# A Robust Algorithm for Automated HER2 Scoring in Breast Cancer Histology Slides Using Characteristic Curves

Ramakrishnan Mukundan (0000-0003-4578-1931)

Department of Computer Science and Software Engineering
University of Canterbury, Christchurch, New Zealand.
mukundan@canterbury.ac.nz

**Abstract.** This paper presents a novel feature descriptor and classification algorithms for automated scoring of HER2 in Whole Slide Images (WSI). Since a large amount of processing is involved in analyzing WSI images, the primary design goal has been to keep the computational complexity to the minimum possible level. We propose an efficient method based on characteristic curves which encode all relevant information in a smooth polynomial curve with the percentage of stained membranes plotted against variations in intensity/saturation of the colour thresholds used for segmentation. Our algorithm performed exceedingly well at a recent online contest held by the University of Warwick [1], obtaining the second best points score of 390 out of 420 and the overall seventh position in the combined leaderboard [2]. The paper describes three classification algorithms with features extracted from characteristic curves and provides experimental results and comparative analysis.

**Keywords:** Whole Slide Image processing. Automated HER2 scoring. Medical image classification. Characteristic curves. Digital pathology.

## 1 Introduction

The most commonly used method for breast cancer grading is the ImmunoHistoChemistry (IHC) test which is a staining process performed on biopsy samples of breast cancer tissues [3]. The IHC stained slides are normally observed under a microscope by pathologists to determine the level of over-expression of Human Epidermal Growth factor Receptor 2 (HER2) protein in cancer cells. The tissue sample is then assigned a HER2 score of 0 to 3+ representing the grade of cancer present in the sample [4]. Manual grading and annotations of breast cancer slides are time consuming, and there are huge maintenance costs associated with collecting, archiving, and transporting tissue specimens. It is also well documented that manual grading can have significant variability in pathologist assessments due to the subjective process of determining the intensity and uniformity of staining in the presence of variable staining patterns and heterogeneity of tumor grade [5]. Automated methods can also suffer

from errors due to inaccuracies in the training algorithm and its inability to segment faint and complex tissue structures [6].

In the rapidly growing field of digital pathology, several Whole Slide Image (WSI) processing algorithms are currently being developed as diagnostic tools to help pathologists in the assessment of disease patterns [7]. WSIs have a pyramidal structure to enable optimized viewing across multiple magnification levels, and they provide a high resolution overview of the entire slide [7,8]. Typically, at 40x magnification, the images have a resolution of approximately 0.25 microns per pixel. At this resolution, a slide region of size 15mm x 15mm could correspond to 60,000 x 60,000 pixels. WSIs were originally used as a computer aided digital microscopy tool, where pathologists could view different parts of a sample at different magnifications to improve the accuracy of their scores [5]. Powerful computational algorithms are being developed to automatically extract features related to cytological and protein structures in the image for accurately quantifying biomarkers like HER2 [9]. In the past, similar studies for quantitative IHC were performed using images of lower resolution [10].

Recently, an online contest was organized by the University of Warwick in conjunction with the UK/Ireland Pathology Society annual meeting 2016, with the aim of advancing research in the field of WSI-based automated HER2 scoring algorithms [1]. This contest was the primary motivation for our research work presented in this paper. Our algorithm (registered with team name UC-CSSE-CGIP) performed exceedingly well in the contest, obtaining the second best points score of 390 out of 420 and the overall seventh position in the combined leader board [2]. The teams that were on the top of the leader board, including our team, were invited to submit a very brief (one paragraph) summary of the algorithms used for inclusion in a journal paper prepared by the contest organizers [12].

WSIs contain voluminous amounts of data. One of the primary design goals has been to keep the computational complexity to the minimum possible level and to develop an efficient method that can process relevant tiles of an input WSI image quickly and classify the image into one of the four classes corresponding to the four HER2 scores. The second design goal was to have a feature set whose correlation to the percentage of membrane staining in the given sample could be easily visualized and interpreted by pathologists. The third design goal was to reduce the amount of information redundancy in the feature set by extracting a minimal set of characteristic features that would adequately represent the staining pattern. This paper presents classification algorithms using characteristic curves, providing detailed descriptions of the processing stages, development and selection of features, and the experimental analysis performed. We hope that the methods presented in this paper will contribute significantly to the development of faster and accurate automatic HER2 scoring techniques in the area of breast cancer histopathology.

