

Her2Net: A Deep Framework for Semantic Segmentation and Classification of Cell Membranes and Nuclei in Breast Cancer Evaluation

Monjoy Saha, *IEEE Member*, Chandan Chakraborty*, *IEEE Member*

Abstract—We present an efficient deep learning framework for identifying, segmenting and classifying cell membranes and nuclei from human epidermal growth factor receptor-2 (HER2)-stained breast cancer images with minimal user intervention. This is a long-standing issue for pathologists because the manual quantification of HER2 is error-prone, costly and time-consuming. Hence, we propose a deep learning-based HER2 deep neural network (*Her2Net*) to solve this issue. The convolutional and deconvolutional parts of the proposed *Her2Net* framework consisted mainly of multiple convolution layers, max-pooling layers, spatial pyramid pooling layers, deconvolution layers, up-sampling layers and trapezoidal long short-term memory (TLSTM). A fully connected layer and a softmax layer were also used for classification and error estimation. Finally, HER2 scores were calculated based on the classification results. The main contribution of our proposed *Her2Net* framework includes the implementation of TLSTM and a deep learning framework for cell membrane and nucleus detection, segmentation, and classification and HER2 scoring. Our proposed *Her2Net* achieved 96.64% precision, 96.79% recall, 96.71% F-score, 93.08% negative predictive value, 98.33% accuracy and a 6.84% false-positive rate. Our results demonstrate the high accuracy and wide applicability of the proposed *Her2Net* in the context of HER2 scoring for breast cancer evaluation.

Index Terms—Breast cancer, HER2, LSTM, deep learning, cell membrane, nuclei

I. INTRODUCTION

Breast cancer (BC) is very common (the second-most prevalent cancer after lung cancer) among females globally [1]–[4]. Approximately 10% of women in the United States and the European Union are diagnosed with BC during at some point in their lifetimes [5]. BC often affects women aged 20–59 years [6]. In 2012, BC was estimated to affect 1.7 million women, and it caused nearly 521,900 deaths [7]. According to U.S. breast cancer statistics (2017), approximately 252,710 invasive and 63,410 non-invasive new BC cases have yet to be diagnosed [8], [9]. BC screening and grading are mostly performed through visual inspection, which is susceptible to subjectivity, excess time-consumption and ambiguities.

BC grading is mostly done following the Nottingham grading system [10]. Immunohistochemical (IHC) study is always recommended by clinicians and pathologists to measure the severity of BC. IHC screening analyses proteins in the cells and on cell surfaces in biopsied tissue sections [11] to determine the molecular phenotype and histologic subtype

M. Saha and C. Chakraborty are with School of Medical Science and Technology, Indian Institute of Technology, Kharagpur, West Bengal, India.
*e-mail:chandanc@smst.iitkgp.ernet.in

of a tissue section. The IHC diagnostic markers, such as oestrogen receptor, progesterone receptor, Ki-67, and human epidermal growth factor receptor 2 (HER2), are widely used to identify metastatic and benign tumours, to grade tumours and to determine the origin of cancerous tissue.

One of the most important IHC examinations for BC evaluation is the assessment of HER2. HER2 is a membrane-bound receptor that regulates cell proliferation and cell growth [12]. Approximately 20–25% of BCs are HER2 positive [13], [12]. The overexpression of HER2 oncprotein in malignant BC cell membranes is associated with poor survival and advanced tumour stage [6], [14], [15]. In clinical practice, the HER2 status in BC is mostly evaluated by gene amplification and IHC. Gene amplification is mainly done by using fluorescence in-situ hybridization (FISH), whereas IHC is basically an immunohistochemical imaging technique [16].

In routine clinical practice, pathologists visually analyse the biopsy tissue slides under the microscope. Such visual inspection is often subjective and error-prone. Eventually, in resource-limited areas where expert pathologists are not always available quickly, HER2 assessment becomes a challenging task [17]. To address both these issues, digitization and quantitative image analysis of IHC slides have become essential for image preservation and reproducible diagnosis. Thus, pathologists are mostly performing quantitative image analysis using computer-aided digital techniques. The main reasons for adopting digital pathology are its reproducibility and simplicity. Additionally, the digital imaging technology is a pixel-based technology, which decreases the inter-observer variability and false positivity by improving detection, segmentation accuracy, and other factors. The U.S. Food and Drug Administration (FDA) and the World Health Organization (WHO) have placed more emphasis on computer-assisted HER2 scores [18]. As per the American Society of Clinical Oncology, HER2 scoring is based on the HER2 staining expression in a cell membrane and a population of nuclei. The details of the HER2 scoring calculation are shown in table I [18]–[21].

A few machine learning algorithms have been developed using conventional techniques for the assessment of HER2-stained images [22], but their success rate and accuracy are not up to the mark. Henceforth, there is an urgent need to develop an advanced machine learning technique to solve the current issues related to HER2 image assessment. Thus, we suggest the use of an efficient deep learning framework for cell membrane and nucleus detection, segmentation, classification and scoring using HER2-stained IHC images. The main advantages of using

Table I: HER2 SCORING SYSTEM

Score	Expression	Membrane staining
0	Negative	cell membrane stain is NIL in >10% of cells
1+	Negative	cell membrane stain is faint in <10% of cells
2+	Equivocal or, weakly positive	cell membrane stain is moderate/weak staining in <10% of cells
3+	Strongly positive	cell membrane stain is complete and strong in <10% of cells

deep learning for medical imaging are its capability of handling and maintaining best-in-class performance on a particular task which significantly outperforms other solutions in multiple domains. For HER2 scoring, none or unfinished work has been done based on deep learning.

The manuscript is organized as an introduction, literature review, experimental setup, results, discussion and conclusion. Figure 1 illustrates HER2-stained IHC images of BC. Figure 2 (a-d) shows the HER2-stained BC images according to their score, and their corresponding colour distribution map in RGB colour space is shown in figure 2 (e-h). In figure 2 (e-h), it is noticeable that the ranges of different shades of brown and blue pixels are different. These ranges vary greatly across different grades of BC. The blue pixels, representing normal nuclei, are only visible when the HER2 score is negative. The brown pixels, representing malignant nuclei, range from score 1+ to 3+, where 1+ and 3+ indicate low and high aggressiveness, respectively.

II. LITERATURE REVIEW

The frequently acquired microscopic images of HER2-stained slides are found to be complex due to improper staining, overlapped nuclei, non-homogeneous backgrounds, and other factors. A number of conventional approaches have already been offered to solve these issues. Here, we summarize the overall HER2 scoring techniques that have been published by various interdisciplinary biomedical research groups. Table II shows the summary of different HER2 scoring methods used so far.

Yaziji et al. (2004) suggested a parallel tissue-based technique for determining the gene status using FISH and IHC images [13]. The authors reported that the performance of IHC was better than that of FISH in terms of sensitivity, specificity, and positive and negative predictive values. The FISH test failed to provide augmented molecular expression level due to a higher false-positive rate, longer interpretation, more testing time and high reagent cost. Hence, the authors proposed IHC image quantification technique for improved HER2 scoring. Lloyd et al. (2010) focused on the reliability-based imaging method for the prognostic and predictive evaluation of BC [32]. The authors used 33 whole-slide images (WSIs) (10 oestrogen receptor and 23 HER2 WSIs) for their analysis. The authors suggested the use of commercially available software, i.e., Leica Aperio Image Analyzer and Quantification Cellular Analyzer, for cell membrane and nucleus evaluation. The authors of

