








Article

# Weakly-Supervised Classification of HER2 Expression in Breast Cancer Haematoxylin and Eosin Stained Slides

Sara P. Oliveira <sup>1,2,\*</sup> , João Ribeiro Pinto <sup>1,2</sup> , Tiago Gonçalves <sup>1</sup> , Rita Canas-Marques <sup>3,4</sup> ,  
Maria-João Cardoso <sup>4,5</sup> , Hélder P. Oliveira <sup>1,6</sup>  and Jaime S. Cardoso <sup>1,2</sup> 

<sup>1</sup> INESC TEC, 4200-465 Porto, Portugal; joao.t.pinto@inesctec.pt (J.R.P.); tiago.f.goncalves@inesctec.pt (T.G.); hfpo@inesctec.pt (H.P.O.); jaime.cardoso@inesctec.pt (J.S.C.)

<sup>2</sup> Faculty of Engineering (FEUP), University of Porto, 4200-465 Porto, Portugal

<sup>3</sup> Anatomic Pathology Service, Champalimaud Clinical Centre, Champalimaud Foundation, 1400-038 Lisbon, Portugal; rita.marques@fundacaochampalimaud.pt

<sup>4</sup> Breast Unit, Champalimaud Clinical Centre, Champalimaud Foundation, 1400-038 Lisbon, Portugal; maria.joao.cardoso@fundacaochampalimaud.pt

<sup>5</sup> NOVA Medical School, 1169-056 Lisbon, Portugal

<sup>6</sup> Faculty of Sciences (FCUP), University of Porto, 4169-007 Porto, Portugal

\* Correspondence: sara.i.oliveira@inesctec.pt

Received: 3 June 2020; Accepted: 1 July 2020; Published: 9 July 2020



**Featured Application:** This work finds its key application in medical diagnosis and prognosis. It paves the way to robust automatic HER2 classification using only H&E slides. This approach possibly avoids the additional costs and time spent IHC testing, providing an indication of the IHC result.

**Abstract:** Human epidermal growth factor receptor 2 (HER2) evaluation commonly requires immunohistochemistry (IHC) tests on breast cancer tissue, in addition to the standard haematoxylin and eosin (H&E) staining tests. Additional costs and time spent on further testing might be avoided if HER2 overexpression could be effectively inferred from H&E stained slides, as a preliminary indication of the IHC result. In this paper, we propose the first method that aims to achieve this goal. The proposed method is based on multiple instance learning (MIL), using a convolutional neural network (CNN) that separately processes H&E stained slide tiles and outputs an IHC label. This CNN is pretrained on IHC stained slide tiles but does not use these data during inference/testing. H&E tiles are extracted from invasive tumour areas segmented with the HASHI algorithm. The individual tile labels are then combined to obtain a single label for the whole slide. The network was trained on slides from the HER2 Scoring Contest dataset (HER2SC) and tested on two disjoint subsets of slides from the HER2SC database and the TCGA-TCIA-BRCA (BRCA) collection. The proposed method attained 83.3% classification accuracy on the HER2SC test set and 53.8% on the BRCA test set. Although further efforts should be devoted to achieving improved performance, the obtained results are promising, suggesting that it is possible to perform HER2 overexpression classification on H&E stained tissue slides.

**Keywords:** weakly-supervised learning; HER2; breast cancer

## 1. Introduction

Breast cancer (BCa) is the most commonly diagnosed cancer and the leading cause of cancer-related deaths among women worldwide. However, over the most recent years, despite the

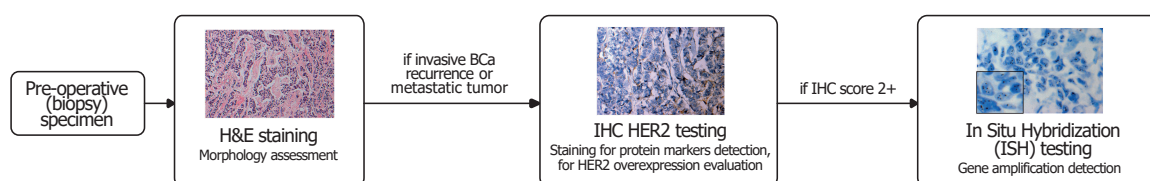
increasing incidence trends, the mortality rate has significantly decreased. Among other factors, this improvement results from better treatment strategies that can be delineated from the assessment of histopathological characteristics [1,2].

The analysis of tissue sections of cancer specimens (Figure 1) obtained by preoperative biopsy, commonly starts with haematoxylin and eosin (H&E) staining, which is usually followed by immunohistochemistry (IHC), a more advanced staining technique, used to highlight the presence of specific protein receptors [3]. In fact, according to the current guidelines [4], Human epidermal growth factor receptor 2 (HER2) quantification must be routinely tested in all patients with invasive BCa, recurrence cases, and metastatic tumours. The overexpression of this receptor is observed in 10–20% [4] of BCa cases and has been associated with aggressive clinical behaviour and poor prognosis [5]. However, patients diagnosed with HER2-positive BCa have a better response to targeted therapies and consequent improvements in healing and overall survival, which emphasizes the importance of an accurate evaluation of HER2 status [5,6].