The paper is organized as follows: The next section gives a description of the dataset used, an outline of HER2 assessment scheme and an overview of the stages of the processing pipeline. Section 3 provides an introduction to a novel set of features called characteristic curves. Section 4 gives a description of the classification algorithms using characteristic curves, and Section 5 presents experimental results and

comparative analysis. Section 6 gives a summary of the work reported in this paper and outlines future directions.

## 2 Materials and Methods

### 2.1 HER2 Assessment

The assessment of HER2 protein over-expression is done based on the percentage of membrane staining observed in tumor cells as well as the intensity of staining [4]. The mapping between the level of membrane staining and the reported HER2 score is shown in Table 1.

**Table 1.** Correspondence between HER2 scores and membrane staining [4].

| HER2 Score | Assessment | Staining Pattern |
|---|---|---|
| 0 | Negative | No staining is observed, or membrane staining is observed in less than 10% of tumor cells |
| 1+ | Negative | A faint/barely perceptible membrane staining is detected in greater than 10% of tumor cells. The cells exhibit incomplete membrane staining. |
| 2+ | Weakly Positive | A weak to moderate membrane staining is observed in greater than 10% of tumor cells. |
| 3+ | Positive | A strong complete membrane staining is observed in greater than 10% of tumor cells. |

The WSI image segments of Immunohistochemical (IHC) stained slides given in Fig. 1 correspond to different HER2 scores and show the variations in the level of membrane staining.

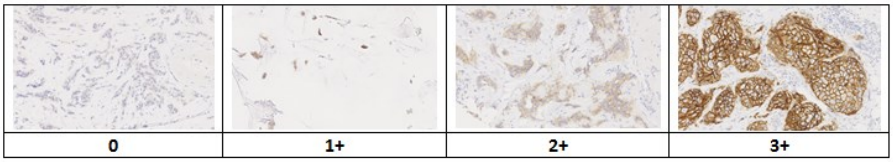


**Fig. 1.** WSI tiles showing different levels of staining and corresponding HER2 scores.

### 2.2 Dataset

The dataset used in this research work was provided by the University of Warwick as part of the online HER2 scoring contest [1]. Permission was granted by the contest organizers to participating teams for the use of the dataset for research and academic purposes. The dataset consisted of a total of 172 whole slide images in Nano-zoomer Digital Pathology (NDPI) format. These WSIs were extracted from 86 cases of pa-

tients with invasive breast carcinomas [12]. For each case, WSIs of both Hematoxylin and Eosin (H&E) stained and Immunohistochemical (IHC) stained slides were provided. There were two HER2 scoring contests, and the number of WSIs provided for training and testing the classification algorithm is given in Table 2. The training data included ground truth provided by expert pathologists and consisted of the HER2 score assigned for each case and also the percentage of membrane staining in the tissue sample.

**Table 2.** Number of WSIs provided for training and testing the classification algorithm.

| **Training Set** | | **Test Set** | |
|---|---|---|---|
| HER2 Score (Ground Truth) | Number of WSIs | Contest-1 No. of WSIs | Contest-2 No. of WSIs |
| 0 | 13 | **28** | **6** |
| 1+ | 13 | | |
| 2+ | 13 | | |
| 3+ | 13 | | |
| Total | **52** | | |

### 2.3 Processing Stages

Various stages of the processing pipeline are shown in Fig. 2. We used the OpenSlide API [11] to read WSIs of IHC stained slides, and a region of interest (ROI) containing a significant portion of the imaged tissue is extracted from the middle segment of the image. Rectangular tiles of size 1800×1200 pixels at 20x magnificaiton that contain at most 20% background pixels are then created and used as inputs for the method that computes characteristic curves. At least six tiles at randomly selected locations within the ROI are generated for each WSI. The remaining part of the pipeline computes the percentage of staining in the tissue sample to obtain the characteristic curve as detailed in the next section.