[33] employed IQFISH pharmDx assay software for the assessment of HER2 amplification. In [34], the authors reported a semi-quantitative method for the detection and classification of cell membranes based on morphological features. The authors of [35]–[37] compared computerized analysis and manual evaluation for HER2 quantification and its scoring. The authors recommended computerized imaging and machine learning techniques as reliable substitutes for the manual technique. Raimondo et al. (2005) focused on the automatic screening of HER2 status from FISH breast tissue images [38]. The authors suggested the use of top-hat transformation, cross-correlation, gradient evaluation, brightness balancing, distance transformation, marker-controlled watershed and spot index calculation for cell membrane and cell detection. The authors achieved very low accuracy (90%) in comparison with the state of the art. Tuominen et al. (2012) developed ImmunoMembrane, a web application for evaluating HER2 score using BC IHC images [26]. The software includes colour deconvolution, blank field correction, stain separation, and cell membrane segmentation and classification. The authors developed a web application that uses the conventional machine learning techniques and ImageJ. The authors of [39] proposed METASYSTEMS software for automated HER2 scoring using FISH images. The authors achieved an overall 92.8% classification accuracy. In [40], a membrane isolation algorithm (MIA) consisting of colour decomposition, Otsu's thresholding, Gaussian filtering, feature extraction and scoring was proposed. In [41], an active learning technique-based deep network for membrane detection was proposed. The authors used 301×301 single plane illumination microscopy image patches for their analysis. The method was developed using conventional imaging techniques and convolution neural networks. In [42], the authors proposed lymphocyte infiltration detection and grading using HER2-positive histology, and 90% classification accuracy was achieved. In [43], the authors suggested the use of colour deconvolution, thresholding, maxima finding, and other protocols for HER2 image analysis. One of the most recently published research articles on HER2 scoring using conventional and deep learning was by [18]. The authors used segmentation, feature extraction and classification techniques for cell/nucleus detection. Next, the detected nuclei were cropped into 44×44 image patches to feed into the convolutional neural network (ConvNets). Their proposed ConvNets consists of only three convolution layers followed by ReLu, dropout and fully connected layers, and 78% accuracy was achieved. The drawbacks of the ConvNet based technique are that (a) the method very often leads to lower accuracy because it does not consider a full deep learning model; and (b) only the cell population is considered for HER2 scoring, whereas cell membrane staining expression is used in the international guidelines for HER2-positive image scoring and grading. The classification of HER2 images using deep learning has been discussed in [44], where image patches were fed directly to the deep convolution network for classification of scoring. They achieved 97.7% testing accuracy. In comparison with ConvNets, our proposed *Her2Net* possess a much deeper network along with TLSTM.

From an image-processing perspective, the authors of [45]

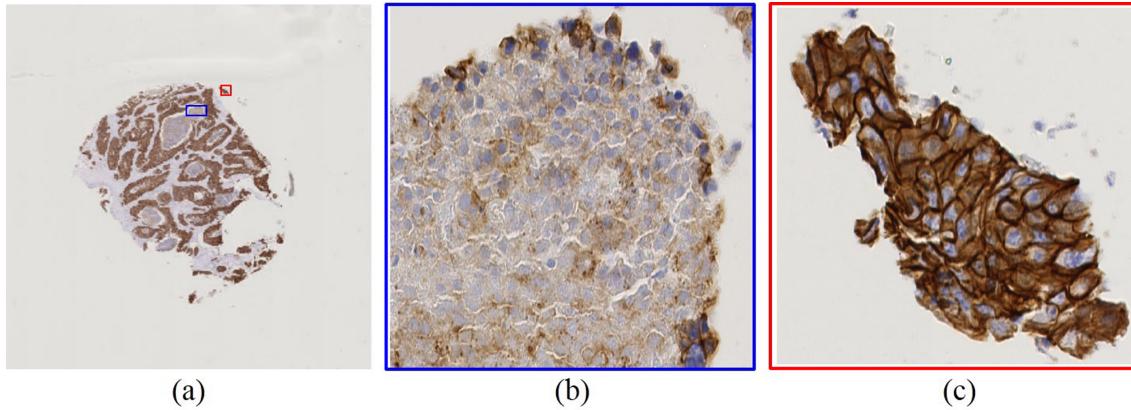


Figure 1: HER2-stained IHC image of breast cancer. (a) Whole-slide image; (b) low/moderate HER2 expression (blue rectangle); (c) strong HER2 expression (red rectangle)

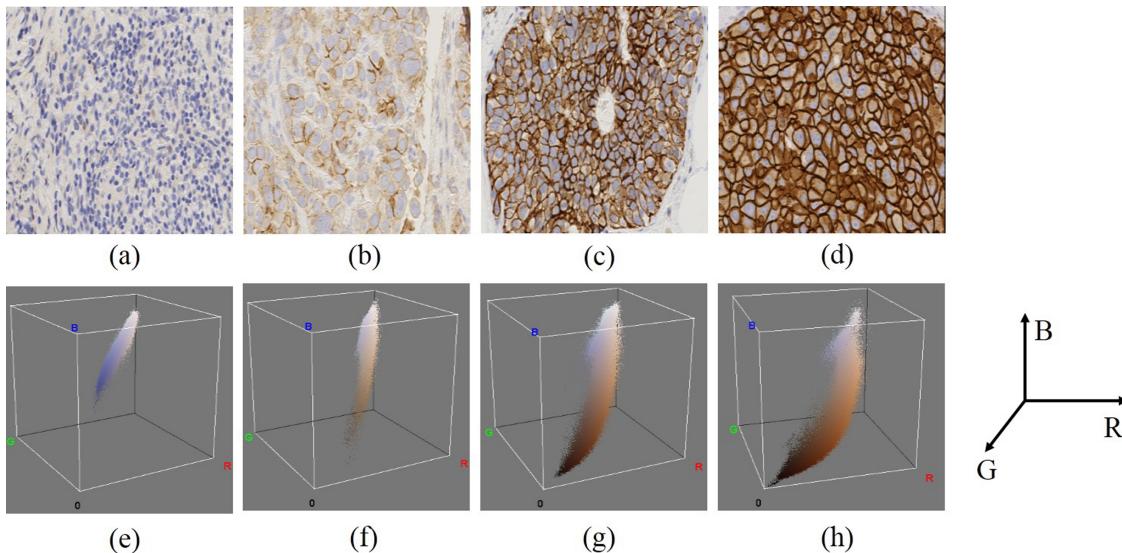


Figure 2: HER2 scoring by the pathologists with respect to differential colour distribution in RGB colour space: four input HER2-stained images (size 2048×2048) of breast cancer at $40\times$ with the scores (a) = 0; (b) = 1+; (c) = 2+; (d) 3+ and colour spectrum visualization of the input images (e-h) [axes R, G and B denote red, green and blue channels, respectively]

contributed significantly to restoring images using the edge-guided interpolation technique. Their method consists of directional interpolation, multi-resolution transforms and edge-guided interpolation, which is capable of restoring missing blocks and edges in an image. The authors obtained improved image quality of the reconstructed images. In [46], multiplicative noise removal and image optimization algorithms were proposed using a sparse analysis model. Their experimental results showed an improved performance of the proposed method.

The main contributions of our proposed *Her2Net* framework are listed below:

- The proposed *Her2Net* is a fully deep learning-based technique.
- The concept of developing a trapezoidal LSTM connection topology (TLSTM) structure is novel and unique. The implementation of TLSTM improves the performance of the proposed method.

- The proposed *Her2Net* introduces a new way of deep learning-based semantic segmentation of cell membrane and nucleus detection, segmentation and scoring.
- It is a value addition in terms of the primary quantification in the existing techniques for HER2 scoring.

III. EXPERIMENTAL SETUP

A. Dataset

The proposed method was assessed based on the online HER2 image database of the Department of Computer Science, University of Warwick, United Kingdom [47], [48]. The authors thank Prof Nasir Rajpoot, Department of Computer Science, University of Warwick, for providing the necessary approval for downloading the dataset and using it in this study. The dataset mainly consisted of 158 WSIs, which were acquired using a Hamamatsu NanoZoomer C9600 scanner. Out of 158 WSIs, 79 were stained using haematoxylin & eosin (H&E), and the

Table II: SUMMARY OF HER2 SCORING APPROACHES

Category	Year	Image Type	Method Used	Remarks	Reference	
Conventional Techniques	2012	IHC	Colour channel extraction; feature extraction (e.g., morphology, texture, intensity, etc.); support vector machine (SVM) classifier	90.69% accuracy	[23]	
			Semi-quantitative method	95% concordance rate	[24]	
			HER2-CONNECTTM software	99.2% sensitivity; 100% specificity	[25]	
			ImmunoMembrane software	0.80 Kappa value	[26]	
	2011		Delaunay triangulation; Voronoi diagram	0.98 correlation coefficient; 0.14 average error	[22]	
			Cell membrane segmentation; watershed for nucleus segmentation	90% accuracy	[27]	
			Histogram; morphological operations; adaptive thresholding;	99% classification accuracy	[28]	
	2010		Semi-quantitative method	78% sensitivity; 88% specificity	[29]	
			Colour pixel classifier; nucleus segmentation; membrane staining assessment; classification	90% accuracy	[30]	
	2005	IHC and FISH	Semi-quantitative method	63%, 19% and 18% classification accuracy for score 1, 2+ and 3+ respectively	[31]	
Deep Learning	2017	IHC	Convolution network	83% concordance rate	[18]	

other 79 were stained using HER2 monoclonal antibody. The size of each WSI was $100,000 \times 80,000$ pixels (width×height). The WSIs could be viewed at $4\times$ to $40\times$ magnification.