The current guidelines [7], revised by the American Society of Clinical Oncology/College of American Pathologists (ASCO/CAP), in 2018, indicate the following scoring criteria for HER2 IHC:

- IHC 0+: no staining or incomplete, barely perceptible membrane staining in 10% of tumour cells or less;
- IHC 1+: incomplete, barely perceptible membrane staining in more than 10% of tumour cells;
- IHC 2+: weak to moderate complete membrane staining in more than 10% of tumour cells;
- IHC 3+: circumferential, complete, intense membrane staining in more than 10% of tumour cells.

Moreover, cases scoring 0+ or 1+ are classified as HER2 negative, while cases with a score of 3+ are classified as HER2 positive. Cases with score 2+ are classified as equivocal and are further assessed by in situ hybridization (ISH), to test for gene amplification (see Figure 1). In these cases, the HER2 status is given by the ISH result [7].

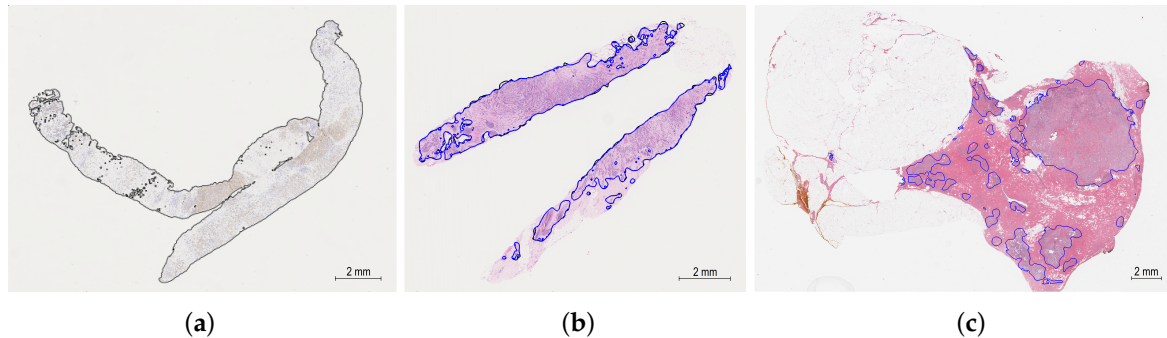


**Figure 1.** Schema of the process of BCa HER2 evaluation, involving H&E staining, IHC testing and, in specific cases, ISH testing. The proposed method aims to evaluate HER2 using only H&E stained slide images. Image examples were adapted from [8].

At the moment, besides very well differentiated tumours, with low nuclear/cytoplasm area ratio, which typically are hormonal driven and therefore generally not positive for HER2, there are no morphological features on H&E slides that allow a reliable prediction of the HER2 status. Therefore, the standard procedure is to perform an additional immunohistochemical study, with additional molecular study in case of equivocal results. Despite the efficiency of IHC and ISH, the additional cost and time spent on these tests might be avoided if all the information needed to infer the HER2 status could be extracted only from H&E whole slide images (WSI), as a preliminary indication of the IHC result. However, to the extent of our knowledge, the task of predicting HER2 status on H&E stained slides has not yet been addressed in the literature, except for a recent challenge: ECDP2020's HEROHE Challenge [9].

In this paper we propose a method using a convolutional neural network (CNN), inspired by multiple-instance learning (MIL), to automatically identify the HER2 status on BCa H&E stained slides. To deal with the sheer dimensions of the slides, tiles are extracted from the original images and separately processed by the model, which learns to aggregate the individual tile predictions into a single, image-wide label. Moreover, to introduce some prior knowledge about the morphology of

tissue structures into the model, the CNN has been pre-trained with HER2 IHC stained slides (example in Figure 2a) from the HER2 Scoring Contest (HER2SC) training set [10]. The final architecture was trained with the H&E stained slides of HER2SC (example in Figure 2b) and tested with a disjoint subset of H&E stained slides (example in Figure 2c) from the TCIA-TCGA-BRCA (BRCA) collection [11,12]. The code is publicly available in a GitHub repository [13].



**Figure 2.** Image examples from used datasets: HER2SC [10] IHC stained slides (a), HER2SC [10] H&E stained slides (b), BRCA [11,12] H&E stained slides (c). The tile extraction was solely done on tissue, here denoted by the delineated regions.

## 2. Related Work

With the advent of WSI over the last decade, a huge amount of tissue slides are routinely scanned in clinical practice, thereby increasing data availability. Consequently, and along with its important role in oncological clinical routine, more research opportunities are raised in computer-assisted imaging analysis, with this new source of “big data” [14]. In fact, due to the high-resolution and complex nature of this imaging technique, advances in image analysis are required, providing an opportunity to apply/develop more advanced image processing techniques, as well as machine and deep learning algorithms. [15].

The analysis of digital images of digital breast pathology can be applied to tackle several clinical and pathology tasks such as, for example, mitosis detection [16,17], tissue type classification/segmentation [18–20], cancer grading [21] or histological classification [22,23]. These approaches commonly use H&E stained slides and, in recent years, have been focused on applying deep learning techniques to improve the performance of the models and also to take advantage of the increasing availability of medical data.

