An Automatic Patch-based Approach for HER-2 Scoring in Immunohistochemical Breast Cancer Images Using Color Features

Caroline Quadros Cordeiro[1], Sergio Ossamu Ioshii[2], Jeovane Honório Alves[1], Lucas Ferrari de Oliveira[1]

[1]Universidade Federal do Paraná (UFPR)
Laboratório de Visão, Robótica e Imagem (VRI)
Curitiba - PR, Brasil

[2]Pontifícia Universidade Católica do Paraná (PUCPR)
Pós-Graduação de Tecnologia em Saúde da PUCPR
Curitiba - PR, Brasil

cqcordeiro@inf.ufpr.br, sergio.ioshii@pucpr.br

{jhalves, lferrari}@inf.ufpr.br

Abstract. *Breast cancer (BC) is the most common cancer among women worldwide, approximately 20-25% of BCs are HER-2 positive. Analysis of HER-2 is fundamental to defining the appropriate therapy for patients with breast cancer. Inter-pathologist variability in the test results can affect diagnostic accuracy. The present study intends to propose an automatic scoring HER-2 algorithm. Based on color features, the technique is fully-automated and avoids segmentation, showing a concordance higher than 90% with a pathologist in the experiments realized.*

1. Introduction

Cancer is a disease with a high mortality rate that has increasingly reached the world's population, especially the female population. In Brazil, Breast Cancer (BC) is the most common tumor among women, affecting almost 60,000 patients in 2014 [INCA 2014]. BC is the second most common tumor worldwide. In US one out of eight women are affected by BC during their lifetime [DeSantis et al. 2014]. In the last decade, incidence of cancer has grown 20% in the world. In Brazil, *National Institute of Cancer José Alencar Gomes da Silva* (INCA) estimates 59,700 new BC cases in 2018 [INCA 2018]. According to the International Agency for Research on Cancer (IARC), while cancer mortality rate increased by 8% in 2012, mortality rate of BC was 14% in the same period [Jacques et al. 2015].

In breast cancer patients, the amplification of the *Her2* (Human Epidermal growth factor Receptor-type 2) gene is an individual prognosticator and a predictive marker of response to targeted treatment with trastuzumab and adjuvant chemotherapy [Slamon et al. 2001]. Approximately 20-25% of BCs are HER-2 positive [Yaziji et al. 2004].

For HER-2 score determination, immunohistochemical tests (IHC) are performed. HER-2 test indicates whether this protein is carrying some role in the development of

breast cancer, since with many HER-2 receptors, the cells receive many signals to grow and split. The amount of HER-2 is scored as 0, 1+, 2+ or 3+. If the score is 0 or 1+, it is called "HER-2 negative"; if the score is 2+, then it is called "limit"; and a 3+ score is called "HER-2 positive" [Kumar et al. 2013].

HER-2 scoring still has a visual and manual analysis of histological tissues as a standard method. Such method is strongly dependent on the expertise and experience of histopathologists and has the disadvantages of being time-consuming and non-replicable [Aktan et al. 2016]. Some HER-2 tests may present different results, indicating the existence of variations within and between specialist observation [Kumar et al. 2013].

In the past few years, several works were developed for HER-2-assisted computer classification. Most are commercial, depend on specific materials and are financially costly [Brügmann et al. 2012, Jeung et al. 2012, Dobson et al. 2010, Viale et al. 2016]. Also, methods developed in other works did not show much agreement with pathologists [Aktan et al. 2016, Tuominen et al. 2012, Skaland et al. 2008, Masmoudi et al. 2009, Hall et al. 2008, Joshi et al. 2007]. Currently, some HER-2 scoring software are available on the market. Among then are the Automated Cellular Imaging System III (ACIS III) (Dako) and the HER2-CONNECT (Visiopharm).

The Automated Cellular Imaging System III (ACIS III) (Dako) was evaluate about correlation between manual HER-2 scoring and HER-2 image analysis in gastroesophageal (GE) adenocarcinomas in [Jeung et al. 2012]. They achieved an overall correlation of 84%.

HER2-CONNECT presented a 92.3% agreement between the HER2-CONNECT software and the pathologists according to [Brügmann et al. 2012]. This software exploits the ability of computer image analysis to quantify the standard HER-2 IHC "wire mesh" pattern by measuring the connectivity and size distribution of colored membranes. Their approach is based on brown segmentation, membrane skeletonization and elimination of noise. HER-2 score is defined based on the size of the membrane distribution and the area it occupies.

A comparison of slidePath's tissue IA system with other commercially available systems for HER-2-analysis are presented in the study conducted by [Dobson et al. 2010], which determined HER-2 score as 0/1+, 2+ or 3+ (negative, limit and positive). The concordance with manual review are: Slidepath: 91%, Aperio: 86%, BioImagene: 81%, Dako (Chromavision): 75% and Ventana (TriPath Imaging): 86% and 77%.

