Computers in Biology and Medicine 110 (2019) 164–174

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Automated segmentation of cell membranes to evaluate HER2 status in whole slide images using a modified deep learning network

Fariba Damband Khameneh[a], Salar Razavi[b,*], Mustafa Kamasak[c]

[a] Informatics Institute, Istanbul Technical University, Istanbul,Turkey
[b] Department of Electronics and Communication, Istanbul Technical University, Istanbul, Turkey
[c] Department of Computer and Informatics, Istanbul Technical University, Istanbul, Turkey

ARTICLE INFO

Keywords:
CNN
Digital pathology
Whole slide image
Deep learning
HER2 assessment
Membrane segmentation

ABSTRACT

The uncontrollable growth of cells in the breast tissue causes breast cancer which is the second most common type of cancer affecting women in the United States. Normally, human epidermal growth factor receptor 2 (HER2) proteins are responsible for the division and growth of healthy breast cells. HER2 status is currently assessed using immunohistochemistry (IHC) as well as in situ hybridization (ISH) in equivocal cases. Manual HER2 evaluation of IHC stained microscopic images involves an error-prone, tedious, inter-observer variable, and time-consuming routine lab work due to diverse staining, overlapped regions, and non-homogeneous remarkable large slides. To address these issues, digital pathology offers reproducible, automatic, and objective analysis and interpretation of whole slide image (WSI). In this paper, we present a machine learning (ML) framework to segment, classify, and quantify IHC breast cancer images in an effective way. The proposed method consists of two major classifying and segmentation parts. Since HER2 is associated with tumors of an epithelial region and most of the breast tumors originate in epithelial tissue, it is crucial to develop an approach to segment different tissue structures. The proposed technique is comprised of three steps. In the first step, a superpixel-based support vector machine (SVM) feature learning classifier is proposed to classify epithelial and stromal regions from WSI. In the second stage, on classified epithelial regions, a convolutional neural network (CNN) based segmentation method is applied to segment membrane regions. Finally, divided tiles are merged and the overall score of each slide is evaluated. Experimental results for 127 slides are presented and compared with state-of-the-art handcraft and deep learning-based approaches. The experiments demonstrate that the proposed method achieved promising performance on IHC stained data. The presented automated algorithm was shown to outperform other approaches in terms of superpixel based classifying of epithelial regions and segmentation of membrane staining using CNN.

1. Introduction

Breast cancer is the most prevalent form of cancers among women. Currently, based on American society of clinical oncology/college of American pathologists ASCO/CAP guideline, there is a 1 in 8 chance American women would develop breast cancer in their life. HER2 could play an important role in the development of breast cancer. HER2 proteins are responsible for how cells grow and divide. Therefore, HER2 therapy in combination with chemotherapy or/and endocrine therapy could be the most effective treatment. Furthermore, in breast cancer treatment, trastuzumab and lapatinib therapies are implied to be effective in HER2 amplified cases. The traditional system of HER2 breast cancer assessment deficits from accuracy on detection of correct patients overexpressing HER2. The prevalence of HER2 overexpression is associated with invasive breast cancer in about 20% of breast cancers. Admittedly, precise and fast HER2 assessment is crucial to consider the appropriate action for patients. Quantitative image analysis (QIA) of digitalized slides decreases human error, increases the accuracy of diagnosis, reduces the workload of pathologists, and standardizes scoring systems. In a HER2 assessment of IHC slides, to address ambiguities and subjectivities of manual scoring, computer-aided solutions are provided to simplify the overall progress. With the advent of image analysis in digital pathology, a huge interest has received to digital slide scanners to process and evaluate typical pathology lab workload in a digital, fast, accurate, and efficient way. According to ASCO/CAP guidelines [1] shown in Table 1, in IHC slides if more than 10% of the whole tissue comprises strong tumor cells the case displays 3+, which is accepted as positive and are allowed for therapies. If the ratio for moderate tumor

[*] Corresponding author.
E-mail address: razavi15@itu.edu.tr (S. Razavi).

https://doi.org/10.1016/j.compbiomed.2019.05.020
Received 29 January 2019; Received in revised form 24 May 2019; Accepted 25 May 2019

Table 1
Evaluation criteria for HER2 (ERBB2) protein expression by IHC assay of the invasive component of a breast cancer specimen.

Specimen Staining Pattern	Score	Classification	
Incomplete membrane staining that is faint or barely perceptible and within ≤ 10% of the invasive tumor cells or no staining observed	0	Negative	
Incomplete membrane staining that is faint/barely perceptible in ≥10% of tumor cells	1+	Negative	
Weak to moderate complete membrane staining observed in ≥10% of tumor cells	2+	Equivocal (Observer blinded to previous results recounts ISH to Evaluate HER2/CEP17 ratio and average HER2 signals/cell)	
Circumferential membrane staining that is complete, intense and in ≥10% of tumor cells		3+	Positive

cells is more than 10% the case is considered as equivocal 2 + and reflexed to ISH test to assess HER2 status [2]. In no staining or weak conditions, the case is HER2 negative.

Recently, machine learning methods have considerably enhanced the ability of computers to automatically diagnose various components in biomedical images. Among pixel-based clustering methods, simple linear iterative clustering (SLIC) [3] is studied in histopathological segmentation tasks. SLIC is a clustering method that agglomerates similar and nearby pixels and is accepted as a superior method in terms of accuracy and efficiency. Jiří Borovec et al. [4] employed SLIC as a preliminary step in histopathological images to increase the efficiency of Graph-cut method. In Ref. [5], epithelium and stromal regions in histopathological images are segmented using a hierarchal fuzzy c-means method. Babak et al. [6], detected regions of interest in WSIs using a multi-scale superpixel classification approach that classifies at different scales based on the acquired details of the region of interest (ROI). Belsare et al. [7] also used texture features to classify malignant and benign breast histopathology images. Most of the developed approaches are related to the automatic classification of hematoxylin and eosin (H&E) tissue images [8]. In automated HER2 and H&E assessment from pathology tissue slides, several classical and handcraft approaches are presented [9–11]. Most of these methods are about threshold-based approaches as in Ref. [12] by using an optimal threshold value, the percentage of the stained area and the score is evaluated. The effectiveness of some segmentation approaches [13,14] is limited by the remarkable intensity and color variation in nuclear and nuclear membranes. In a work by Morteza et al. [15], a WSI based classifier using robust local binary pattern (LBP) and characteristic features is provided. The extracted characteristic is scored through a naïve rule-based classifier. The rotation invariant LBP features are also converted to percentages that are used to classify selected ROIs. The proposed method is evaluated on various stainings. A membrane connectivity based algorithm that automatically specifies the HER2 status in preselected sections of the tissues is presented in Ref. [16]. The method segmented brown pixels and scored each slide using the skeletonized connected membrane. In Ref. [17], basolateral membranous activity and neoplastic cell count were evaluated using segmentation and thresholding methods and the results of the computer-aided analysis were compared with manual evaluation. In another study [18], manual outlined ROIs were transformed into HIS color space and various features were extracted to train an SVM classifier. In the test step, the image was classified through a voting system. In IHC membrane staining, one of the most challenging issues is the reconstruction of the membranes that are not revealed and then not visible. A method based on nuclear membranes and approximating cellular membranes to automatically detect the bounding membrane of each cell were presented by Ficarra et al. [19]. In Immunumembrane [20], as a public domain and developed using ImageJ, a set of thresholding, morphology, segmentation, and a point-based membrane evaluation named IM-score was presented. The proposed method in Ref. [21], showed that the characteristic curves and uniform rotation invariant LBP feature curves could be useful in automated HER2 scoring. Lately, state-of-the-art approaches like deep learning have intensely attracted the attention of researchers. Due to the rapid growth in large medical images, new interesting machine learning challenges rise which are supposed to give promising results under uncertain conditions. Deep learning is a computational model resembled from human cognition system that can be used efficiently in different applications [22]. In practice, an artificial neural network that has more than one hidden layer could be considered as a deep learning architecture. Currently, medical imaging and digital pathology show increasing interest in deep learning as demonstrated by various studies. CNNs that are one form of deep learning, have been well suited to medical data and have been incorporated successfully in different segmentation [23–25], classification [24,26–28], and detection [26,29,30] tasks.