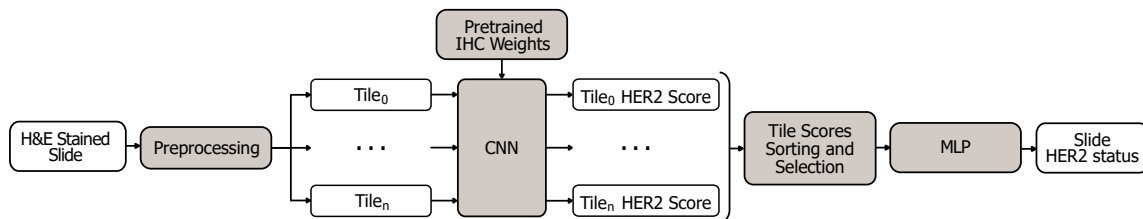
Besides H&E staining slides, some authors, such as Oscanoa et al. [24], Mojjoy et al. [25], Jamaluddin et al. [26] have addressed diverse tasks using IHC slides. On the specific task of automatic breast cancer HER2 classification, the focus of this work, the literature is limited to the work by Rodner et al. [27], Mukundan [28], and other studies related with the 2016 HER2 Scoring Contest [10]. The common disadvantage of these prior works is requiring IHC staining images to perform HER2 classification, since this modality requires additional cost and time. In contrast, our approach aims at using the H&E staining slides as the only source of information to obtain the IHC status for HER2 overexpression.

In order to complete the proposed task, we followed the idea of using a data source as initialization for the framework, transferring some domain knowledge to the final training. This is a recent trend that has been applied to medical imaging processing for different purposes, such as cardiac structures segmentation [29], Alzheimer disease classification [30], radiological breast lesions classification [31] and even digital pathology classification/segmentation [32].

Despite the growing popularity of digital pathology and the increasing number of publications in this area, to the extent of our knowledge, the task of predicting HER2 status on H&E stained slides has not yet been addressed in the literature.

### 3. Methodology

The proposed method (Figure 3) comprises a convolutional neural network (CNN), which is pre-trained for the task of HER2 scoring of tiles extracted from IHC stained slides. The pre-trained parameters are then transferred to the task of HER2 status prediction on H&E staining slide tiles, to provide the network with some knowledge of the tissue structures' appearance. Individual tile scores are then combined to obtain a single label for the respective slide. The data preprocessing methodology and the implemented networks are described below.



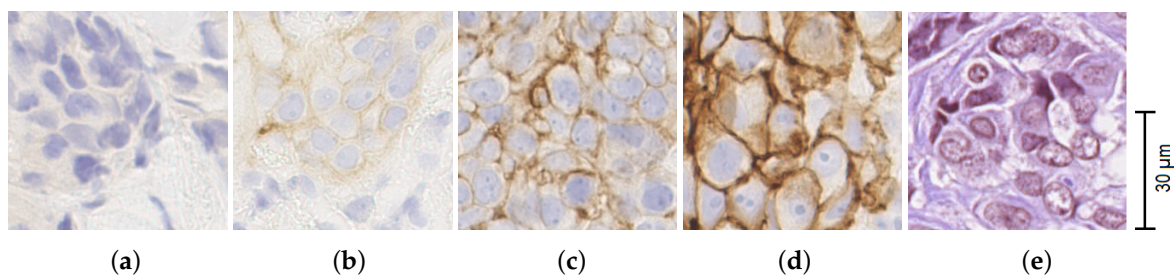
**Figure 3.** The proposed approach for weakly-supervised HER2 status classification on BCa H&E stained slides.

#### 3.1. Data Preprocessing

##### 3.1.1. IHC Stained Slides

For the IHC stained slides of classes 2+ and 3+, the preprocessing begins with automatic tissue segmentation with Otsu's thresholding on the saturation (S) channel of the HSV colour space, obtaining the regions with more intense staining, that correspond to the HER overexpression areas. For slides of classes 0+ and 1+, the segmentation consists of simple removal of pixels with the greatest HSV value (V) intensity, corresponding to background pixels, which do not contain essential information to the problem. These processes, which are performed at  $32\times$  downsampled slides, return the masks used in tile extraction.

Tiles with size  $256 \times 256$  are extracted from the slide with original dimensions (without downsampling), provided they are completely within the mask region. These tiles are converted from RGB to HSL colour space, of which only the lightness (L) channel is used. Each tile inherits the class from the respective slide (examples in Figure 4a–d), turning the learning task into a weakly-supervised problem.



**Figure 4.** Tile examples extracted from IHC 0+ (a), IHC 1+ (b), IHC 2+ (c), IHC 3+ (d), H&E (e) slides. Tiles from IHC 2+ and 3+ and H&E slides were obtained by Otsu's thresholding and the remaining were obtained by simply removing the pixels with background value. The IHC tiles were obtained from slides of the HER2SC dataset [10] and the H&E tile was obtained from a slide of the BRCA dataset [11,12].