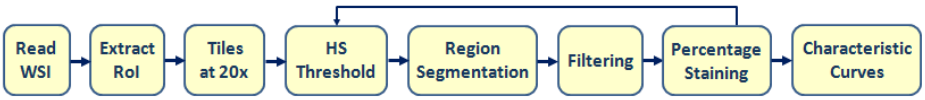


**Fig. 2.** Processing stages in the extraction of characteristic curves.

## 3 Characteristic Curves

In this section, we introduce a novel feature vector called a characteristic curve. An important parameter in HER2 assessment is the percentage of membrane staining perceived in an image segment. Assuming that we can compute the percentage of membranes stained in a particular colour range (this computation will be discussed in detail below), we can analyse the variations in this percentage value with respect to

changes in the colour saturation threshold. Specifically, if [*h, s, v*] represent the stain colour components in HSV space, and if $p(s_{low})$ denotes the pecentage of staining with colour in the range given by the following inequalities:

$$h_1 \leq h < h_2$$

$$s > s_{low}$$

$$v_1 \leq v < v_2 \tag{1}$$

then, the variation of $p(s_{low})$ plotted against $s_{low}$ gives the characteristic curve (or the percentage-saturation curve) of the image. In eq.(1), [$h_1$, $h_2$] denote fixed hue thresholds specifying allowable variations in the hue value, and similarly [$v_1$, $v_2$] denote value thresholds. Since we specify only the lower bound for saturation, progressively increasing $s_{low}$, typically from 0.1 to 0.5, produces a non-increasing characteristic curve (Fig. 3). In our experiments, we used the following threshold values: $h_1 = 0$, $h_2 = 0.1$, $s_{low} = 0.1$, $v_1 = 0$, $v_2 = 1$.

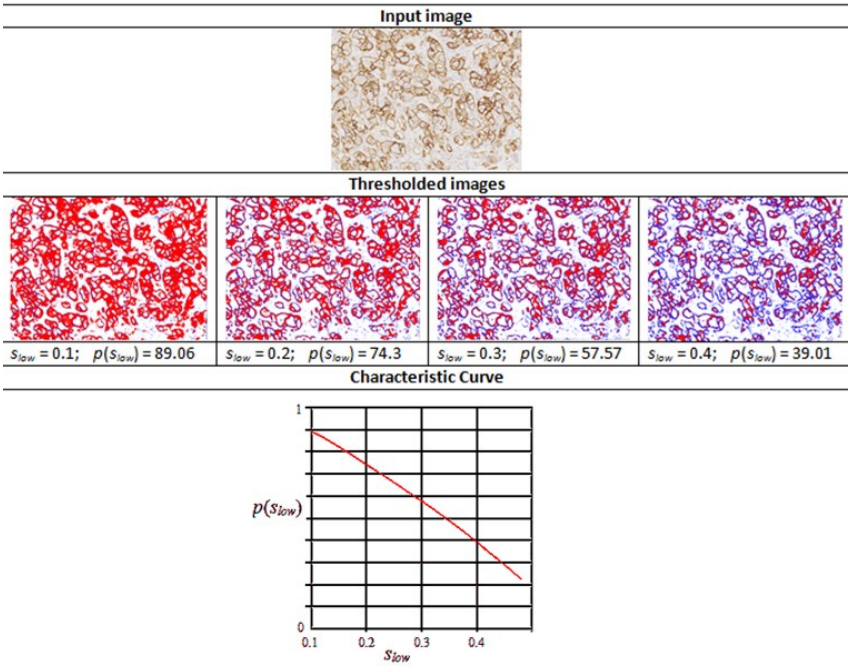


**Fig. 3.** Intermediate stages in the generation of a characteristic curve.

The base components of the stain colour [*h*, *s*, *v*] are computed using the training set where the given percentage of staining is above 80%. While computing the percentage of staining for the test (or cross-validation) sets, it is important to eliminate not only the background region but also other segments that are not part of the membrane