In our study, we used only the 79 HER2 monoclonal antibody-stained WSIs to divide the dataset into two subsections, i.e., training (51 WSIs) and testing (28 WSIs). A total of 752 images (188 for each score, i.e., 0, 1+, 2+, 3+) at $40\times$ magnification were cropped (using MATLAB *imcrop()* function) from the 79 WSIs. The size of the each cropped image was kept as 2048×2048 pixels (width×height). Noise during image grabbing was automatically removed by the Hamamatsu NanoZoomer image-grabbing software. The rest of the biological and acquisition-related noise was removed using morphological operations and an adaptive filter, respectively.

B. Patch Selection

In this section, the HER2 image patches of size 251×251 were cropped from images of size 2048×2048 pixels based on the sequential window movement technique. In this technique, the window will move from left to right and up and down over the images. We considered zero or very negligible overlapping regions in our image patches. Around the boundary areas, where the image size did not match the patch size, patches were not included in our experiment.

C. Convolutional Layer

The convolution layer was used to convolve the image patches with the kernels. Let a bank of kernel/filter is denoted by [49], [50]:

$$F^n = (F_1^n, F_2^n, F_3^n, \dots F_{e_h^n}^n) \quad (1)$$

Here, e_h^n indicates the number of kernels, and F_k^n , $k \in \{1, 2, 3, \dots, e_h^n\}$ is a linear filter of size $m \times m$. An input of size $\omega \times \omega$ patch I_p^{n-1} is convolved with a $m \times m$ local receptive

field in image I_p^{n-1} surrounded by the filter F_k^n . The filter F_k^n moves over the input patch I_p to perform the local convolution operation. The convolution operation produces e_h^n different feature maps. The output of the convolution operation on I_p^{n-1} by F_k^n can be written as $Conv_{out} = F_k^n * I_p^{n-1}$. Here, the resultant image's pixel value is the sum of products of the image pixels of local receptive fields and filter coefficients. The convolution operation is done across $(\omega - m + 1) \times (\omega - m + 1)$ pixels. Now the resultant image $Conv_{out}$ size is $(\omega - m + 1) \times (\omega - m + 1)$. The $Conv_{out}$ is also called the activation function. The output of $Conv_{out}$ is calculated as follows:

$$f(Conv_{out}) = \frac{1}{1 + \exp(-Conv_{out})} \quad (2)$$

D. Pooling Layers

The principle of pooling layers reduces the feature map dimensionality for computational efficiency. Hence, pooling layers are often called down-sampling layers [51]. The max-pooling and spatial pyramid pooling layers are the most used down-sampling layers. Down-sampling layers make features transition-independent. Three types of pooling operations were used in our proposed framework, i.e., max-pooling, spatial pyramid pooling and up-sampling/up-pooling. The max-pooling layer was used to perform spatial invariance using feature map resolution reduction. The max operation of max-pooling layer helped in aggregating the features from the spatial regions. The spatial pyramid sampling/pooling was used to divide the image patches into areas to discern local information of the image. In this pooling operation, the partition was considered in itself, and its features were combined in some way, first for each partition and then for the combination of those partitions as a global representation of the image [50], [52], [52].

The properties of up-sampling layers are completely opposite to down-sampling layers [53]. It is a reverse operation of max-pooling layers. Up-sampling of layers was used to enlarge the

input image and densify sparse activations. In the deconvolution layer, the up-sampling helped in reconstructing the original image and preserved the structure of a stimulus.

E. Deconvolution Layer

In our proposed *Her2Net*, a deconvolution layer was used after an up-sampling layer. This layer densified the sparse activations through multiple learned filters and convolution-like operations [54]. Unlike a convolution layer, the deconvolution layer connected single-input activation with multiple outputs. The deconvolution layer gave a dense and enlarged action map of an input image. The learned filters in this layer were mainly responsible for reconstructing the shape of an input image. A hierarchical structure of deconvolution layers was formed to save the shape information of the input image. The lower-layer filters were responsible for capturing shape information. Through this procedure, deconvolution layers assisted in semantic segmentation [55].

F. Fully Connected (FC) Layer

The FC layer was used as a final layer or classification layer. In our *Her2Net* framework, the FC layer's height and width of each blob was assigned to 1.

G. Dropout Layer

Dropout is a regularization technique which increases the accuracy of a model by reducing overfitting and preventing complex co-adaption on training data. The dropout neurons do not contribute in the forward pass or back-propagation.

H. Softmax Layer

The softmax layer was used to reduce cross-entropy in a multi-class problem.

I. Long Short-Term Memory (LSTM)

LSTM, a recurrent network, computes a new hidden state where new input and previously hidden states are given. Unlike other recurrent neural networks (RNNs), LSTM does not suffer from vanishing gradients or the phenomena of exploding [56]. LSTM is widely used in speech recognition and handwriting recognition [57], [58]. The goal of using LSTM in *Her2Net* was to employ pixel details from the multiple frames in making the semantic segmentation and classification cases. The memory cells of LSTM were used for maintaining long-range relationships with the rest of the network. Our single-cell LSTM connection is shown in figure 3. The one-cell LSTM consisted of input, memory cell, output and the three gates (i.e., input, forget and output gates). These gates used logistic functions for computing the activation of the LSTM. In our method, these gates were considered as conventional artificial neurons. Each gate possessed its own weight and bias values, which were basically the output of the previous layers outside LSTM. The input gate controlled a mechanism on which a value flowed into a memory. The forget gate was used to keep a value in memory. Finally, the output gate was used to control the activation of LSTM based on the total parameters.

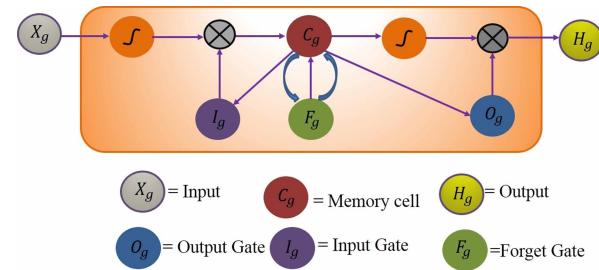


Figure 3: One cell of the LSTM memory block

The activations and gates were calculated as follows, where $g=1, 2, 3, \dots$ [59]:

$$I_g = \xi(W_{XI}X_g + W_{HI}H_{g-1} + W_{CI}C_{G-1} + B_I) \quad (3)$$

$$F_g = \xi(W_{XF}X_g + W_{HF}H_{g-1} + W_{CF}C_{G-1} + B_F) \quad (4)$$

$$C_g = F_g C_{g-1} + I_g \tanh(W_{XC}X_g + W_{HC}H_{g-1} + B_C) \quad (5)$$

$$O_g = \xi(W_{XO}X_g + W_{HO}H_{g-1} + W_{CO}C_g + B_O) \quad (6)$$

$$H_g = O_g \tanh C_g \quad (7)$$

where W , ξ and \tanh indicate the weight, activation function and non-linear function, respectively.

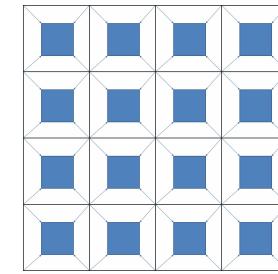


Figure 4: Trapezoidal LSTM connection

1) *TLSTM Connection Topology*: As shown in figure 4, TLSTM recursively accessed the pixel context from the four-corner axis by 45° and formed a shape like a trapezoid. The TLSTM grids were required to process the image pixel information from the four-corner axis. The TLSTM convolution was formed using four LSTMs (one memory cell) as described in figure 3. A single LSTM processed an entire volume in two directions. The overall workflow diagram of the proposed method is shown in figure 5. The proposed framework consisted of two main parts: (a) convolution and (b) deconvolution. The convolution part consisted of sixteen convolution layers, three max-pooling layers, two spatial pyramid pooling layers and one TLSTM. The deconvolution part included sixteen deconvolution layers, three up-sampling layers and one TLSTM. The two TLSTMs were used in the transition phase, i.e., at the end of the convolution part and the beginning of the deconvolution part. These TLSTMs helped in preserving cellular and textural structures to restore the pixel distortion. The TLSTM also made the training process much faster and more accurate. Two other layers, i.e., a fully connected layer and a softmax layer, were used mainly for classification and loss estimation. An image patch size of 251×251 was fed into the deep framework, and the output image was the same size.