##### 3.1.2. H&E Stained Slides

According to the ASCO/CAP guidelines for IHC evaluation, the diagnosis is performed based only the tumoral region of the slides. Hence, the preprocessing of H&E stained slides begins

with an automatic invasive tissue segmentation with the HASHI method [18,33] (High-throughput Adaptive Sampling for whole-slide Histopathology Image analysis), which consists of an adaptive gradient-based sampling approach that iteratively refines an initial coarse invasive BCa probability map, from CNN inference.

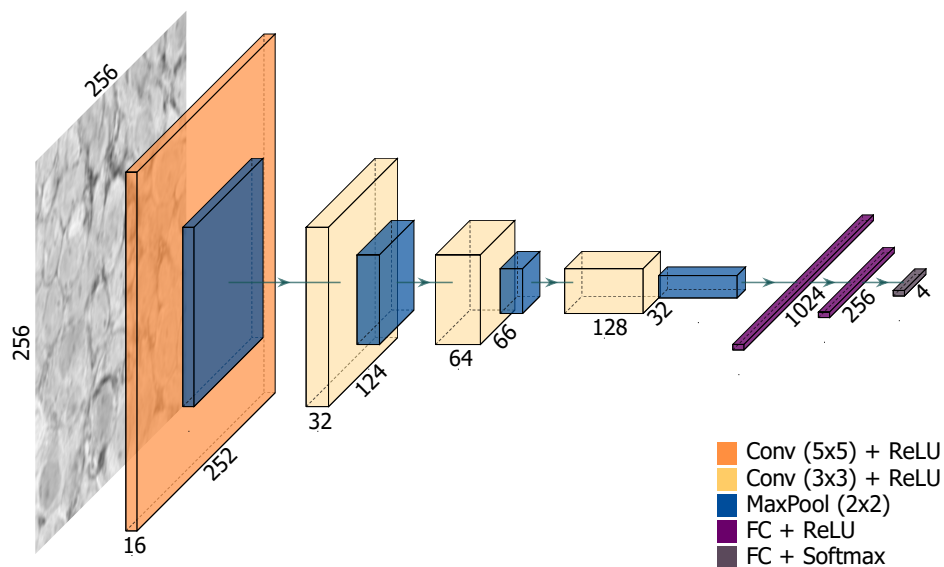
The algorithm begins with a WSI as input, that is sampled in 100 tiles, each of them being classified using a CNN-trained model, to obtain the probability of invasive BCa presence. By interpolation of each tile probability, a heatmap is generated for the entire WSI. Then, the gradient of the map is calculated and used to prioritize the sampling selection on the next iteration. The process is repeated during 20 iterations [18].

The method was implemented in the images referred by Cruz-Roa et al. [18] as the test set, using the original magnification and extracting squared  $512 \times 512$  tiles. Moreover, to exclude eventual small background zones included in HASHI segmentation, this mask region was intersected with the segmentation obtained using Otsu's thresholding on the saturation (S) channel of the HSV colour space.

The final segmentation mask was then used to generate H&E tiles (example in Figure 4e), extracted and processed accordingly to the methodology described for IHC slides. The number of tiles per slide varies according to the extent of the tissue region.

### 3.2. CNN for IHC Tile Scoring

The CNN architecture (Figure 5) consists of four convolutional layers (16, 32, 64 and 128 filters, respectively, with ReLU activation). The first layer has  $5 \times 5$  square kernels, while the remaining have  $3 \times 3$  square kernels. Each convolutional layer is followed by one pooling layer (a max-pooling function without overlap, with kernel  $2 \times 2$ ). The network is topped with three fully-connected layers, with 1024, 256, and 4 units, respectively. The first two have ReLU activation, while the third is followed by softmax activation for the output of probabilities for each class.



**Figure 5.** Architecture of the implemented convolutional neural network.

### 3.3. CNN for H&E Stained Slide Classification

The network parameters pre-trained with IHC stained slides were used as initial network weights for HER2 status classification on H&E stained slides. It is worth mentioning that IHC data is only used for the network pre-train, and not on the inference/test phase. To achieve a single prediction per

tile, instead of four (as it was initially trained for on the IHC setting), a soft-argmax activation [34,35] replaces the softmax activation, following the equation

$$\text{soft-argmax}(s) = \sum_i \text{softmax}(\beta s_i) i, \quad (1)$$

where  $\beta$  is an adjustment factor which controls the range of the probability map given by the softmax,  $s$  is the tile score array, and  $i$  is the index that corresponds to each class.

Having a single value per tile enables the easy sorting of tiles, which is performed before the aggregation into a single HER2 label. With the HER2 scores of each tile, output by the soft-argmax activation, tiles are sorted from 3+ to 0+. Then, the 15% highest scores are selected to serve as input to the aggregation process. This percentage was chosen to limit the information given to the aggregation network, while still including and barely exceeding the reference 10% of tumour area considered in the HER2 scoring guidelines.