Among deep learning models, CNNs are the most researched methods in medical image understanding tasks. One of the implementations of CNN in HER2 assessment was presented in Ref. [31], where detected cells from IHC stained tissues were classified after some morphological operations. The proposed method was based on whole slide cell classification using CNN and the results were considerably higher than classical machine classification methods. In Ref. [32], 128×128 blocks of four labels at a low resolution were considered as the training data. For each slide, the ratio of blocks with each label to total blocks was considered to determine the HER2 score for a WSI. The proposed methods in DL for HER2 assessment are all about cell [31] or tile-based classification [33]. To the best of our knowledge, for HER2 scoring, none DL based work has been done based on segmentation of cell membranes. However, there are some other applications of DL in this area. Ideally, deeper architectures in CNN represent better results. In Ref. [33], long short term memory (LSTM) architecture was proposed to detect cell membrane and nucleus in small patches. However, in HER2 assessment WSI is usually considered to evaluate the overall result. In this paper, we propose a novel architecture for HER2 assessment of IHC biomarker. The proposed architecture exploits a SLIC clustering, SVM classifier, and CNN segmentation. We would investigate different variables which take into account segmentation of epithelial area as well as classifying WSI as positive, equivocal, or negative. The main goals of this study are (1) to segment and classify epithelium areas from stromal parts of slides correctly and (2) to apply a precise membrane segmentation using a convolutional auto-encoder that unlike sliding-window convolutional networks, relies on a strong data augmentation that efficiently trains with a few annotated samples and leads to fast and precise segmentation. The rest of this paper is organized as follows. The details of the proposed method are provided in section 2. The experimental results and comparative discussions are reported in Section 3. Further discussion of the results and the method is provided in Section 4. The Conclusion is presented in Section 5.

2. Proposed method

The HER2 IHC assessment is defined by ASCO/CAP guidelines [1]. In HER2 IHC images, intensity and completeness of epithelial brown (DAB) areas are two main factors. As described in Table 1, each slide is classified into 0/1 +, 2 +, 3 + cases based on these parameters. Each case specifies the amount of receptor protein on the surface of cells. Due to some factors, these test results are not always correct. The inter-observer and intra-observer variations are the main reasons for this probably wrong classification. Each pathologist and laboratory use slightly different rules and criteria to classify the breast cancer tissue sample. This inaccurate HER2 score would cause the breast cancer patient to get the wrong treatment and may be exposed to risks. The main advantage of automatic HER2 scoring is the ability to give consistent results on similar slides in a short time compared with the manual scoring performed by pathologists. In Fig. 1 the overview of the proposed methodology is shown.

The proposed methods in HER2 scoring are developed based on a pipeline consisted of preprocessing, patch-based processing, including feature detection, and techniques to score the WSI. In histological images, to automatically analyze the WSI, it is important to discriminate between epithelial and stromal tissues. In the first part of this research, by developing traditional machine learning approaches based on SVM (RBF kernel with Sigma 6), LBP (with 1-pixel distance of 8-pixel neighborhood), and color histogram (with 224 bins), HER2 stained WSI is classified to epithelial and non-epithelial areas. An illustration of the tissue segmentation part and its impact on the overall result are presented in Fig. 1.

The proposed approach is based on handcraft features that contribute to a supervised classifier. Classifying the regions aims to specify regions of interest for each WSI. Instead of working on all pixels, we used superpixels to compute features on meaningful and similar regions not only to increase the performance but also decrease the input variable for the subsequent classification step. All of the extracted tiles are normalized by matching histograms of each tile with a reference tile, which is acquired with the statistical analysis of deconvolved H and E color spaces. In addition to the reasons raised above, considering other factors of the efficiency beside consistent magnification level could also substantiate the justifiability of the claim made at the outset of this research. To delineate, CNN based solution reaches to higher accuracy in proportion to traditional image processing methods. To be impartial, working with various datasets would result in a rich trained model which would segment membranes with various zoom level and size.

2.1. Superpixel

Investigating and processing microscopic WSIs are an extremely tedious and time-consuming procedure. Computer-based machine learning tools assist pathologists to work on stained tissues to segment, recognize, classify, and reveal important information about the samples. In HER2 membrane scoring, it is important to count the number of closed and open membranes that are in the epithelial area. In this paper, we combined both local color histogram and LBP features together to build an automatic classification solution to reduce inter-observer variability and increase the accuracy of the procedure. The main reason to do this is to remove tiles that contain white pixels and to decrease the computational complexity. In a normal slide, about 30% of tiles with size 512×512 are redundant and empty. Therefore, we can increase the performance of the DL task by putting aside redundant tissue blocks. SLIC [3] is a method that clusters pixels in color and distance spaces to generate uniform areas named as superpixels. The main contributions of SLIC comparing other local feature extraction methods are rich quality segmentation, consistent size, and low computational cost. This simple approach performs based on the 5-dimensional feature space defined by [*labxy*] space in which l, a, b are the values of CIELAB color space and x, y stand for coordination of the point. This method is a special form of k-means clustering adapted to the local uniform group of pixels without redundant distance calculations.

2.2. Color features

In patients with breast cancer, the HER2 state is critical as the variations would specify the type of therapies. IHC is a special staining method to discover HER2 protein in the cancer cells that works based on detecting particular antigens in tissues. The IHC stained slides are comprised of the brown channel (diaminobenzidine DAB signal) and counterstain blue-violet channel (hematoxylin signal). This color-based reaction produces different structures like the nucleus, membrane or cytoplasm. In IHC HER2, the membrane and nuclei detection methodologies are performed on the brown and blue channels, respectively. Therefore, extracting ROI depends on the color information of the superpixels. Local color histogram and distribution of superpixels in clustered uniform patterns is used. The color bin of each histogram is computed based on areas contain pixels in superpixels.

2.3. Texture features

The LBP algorithm is a robust and powerful descriptor that thresholds the neighboring pixels based on a center pixel. Here, we would consider each superpixel separately and evaluate the histogram of each pixel. This uniform LBP archives rotation invariant descriptor of the pattern. The histogram of the pixels inside a superpixel is concatenated and for each superpixel, a unique feature and label are considered.

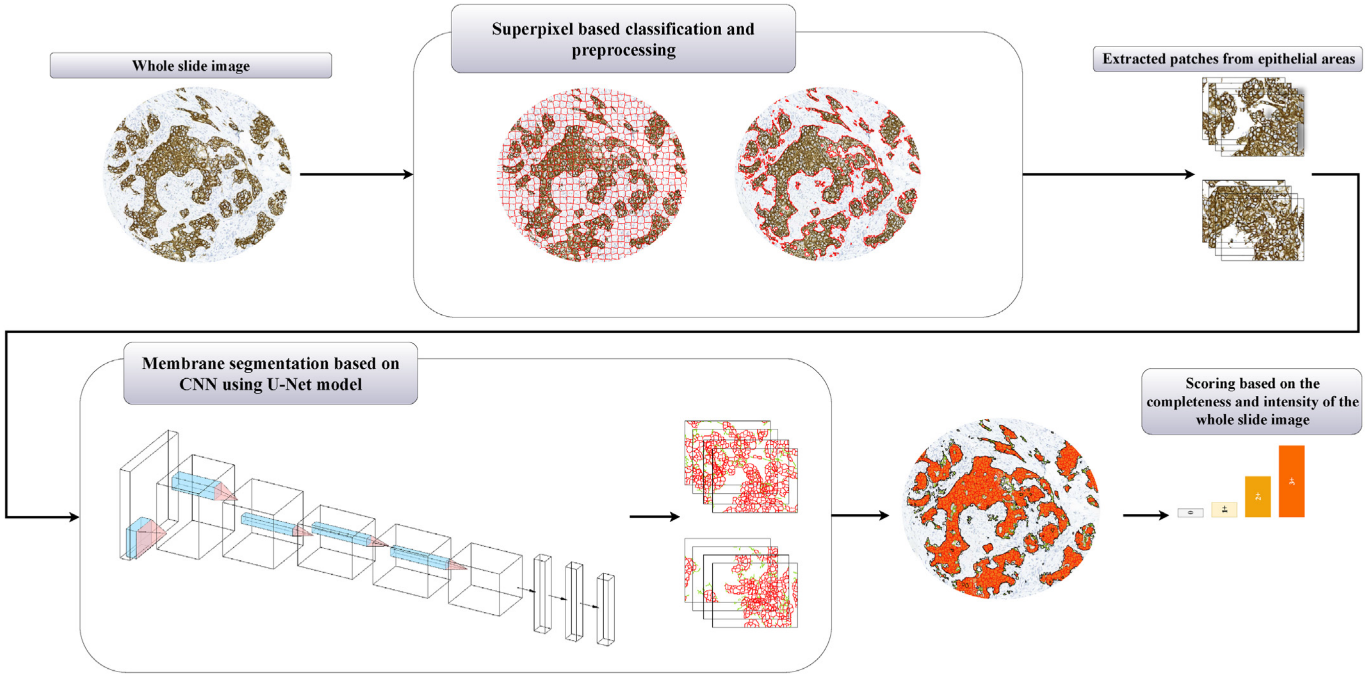
Whole slide image
Superpixel based classification and preprocessing
Extracted patches from epithelial areas
Membrane segmentation based on CNN using U-Net model
Scoring based on the completeness and intensity of the whole slide image

Fig. 1. The diagram of the applied method to segment cell membrane in HER2 stained IHC specimen which is the combination of deep learning and traditional machine learning algorithms. First, superpixel breaks the image to manageable parts. Handcrafted features are extracted to classify each superpixel to epithelium or stroma. The deep learning part is an end-to-end method that takes images as input and learns the modified U-Net model to produce segmentation result. The WSI merged from all tiles is obtained to get the overall score of the specimen.