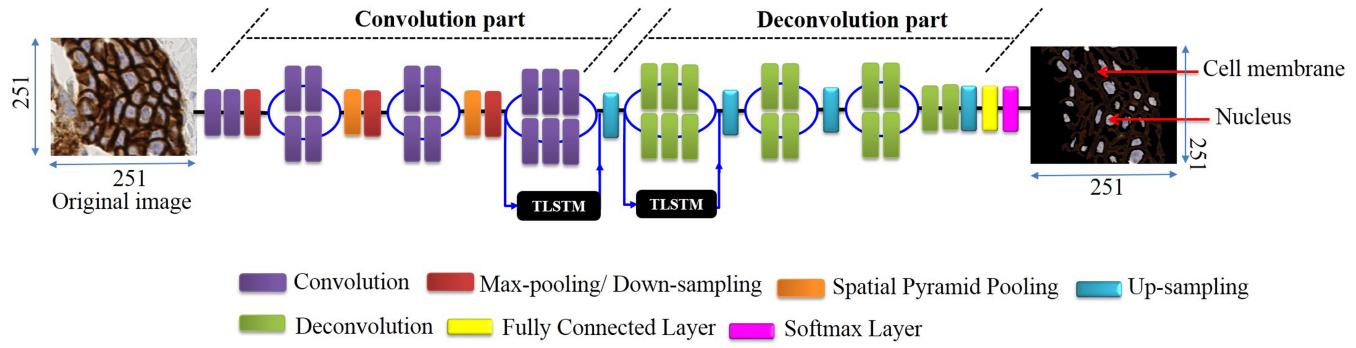


Figure 5: Workflow diagram of the proposed method

J. Parameter Initialization

In our proposed method, a patch (RGB image) of size $\omega \times \omega = 251 \times 251 = 63001$ pixels was used. The training and validation batch sizes were set to 128. The testing interval and maximum iteration were assigned to 4500 and 500,000, respectively. The other significant parameters were the learning rate ($= 0.01$), weight decay ($= 0.046$) and momentum ($= 0.5$).

K. Training Her2Net

We computed a weight and batch normalization layer across the training dataset during validation. As per our observations, these normalizations provided a better result. We trained the whole model using the end-to-end stochastic gradient descent method.

IV. EXPERIMENTAL RESULTS

We tested the performance of the proposed *Her2Net* deep learning model on the testing data of the HER2 scoring contest at the University of Warwick [47]. HER2 scoring was done based on the formula described in [15], [19], [20], [60]. The segmentation results were validated by the pathologists' manual annotations. Additionally, the *qualitative* and *quantitative* validation measures were performed to measure the efficiency and efficacy of the proposed *Her2Net* framework.

A. Qualitative Results

The qualitative results of *Her2Net* on IHC images are shown in figure 6. In figure 6(b), red and green regions represent cell membranes and nuclei, respectively. The black regions in figure 6(b) were detected as background. Figure 6(c) shows post-processed images of (b) by retaining the original colour of (a). Figure 6(d) illustrates the detected cell membrane and nucleus contour in (a). The ground truth and the corresponding *Her2Net* output are shown in figure 7.

B. Quantitative Results

The quantitative results were also essential for the statistical evaluation of our proposed framework. The quantitative results are shown in table III and IV. Table III mainly deals with the performance measures using various combinations of the training and testing datasets. From table III, it is clear that

the proposed deep learning network provided consistent and efficient results, as evident from the similar performances. As the training image size was easily large enough, there was a higher chance of having more or less similar performances, though the combinations were different. Table IV shows the validation technique based on pixel-wise distance metrics, e.g., Dice's coefficient (DC), correlation coefficients (CC) and segmentation accuracy (SA) [61]. *Her2Net* yielded an accurate result in terms of precision (Pre), recall (Rec), F-score, negative predictive value (NPV) and accuracy. ROC curves of *Her2Net*, SegNet, Bayesian SegNet and U-Net on the same dataset are shown in figure 8(a). *Her2Net* provided an area under the curve (AUC) of 94%, AUC of SegNet of 66%, AUC of Bayesian SegNet of 73% and U-Net as 88%. From these results, we concluded that *Her2Net* provided better results than U-Net, SegNet and Bayesian SegNet. The equation of evaluation measures are shown below [50], [62]:

$$\text{Pre (\%)} = \frac{\text{TPOS}}{\text{TPOS} + \text{FPOS}} \times 100 \quad (8)$$

$$\text{Rec (\%)} = \frac{\text{TPOS}}{\text{TPOS} + \text{FNEG}} \times 100 \quad (9)$$

$$\text{F-score (\%)} = 2 \times \left(\frac{\text{Pre} \times \text{Rec}}{\text{Pre} + \text{Rec}} \right) \times 100 \quad (10)$$

$$\text{NPV (\%)} = \frac{\text{TNEG}}{\text{TNEG} + \text{FNEG}} \times 100 \quad (11)$$

$$\text{Accuracy (\%)} = \frac{\text{TPOS} + \text{TNEG}}{(\text{TPOS} + \text{TNEG} + \text{FPOS} + \text{FNEG})} \times 100 \quad (12)$$

$$DC(SI, GTI) = \frac{|SI \cap GTI|}{|SI| + |GTI|} \quad (13)$$

$$SA = [1 - \frac{\text{Total no. of misclassified pixels}}{\text{Total no. of pixels in the segmented regions}}] \quad (14)$$

$$CC(SI, GTI) = \frac{\sum_{i=1}^M \sum_{j=1}^N GT_{new} \times SI_{new}}{\sqrt{\left(\sum_{i=1}^M \sum_{j=1}^N (GT_{new})^2 \right) \times \left(\sum_{i=1}^M \sum_{j=1}^N (SI_{new})^2 \right)}} \quad (15)$$

where $(GT_{i,j} - \mu GTI) = GT_{new}$ and $(SI_{i,j} - \mu SI) = SI_{new}$. True positive, false positive, true negative and false negative

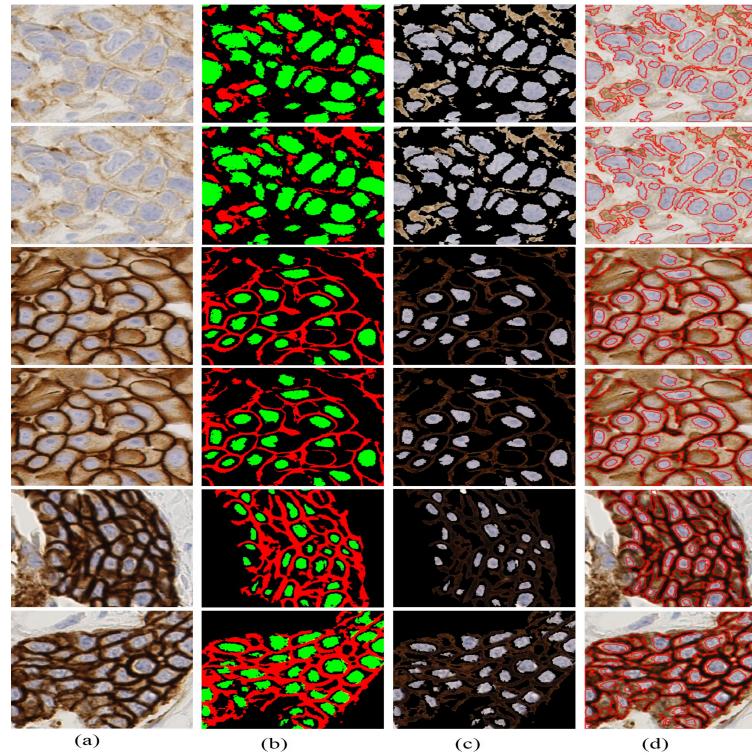


Figure 6: Breast cancer HER2-stained IHC image patches at 40 \times : (a) original image; (b) image segmented by *Her2Net* framework; (c) post-processed image of (b) by retaining the original colour of (a); (d) detected cell membrane and nuclei in (a)