region such as connective tissues, lobules and nuclei. These regions can be segmented using colour (nuclei are stained in a distinctly different colour) or using a distance measure evaluated in colour space over a neighbourhood mask around each pixel (for identifying regions of nearly constant colour value). Fig. 3 shows thresholded images with stained regions in red colour as the value of $s_{low}$ is increased from 0.1 to 0.4. The resulting characteristic curve is also shown. The characteristics curves have the property that they are always monotonically decreasing smooth curves. They allow accurate polynomial approximations using cubic curves. The shape of the curve can be directly matched with the staining patterns given in the HER2 assessment guidelines (Table 1) for a straightforward interpretation of the derived score (Fig. 4). For example, the characteristic curve always lies below the 10% threshold when the score is 0, and only a small initial segment of the curve lies above the 10% mark when the score is 1. If the score is 3+, the curve lies completely above the 30% mark showing a strong and complete membrane staining. As seen in Fig. 4, the curve passes through a much wider range of values of percentage staining when the score is 2+.

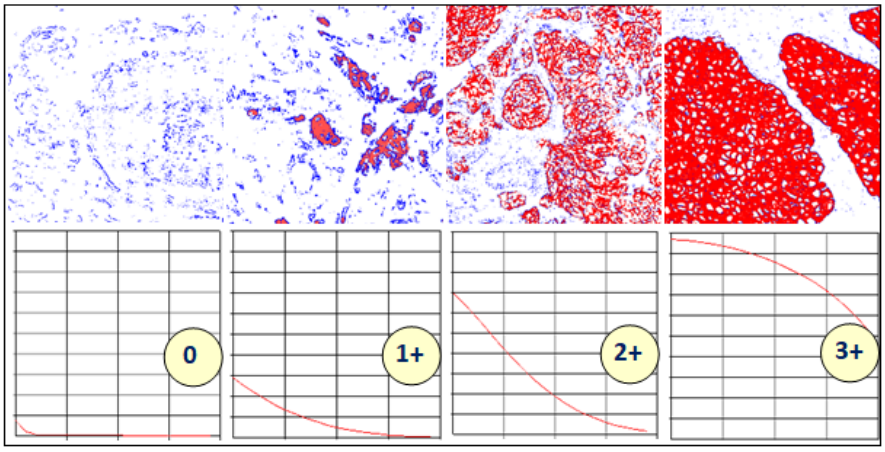


**Fig. 4.** Variations in the shapes of the characteristic curves with different levels of staining.

## 4 Classification

The properties of the characteristic curve outlined in the previous section, particularly the fact that the curve is non-increasing, can be effectively used for developing a rule-based classification algorithm as follows.

- if $z_0$ ( = $p(0.1)$) < 10%, then the whole curve lies below 10%, and the score is 0 (rule 1)
- else if $z_{n-1}$ ( = $p(0.5)$) > 30%, then the whole curve lies above 30%, and the score is 3+ (rule 2)
- else if 10% ≤ $z_0$ ( = $p(0.1)$) < 40% and $p(0.2)$ < 15%, the score is 1+ (rule 3)

- else if $p(0.4) < 15\%$, then the score is 2+ (rule 4)
- else, the score is 3+ (rule 5)

The rules were formed by analyzing the shapes of characteristic curves for several image tiles with ground truth values of HER2 scores assigned by pathologists. Note that for the above simple classification algorithm, we sample the curve at only four key points $p(0.1)$, $p(0.2)$, $p(0.4)$, and $p(0.5)$. As discussed in the next section on experimental results, the rule based algorithm is primarily used to assess the feature representation capability of the characteristic curves.