2.4. *Classifier*

To discriminate epithelial and stromal areas using the local color histogram and LBP superpixel-level features, an SVM classifier is applied to predict the label of superpixels. To build the SVM classifier, labeled training data provided by an expert through a GUI is desired. The trained model for SVM classifier would be used for future tests. Due to the complexity of the whole IHC image, a pathologist should label small and descriptive epithelial and stromal patterns to segment the whole slide. This process is performed once as we would use this trained model for further samples.

2.5. *Modified U-Net model*

CNNs are typically used for classification tasks; however, the U-Net architecture extracts localized features recognizable by the human visual system to classify and segment complex structures. This is provided by the downsampling feature of the U-Net that makes it possible to assign a label to each pixel. Furthermore, the success of convolutional networks is usually dependent on the size of the training dataset. In biomedical applications, due to the complexity and the high cost of data collection, machine learning methods would work elegantly with small training samples. In the U-Net model, by acquiring upsampling layers instead of pooling layers the resolution of the input image is increased which enables the successive convolution layer to learn the more accurate result. In this type of network architecture, low-level feature maps are combined with higher-level ones to precisely locate the interest regions. We considered training images with corresponding annotations to train the proposed network. To use GPU memory efficiently, we favor large input tiles over and discard white empty tiles eliminated from the classification part. The segmentation procedure for HER2 is assigning one and zero pixels to each pixel in each tile. The network architecture is illustrated in Fig. 2.

It consists of two encodings (left) and decoding (right) sides. The encoding path is comprised of repeated convolutions, followed by the rectified linear unit (ReLU) and max-pooling layers. In right side after upsampling feature maps, a cropped feature map from encoding part is concatenated which is followed by the convolution and ReLU layers. At last, a convolution layer with 1×1 size is used to map 64 component feature maps to the number of output class. To reduce internal covariance shift, after two consecutive activation layers, batch normalization is employed in the proposed model. The interdependent layers of a CNN that are connected to each other may lead to over-fitting. Regularization is the best way to overcome over-fitting. Here, after each max pooling layer, dropout is used to address the over-fitting. In general, the number of foreground membrane areas is considerably smaller than the background ones. The introduced modified U-Net segmentation model achieves high accuracy by adopting a jointly trained edge and semantic losses. The cross entropy loss for N_{tot} inputs could give more importance by incorporating the number of foreground and background pixels. We adopted a weighted loss function [35] where the calculation of trade-off weight for biased sampling:

$$\mathscr{L}_{CE} = \frac{-1}{N_{tot}}\left(\sum_{i=1}^{P} \beta y_i \log(\hat{y}_i) + (1-\beta)(1-y_i)\log(1-\hat{y}_i)\right) \tag{1}$$

Where P is the number of pixels, y_i indicates whether a membrane is correctly predicted, $\hat{y}_i$ is the probability of model prediction to be membrane and β is the fraction of foreground (membrane pixels) to background in the ground-truth images.

The dice coefficient (DC) is widely used for evaluation image segmentation to evaluate spatial overlap with ground-truth. Considering the output of the soft layer, the loss function for the edge parts is:

$$\mathscr{L}_{DC} = \frac{-1}{N_{tot}}\left(1 - 2\times\left(\frac{\sum_{i=1}^{P} y_i^{edge} \times \hat{y}_i^{\,edge}}{\sum_{i=1}^{P} y_i^{edge} + \sum_{i=1}^{P} \hat{y}_i^{\,edge}}\right)\right) \tag{2}$$

where y_i^{edge} and $\hat{y}_i^{\,edge}$ represent the binary map for edge and the predicted output of the edge map respectively. Fig. 3 is an example of the target output that we aim to predict at the output of the edge label to optimize the loss function. These losses are added to minimize the overall loss.

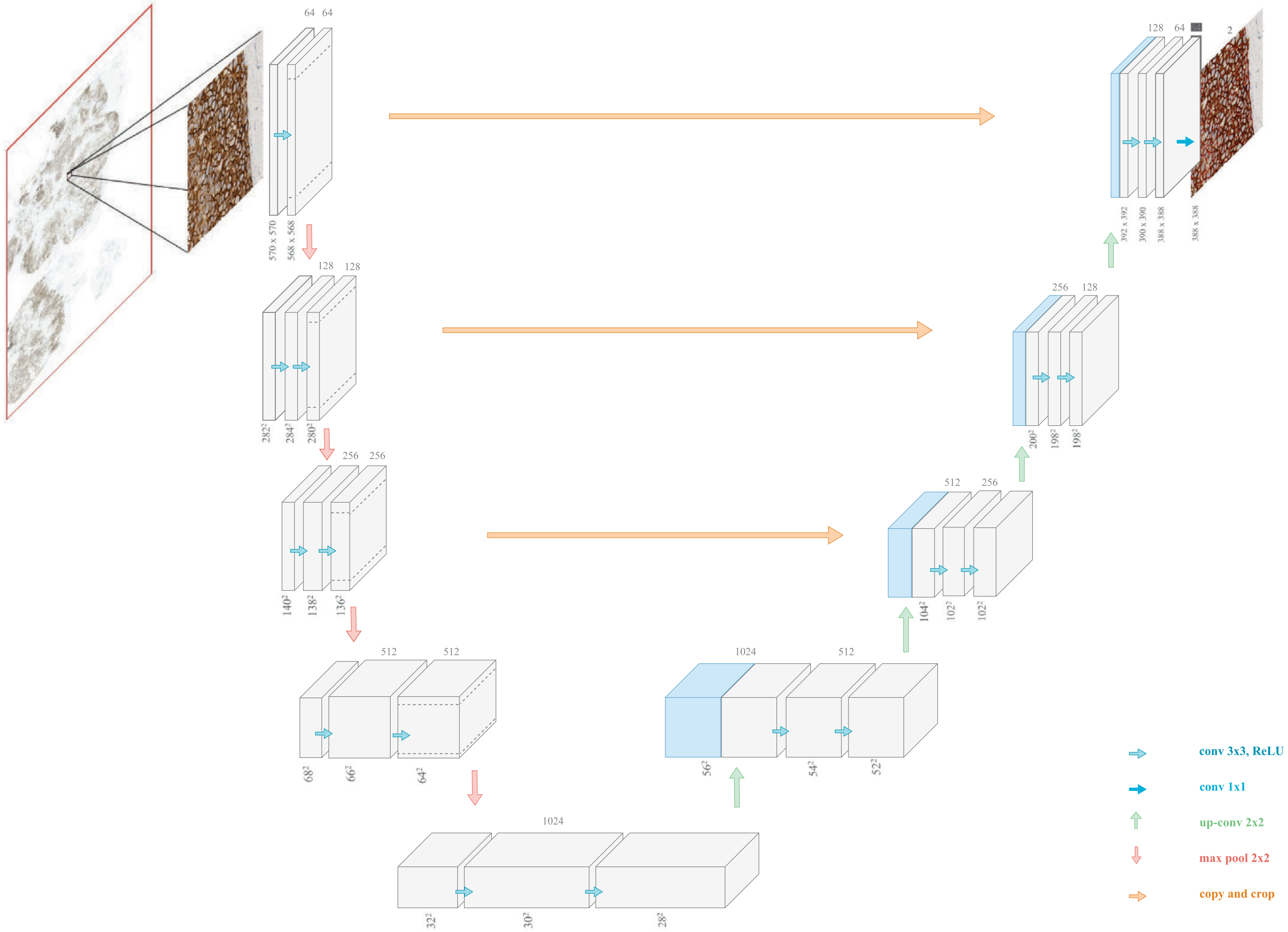

Fig. 2. Architecture of the U-Net model [34]. Combination of the information from the downsampling to upsampling path gives general knowledge about the localization and context. To segment cells with different membrane size, the dense layer is not used here.

$$\mathscr{L} = \mathscr{L}_{CE} + \lambda \mathscr{L}_{DC} \tag{3}$$

where λ represent the weighting coefficient which is set 1.1.