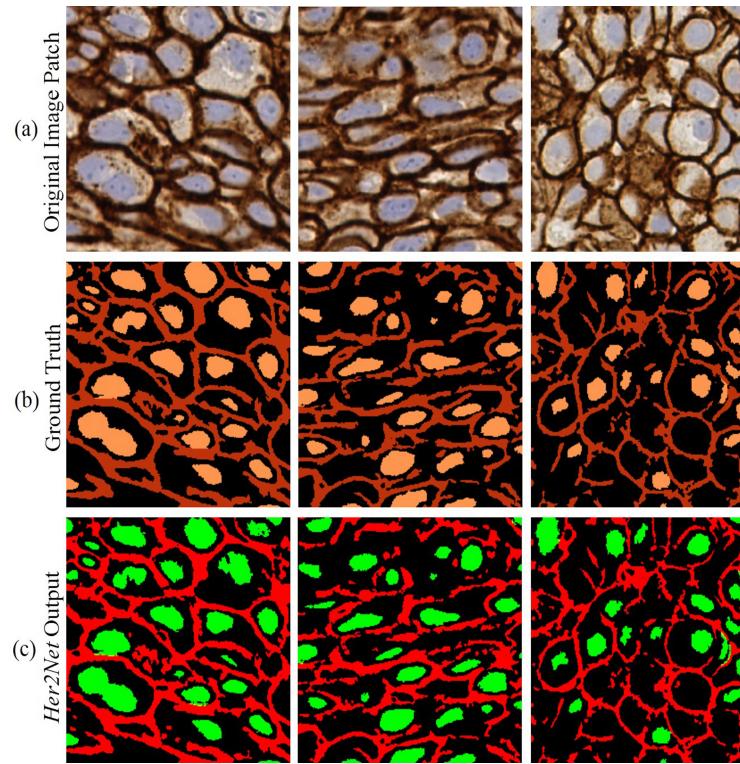


Figure 7: Results generated by the proposed *Her2Net* with respect to the ground truths: (a) original image patch (251 \times 251); (b) ground truth image; (c) *Her2Net* output

Table III: PERFORMANCE BASED ON VARIOUS COMBINATIONS OF TRAINING AND TESTING DATASETS

<i>Training Images (%)</i>	<i>Testing Images (%)</i>	<i>Pre (%)</i>	<i>Rec (%)</i>	<i>F-score (%)</i>	<i>NPV (%)</i>	<i>Accuracy (%)</i>	<i>FPR (%)</i>
0	100	97.20	96.57	96.88	85.49	97.95	12.10
25	75	95.33	96.38	95.85	94.35	97.95	7.24
50	50	96.22	96.56	96.39	94.59	98.19	5.92
75	25	97.02	97.36	97.19	95.85	98.56	4.67
100	0	97.43	97.08	97.26	95.12	99.00	4.30
Average		96.64	96.79	96.71	93.08	98.33	6.84

Table IV: QUANTITATIVE PIXEL-WISE EVALUATION OF *Her2Net* SEGMENTATION USING BREAST CANCER IHC IMAGES

<i>Score</i>	<i>Pathologist's Evaluation</i>	<i>DC</i>	<i>CC</i>	<i>SA (%)</i>
0	Pathologist 1	0.9153±0.0578	0.8185±0.0371	98.33
	Pathologist 2	0.8929±0.0603	0.8185±0.0240	97.97
	Average	0.9041±0.0590	0.8185± 0.0305	98.15
1+	Pathologist 1	0.8819±0.0321	0.8802±0.0719	99.01
	Pathologist 2	0.8779±0.0441	0.9081±0.0625	99.00
	Average	0.8799±0.0381	0.8941±0.0672	99.00
2+	Pathologist 1	0.9365±0.0010	0.9655±0.0337	100.00
	Pathologist 2	0.9360±0.0110	0.9666±0.0065	98.95
	Average	0.9362±0.0006	0.9660±0.0201	99.47
3+	Pathologist 1	0.9743±0.0126	0.9765±0.0432	97.77
	Pathologist 2	0.9689±0.0105	0.9820±0.0020	99.01
	Average	0.9717±0.0115	0.9792±0.0226	98.39

are denoted by TPOS, TNEG, FPOS, and FNEG respectively. SI means a segmented image, and GTI indicates a ground truth image. The μ_{GTI} and μ_{SI} represent the mean of *GTI* and *SI*, respectively. The DC was used to measure the similarity between *SI* and *GTI*. A higher DC value ($DC \geq 0.8$) indicated better segmentation. The similarity between *SI* and *GTI* based on pixel intensity was calculated using the CC ($CC \geq 0.8$).

C. Sensitivity Analysis

Figure 8(b) illustrates the sensitivity analysis. The X and Y axes represent window size and the AUC of ROC curves, respectively, for the *Her2Net* model. From the figure, it can be seen that *Her2Net* provided a better AUC value (94%) for the window size 251×251 . Thus, we chose the 251×251 image patch size for all our experiments.

D. Comparative Strategies

From an extensive literature survey, we found that evaluating cell membranes, nuclei and HER2 scoring of HER2-stained BC or other cancer IHC images using a deep learning approach had not been attempted. Few articles, e.g., [18], [44], have attempted deep learning for HER2-stained images, but gaps remain for future work. Our method is compared with the existing methods in table V. From table V, we conclude that our proposed *Her2Net* performed better than the current methods used in similar works.

The performance of *Her2Net* was also compared with the most popular and recently published deep learning segmentation

frameworks, including SegNet [63], Bayesian SegNet [64] and U-Net [65]. The comparison results using all the frameworks are shown in table VI. From the table, it is evident that SegNet, Bayesian SegNet and U-Net performed satisfactorily in comparison with *Her2Net*. Figure 9(a) shows graphs comparing various methods and their performances.

E. HER2 Scoring

The overall HER2 score of the proposed *Her2Net* with respect to the two pathologists' observations is shown in Table VII. When the score was zero, there was no error between manual observation and *Her2Net*. However, error was observed as the score increased. The small fluctuation of the error rate suggests that our *Her2Net* performed well and showed similar results to those of the pathologists. These results suggest that the proposed method is reliable to use for HER2 scoring.

F. Computational Time

Computational time is one of the significant factors for evaluations of deep learning models. The time and image patch size are directly proportional; i.e., if the image patch size increases, the computational time will increase as well. Small patch size increases the detection accuracy. Our proposed *Her2Net* framework was run for almost 7 days ($24 \times 7 = 168$ hours) for training (training done on a parallel computing platform), and on average, it took 1.06 seconds for cell membrane and nucleus segmentation, classification and HER2 scoring on test data. Figure 9(b) shows the loss (= 0.000434) and accuracy (= 0.9912) curves with respect to iteration.

V. DISCUSSION

Our results show that the proposed *Her2Net* can segment cell membranes and nuclei on IHC images with a high degree of recall, precision, F-score, negative predictive value, and classification accuracy. The most critical cases for the experiment were images where cytoplasm was coloured with HER2 monoclonal antibodies. In such situations, it was difficult for pathologists to annotate the cell membrane manually. A vital part of the proposed *Her2Net* was to measure the performances using pixel-wise validation techniques such as DC, CC and segmentation accuracy. The performances of the data cohorts exhibit highly correlated performance measures with respect to the training and testing datasets. The proposed *Her2Net* shows a very low false-positive rate. The performance of *Her2Net* improved as the number of training image patches increased.

Her2Net has been used as a patch-based segmentation, classification, and scoring tool. We addressed the cell membrane and nucleus segmentation task through a deep neural network, which consists of convolutional and deconvolutional components. The most interesting aspect of our proposed *Her2Net* is that it can be integrated into other computational frameworks for segmentation, classification and scoring. Thus, our proposed method could pave the way for the creation of decision support tools for BC IHC screening, diagnosis and prognosis.

