-848.
- [61] **Fan, R. E., Chang, K. W., Hsieh, C. J., Wang, X. R., & Lin, C. J.** (2008). LIBLINEAR: A library for large linear classification. *Journal of machine learning research*, 9(Aug), 1871-1874.
- [62] **Gong, Y., Jia, Y., Leung, T., Toshev, A., & Ioffe, S.** (2013). Deep convolutional ranking for multilabel image annotation. *arXiv preprint arXiv:1312.4894*.

- [63] **Wu, Y., & Liu, Y.** (2007). Robust truncated hinge loss support vector machines. *Journal of the American Statistical Association*, 102(479), 974-983.
- [64] **Friedman, J., Hastie, T., & Tibshirani, R.** (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2), 337-407.
- [65] **Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.** (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).
- [66] **Gkioxari, G., Girshick, R., & Malik, J.** (2015). Contextual action recognition with r* cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1080-1088).
- [67] **Uijlings, J. R., Van De Sande, K. E., Gevers, T., & Smeulders, A. W.** (2013). Selective search for object recognition. *International journal of computer vision*, 104(2), 154-171.
- [68] **Girshick, R.** (2015). Fast r-cnn. arXiv preprint arXiv:1504.08083.
- [69] **Ren, S., He, K., Girshick, R., & Sun, J.** (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In *Advances in neural information processing systems*(pp. 91-99).
- [70] **Zitnick, C. L., & Dollár, P.** (2014, September). Edge boxes: Locating object proposals from edges. In *European Conference on Computer Vision* (pp. 391-405). Springer, Cham.
- [71] **Url-2** <<https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006>> citation date: 01.05.2018.
- [72] **Redmon, J., & Farhadi, A.** (2018). YOLOv3: An Incremental Improvement. arXiv preprint arXiv:1804.02767.
- [73] **Redmon, J., & Farhadi, A.** (2017). YOLO9000: better, faster, stronger. *arXiv preprint..* <https://pjreddie.com/publications/yolo/>.
- [74] **Url-3** < <https://github.com/aleju/imgaug> > citation date: 30.04.2018.
- [75] **Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., & LeCun, Y.** (2013). Overfeat: Integrated recognition, localization and detection using convolutional networks. *arXiv preprint arXiv:1312.6229*.
- [76] **Tataroğlu, G. A., Genç, A., Kabakçı, K. A., Çapar, A., Töreyn, B. U., Ekenel, H. K., ... & Çakır, A.** (2017, May). A deep learning based approach for classification of CerbB2 tumor cells in breast cancer. In *Signal Processing and Communications Applications Conference (SIU), 2017 25th* (pp. 1-4). IEEE.
- [77] **Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z.** (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2818-2826).

CURRICULUM VITAE

Name Surname: Gözde Ayşe TATAROĞLU

E-Mail : gozdeayse.t@gmail.com

B.Sc. : 2014, Computer Engineering

M.Sc. : 2018, Computer Science

Professional Experience :

I have been working as a research engineer on deep learning, machine learning, computer vision.

I am pursuing both academical and industrial studies in these fields.

PUBLICATIONS/PRESENTATIONS ON THE THESIS

- **Tataroğlu G.A.**, Kabakçı K., et al. Töreyin B.U., 2018 :Scoring of CerbB2 Receptors Using Histogram Based Analysis of IHC-Stained Breast Cancer Tissue Images, *Computer Methods and Programs in Biomedicine*, Elsevier. (under second round of review)
- **Tataroğlu, G. A.**, Genç, A., Kabakçı, K. A., Çapar, A., Töreyin, B. U., Ekenel, H. K., ... & Çakır, A., 2017: A deep learning based approach for classification of CerbB2 tumor cells in breast cancer. *Signal Processing and Communications Applications Conference (SIU)*, 2017 25th (pp. 1-4). IEEE.