2.6. Dataset

The dataset consists of 127 WSIs of breast tumor patients which are gathered from Acıbadem hospital and Warwick competition. The dataset from Acıbadem hospital was acquired using a 3DHISTECH scanner while the dataset from the University of Warwick was scanned using a Hamamatsu NanoZoomer C9600. Out the total 127 WSIs, 79 were from Warwick dataset. The other 48 were acquired from Acıbadem Hospital for training. The size of the slides were about 150000×100000 pixels. The datasets were divided into two test and training parts. The dataset from Warwick is only used for testing. The size of each tile was 512×512 which was automatically given to modified U-Net architecture to get output results. The overall result was evaluated after

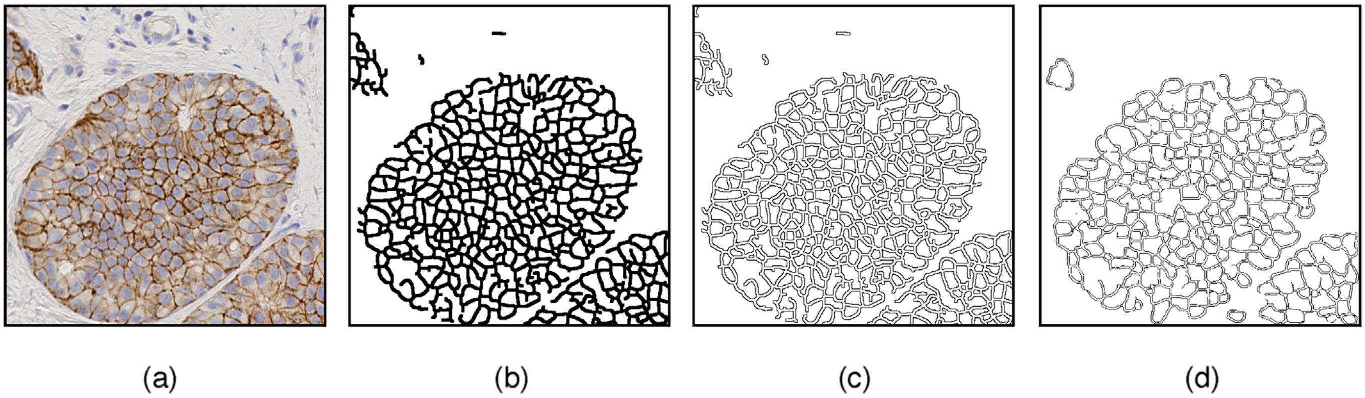

Fig. 3. (a) Original image; (b) Ground-truth map label; (c) Edge map label; (d) Prediction. Examples of the edge map label and the predicted output to evaluate DC loss.

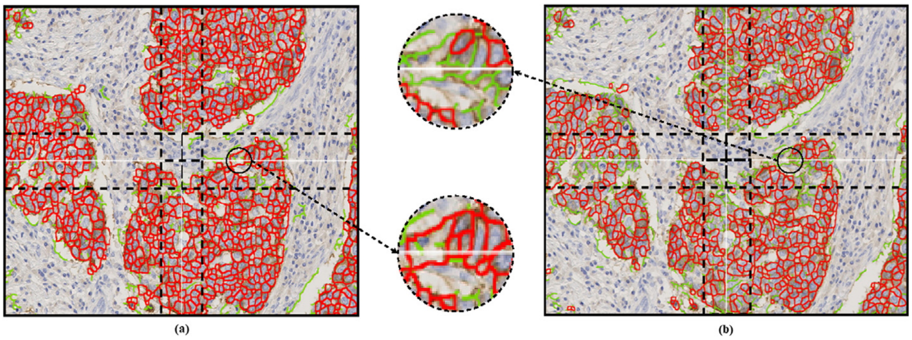

Fig. 4. Merging in edge parts; (a) after merging; (b) before merging. In order to find edge cell membranes, for each tile, we have to consider connecting and edge parts once more to combine the result of right, below, top, corner, and bottom tiles. To apply the merging procedure, only regions inside black dotted lines are sufficient.

merging tiles with corresponding neighbors (upper, left, right, bottom, and corner if available) [36]. For membrane in the edge parts, a similar approach is employed. This process is important as the membrane in the edge part would be counted as open membrane and the overall results would be wrong. The main idea in the merging process is to evaluate edge parts and copy parts that cross the edge parts as shown in Fig. 4.

In the training part, the batch size for training was set to 2 while the maximum iteration for training was assigned to 50 to minimize the training error. The modified U-Net model was fit and evaluated within the C++ programming environment using Dlib library [37]. To minimize the loss function, initial stochastic gradient descent of 0.01 with momentum 0.9 is considered.

3. Experiments and evaluation

3.1. Network architecture selection

To achieve higher accuracies and lower training losses we analyzed various CNN architectures. Different combinations of kernel sizes and convolutional layers are tested. For our experiments, we considered three models. In the first model, convolutional layers are used with filter sizes 3×3 and 3×3. In the second model, convolutional layers with 3×3 and 5×5 sizes are employed. In the third model, the convolutional kernels of size 5×5 are followed by 3×3. As shown in Table 2, an architecture with higher convolutional kernel size followed by a smaller kernel size leads to better results presumably because of the importance of a larger neighborhood in pathological images.

Fig. 5 shows the value of the loss function and accuracy on the training and validation sets of the proposed model.

The model converges in the training process only after 20 epochs. This architecture enables the employment of CNN models with superior accuracy for the scoring of IHC stained images to evaluate HER2 score. From the values of false positive rate and true positive rate of the ROC curve shown in Fig. 6, it can be calculated that the proposed model for cell segmentation provides a high AUC value (99%).

Considering the fact that in order to score HER2 IHC slides, we have to consider WSI and evaluate about 3000 tiles in a short time. The computing time of an average WSI is less than 500 s, which ensures the high speed of the proposed methodology. The performance of the proposed architecture for different training and validation splits based on the DC is evaluated in Table 3. In all evaluations, the same images from the test dataset are used. Furthermore, in Table 4, results of other models are compared with the proposed architecture which shows higher accuracy and lower loss. To have a better view of the results of the model, in Fig. 7, the predicted cell membranes are overlaid on the original images. The black polygons in Fig. 7 (b) depicts the classified epithelial areas. Because of high resolution of the WSI, the constituent region containing image tiles is shown in Fig. 7 (c). The results obtained in Fig. 7 (d) are by and large consistent with the expert annotation of cell membranes. Here, alleged cell membranes are in red color.

Table 2
Validation results after 20 epochs for various configurations of the evaluated CNN architectures for the segmentation of cell membranes.

Model	Precision	Recall	Accuracy	Loss	F1
1	0.9198	**0.9970**	0.9205	0.1751	0.9552
2	0.9245	0.9947	0.9258	0.1650	0.9569
3	**0.9283**	0.9922	**0.9482**	**0.1167**	**0.9580**

3.2. Confusion matrix of the trained model

The dataset and ground-truth provided by Acıbadem hospital are used for the training of the model. However, the performance of the proposed method is evaluated on two different datasets. For dataset provided by the University of Warwick, each training specimen is provided with fluorescence in situ hybridization (FISH) and HER2 IHC scores. As the Warwick dataset only provides a quantitative result of each slide, we used Acıbadem hospital dataset to train the segmentation model. In Fig. 8 (b) some masks annotated by experts are shown. After training the segmentation model, it is used to classify WSIs in the Warwick dataset. The quantitative comparison of the proposed method and the provided results from Warwick dataset are shown in Table 5. Here, Cohen's Kappa expresses the level of agreement and is defined as:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{4}$$

where p_o and p_e are the observed and expected agreements. In all of the 3 equivocal cases, the result from the proposed method matches with values from FISH in 2 + cases. In these 3 discordant cases, pathologists diagnosed equivocal case where they were scored as positive by automated image analysis solution. The main reason for these equivocal cases can be listed as the heterogeneity of staining and insufficient tissue blocks that are acquired to define the HER2 status. Correct classification of 2 + cases are the most challenging problem in HER2 IHC test and the promising results show the efficacy of the methodology. The HER2 status determined by the combination of the machine learning method and by pathologists confirmed the accuracy of the automated image analysis solution. Despite other methods that are scoring patches solely based on their staining intensity [40], the proposed method explicitly learns membrane intensity as a feature to distinguish between diverse tumor cells. The high agreement between the results of automated scoring using CNN and manual scores by pathologies represents the feasibility and reliability of the HER2 scoring approaches.