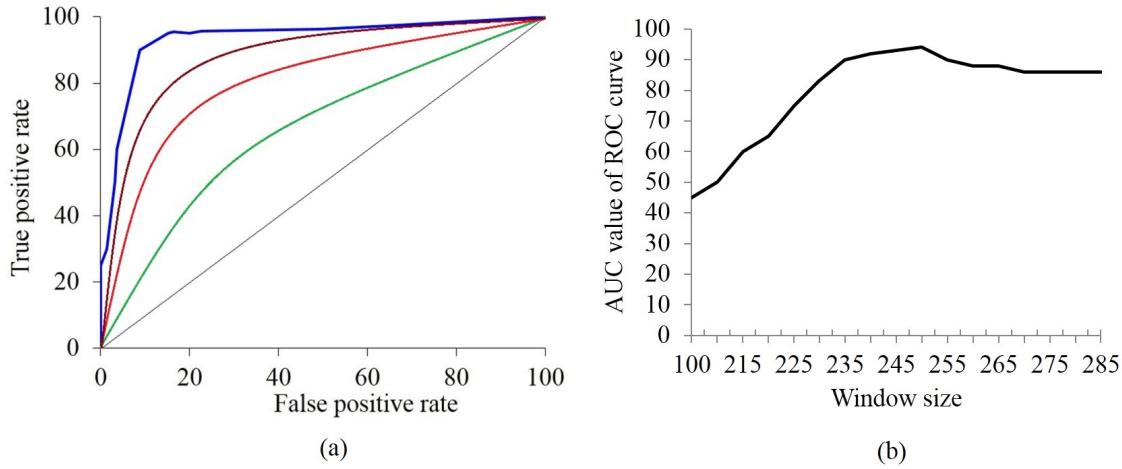


Figure 8: Performance evaluation: (a) ROC curves [blue: *Her2Net*; green: *SegNet*; red: Bayesian *SegNet*; brown: *U-Net*]; (b) plot of window size vs AUC for the *Her2Net* model

Table V: COMPARISON WITH EXISTING METHODS

<i>Comparison parameters</i>	<i>Vandenbergh et al. (2017)</i> [18]	<i>Pitaaho et al. (2016)</i> [44]	<i>Ficarra et al. (2011)</i> [22]	<i>Proposed Methodology</i>
Image Type	IHC	IHC	IHC	IHC
Sample Size	74 WSIs	28 WSIs	8 dataset	79 WSIs
Image Patch Size	44 × 44	128 × 128	Not Mentioned	251 × 251
Magnification	resolution 0.49 $\mu\text{m}/\text{pixel}$	40	40	40
Methodology Used	CovNet	Convolutional network	Convolutional network	<i>Her2Net</i>
Accuracy (%)	78.00	97.70	Not mentioned	98.33
Computation Time	45 minute	Not mentioned	Not mentioned	1.06 seconds
CPU or GPU	Not mentioned	GPU	CPU	GPU
Error Rate	Not mentioned	0.081	0.14	0.03
<i>Segmentation</i>				
Cell Membrane	No	No	Yes	Yes
Nuclei	No	No	Yes	Yes
<i>HER2 Score Calculation</i>				
	Yes	No	No	Yes

Table VI: COMPARISON OF *Her2Net* WITH THE EXISTING DEEP FRAMEWORKS

<i>Deep Frameworks</i>	<i>Pre</i>	<i>Rec</i>	<i>NPV</i>	<i>F-score</i>	<i>Accuracy</i>	<i>FPR</i>
SegNet	0.8109	0.8671	0.8067	8688	0.8361	0.3484
Bayesian SegNet	0.8632	0.8215	0.8466	0.8985	0.8439	0.3004
U-Net	0.9452	0.9612	0.8974	0.9331	0.9055	0.0910
<i>Her2Net</i>	0.9664	0.9679	0.9308	0.9671	0.9833	0.0684

VI. CONCLUSION

Our results indicate that the proposed *Her2Net* is feasible for quantifying and scoring HER2 status using BC IHC images. The proposed deep neural framework achieved good agreement with two expert pathologists' assessments. Furthermore, it had

an AUC of 94% and an F-score of 96.71% in classifying the cell membrane staining pattern (i.e., no staining, faint staining, moderate staining and strong complete staining). In addition, we present an automated HER2 status scoring method using deep learning.

This unique HER2 scoring tool for BC IHC images is better than current methods. Our approach is unique with respect to the deep learning framework, scoring layer, and implementation of the work using deep learning. Independent testing results revealed that *Her2Net* is efficient and robust. This framework can be useful for junior and senior pathologists to detect cell membranes and nuclei and score them with improved accuracy using HER2-stained images.

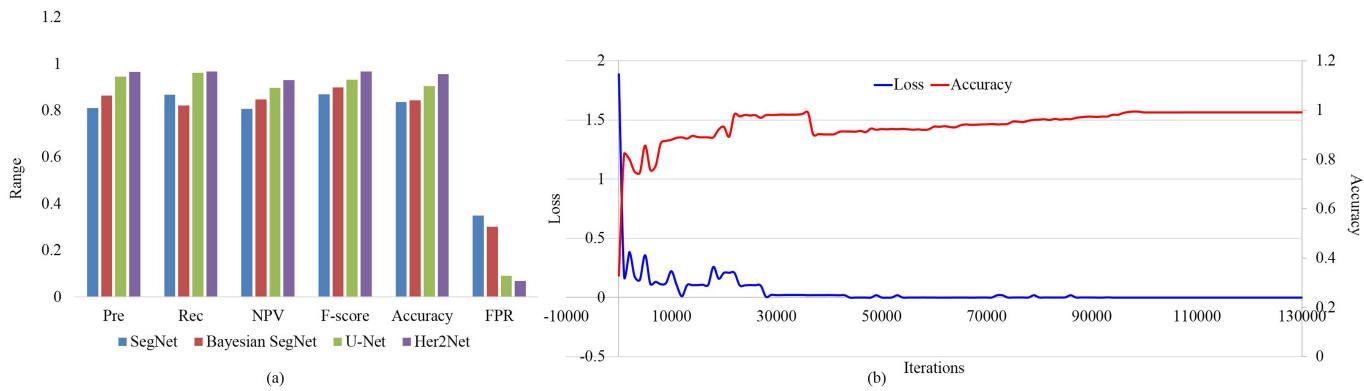


Figure 9: (a) Comparison graphs of various methods and their performances; (b) loss and accuracy curves with respect to iteration (loss=0.000434 and accuracy= 0.9912)

Table VII: OVERALL HER2 SCORE, MANUAL vs *Her2Net*

Score	Staining pattern	Pathologists	Manual (%)	Her2Net (%)	Error (%)
0	No membrane staining	Expert-1	100.00	100.00	0.00
		Expert-2	100.00	100.00	0.00
		Average	100.00	100.00	0.00
1+	Faint membrane staining	Expert-1	98.15	98.10	0.05
		Expert-2	98.10	98.10	0.00
		Average	98.13	98.10	0.03
2+	Moderate membrane staining	Expert-1	97.95	98.00	0.05
		Expert-2	98.00	98.00	0.00
		Average	97.98	98.00	0.03
3+	Strong membrane staining	Expert-1	99.93	100	0.07
		Expert-2	100	100	0.00
		Average	99.97	100	0.04

CONFLICTS OF INTEREST

No conflict of interest

ACKNOWLEDGMENT

M. Saha would like to acknowledge the Department of Science and Technology (DST), Govt. of India, for providing the DST-Inspire fellowship (IVR Number: 201400105113). C. Chakraborty acknowledges the Ministry of Human Resource Development (MHRD), Govt. of India, for financial support to carry out this work (grant no: 4-23/2014 T.S.I. date: 14-02-2014) under Signals & Systems for Life Science, a mega-initiative of IIT Kharagpur, India. The authors would like to acknowledge Dr I. Arun, Dr S. Chatterjee, Dr S. Agarwal and Dr R. Ahmed of Tata Medical Center, Kolkata, India, for their clinical advice. The authors are thankful to Prof N. Rajpoot, Dept. of Computer Science, University of Warwick, UK for providing a dataset for this study.