For more accurate classification, we use the 'one-vs-all' multi-class classification algorithm using logistic regression [13]. For a given training example with index $j$, the points sampled along its characteristic curve $x_i^{(j)} = p(s_i)$, $i = 1..n$, $j = 1..m$ are used as features. The class labels are denoted by $y_j \in [0, 3]$, $j = 1..m$. We denote the feature matrix by $X \in \Re^{m \times (n+1)}$, the output vector of labels by $Y \in \Re^{m \times 1}$, and the classifier parameter vector for each class by $\theta_k \in \Re^{(n+1) \times 1}$, $k = 1..4$. Here, class-1 corresponds to the set of training examples with HER2 score 1+, class-2 with HER2 score 2+, class-3 with HER2 score 3+ and class-4 with HER2 score 0. The hypothesis function vector $H \in \Re^{m \times 1}$ is given by $H = g(X\theta_k)$, where $g()$ denotes the sigmoid function. For prediction, the points $x_i$ on the characteristic curve of given sample are combined with the trained values of class parameters $\theta_k$ for each class $k = 1..4$, and the class that gives the maximum value for $g(x_i'\theta_k)$ is chosen. In the next section, we provide the result of classification experiments using the above methods.

## 5 Experimental Results and Analysis

First, we provide the results for the rule-based classification algorithm. The percentage of staining values $p()$ obtained from the characteristic curves computed for some of the WSI images in the training set are given below. The table also gives the ground truth values and the predicted scores computed using the five rules given in the previous section. For each case, three segments of the WSI (tiles) at 20x magnification were used. The incorrect predictions are highlighted in red colour.

The overall performance of the rule-based classification algorithm can be seen in the confusion matrix below. 52 WSIs with 3 tiles at 20x from each image (comprising of 156 images) were used in this experiment as the training data. Another set of 3 tiles from each of the 52 cases formed the cross-validation set. Out of the total of 156 image tiles in the cross-validation set, 39 belonged to each of the four classes corresponding to four HER2 scores. As seen in Table 4, a few images for cases with score 1+ were wrongly classified as having either 0 or 2+ scores, while all images with score 3+ were correctly classified. These results of the rule based algorithm are presented here only to show that one could roughly estimate the HER2 scores directly from the shapes of the characteristic curves.

**Table 3.** Sampled values of the characteristics curves and the predicted scores obtained a set of WSI images in the training set.

| CaseNo (tileNo) | $p(0.1)$ | $p(0.2)$ | $p(0.4)$ | $p(0.5)$ | Ground Truth | Predicted | Rule |
|---|---|---|---|---|---|---|---|
| 1 (1) | 0.72 | 0 | 0 | 0 | 0 | 0 | Rule-1 |
| 1 (2) | 7.16 | 0.01 | 0 | 0 | 0 | 0 | Rule-1 |
| 1 (3) | 7.21 | 0.01 | 0 | 0 | 0 | 0 | Rule-1 |
| 12 (1) | 14.31 | 4.09 | 0.10 | 0 | 1 | 1 | Rule-3 |
| 12 (2) | 35.81 | 13.02 | 0.61 | 0.04 | 1 | 1 | Rule-3 |
| 12 (3) | 28.76 | 13.44 | 1.25 | 0.19 | 1 | 1 | Rule-3 |
| 15 (1) | 76.16 | 22.07 | 0.02 | 0.00 | 1 | 2 | Rule-4 |
| 15(2) | 74.68 | 22.6 | 0.18 | 0 | 1 | 2 | Rule-4 |
| 15 (3) | 17.97 | 0.33 | 0 | 0 | 1 | 1 | Rule-3 |
| 16(1) | 8.09 | 0.91 | 0 | 0 | 1 | 0 | Rule-1 |
| 16(2) | 11.44 | 0.63 | 0 | 0 | 1 | 1 | Rule-3 |
| 16(3) | 1.98 | 0.09 | 0 | 0 | 1 | 0 | Rule-1 |
| 25 (1) | 75.79 | 36.06 | 1.21 | 0.08 | 2 | 2 | Rule-4 |
| 25 (2) | 48.12 | 15.33 | 0.52 | 0.03 | 2 | 2 | Rule-4 |
| 25 (3) | 61.18 | 24.24 | 0.56 | 0.02 | 2 | 2 | Rule-4 |
| 33 (1) | 88.64 | 83.33 | 67.65 | 46.96 | 3 | 3 | Rule-2 |
| 33 (2) | 90.72 | 87.27 | 70.15 | 50.13 | 3 | 3 | Rule-2 |
| 33 (3) | 86.22 | 82.62 | 68.48 | 50.39 | 3 | 3 | Rule-2 |
| 84 (1) | 77.53 | 66.29 | 38.37 | 16.4 | 3 | 3 | Rule-5 |
| 84 (2) | 75.20 | 63.42 | 39.64 | 20.35 | 3 | 3 | Rule-5 |
| 84 (3) | 57.81 | 44.88 | 21.42 | 6.65 | 3 | 3 | Rule-5 |