In Fig. 8, detected cell membrane and corresponding ground-truth of some sample tiles from Acıbadem dataset are illustrated. In Fig. 8, the first column shows six tiles corresponding scores, with the second

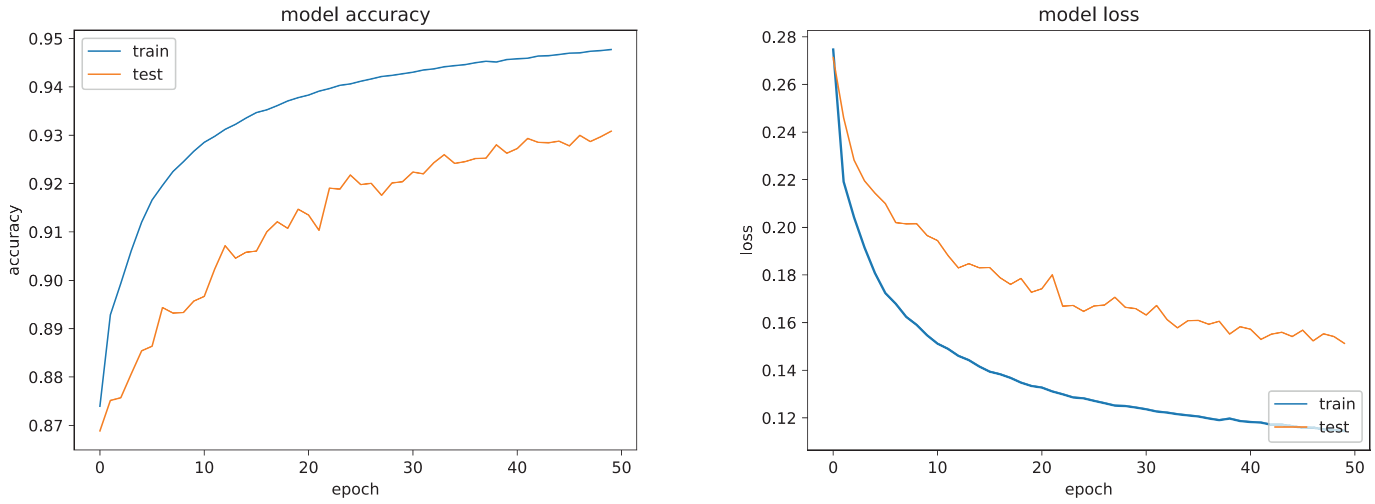

Fig. 5. Accuracy and learning curves for training and validation steps for the segmentation of cell membranes. The model is generalized well and it is not overfitted. The appropriate learning rate increases the accuracy curve of training is a sustained way.

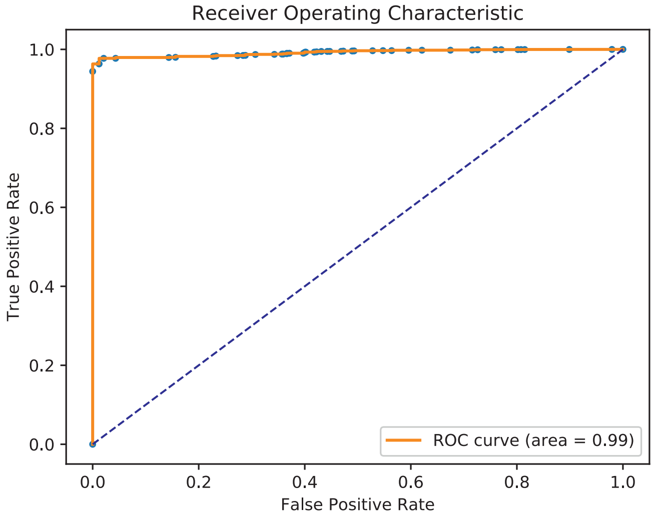

Fig. 6. The ROC curve and AUC give an estimation of the cell segmentation model's performance. The closer the curve to the upper left-hand corner and the larger the AUC value, the better segmentation of membrane from not membrane pixels.

column showing the ground-truth annotated by expert pathologists. The third column shows the binary images obtained by morphology operators. The outputs of the proposed model are shown in the last column. Here, results of the proposed model are consistent with the ground-truth and show similarity in closed cell membranes.

3.3. Quantitative comparison with comparative methods

To evaluate the performance of the method, statistical evaluation is essential. DC and accuracy are important metrics in pixel-wise segmentation tasks. By denoting the sum of elements in the predicted region (X) and ground truth region (Y) the DC is considered as:

$$\text{Dice coefficient} = \frac{2\cdot|X \cap Y|}{2\cdot|X \cap Y| + |Y\backslash X| + |X\backslash Y|} = \frac{2\cdot TP}{2\cdot TP + FN + FP} \qquad (5)$$

where accuracy has the form as following:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \qquad (6)$$

DC mean values for 1000 tiles with 512×512 pixels between pathologists and the proposed method are 0.98 with a standard deviation of 0.06 which indicates a good segmentation. The DC values range between 0 (not overlapped) and 1 (perfect overlapped).

In segmentation tasks, true positive (TP), false positive (FP), false negative (FN), and true negative (TN) are defined as the intersection between segmentation and ground-truth, segmented parts not covering the ground-truth, missed parts of the ground-truth, and parts of the image beyond the union segmentation plus ground-truth, respectively.

The results and details of the proposed and existing methods are provided in Table 6. Most of the methods are considering representative patches and ROI from WSI to train the classifier [31–33,40–44]. Qaiser et al. [45] trained a WSI-based model to classify the image patches. However, this method selects random ROI to test the score of the WSI and the patches to 0/1+, 2+, or 3 + classes. The proposed method allows pixel-wise prediction over an entire WSI which allows an accurate comprehension of the image. This process offers advantages over other methods as the scoring criteria are accepted by pathologies and the results provide significant information e.g., number of closed membranes, the ratio of closed membranes to all membranes, and intensity of membrane intensity for the WSI.

3.4. Comparative results of the Warwick contest

The test data for Warwick dataset is unlabeled and participants are acquired to provide HER2 score and confidence values for each WSI. The proposed model is compared with other methods in Table 7. The details about the evaluation criteria are available in Ref. [47]. The proposed method gets a high point in the classification of each class. The methods proposed in the contest are determining the HER2 score

Table 3
Comparison of the cell segmentation results from various combination of training and test dataset.

Training images (%)	Test images (%)	Train accuracy	Train loss	Validation accuracy	Validation loss
75	25	0.9541	0.1003	0.9308	0.1512
50	50	0.9396	0.1357	0.9187	0.1750
25	75	0.9132	0.1446	0.8797	0.1905

Table 4
Comparison of cell segmentation results from various models trained based on the default provided parameters.

	Train accuracy	Train loss	Validation accuracy	Validation loss	Time for training each epoch(s)
DenseNet [38]	0.8908	1.7403	0.8669	2.1226	198
SegNet [39]	0.8428	0.1583	0.8327	0.2076	**148**
Modified U-Net	**0.9541**	**0.1003**	**0.9308**	**0.1512**	488