The score aggregation is performed by a multilayer perceptron (MLP), composed of four layers, with 256, 128, 64, and 2 neurons, respectively. All layers are followed by ReLU activation, except the last layer, which is followed by softmax activation. Since the input dimension  $M$  of the MLP is fixed (we set  $M = 300$  in our experimental analysis, to limit memory cost), for images where 15% of the number of tiles exceeds  $M$ , we downsample to 300 using evenly distributed tile selection. In cases where 15% of the number of tiles is lower than  $M$ , tiles are extracted with overlap, to guarantee that  $M$  tiles can be selected. The MLP will process these 300 HER2 scores and output a single HER2 status label for the respective slide.

## 4. Experimental Settings

### 4.1. Data

The dataset is composed of subsets of WSI from two public datasets: the HER2 Scoring Contest (HER2SC) training set [10] and the TCGA-TCIA-BRCA (BRCA) collection [11,12]. The HER2SC training set (the subset with available labelling) comprises WSI of sections of 52 cases of invasive BCa stained with both IHC and H&E (example in Figure 2a,b). From this set, all IHC and H&E stained slides were used, except 4 H&E excluded because of manual ink markings. The subset from the BRCA dataset includes 54 H&E stained WSI (example in Figure 2c). All slides have the same original resolution and are weakly annotated with HER2 status (negative/positive) and score (0+, 1+, 2+, 3+), obtained from the corresponding histopathological reports.

The IHC stained slides were manually segmented into regions of interest (ROI), using the Sedeen Viewer software [36]. However, it is noteworthy that these slides were only used for training and, thus this step is not needed for testing.

The training and validation sets, used for model parameter tuning and optimization, have 40 and 12 IHC slides, respectively. A total of 7591 tiles per class have been extracted for training (30,364 tiles total) and 624 tiles per class extracted for validation (2496 tiles total), to keep a class balance.

### 4.2. Training Details

The hyperparameters used during training were empirically set to maximize performance. The CNN model for IHC tile scoring was randomly initialized and trained using the Adaptive Moment Estimation (Adam) [37] optimizer (learning rate of  $1 \times 10^{-5}$ ), to minimize a cross-entropy loss function, during 200 epochs, with mini-batches of 128 tiles. The soft-argmax used a parameter  $\beta = 1000$ . The aggregation MLP was trained using the Adam optimizer, with learning rate of  $10^{-5}$  for 150 epochs and mini-batches of 1 WSI (consisting of soft-argmax scores of the respective 300 tiles), saving the best considering validation accuracy.

## 5. Results and Discussion

### 5.1. Individual IHC Tile Scoring Results

After training, the model offered 76.8% accuracy (see Table 1). This indicates that the model was able to adequately discriminate against the IHC tiles between the four classes. This model was subsequently transferred for HER2 scoring in tiles from H&E slides.

**Table 1.** Confusion matrix of the CNN for HER2 scoring in IHC tiles.

		True Class			
		0	1	2	3
Prediction	0	490	132	2	0
	1	176	384	64	0
	2	45	159	419	1
	3	0	0	1	623

### 5.2. Invasive Tumor Tissue Segmentation

Tiles from H&E WSI are extracted from the intersection area between the HASHI-based invasive tumour segmentation and the Otsu-based tissue segmentation. The HASHI segmentation method was trained on the BRCA data reported as the test set by Cruz-Roa et al. [18], with 179 WSI on their original magnification. The results were comparable to the original paper (see examples in Figure 2) and were further evaluated by a pathology specialist, who confirmed the adequacy of the invasive tumour segmentation results.

### 5.3. Slide Scoring

On the HER2SC test set, this method achieved an F1-score of 86.7% and a weighted accuracy of 83.3% (see Table 2). Despite the small size of this test set, the proposed method was able to correctly classify all positive WSI and only misclassify one negative sample as positive. In this context, one might consider this a desirable behaviour, as false positives are less impactful than false negatives.

When tested on the BRCA test set, this method achieved an F1-score of 21.5% and a weighted accuracy of 53.8% (see Table 2). The method retains the behaviour presented in HER2SC, preferring to err on the side of false positives than the alternative. On the other hand, the performance metrics on BRCA differ considerably from those obtained on HER2SC. While the method was trained on HER2SC data, which is expected similar to the test data, the WSI of the BRCA dataset presents some notable differences. These slides have a greater extent of tissue, which generates more tiles per image and impacts the distribution of the scores, which may influence the method's behaviour.

**Table 2.** H&E HER2 status classification results of the proposed method on the HER2SC and BRCA test sets.

	Accuracy	F1-Score	Precision	Recall
HER2SC	83.3%	86.7%	89.6%	87.5%
BRCA	53.8%	21.5%	81.2%	31.5%

The evaluation results in single-database (HER2SC) and cross-database (BRCA) settings show the potential of the proposed method in standard and more challenging situations. However, the method appears to be dataset-dependent: it performed much better in conditions similar to the training. This should be addressed with additional efforts regarding domain adaptation.

The other shortcomings of the method appear to be related to the invasive tumour tissue segmentation and the individual tile scoring network, which could be improved with additional data and more accurate ground truth information. With these additional efforts, the proposed method could offer robust weakly-supervised WSI HER2 classification without IHC information.