**Table 4.** Confusion matrix showing the performance of the rule based classification method.

| | | Predicted | | | | Accuracy = 80.12% | |
|---|---|---|---|---|---|---|---|
| | | **0** | **1+** | **2+** | **3+** | **Precision** | **Recall** |
| Actual | **0** | 30 | 7 | 2 | 0 | 0.81 | 0.77 |
| | **1+** | 7 | 23 | 9 | 0 | 0.74 | 0.59 |
| | **2+** | 0 | 1 | 33 | 5 | 0.75 | 0.84 |
| | **3+** | 0 | 0 | 0 | 39 | 0.88 | 1.0 |

The results given above show that even a minimal set of four points derived from the characteristic curves can have a good discriminating power. The accuracy can be further improved by including the slope information at the key points also in the classification rules (Table 5). The slope at the point $p(s_i)$ is computed as

$$p'(s_i) = \frac{1}{0.02}\left(p(s_i) - p(s_i + 0.02)\right) \tag{2}$$

**Table 5.** Confusion matrix for the rule based classification method augmented by the slope information..

| | | Predicted | | | | Accuracy = 85.25% | |
|---|---|---|---|---|---|---|---|
| | | **0** | **1+** | **2+** | **3+** | **Precision** | **Recall** |
| Actual | **0** | 35 | 4 | 0 | 0 | 0.83 | 0.89 |
| | **1+** | 7 | 25 | 7 | 0 | 0.83 | 0.64 |
| | **2+** | 0 | 1 | 34 | 4 | 0.82 | 0.87 |
| | **3+** | 0 | 0 | 0 | 39 | 0.9 | 1.0 |

For generating feature vectors for classification using logistic regression, it was found that a step size of 0.02 for the saturation threshold would provide an adequate number of 20 points (features) within the saturation range [0.1, 0.5]. The feature matrix $X$ in Eq.(2) therefore had the dimension 156×20. The gradient descent algorithm used 100 iterations to converge to the solution with a learning rate of 0.001 (Fig. 5).

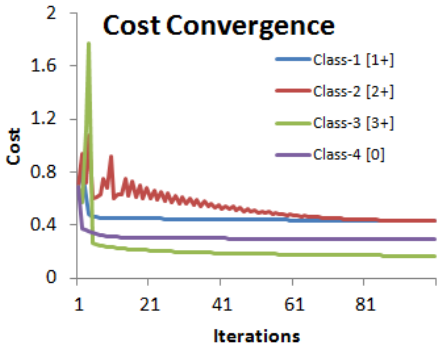


**Fig. 5.** Convergence of the cost functions of the multi-class logistic regression algorithm.

The confusion matrix showing the improvement of accuracy on the rule-based method is given in Table 6. Note that this method had 20 features for each sample, while the rule-based method used only four points from the characteristic curve.

**Table 6.** Confusion matrix for the multi-class logistic regression algorithm.

| | | Predicted | | | | Accuracy = 88.46% | |
|---|---|---|---|---|---|---|---|
| | | **0** | **1+** | **2+** | **3+** | **Precision** | **Recall** |
| Actual | **0** | 37 | 2 | 0 | 0 | 0.86 | 0.95 |
| | **1+** | 6 | 29 | 4 | 0 | 0.83 | 0.74 |
| | **2+** | 0 | 4 | 34 | 1 | 0.87 | 0.87 |
| | **3+** | 0 | 0 | 1 | 38 | 0.97 | 0.97 |

The smoothness and monotonically decreasing properties of the characteristic curve can be effectively made use of in reducing the dimensionality of the features in the

logistic regression algorithm. As in the case of the rule based classification method, we can sample the curve at only four key points $p(0.1)$, $p(0.2)$, $p(0.4)$, and $p(0.5)$, and also use the slope information at those points $p'(0.1)$, $p'(0.2)$, $p'(0.4)$, and $p'(0.5)$ to get a feature vector of size 8 instead of 20. The cost functions converge to almost similar values with only a slight increase in the magnitudes (Fig. 6).