A limitation of commercial systems is that they require manual intervention, in the sense that they are trained for a particular biomarker set and need to be manually optimized. Such adjustments introduce subjective criteria and become sources of inter-laboratory variability [Masmoudi et al. 2009]. Since the systems have these limitations and are expensive, alternatives to this problem are still been developed. Also, a segmentation step is generally necessary.

The method described in [Masmoudi et al. 2009] is a multi-stage algorithm, with an agreement of 81%-83%. The algorithm steps are color pixel classification, nuclei segmentation, and cell membrane modeling, and extracts quantitative, continuous measures of cell membrane staining intensity and completeness. A minimum cluster distance classifier merges the features to classify the slides into HER-2 categories.

The study presented in [Hall et al. 2008] used color decomposition based on polar transform, threshold and gaussian filters, resulting in an AUC (Area Under the Curve) of 87%. A correlation of 84% was obtained in [Joshi et al. 2007] by preprocessing image and RGB channels segmentation. More recently, a study about deep learning applied on this topic obtained a concordance of 83% with a pathologist [Vandenberghe et al. 2017]. They used ConvNets for segmentation, feature extraction and classification techniques for cell/nucleus detection. In [Gaur et al. 2016], a transfer learning mechanism based on active learning was applied to segment membrane in FISH images.

[Saha and Chakraborty 2018] developed a deep learning framework for detection, segmentation and classification of cell membranes and nuclei from HER-2 stained breast cancer images, achieving 98.33% accuracy. The proposed method was assessed based on the HER-2 challenge contest image database of University of Warwick [Qaiser et al. 2018].

This challenge received 18 submissions, which most applied a supervised patch-based classification approach to handle the problem. A common pipeline was based on three main components: 1) preprocessing, including identification of regions of interest, 2) patch classification based on handcrafted or neural network learned features and 3) techniques to define the HER-2 score at WSI (Whole Slide Image) level. The best result on this competition built a handcrafted sub-dataset. For this purpose, a set of 68x68 patches was extracted from training. GoogLeNet and a percent-based rule were used for HER-2 score classification.

Aiming to bring a method without manual intervention and segmentation, we propose a fully-automated classification based on color features, thus reducing the complexity in this analysis. Section 2 describes the dataset and methods applied in the classification. Experimental results are reported in Section 3, then proposing future works and a conclusion in Section 4.

2. Materials and Methods

2.1. Dataset

The proposed method was developed based on the HER-2 image database of the Department of Computer Science, University of Warwick, United Kingdom [Qaiser et al. 2018]. The dataset entailed 172 whole slide images (WSI) extracted from 86 cases of invasive breast carcinomas and included both the H&E (Haematoxylin & Eosin) and HER-2 stained slides. Images stained with H&E are used in routine diagnostic practice of BC to identify tumour regions. Our approach only uses the HER-2 stained slides, being composed of 52 images for training and 34 for testing.

The histology slides for this contest were scanned on a Hamamatsu NanoZoomer C9600, enabling the image to be viewed from a ×4 to a ×40 magnification. Each WSI was cropped in patches, at ×40 magnification, by a OpenSlide [Goode et al. 2013] function, each one in size of 250x250 pixels. The patches with more tissue information were automatically selected by analyzing their histogram. Figure 1 illustrates examples of classes' patches.

The authors of this dataset only provided ground truth for training images. It is required to submit the algorithm to evaluate on testing images. Therefore, we present on
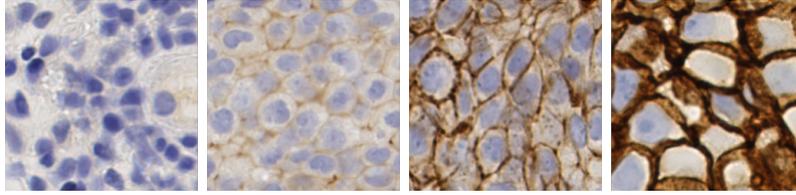
Figure 1. Example of classes 0, 1+, 2+, 3+

this paper only evaluation in the training subset. We might report test evaluation in future works.

2.2. Proposed approach

Our approach is divided into two levels: image and patient. In the first level, we analysis the classification of individual patches. Then, based on the analysis done in the previous level, occurrence of each class' patches is employed to predict HER-2 score. Figure 2 illustrates our algorithm pipeline.

Image level: The purpose of this step is to state the features which best represent the relevant patches. Assisted by a pathologist, the most representative patches (generally around 30) were selected out of each WSI. This amount was decided to balance relevant ones among total patches of each slide. Firstly, color and texture features were extracted. For color features, histograms in RGB and HSV models, additionally with mean and standard deviation of each channel, were experiment. For texture, we employed the LBP (Local Binary Pattern) [Ojala et al. 1996] and PFTAS (Parameter-free Threshold Adjacency Statistic) [Coelho et al. 2010] descriptors. Then, SVM (Support Vector Machine) [Vapnik and Cortes 1995], KNN (K-Nearest Neighbor) [Dasarathy 1991], MLP (Multi-layer Perceptron) [LeCun et al. 1998] and Decision Tree [Breiman 1984] classifiers' accuracy were evaluated using leave-one-patient-out validation. In this step, we also trained classifiers to distinguish noise patches.