REFERENCES

- [1] R. H. Engel and V. G. Kaklamani, "HER2-positive breast cancer," *Drugs*, vol. 67, no. 9, pp. 1329–1341, 2007.
- [2] M. Saha, S. Agarwal, I. Arun, R. Ahmed, S. Chatterjee, P. Mitra, and C. Chakraborty, "Histogram based thresholding for automated nucleus segmentation using breast imprint cytology," in *Advancements of Medical Electronics*. Springer, 2015, pp. 49–57.
- [3] M. Saha, I. Arun, S. Agarwal, R. Ahmed, S. Chatterjee, and C. Chakraborty, "Imprint cytology-based breast malignancy screening: an efficient nuclei segmentation technique," *Journal of Microscopy*, vol. 268, no. 2, pp. 155–171, 2017.
- [4] S. Banerjee, M. Saha, I. Arun, B. Basak, S. Agarwal, R. Ahmed, S. Chatterjee, L. B. Mahanta, and C. Chakraborty, "Near-set based mucin segmentation in histopathology images for detecting mucinous carcinoma," *Journal of Medical Systems*, vol. 41, no. 9, p. 144, 2017.
- [5] Y. Yamamoto, A. Saito, A. Tateishi, H. Shimojo, H. Kanno, S. Tsuchiya, K.-i. Ito, E. Cosatto, H. P. Graf, R. R. Moraleda *et al.*, "Quantitative diagnosis of breast tumors by morphometric classification of microenvironmental myoepithelial cells using a machine learning approach," *Scientific Reports*, vol. 7, 2017.
- [6] M. Tabakov and P. Kozak, "Segmentation of histopathology HER2/neu images with fuzzy decision tree and takagi–sugeno reasoning," *Computers in biology and medicine*, vol. 49, pp. 19–29, 2014.
- [7] L. A. Torre, F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent, and A. Jemal, "Global cancer statistics, 2012," *CA: a cancer journal for clinicians*, vol. 65, no. 2, pp. 87–108, 2015.
- [8] C. E. Desantis, J. Ma, A. Goding Sauer, L. A. Newman, and A. Jemal, "Breast cancer statistics, 2017, racial disparity in mortality by state," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 6, pp. 439–448, 2017.
- [9] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2017," *CA: A Cancer Journal for Clinicians*, vol. 67, no. 1, pp. 7–30, 2017.
- [10] C. W. Elston and I. O. Ellis, "Pathological prognostic factors in breast cancer. I. the value of histological grade in breast cancer: experience from a large study with long-term follow-up," *Histopathology*, vol. 19, no. 5, pp. 403–410, 1991.
- [11] D. C. Zaha, "Significance of immunohistochemistry in breast cancer," *World journal of clinical oncology*, vol. 5, no. 3, p. 382, 2014.
- [12] M. A. Kaplan, H. Ertugrul, U. Firat, M. Kucukoner, A. İnal, Z. Urakci, Z. Pekkolay, and A. Isikdogan, "Brain metastases in HER2-positive metastatic breast cancer patients who received chemotherapy with or without trastuzumab," *Breast cancer*, vol. 22, no. 5, pp. 503–509, 2015.
- [13] H. Yaziji, L. C. Goldstein, T. S. Barry, R. Werling, H. Hwang, G. K. Ellis, J. R. Gralow, R. B. Livingston, and A. M. Gown, "HER-2 testing in breast cancer using parallel tissue-based methods," *Jama*, vol. 291, no. 16, pp. 1972–1977, 2004.
- [14] N. T. Foged, A. Brügmann, and J. T. Jørgensen, "The HER2 CISH pharmDx™ kit in the assessment of breast cancer patients for anti-HER2 treatment," *Expert review of molecular diagnostics*, vol. 13, no. 3, pp. 233–242, 2013.
- [15] E. A. Rakha, S. E. Pinder, J. M. Bartlett, M. Ibrahim, J. Starczynski, P. J. Carder, E. Provenzano, A. Hanby, S. Hales, A. H. Lee *et al.*, "Updated UK Recommendations for HER2 assessment in breast cancer," *Journal of clinical pathology*, vol. 68, no. 2, pp. 93–99, 2015.
- [16] P. Hagen, K. Biermann, J. Boers, O. Stoss, H. Sleddens, J. Lanschot, W. Dinjens, J. Rueschhoff, and B. Wijnhoven, "Human epidermal growth factor receptor 2 overexpression and amplification in endoscopic biopsies and resection specimens in esophageal and junctional adenocarcinoma," *Diseases of the Esophagus*, vol. 28, no. 4, pp. 380–385, 2015.
- [17] M. Saha, R. Mukherjee, and C. Chakraborty, "Computer-aided diagnosis of breast cancer using cytological images: A systematic review," *Tissue and Cell*, vol. 48, no. 5, pp. 461–474, 2016.