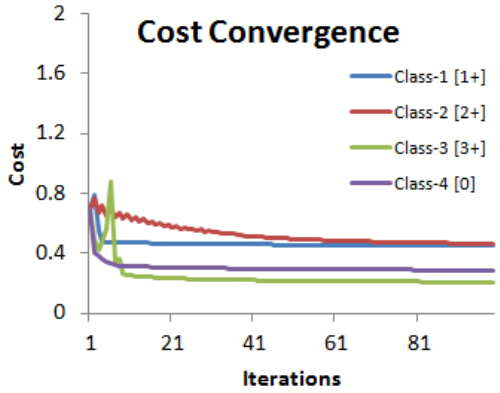


**Fig. 6.** Convergence of the cost functions with reduced feature set.

The confusion matrix obtained by running the algorithm with the reduced set of features of the characteristic curve is shown in Table 7.

**Table 7.** Confusion matrix for the multi-class logistic regression algorithm with reduced feature set.

| | | Predicted | | | | Accuracy = 83.3% | |
|---|---|---|---|---|---|---|---|
| | | **0** | **1+** | **2+** | **3+** | **Precision** | **Recall** |
| Actual | **0** | 37 | 2 | 0 | 0 | 0.80 | 0.95 |
| | **1+** | 8 | 24 | 7 | 0 | 0.75 | 0.61 |
| | **2+** | 1 | 6 | 31 | 1 | 0.79 | 0.79 |
| | **3+** | 0 | 0 | 1 | 38 | 0.97 | 0.97 |

The rule-based classification algorithm with augmented slope information is computationally fast, and provides very good results for all classes except class-1 (corresponding to HER2 score 1+) where the recall rate is 0.64 (Table 5). All methods gave the lowest recall rate for this class. This is because the characteristic curves for several tiles with HER2 score 1+ grossly resembled the shape of curves with score 0 or 2+. However, the performance of the rule-based method for this class is even better than logistic regression with reduced feature set (Table 7). Multi-class logistic regression gave better results in all remaining classes. Reducing the dimensionality of the feature set from 20 to 8 only affected the recall rates of classes 1 and 2. Overall, logistic regression with 20 feature points gave the highest accuracy of 88.5%.

Analysing the staining patterns in tiles that were wrongly classified revealed a common problem in the automatic extraction of tiles from WSIs (see Fig. 2). Some of the samples with scores 1+ and 2+ had large tissue regions without any staining. The example shown in Fig. 7 contains a tissue sample at 10x magnification with an assigned score of 2+.

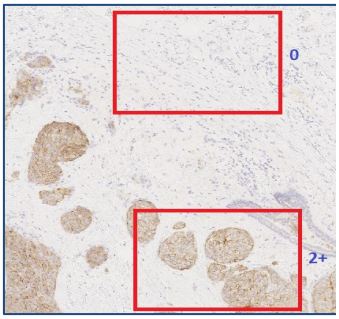


**Fig. 7.** An example showing two tile positions with varying image characteristics within the same WSI.

In Fig.7, the tile on the top didn't contain any stained membrane regions and was assigned a ground truth value of 2+ at the training stage, and a predicted value of 0 at the cross-validation stage. This tile could have been a valid part of any WSI with a score 0, and therefore there is no way by which such tiles can be identified and discarded by the automatic tile extraction method. Manually identifying such tiles from the training and cross-validation sets significantly improved the scores of the classification algorithms. The tile on the bottom half of Fig. 7 was assigned the correct score of 2+.