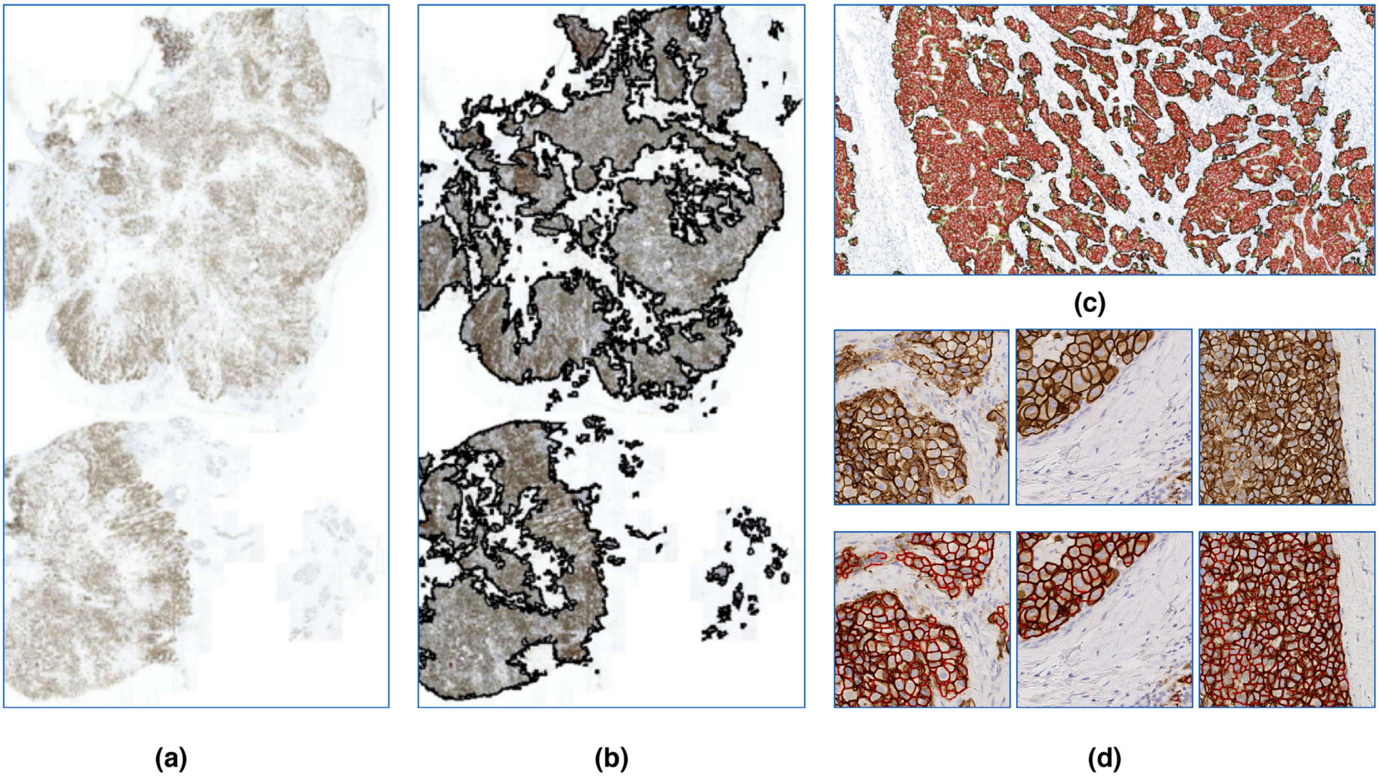

Fig. 7. (a) Whole slide images; (b) Segmented tissue; (c) Classified Epithelial areas; (d) Closer view of the input images and overlapped results from the model's prediction. Result images with different image characteristics from the test dataset. The membrane borders (red color) of each cell is enclosing a specific nucleus.

for a WSI by the percentage of each class in the patches extracted from the WSI. It should be considered that in the patch-based methods the intensity and completeness of each membrane are not considered intuitively and the HER2 score is not based on the connectivity or intensity of membranes. The method proposed by Team Indus considered directly calculated completeness for the combined score. However, this process is not automatic and to detect tumor areas from H&E slides human intervention is necessary. In the proposed methods no information about the membranes in the corners and edge is given. Besides, as the dataset doesn't contain label maps for membrane areas, the training images are different in our case. Our method automatically discards stromal areas and segments membranes to enable WSI-based HER2 score assessment based on the criteria accepted by the ASCO/CAP [1].

4. Discussion

In this paper, we introduced a computer-aided whole slide based deep learning method to automatically evaluate the IHC score of HER2 in breast cancer images. Because of the subjectivity of interpreting histopathology images and inter-observer disagreement between pathologists, reliable methods replicable of the manual annotation is necessary. In contrast to previous methods in HER2 assessment, in this research classification of HER2 IHC using deep learning-based segmentation of cell membrane in WSIs is evaluated. The analysis of results in Fig. 7 indicates that the network succeeds to distinguish between membranes and cytoplasmic or wrong staining. As test dataset is not used for the training, the noticeable high accuracy indicates that the model has learned well to segment cell membranes correctly. Despite other deep learning-based methods that score and classify slides based on the patches [32,33,45], we assessed the WSIs considering the guideline that is accepted by the pathologists [1]. The essential acquisition of information about the studied samples are summarized in Table 6. Most of these methods have considered manually extracted ROI to train and test the model. The proposed method is capable of WSI based HER2 scoring with high accuracy. Extraction of these ROIs is time-consuming and needs a good knowledge or would result in inconsistency in the results. From another perspective, cell membrane segmentation outperforms patch-based classification in explicitly considering membrane staining intensity by providing ground-truth cell membranes of each image. This helps to increase the performance of the model by learning precise features related to cell membranes. Traditional image processing methods [21,31,40,41,43,44] rely on color thresholding which is highly dependent on staining and intensity variations between slides and the values for each slide should be maintained manually to achieve acceptable results. The proposed DL based method outcomes this problem by training on multi-resolution slides. Experimental assessment in Table 7 with respect to the points for correct classification on the test WSI from Warwick contest demonstrates the efficacy of the proposed framework. This is an inevitable fact that the experts are not generally available for tedious ROI selection task and the performance of our proposed method would increase if we

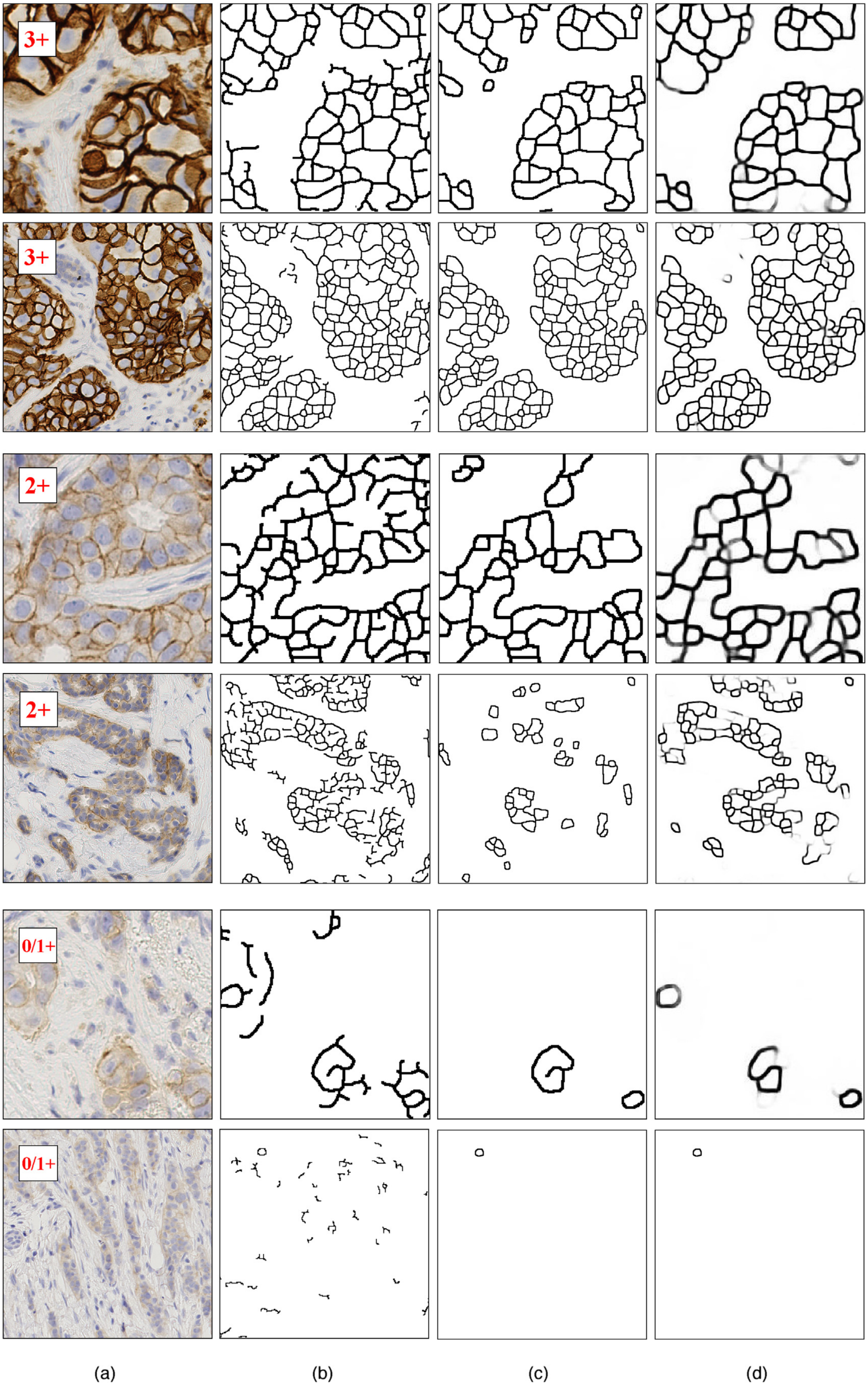

Fig. 8. (a) HER2/neu image fragments from different multi-resolution 40× and 20× regions and corresponding segmentation results of DL output compared to ground-truth. Here column (a) represents the original images from different slides with various scores. Column (b) illustrates the ground-truth cell membranes acquired by pathologies and confirmed with conventional image processing methods. In column (c) the uncompleted cell membrane is dissolved using morphology pruning operations. In column (d) detected cell membrane by the proposed method is shown.