#### 5.4. Ablation Study

Considering the lack of literature methods, to perform a benchmark, an ablation study was performed to confirm the capabilities of the proposed method. Experiments were conducted without IHC individual tile scoring CNN initialization, and using alternative statistical methods for individual tile score aggregation instead of MLP (median and mean), as can be seen in Tables 3 and 4. The results show that these alternatives are, in most settings, less adequate for the task at hand.

It is noteworthy that the median and mean-based aggregation are followed by a conversion to binary classes (0+ and 1+ are considered negative, while 2+ and 3+ are considered positive), since tiles have four possible labels. According to the guidelines, 2+ cases can be either negative or positive, but in an uncertain diagnosis scenario, it is preferable to classify them as positive.

**Table 3.** Results on the HER2SC test set.

Method	Accuracy	F1-Score	Precision	Recall
<b>MLP Aggregation:</b>				
proposed method	83.3%	86.7%	89.6%	87.5%
w/out pretrained CNN weights	62.5%	48.1%	39.1%	62.5%
<b>Median Aggregation:</b>				
w/pretrained CNN weights	50.0%	43.3%	78.6%	50.0%
w/out pretrained CNN weights	62.5%	48.1%	39.1%	62.5%
<b>Mean Aggregation:</b>				
w/pretrained CNN weights	50.0%	43.3%	78.6%	50.0%
w/out pretrained CNN weights	62.5%	48.1%	39.1%	62.5%

**Table 4.** Results on the BRCA test set.

Method	Accuracy	F1-Score	Precision	Recall
<b>MLP Aggregation:</b>				
Proposed method	53.3%	21.5%	81.2%	31.5%
w/out pretrained CNN weights	50.0%	60.3%	51.8%	72%
<b>Median Aggregation:</b>				
w/pretrained CNN weights	50.0%	12.3%	7.8%	28.0%
w/out pretrained CNN weights	52.2%	63.5%	66.5%	72.0%
<b>Mean Aggregation:</b>				
w/pretrained CNN weights	50.0%	12.3%	7.8%	28.0%
w/out pretrained CNN weights	52.2%	63.5%	66.5%	72.0%

## 6. Conclusions

In this work, a framework is proposed for the weakly supervised classification of HER2 overexpression status on H&E stained BCa WSI. The proposed approach integrates a CNN trained for HER2 scoring of individual H&E stained slide tiles, initialized with the network parameters pre-trained with data from IHC stained images. The objective of this initialization is to transfer some domain knowledge to the final training. The individual scores are aggregated on a single prediction per slide, returning the HER2 status label.

Tested with the BRCA data subset, the proposed method attained suitable performance. These preliminary results indicate that it is possible to accurately infer BCa HER2 status solely from



H&E stained slides. The results of an ablation study suggest that the proposed method with MLP tile score aggregation is more promising than simpler aggregation methods (mean or median).

Despite these results, further efforts should be devoted to performance improvements in this task. Firstly, the training of the tile HER2 scoring CNN and the aggregation MLP could be integrated into a single optimization process to achieve better performance. On the other hand, the aggregation of individual scores could use tile locations to take spatial consistency into account. Finally, the knowledge embedded in the networks through the pre-trained parameters could be better seized if input H&E tiles could be previously converted into IHC (possibly using generative adversarial models).

**Author Contributions:** Conceptualisation, S.P.O., J.R.P., T.G., J.S.C. and H.P.O.; Methodology, S.P.O., J.R.P., T.G. and J.S.C.; Data Curation, S.P.O. and R.C.-M.; Writing, S.P.O., J.R.P., T.G. and R.C.-M.; Review, J.S.C. and H.P.O.; Clinical Supervision, R.C.-M. and M.-J.C.; Technical Supervision, J.S.C. and H.P.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially funded by the Project “CLARE: Computer-Aided Cervical Cancer Screening” (POCI-01-0145-FEDER-028857), financially supported by FEDER through Operational Competitiveness Program—COMPETE 2020 and by National Funds through the Portuguese funding agency, FCT—Fundação para a Ciência e a Tecnologia, and also the FCT PhD grants “SFRH/BD/139108/2018” and “SFRH/BD/137720/2018”.

**Acknowledgments:** The results published here are in whole or part based upon data generated by the TCGA Research Network: <https://cancergenome.nih.gov/> and the HER2 Scoring Contest: <https://warwick.ac.uk/fac/sci/dcs/research/tia/her2contest>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

Adam	Adaptive Moment Estimation
ASCO/CAP	American Society of Clinical Oncology/College of American Pathologists
BCa	Breast cancer
BRCA	TCGA-TCIA-BRCA
CNN	Convolutional Neural Network
H&E	Haematoxylin and Eosin
HER2	Human Epidermal growth factor Receptor 2
HER2SC	HER2 Scoring Contest dataset
IHC	Immunohistochemistry
ISH	In situ Hybridization
MIL	Multiple Instance Learning
MLP	Multilayer Perceptron
ROI	Regions of Interest
WSI	Whole Slide Images