## 6 Conclusions

This paper has introduced a novel feature descriptor called a characteristic curve that could be effectively used in classification algorithms for automated scoring of HER2 in breast cancer histology slides. The computational aspects of characteristic curves and their shape features that embed information on the staining patterns for different HER2 scores have been discussed in detail. The usefulness of features based on characteristic curves and their applications in classification algorithms have been demonstrated through experimental results obtained using a comprehensive WSI dataset provided by the University of Warwick[1]. The results show that the features used with a multi-class classification algorithm such as logistic regression can provide very good levels of accuracy. The paper also outlined computational stages in the overall processing pipeline for automatic HER2 scoring using WSI files as inputs.

Experimental results showed the need for further improving the discriminating power of the characteristic curves by developing methods for accurate identification of membrane morphology and region segmentation, particularly for samples with an assigned HER2 score 1+. It is also necessary to assess the reproducibility of results,

specifically inter-scanner variability [14] of the rule-based classification algorithm as the rules were formed using data produced by a single scanner.

## 7 References


1. Department of Computer Science, University of Warwick: Her2 Scoring Contest. http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/her2contest/
2. Department of Computer Science, University of Warwick: Her2 Contest Results. http://www2.warwick.ac.uk/fac/sci/dcs/research/combi/research/bic/her2contest/outcome.
3. Hicks DG, Schiffhauer L (2015) Standardized assessment of the HER2 status in breast cancer by immunohistochemistry. Lab. Med. vol 42(8) pp 459-467. doi: 10.1309/LMGZZ58CTS0DBGTW
4. Rakha EA, et.al. (2015) Updated UK recommendations for HER2 assessment in breast cancer. J. Clin. Pathol.. vol 68. pp 93-99. doi: 10.1136/jclinpath-2014-202571
5. Gavrielides MA, Gallas BD, Lenz P, Badano A, Hewitt SM (2011) Observer variability in the interpretation of HER2 immunohistochemical expression with unaided and computer aided digital microscopy. Arch Pathol Lab Med. vol 135(2). pp 233-242. doi: 10.1043/1543-2165-135.2.233
6. Akbar S, Jordan LB, Purdie CA, Thompson AM, McKenna SJ (2015) Comparing computer-generated and pathologist-generated tumor segmentations for immunohistochemical scoring of breast tissue microarrays. Br J Cancer 113(7) pp 1075-1080. doi: 10.1038/bjc.2015.309
7. Hamilton PW et.al. (2014) Digital pathology and image analysis in tissue biomarker research. Methods 70(1) pp 59-73. doi: 10.1016/j.ymeth.2014.06.015
8. Farahani N, Parwani AV, Pantanowitz L (2015) Whole slide imaging in pathology: advantages, limitations and emerging perspectives. Path. Lab. Med. Intl. vol 7 pp 23-33. doi: 10.2147/PLMI.S59826
9. Ghaznavi F, Evan A, Madabhushi A, Feldman M (2013) Digital imaging in pathology: Whole-slide imaging and beyond. Annul. Rev. Pathol. Mech. Dis. vol 8 pp 31-59. doi: 10.1146/annurev-pathol-011811-120902
10. Matkowskyj KA, Cox R, Jensen RT, Benya RV (2003) Quantitative immunohistochemistry by measuring cumulative signal strength accurately measures receptor number. Journal of Histochemistry & Cytochemistry. vol 51(2). 205-214. doi: 10.1177/002215540305100209
11. Goode A, Gilbert B, Harkes J, Jukie D, Satyanarayanan M (2013) OpenSlide: A vendor-neutral software foundation for digital pathology. J. Pathol. Inform. vol 4(27). doi: 10.4103/2153-3539.119005.
12. Qaiser T, et.al. (2017) HER2 Challenge Contest: A detailed assessment of HER2 scoring algorithms and man vs machine in whole slide images of breast cancer tissues. Submitted to Histopathology (Wiley).
13. Watt J, Borhani R, Katsaggelos AK (2016) Machine Learning Refined: Foundations, Algorithms and Applications. Cambridge Uni. Press.
14. Keay T, et.al. (2013) Reproducibility in the automated quantitative assessment of HER2/neu for breast cancer. J. Pathol Inform. vol 4(19). doi: 10.4103/2153-3539.115879.