- [18] M. E. Vandenberghe, M. L. Scott, P. W. Scorer, M. Söderberg, D. Balcerzak, and C. Barker, "Relevance of deep learning to facilitate the diagnosis of HER2 status in breast cancer," *Scientific Reports*, vol. 7, 2017.
- [19] A. C. Wolff, M. E. H. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer *et al.*, "American society of clinical oncology/college of american pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer," *Journal of clinical oncology*, vol. 25, no. 1, pp. 118–145, 2006.
- [20] A. C. Wolff, M. E. H. Hammond, J. N. Schwartz, K. L. Hagerty, D. C. Allred, R. J. Cote, M. Dowsett, P. L. Fitzgibbons, W. M. Hanna, A. Langer *et al.*, "American society of clinical oncology/college of american pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer," *Archives of pathology & laboratory medicine*, vol. 131, no. 1, pp. 18–43, 2007.
- [21] M. E. H. Hammond, D. F. Hayes, M. Dowsett, D. C. Allred, K. L. Hagerty, S. Badve, P. L. Fitzgibbons, G. Francis, N. S. Goldstein, M. Hayes *et al.*, "American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version)," *Archives of pathology & laboratory medicine*, vol. 134, no. 7, pp. e48–e72, 2010.
- [22] E. Ficarra, S. Di Cataldo, A. Acquaviva, and E. Macii, "Automated segmentation of cells with IHC membrane staining," *IEEE Transactions on Biomedical Engineering*, vol. 58, no. 5, pp. 1421–1429, 2011.
- [23] C.-Y. Chang, Y.-C. Huang, and C.-C. Ko, "Automatic analysis of HER-2/neu immunohistochemistry in breast cancer," in *Innovations in Bio-Inspired Computing and Applications (IBICA), 2012 Third International Conference on*. IEEE, 2012, pp. 297–300.
- [24] D. M. Minot, J. Voss, S. Rademacher, T. Lwin, J. Orsulak, B. Caron, R. Ketterling, A. Nassar, B. Chen, and A. Clayton, "Image analysis of HER2 immunohistochemical staining," *American journal of clinical pathology*, vol. 137, no. 2, pp. 270–276, 2012.
- [25] A. Brüggemann, M. Eld, G. Leikaitis, S. Nielsen, M. Grunkin, J. D. Hansen, N. T. Foged, and M. Vyberg, "Digital image analysis of membrane connectivity is a robust measure of HER2 immunostains," *Breast cancer research and treatment*, vol. 132, no. 1, pp. 41–49, 2012.
- [26] V. J. Tuominen, T. T. Tolonen, and J. Isola, "Immunomembrane: a publicly available web application for digital image analysis of HER2 immunohistochemistry," *Histopathology*, vol. 60, no. 5, pp. 758–767, 2012.
- [27] Y.-L. Kuo, C.-C. Ko, and J.-Y. Lai, "Automated assessment in HER-2/neu immunohistochemical expression of breast cancer," in *Computer Communication Control and Automation (3CA), 2010 International Symposium on*, vol. 2. IEEE, 2010, pp. 585–588.
- [28] A. Del Bimbo, M. Meoni, and P. Pala, "Accurate evaluation of HER-2 amplification in fish images," in *Imaging Systems and Techniques (IST), 2010 IEEE International Conference on*. IEEE, 2010, pp. 407–410.
- [29] B. Balluff, M. Elsner, A. Kowarsch, S. Rausser, S. Meding, C. Schuhmacher, M. Feith, K. Herrmann, C. Rocken, R. M. Schmid *et al.*, "Classification of HER2/neu status in gastric cancer using a breast-cancer derived proteome classifier," *Journal of proteome research*, vol. 9, no. 12, pp. 6317–6322, 2010.
- [30] H. Masmoudi, S. M. Hewitt, N. Petrick, K. J. Myers, and M. A. Gavrielides, "Automated quantitative assessment of HER-2/neu immunohistochemical expression in breast cancer," *IEEE transactions on medical imaging*, vol. 28, no. 6, pp. 916–925, 2009.
- [31] C. Ellis, M. Dyson, T. Stephenson, and E. Malthy, "HER2 amplification status in breast cancer: a comparison between immunohistochemical staining and fluorescence in situ hybridisation using manual and automated quantitative image analysis scoring techniques," *Journal of clinical pathology*, vol. 58, no. 7, pp. 710–714, 2005.
- [32] M. C. Lloyd, P. Allam-Nandyala, C. N. Purohit, N. Burke, D. Coppola, M. M. Bui *et al.*, "Using image analysis as a tool for assessment of prognostic and predictive biomarkers for breast cancer: how reliable is it?" *Journal of pathology informatics*, vol. 1, no. 1, p. 29, 2010.
- [33] G. Viale, J. Paterson, M. Bloch, G. Csathy, D. Allen, P. Dell'Orto, G. Kjersgaard, Y. Y. Levy, and J. T. Jørgensen, "Assessment of HER2 amplification status in breast cancer using a new automated HER2 IQFISH pharmDx™(Dako Omnis) assay," *Pathology-Research and Practice*, vol. 212, no. 8, pp. 735–742, 2016.
- [34] P. E. Aktan, G. Hatipoğlu, and N. Arica, "Risk classification for breast cancer diagnosis using HER2 testing," in *Signal Processing and Communication Application Conference (SIU), 2016 24th*. IEEE, 2016, pp. 2133–2136.
- [35] E. Ayad, M. Mansy, D. Elwi, M. Salem, M. Salama, K. Kayser *et al.*, "Comparative study between quantitative digital image analysis and fluorescence in situ hybridization of breast cancer equivocal human epidermal growth factor receptors 2 score 2+ cases," *Journal of pathology informatics*, vol. 6, no. 1, p. 31, 2015.
- [36] N. Fusco, E. G. Rocco, C. Del Conte, C. Pellegrini, G. Bulfamante, F. Di Nuovo, S. Romagnoli, and S. Bosari, "HER2 in gastric cancer: a digital image analysis in pre-neoplastic, primary and metastatic lesions," *Modern Pathology*, vol. 26, no. 6, pp. 816–824, 2013.
- [37] M. C. Cheang, S. K. Chia, D. Voduc, D. Gao, S. Leung, J. Snider, M. Watson, S. Davies, P. S. Bernard, J. S. Parker *et al.*, "Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer," *Journal of the National Cancer Institute*, vol. 101, no. 10, pp. 736–750, 2009.
- [38] F. Raimondo, M. A. Gavrielides, G. Karayannopoulou, K. Lyroudia, I. Pitas, and I. Kostopoulos, "Automated evaluation of HER-2/neu status in breast tissue from fluorescent in situ hybridization images," *IEEE Transactions on Image Processing*, vol. 14, no. 9, pp. 1288–1299, 2005.
- [39] Z. Theodosiou, I. N. Kasampalidis, G. Karayannopoulou, I. Kostopoulos, M. Bobos, G. Bevilacqua, P. Aretini, A. Starita, K. Lyroudia, and I. Pitas, "Evaluation of FISH image analysis system on assessing HER2 amplification in breast carcinoma cases," *The Breast*, vol. 17, no. 1, pp. 80–84, 2008.
- [40] B. H. Hall, M. Ianosi-Irimie, P. Javidian, W. Chen, S. Ganesan, and D. J. Foran, "Computer-assisted assessment of the human epidermal growth factor receptor 2 immunohistochemical assay in imaged histologic sections using a membrane isolation algorithm and quantitative analysis of positive controls," *BMC medical imaging*, vol. 8, no. 1, p. 1, 2008.
- [41] U. Gaur, M. Kourakis, E. Newman-Smith, W. Smith, and B. Manjunath, "Membrane segmentation via active learning with deep networks," in *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 1943–1947.
- [42] A. N. Basavanhally, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot, and A. Madabhushi, "Computerized image-based detection and grading of lymphocytic infiltration in HER2+ breast cancer histopathology," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 642–653, 2010.
- [43] I. Skaland, I. Øvestad, E. A. Janssen, J. Klos, K. H. Kjellevold, T. Helliesen, and J. Baak, "Comparing subjective and digital image analysis HER2/neu expression scores with conventional and modified fish scores in breast cancer," *Journal of clinical pathology*, vol. 61, no. 1, pp. 68–71, 2008.
- [44] T. Pitkäaho, T. M. Lehtimäki, J. McDonald, and T. J. Naughton, "Classifying HER2 breast cancer cell samples using deep learning," in *Irish Machine Vision and Image Processing Conference*. Irish Pattern Recognition and Classification Society, 2016.
- [45] B. Langari, S. Vaseghi, A. Prochazka, B. Vaziri, and F. T. Aria, "Edge-guided image gap interpolation using multi-scale transformation," *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4394–4405, 2016.
- [46] J. Dong, Z. Han, Y. Zhao, W. Wang, A. Prochazka, and J. Chambers, "Sparse analysis model based multiplicative noise removal with enhanced regularization," *Signal Processing*, vol. 137, pp. 160–176, 2017.
- [47] HER2 Scoring Contest. Visited on 2017-11-30. [Online]. Available: <http://www2.warwick.ac.uk/fac/sci/dcs/research/tia/her2contest/>
- [48] T. Qaiser, A. Mukherjee, S. D. Munugoti, V. Tallam, T. Pitkäaho, T. Lehtimäki, T. Naughton, M. Berseth, A. Pedraza, R. Mukundan *et al.*, "HER2 challenge contest: A detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues," *arXiv preprint arXiv:1705.08369*, 2017.
- [49] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A deep convolutional neural network for segmenting and classifying epithelial and stromal regions in histopathological images," *Neurocomputing*, vol. 191, pp. 214–223, 2016.
- [50] M. Saha, C. Chakraborty, I. Arun, R. Ahmed, and S. Chatterjee, "An advanced deep learning approach for Ki-67 stained hotspot detection and proliferation rate scoring for prognostic evaluation of breast cancer," *Scientific Reports*, vol. 7, 2017.
- [51] K. Jarrett, K. Kavukcuoglu, Y. LeCun *et al.*, "What is the best multi-stage architecture for object recognition?" in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2146–2153.
- [52] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904–1916, 2015.
- [53] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation

- model,” in *European Conference on Computer Vision*. Springer, 2016, pp. 34–50.
- [54] H. Noh, S. Hong, and B. Han, “Learning deconvolution network for semantic segmentation,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1520–1528.
- [55] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, “Deconvolutional networks,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2528–2535.
- [56] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [57] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained handwriting recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 31, no. 5, pp. 855–868, 2009.
- [58] A. Graves, A. R. Mohamed, and G. Hinton, “Speech recognition with deep recurrent neural networks,” in *Acoustics, speech and signal processing (icassp), 2013 ieee international conference on*. IEEE, 2013, pp. 6645–6649.
- [59] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, “Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation,” in *Advances in Neural Information Processing Systems*, 2015, pp. 2998–3006.
- [60] A. C. Wolff, M. E. H. Hammond, D. G. Hicks, M. Dowsett, L. M. McShane, K. H. Allison, D. C. Allred, J. M. Bartlett, M. Bilous, P. Fitzgibbons *et al.*, “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,” *Journal of clinical oncology*, vol. 31, no. 31, pp. 3997–4013, 2013.
- [61] M. Saha, I. Arun, B. Basak, S. Agarwal, R. Ahmed, S. Chatterjee, R. Bhargava, and C. Chakraborty, “Quantitative microscopic evaluation of mucin areas and its percentage in mucinous carcinoma of the breast using tissue histological images,” *Tissue and Cell*, vol. 48, no. 3, pp. 265–273, 2016.
- [62] J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, “Stacked sparse autoencoder (ssae) for nuclei detection on breast cancer histopathology images,” *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 119–130, 2016.
- [63] V. Badrinarayanan, A. Kendall, and R. Cipolla, “Segnet: A deep convolutional encoder-decoder architecture for image segmentation,” *arXiv preprint arXiv:1511.00561*, 2015.
- [64] A. Kendall, V. Badrinarayanan, and R. Cipolla, “Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding,” *arXiv preprint arXiv:1511.02680*, 2015.
- [65] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.



Monjoy Saha, DST-Inspire fellow, is a Ph.D. student in the School of Medical Science and Technology, Indian Institute of Technology Kharagpur, India. He received the Master of Technology (M.Tech.) in Biomedical Engineering from the Indian Institute of Technology (BHU), Varanasi, India. His research interests include machine learning, deep learning, robotics, signal processing, natural language processing, medical image analysis, etc.



Chandan Chakraborty is an Associate Professor at the School of Medical Science and Technology, Indian Institute of Technology Kharagpur, India. He completed his Masters from IIT Bombay and PhD from IIT Kharagpur. His major research interests include medical image processing, statistical pattern classification, machine learning, computer/smartphone -aided diagnosis and prognosis. He received awards including ISCA Young Scientist Award, DAE-Young Investigator Award, IBM faculty Award etc.