## References

1. American Cancer Society. Breast Cancer Facts & Figures 2017–2018. Available online: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/breast-cancer-facts-and-figures/breast-cancer-facts-and-figures-2017-2018.pdf> (accessed on 21 June 2020).
2. Gandomkar, Z.; Brennan, P.; Mello-Thoms, C. Computer-based image analysis in breast pathology. *J. Pathol. Inform.* **2016**, *7*. [CrossRef] [PubMed]
3. Veta, M.; Pluim, J.P.W.; van Diest, P.J.; Viergever, M.A. Breast Cancer Histopathology Image Analysis: A Review. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1400–1411. [CrossRef] [PubMed]
4. American Society of Clinical Oncology (ASCO). Breast Cancer Guide. 2005–2020. Available online: <https://www.cancer.net/cancer-types/breast-cancer/introduction> (accessed on 21 June 2020).
5. Rakha, E.A.; Pinder, S.E.; Bartlett, J.M.S.; Ibrahim, M.; Starczynski, J.; Carder, P.J.; Provenzano, E.; Hanby, A.; Hales, S.; Lee, A.H.S.; et al. Updated UK Recommendations for HER2 assessment in breast cancer. *J. Clin. Pathol.* **2015**, *68*, 93–99. [CrossRef] [PubMed]

6. Goddard, K.; Weinmann, S.; Richert-Boe, K.; Chen, C.; Bulkley, J.; Wax, C. HER2 Evaluation and Its Impact on Breast Cancer Treatment Decisions. *Public Health Genom.* **2011**, *15*, 1–10. [CrossRef]
7. Wolff, A.C.; Hammond, M.E.H.; Allison, K.H.; Harvey, B.E.; Mangu, P.B.; Bartlett, J.M.; Bilous, M.; Ellis, I.O.; Fitzgibbons, P.; Hanna, W.; et al. 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.* **2018**, *36*, 2105–2122. [CrossRef]
8. Hanna, W.M.; Barnes, P.J.; Chang, M.; Gilks, C.B.; Magliocco, A.M.; Rees, H.; Quenneville, L.; Robertson, S.J.; Sengupta, S.K.; Nofech-Mozes, S. Human epidermal growth factor receptor 2 testing in primary breast cancer in the era of standardized testing: A Canadian prospective study. *J. Clin. Oncol.* **2014**, *32*, 3967–3973. [CrossRef] [PubMed]
9. HEROHE Challenge. Available online: <https://ecdp2020.grand-challenge.org/> (accessed on 21 June 2020).
10. Qaiser, T.; Mukherjee, A.; Reddy PB, C.; Munugoti, S.D.; Tallam, V.; Pitkäaho, T.; Lehtimäki, T.; Naughton, T.; Berseth, M.; Pedraza, A.; et al. HER2 challenge contest: A detailed assessment of automated HER2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology* **2018**, *72*, 227–238. [CrossRef]
11. Clark, K.; Vendt, B.; Smith, K.; Freymann, J.; Kirby, J.; Koppel, P.; Moore, S.; Phillips, S.; Maffitt, D.; Pringle, M.; et al. The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J. Digit. Imaging* **2013**, *26*, 1045–1057. [CrossRef]
12. Lingle, W.; Erickson, B.J.; Zuley, M.L.; Jarosz, R.; Bonaccio, E.; Filippini, J.; Net, J.M.; Levi, L.; Morris, E.A.; Figler, G.G.; et al. Radiology Data from The Cancer Genome Atlas Breast Invasive Carcinoma [TCGA-BRCA] collection. *Cancer Imaging Arch.* **2016**. [CrossRef]
13. GitHub Repository with Code. Available online: <https://github.com/spoliveira/HERclassHE.git> (accessed on 21 June 2020).
14. Madabhushi, A.; Lee, G. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Med. Image Anal.* **2016**, *33*, 170–175. [CrossRef]
15. Robertson, S.; Azizpour, H.; Smith, K.; Hartman, J. Digital image analysis in breast pathology—From image processing techniques to artificial intelligence. *Transl. Res.* **2017**, *194*, 19–35. [CrossRef] [PubMed]
16. Tellez, D.; Balkenhol, M.; Otte-Höller, I.; van de Loo, R.; Vogels, R.; Bult, P.; Wauters, C.; Vreuls, W.; Mol, S.; Karssemeijer, N.; et al. Whole-Slide Mitosis Detection in H E Breast Histology Using PHH3 as a Reference to Train Distilled Stain-Invariant Convolutional Networks. *IEEE Trans. Med Imaging* **2018**, *37*, 2126–2136. [CrossRef]
17. Cai, D.; Sun, X.; Zhou, N.; Han, X.; Yao, J. Efficient Mitosis Detection in Breast Cancer Histology Images by RCNN. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 919–922. [CrossRef]
18. Cruz-Roa, A.; Gilmore, H.; Basavanahally, A.; Feldman, M.; Ganesan, S.; Shih, N.; Tomaszewski, J.; Madabhushi, A.; González, F. High-throughput adaptive sampling for whole-slide histopathology image analysis (HASHI) via convolutional neural networks: Application to invasive breast cancer detection. *PLoS ONE* **2018**, *13*, e0196828. [CrossRef] [PubMed]
19. Li, X.; Radulovic, M.; Kanjer, K.; Plataniotis, K.N. Discriminative Pattern Mining for Breast Cancer Histopathology Image Classification via Fully Convolutional Autoencoder. *arXiv* **2019**, arXiv:1902.08670.
20. Romero, F.P.; Tang, A.; Kadoury, S. Multi-Level Batch Normalization In Deep Networks For Invasive Ductal Carcinoma Cell Discrimination In Histopathology Images. *arXiv* **2019**, arXiv:1901.03684.
21. Wan, T.; Cao, J.; Chen, J.; Qin, Z. Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features. *Neurocomputing* **2017**, *229*, 34–44. [CrossRef]
22. Cao, H.; Bernard, S.; Heutte, L.; Sabourin, R. Improve the Performance of Transfer Learning Without Fine-Tuning Using Dissimilarity-Based Multi-view Learning for Breast Cancer Histology Images. In *Image Analysis and Recognition*; Campilho, A., Karray, F., Ter Haar Romeny, B., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 779–787. [CrossRef]
23. Vo, D.M.; Nguyen, N.Q.; Lee, S.W. Classification of breast cancer histology images using incremental boosting convolution networks. *Inf. Sci.* **2019**, *482*, 123–138. [CrossRef]
24. Oscanoa, J.; Doimi, F.; Dyer, R.; Araujo, J.; Pinto, J.; Castaneda, B. Automated segmentation and classification of cell nuclei in immunohistochemical breast cancer images with estrogen receptor marker. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 2399–2402. [CrossRef]