Table 5
Confusion matrix for HER2 score compares the results of the deep learning-based classification method with provided HER2 scores from Warwick dataset (pathologist-based scores). These images are only considered to test the trained model.

			Actual	
Predicted		0/1+	2+	3+
	0/1+	23	3	0
	2+	0	10	3
	3+	0	1	12
	κ = 0.79			

consider the intervention of experts. By observing the results for HER2 assessment, the proposed modified U-Net architecture outperforms path-based CNN models and also hand-engineered classical image processing methods. The proposed HER2 assessment method exploits the segmented cell membranes to classify WSIs which presents a robust and meaningful way of combining HER2 scoring criteria with deep convolution features. A useful direction would be to assists pathologists by providing detailed results of WSIs. The plots in Fig. 5 are learning and loss curves generated for the given number of epochs for the proposed architecture. The smooth declining in training error and still incrementation in validation accuracy, indicates that the model is well-generalized. The example results in Fig. 8 depicts the original images with annotations showing cell membranes. The obtained results are showing high agreement with the expert annotation, which indicates the desirable behavior of the proposed CNN for HER2 assessment of IHC stained WSIs.

The proposed methodology, which implements superpixel-based tissue classification and deep learning-based cell membrane segmentation addresses automatic and computer-aided HER2 assessment tasks. The high agreement between automatic assessment and the manual scoring approves the generality and acceptedly of the training data. However, the greatest discordant in our evaluation was due to various staining criteria of different laboratories. This discrepancy was the main reason for misclassifying 2 + cases as 3 + or 0/1+. The increment of training data from various laboratories, as well as a better and accurate histopathology stain-color normalization are two important steps that could easily integrate to the existing segmentation workflow to overcome these problems. Severe overlapping of cytoplasmic staining with cell membrane gives rise to poor segmentation which causes errors in a way that some membrane staining connecting two cells are ignored. This could result in a state that inside a closed membrane more than a nucleus or no nucleus is recognized. This difficulty would directly affect the overall score in some cases, and we overcame this by simply considering cases with no distinct nuclei as artifacts.

Table 7
Comparison of the proposed method with some of the teams participated in the Warwick contest. The aim of this challenge is to assess HER2 status in breast cancer histology images.

Method	Pts [a]	Pts + B [b]	Cf [c]	W.Pts [d]
Qaiser et al. [45]	**405**	419	**24.1**	359.1
MTB NLP (AlexNet [46])	390	405.5	22.94	335.7
FSUJena ([48])	370	392	23	345
Team Indus (LeNet [49])	402.5	425	18.45	321.4
The proposed method	**405**	**428.5**	23.98	**359.7**

[a] Points.
[b] Points with bonus.
[c] Weighted confidence.
[d] Combined score for weighted points and points with bonus.

5. Conclusion

An automatic end-to-end machine learning-based framework for IHC HER2 assessment is presented in this paper. The proposed model considers three main properties: (1) the input WSI is classified into two stroma and epithelium areas using superpixel-based classification; (2) the model should be trained using patches extracted from the epithelial part, enabling the segmentation model to extract cell membrane staining pattern; (3) the scoring part of the architecture would merge the results from the tiles and transfer staining intensities and completeness to results accepted by pathologists. The HER2 scores of the model have a high correlation with the scores provided by experienced pathologists from two different and independent datasets. The generality of this methodology could be considered on other diverse membrane segmentation problems.

Conflict of interest statement

None declared.

Acknowledgment

The proposed method was implemented and tested on the Virasoft integrator system. The authors like to show their gratitude to the Virasoft company and collaborators for providing pathological images and helpful discussion. The authors would like to acknowledge Prof Ümit İnce of Acıbadem University, Istanbul, Turkey, for clinical advice. The authors are thankful to Prof N. Rajpoot, Dept. of Computer Science, University of Warwick, UK for providing the Warwick HER2 dataset for this study.

Table 6
Comparative results and details of the existing methods for the assessment of HER2 receptor status.

Method	Dataset	Base	Remarks
Saha et al. [33]	752 core images cropped from 79 WSIs	fully connected LSTM recurrent network, cell membrane and nuclei detection	0.9833% accuracy
Vandernberghe et al. [31]	ROI from 74 WSIs	color deconvolution, watershed segmentation, SVM, random forest, CNN, HER2 scoring by classifying cells	0.83% accuracy
Brügmann et al. [41]	835 core regions for training and 430 for validation extracted from 253 WSIs	connectivity-based scoring, pixel segmentation, skeletonizing	0.923% agreement
Pitkäaho et al. [32]	119 core regions selected from 81 WSIs	CNN, AlexNet [46] architecture, data augmentation, block-based scoring	0.9770% accuracy
Qaiser et al. [45]	86 WSIs	data augmentation, deep reinforcement learning, ROI-based score prediction	0.794% accuracy
Singh et al. [42]	1345 core regions from 52 WSIs	intensity and color features, Neural Network classifier	0.911% accuracy
Caroline et al. [40]	around 2580 patches from 86 WSIs	LBP and color features, KNN, MLP, decision trees	0.90% accuracy
Izzati et al. [43]	40 core regions	naïve bayes classifier, color deconvolution, color features	0.7513% accuracy for 2 + cases
Masmoudi et al. [44]	77 core regions from 77 WSI	color pixel classifier, epithelial nuclei segmentation, membrane staining assessment	0.72%–0.90% agreement
The proposed method	127 WSIs	modified U-Net, WSI segmentation, superpixel-based tissue classifier, WSI merging and scoring, automated WSI evaluation	0.9482% segmentation and 0.87% classification accuracy

References

[1] A.C. Wolff, M.E.H. Hammond, K.H. Allison, B.E. Harvey, P.B. Mangu, J.M. Bartlett, M. Bilous, I.O. Ellis, P. Fitzgibbons, W. Hanna, R.B. Jenkins, M.F. Press, P.A. Spears, G.H. Vance, G. Viale, L.M. McShane, M. Dowsett, Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/college of american pathologists clinical practice guideline focused update, J. Clin. Oncol. 36 (20) (2018) 2105–2122, https://doi.org/10.1200/JCO.2018.77.8738 pMID: 29846122. arXiv:https://doi.org/10.1200/JCO.2018.77.8738 https://doi.org/10.1200/JCO.2018.77.8738.

[2] S. Razavi, G. Hatipoğlu, H. Yalçın, Automatically diagnosing her2 amplification status for breast cancer patients using large fish images, Signal Processing and Communications Applications Conference (SIU), 2017 25th, IEEE, 2017, pp. 1–4.

[3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, et al., Slic superpixels compared to state-of-the-art superpixel methods, IEEE Trans. Pattern Anal. Mach. Intell. 34 (11) (2012) 2274–2282.

[4] J. Borovec, J. Švihlík, J. Kybic, D. Habart, Supervised and unsupervised segmentation using superpixels, model estimation, and graph cut, J. Electron. Imaging 26 (6) (2017) 061610.

[5] F. Bunyak, A. Hafiane, K. Palaniappan, Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets, Software Tools and Algorithms for Biological Systems, Springer, 2011, pp. 413–424.

[6] B.E. Bejnordi, G. Zuidhof, M. Balkenhol, M. Hermsen, P. Bult, B. van Ginneken, N. Karssemeijer, G. Litjens, J. van der Laak, Context-aware stacked convolutional neural networks for classification of breast carcinomas in whole-slide histopathology images, J. Med. Imaging 4 (4) (2017) 044504.

[7] A.D. Belsare, M.M. Mushrif, M.A. Pangarkar, N. Meshram, Classification of breast cancer histopathology images using texture feature analysis, TENCON 2015 - 2015 IEEE Region 10 Conference, 2015, pp. 1–5, , https://doi.org/10.1109/TENCON.2015.7372809.

[8] J. Xu, X. Luo, G. Wang, H. Gilmore, A. Madabhushi, A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images, Neurocomputing 191 (2016) 214–223.

[9] T. Qaiser, K. Sirinukunwattana, K. Nakane, Y.-W. Tsang, D. Epstein, N. Rajpoot, Persistent homology for fast tumor segmentation in whole slide histology images, Proc. Comp. Sci. 90 (2016) 119–124 20th Conference on Medical Image Understanding and Analysis (MIUA 2016) https://doi.org/10.1016/j.procs.2016.07.033 http://www.sciencedirect.com/science/article/pii/S1877050916312133.

[10] S. Akbar, L.B. Jordan, C.A. Purdie, A.M. Thompson, S.J. McKenna, Comparing computer-generated and pathologist-generated tumour segmentations for immunohistochemical scoring of breast tissue microarrays, Br. J. Canc. 113 (7) (2015) 1075.