25. Saha, M.; Chakraborty, C.; Arun, I.; Ahmed, R.; Chatterjee, S. An Advanced Deep Learning Approach for Ki-67 Stained Hotspot Detection and Proliferation Rate Scoring for Prognostic Evaluation of Breast Cancer. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
26. Jamaluddin, M.F.; Fauzi, M.F.A.; Abas, F.S.; Lee, J.T.H.; Khor, S.Y.; Teoh, K.H.; Looi, L.M. Cell Classification in ER-Stained Whole Slide Breast Cancer Images Using Convolutional Neural Network. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 632–635. [[CrossRef](#)]
27. Rodner, E.; Simon, M.; Denzler, J. Deep bilinear features for Her2 scoring in digital pathology. *Curr. Dir. Biomed. Eng.* **2017**, *3*, 811–814. [[CrossRef](#)]
28. Mukundan, R. A Robust Algorithm for Automated HER2 Scoring in Breast Cancer Histology Slides Using Characteristic Curves. In *Medical Image Understanding and Analysis (MIUA)*; Springer: Cham, Switzerland, 2017; Volume 723, pp. 386–397. [[CrossRef](#)]
29. Dou, Q.; Ouyang, C.; Chen, C.; Chen, H.; Heng, P.A. Unsupervised Cross-Modality Domain Adaptation of Convnets for Biomedical Image Segmentations with Adversarial Loss. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*; AAAI Press: Menlo Park, CA, USA, 2018; pp. 691–697. [[CrossRef](#)]
30. Aderghal, K.; Khvostikov, A.; Krylov, A.; Benois-Pineau, J.; Afdel, K.; Catheline, G. Classification of Alzheimer Disease on Imaging Modalities with Deep CNNs Using Cross-Modal Transfer Learning. In Proceedings of the 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS), Karlstad, Sweden, 18–21 June 2018; pp. 345–350. [[CrossRef](#)]
31. Hadad, O.; Bakalo, R.; Ben-Ari, R.; Hashoul, S.; Amit, G. Classification of breast lesions using cross-modal deep learning. In Proceedings of the 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 109–112. [[CrossRef](#)]
32. Bulten, W.; Bándi, P.; Hoven, J.; van de Loo, R.; Lotz, J.; Weiss, N.; van der Laak, J.; van Ginneken, B.; Hulsbergen-van de Kaa, C.; Litjens, G. Epithelium segmentation using deep learning in H&E-stained prostate specimens with immunohistochemistry as reference standard. *Sci. Rep.* **2019**, *9*. [[CrossRef](#)]
33. Cruz-Roa, A.; Gilmore, H.L.; Basavanthally, A.; Feldman, M.D.; Ganesan, S.; Shih, N.C.; Tomaszewski, J.P.; Gonzalez, F.A.; Madabhushi, A. Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent. *Sci. Rep.* **2017**, *7*. [[CrossRef](#)] [[PubMed](#)]
34. Honari, S.; Molchanov, P.; Tyree, S.; Vincent, P.; Pal, C.; Kautz, J. Improving landmark localization with semi-supervised learning. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 19–21 June 2018; pp. 1546–1555. [[CrossRef](#)]
35. Chapelle, O.; Wu, M. Gradient descent optimization of smoothed information retrieval metrics. *Inf. Retr.* **2010**, *13*, 216–235. [[CrossRef](#)]
36. Seseen Viewer Software. Available online: <https://pathcore.com/seseen/> (accessed on 21 June 2020).
37. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.