[11] A.M. Khan, A.F. Mohammed, S.A. Al-Hajri, H.M.A. Shamari, U. Qidwai, I. Mujeeb, N.M. Rajpoot, A novel system for scoring of hormone receptors in breast cancer histopathology slides, 2nd Middle East Conference on Biomedical Engineering, 2014, pp. 155–158, , https://doi.org/10.1109/MECBME.2014.6783229.

[12] K.R. Choudhury, K.J. Yagle, P.E. Swanson, K.A. Krohn, J.G. Rajendran, A robust automated measure of average antibody staining in immunohistochemistry images, J. Histochem. Cytochem. 58 (2) (2010) 95–107.

[13] T. Markiewicz, P. Wisniewski, S. Osowski, J. Patera, W. Kozlowski, R. Koktysz, Comparative analysis of methods for accurate recognition of cells through nuclei staining of ki-67 in neuroblastoma and estrogen/progesterone status staining in breast cancer, Anal. Quant. Cytol. Histol. 31 (1) (2009) 49–62.

[14] S. Di Cataldo, E. Ficarra, A. Acquaviva, E. Macii, Automated segmentation of tissue images for computerized ihc analysis, Comput. Methods Progr. Biomed. 100 (1) (2010) 1–15.

[15] M. Babaie, S. Kalra, A. Sriram, C. Mitcheltree, S. Zhu, A. Khatami, S. Rahnamayan, H.R. Tizhoosh, Classification and retrieval of digital pathology scans: a new dataset, Cvmi Workshop@ Cvpr, 2017.

[16] H. Holten-Rossing, M.-L.M. Talman, M. Kristensson, B. Vainer, Optimizing her2 assessment in breast cancer: application of automated image analysis, Breast Canc. Res. Treat. 152 (2) (2015) 367–375.

[17] N. Fusco, E.G. Rocco, C. Del Conte, C. Pellegrini, G. Bulfamante, F. Di Nuovo, S. Romagnoli, S. Bosari, Her2 in gastric cancer: a digital image analysis in preneoplastic, primary and metastatic lesions, Mod. Pathol. 26 (6) (2013) 816.

[18] C.-Y. Chang, Y.-C. Huang, C.-C. Ko, Automatic analysis of her-2/neu immunohistochemistry in breast cancer, Innovations in Bio-Inspired Computing and Applications (IBICA), 2012 Third International Conference on, IEEE, 2012, pp. 297–300.

[19] E. Ficarra, S. Di Cataldo, A. Acquaviva, E. Macii, Automated segmentation of cells with ihc membrane staining, IEEE (Inst. Electr. Electron. Eng.) Trans. Biomed. Eng. 58 (5) (2011) 1421–1429.

[20] V.J. Tuominen, T.T. Tolonen, J. Isola, Immunomembrane: a publicly available web application for digital image analysis of her2 immunohistochemistry, Histopathology 60 (5) (2012) 758–767.

[21] R. Mukundan, Image features based on characteristic curves and local binary patterns for automated her2 scoring, J. Imaging 4 (2) (2018) 35.

[22] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, nature 521 (7553) (2015) 436.

[23] P. Kainz, M. Pfeiffer, M. Urschler, Segmentation and classification of colon glands with deep convolutional neural networks and total variation regularization, PeerJ 5 (2017) e3874.

[24] Y. Xu, Y. Li, M. Liu, Y. Wang, M. Lai, I. Eric, C. Chang, Gland instance segmentation by deep multichannel side supervision, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2016, pp. 496–504.

[25] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2017) 18–31.

[26] K. Sirinukunwattana, S.E.A. Raza, Y.-W. Tsang, D.R. Snead, I.A. Cree, N.M. Rajpoot, Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images, IEEE Trans. Med. Imaging 35 (5) (2016) 1196–1206.

[27] E. Hashimoto, M. Ishikawa, K. Shinoda, M. Hasegawa, H. Komagata, N. Kobayashi, N. Mochidome, Y. Oda, C. Iwamoto, K. Ohuchida, et al., Tissue classification of liver pathological tissue specimens image using spectral features, Medical Imaging 2017: Digital Pathology, vol. 10140, International Society for Optics and Photonics, 2017, p. 101400Z.

[28] H. Sharma, N. Zerbe, I. Klempert, O. Hellwich, P. Hufnagl, Deep convolutional neural networks for automatic classification of gastric carcinoma using whole slide images in digital histopathology, Comput. Med. Imag. Graph. 61 (2017) 2–13.

[29] M. Saha, C. Chakraborty, D. Racoceanu, Efficient deep learning model for mitosis detection using breast histopathology images, Comput. Med. Imag. Graph. 64 (2018) 29–40.

[30] Y. Zhou, H. Mao, Z. Yi, Cell mitosis detection using deep neural networks, Knowl. Based Syst. 137 (2017) 19–28.

[31] M.E. Vandenberghe, M.L. Scott, P.W. Scorer, M. Söderberg, D. Balcerzak, C. Barker, Relevance of deep learning to facilitate the diagnosis of her2 status in breast cancer, Sci. Rep. 7 (2017) 45938.

[32] T. Pitkäaho, T.M. Lehtimäki, J. McDonald, T.J. Naughton, Classifying her2 breast cancer cell samples using deep learning, Proc. Irish Mach. Vis. Image Process. Conf. 2016, pp. 1–104.

[33] M. Saha, C. Chakraborty, Her2net: a deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation, IEEE Trans. Image Process. 27 (5) (2018) 2189–2200.

[34] O. Ronneberger, P. Fischer, T. Brox, U-net, Convolutional networks for biomedical image segmentation, International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer, 2015, pp. 234–241.

[35] D. Eigen, R. Fergus, Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture, Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 2650–2658.

[36] S. Razavi, F.D. Khameneh, E.A. Serteli, S. Cayir, S.B. Cetin, G. Hatipoglu, S. Ayalti, M. Kamasak, An automated and accurate methodology to assess ki-67 labeling index of immunohistochemical staining images of breast cancer tissues, 2018 25th International Conference on Systems, Signals and Image Processing (IWSSIP), 2018, pp. 1–5, , https://doi.org/10.1109/IWSSIP.2018.8439184.

[37] D.E. King, Dlib-ml: a machine learning toolkit, J. Mach. Learn. Res. 10 (Jul) (2009) 1755–1758.

[38] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, CVPR 1 (2017) 3.

[39] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, arXiv preprint arXiv:1511.00561.

[40] C. Q. Cordeiro, S. O. Ioshii, J. H. Alves, L. F. Oliveira, An Automatic Patch-Based Approach for Her-2 Scoring in Immunohistochemical Breast Cancer Images Using Color Features, arXiv preprint arXiv:1805.05392.

[41] A. Brügmann, M. Eld, G. Lelkaitis, S. Nielsen, M. Grunkin, J.D. Hansen, N.T. Foged, M. Vyberg, Digital image analysis of membrane connectivity is a robust measure of her2 immunostains, Breast Canc. Res. Treat. 132 (1) (2012) 41–49.

[42] P. Singh, R. Mukundan, A robust her2 neural network classification algorithm using biomarker-specific feature descriptors, 2018 IEEE 20th International Workshop on Multimedia Signal Processing (MMSP), 2018, pp. 1–5.

[43] I. Muhimmah, D. Heksaputra, Color feature extraction of her2 score 2+ overexpression on breast cancer using image processing, MATEC Web of Conferences vol. 154, EDP Sciences, 201803016.

[44] H. Masmoudi, S.M. Hewitt, N. Petrick, K.J. Myers, M.A. Gavrielides, Automated quantitative assessment of her-2/neu immunohistochemical expression in breast cancer, IEEE Trans. Med. Imaging 28 (6) (2009) 916–925.

[45] T. Qaiser, N.M. Rajpoot, Learning where to See: A Novel Attention Model for Automated Immunohistochemical Scoring, arXiv e-prints (2019) arXiv:1903.10762arXiv:1903.10762.

[46] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

[47] T. Qaiser, A. Mukherjee, C. Reddy Pb, S.D. Munugoti, V. Tallam, T. Pitkäaho, T. Lehtimäki, T. Naughton, M. Berseth, A. Pedraza, et al., Her 2 challenge contest: a detailed assessment of automated her 2 scoring algorithms in whole slide images of breast cancer tissues, Histopathology 72 (2) (2018) 227–238.

[48] E. Rodner, M. Simon, J. Denzler, Deep bilinear features for her2 scoring in digital pathology, Curr. Dir. Biomed. Eng. 3. doi:10.1515/cdbme-2017-0171.

[49] Y. LeCun, et al., Lenet-5, Convolutional Neural Networks, URL: http://yann.lecun.com/exdb/lenet 20.