Patient level: The best descriptors in the image level were used to classify all patches in each exam. Although a WSI is scored in only one class, these slides may have patches from different classes. Therefore, we need to set a rule for HER-2 scoring. A threshold rule based on the quantity of patches from each class was experimented, but results were not satisfactory. Then, we used each class occurrence as input for a classifier, creating a feature vector of occurrences. Classification is then applied to determine the HER-2 score of the WSI. The same classifiers from the previous step were employed, also accuracy using leave-one-patient-out validation were implemented.

Since clinical decisions do not differentiate 0 and 1+ classes and only consider tests as negative (0/1+), limit (2+) and positive (3+) [Tuominen et al. 2012], we have developed two approaches: with four (0, 1+, 2+, 3+ and noise) and five classes (negative, limit, positive and noise). Our method evaluates classifiers' accuracy using leave-one-patient-out validation and basically consists in:

1. Crop a WSI in patches of size 250x250;
2. Classify each patch using training patches selected by a pathologist (image level);
3. Create a vector with percentage of patches from each class;
4. Test classifiers to define HER-2 score based on these percentages (patient level).
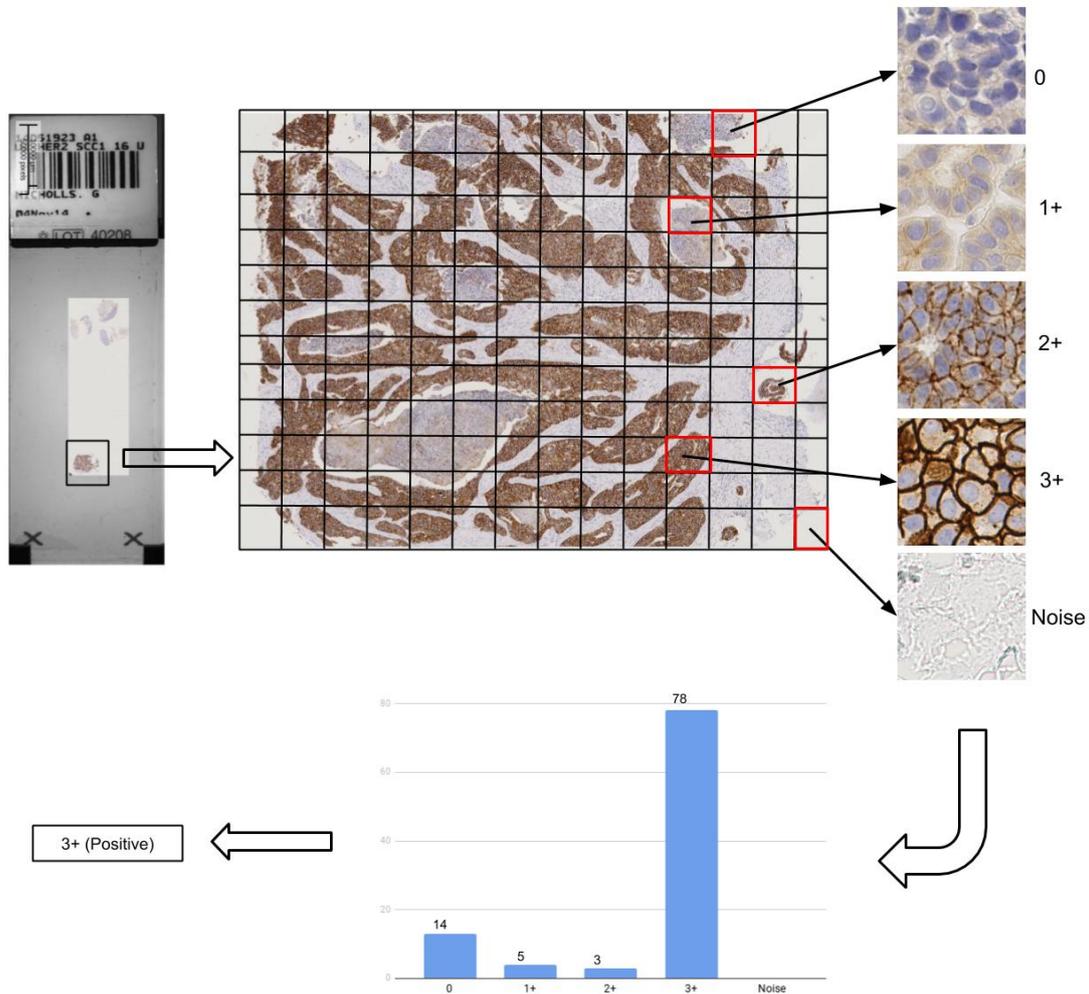
Figure 2. An illustration of our method

2.2.1. Descriptors and Classifiers Parameters

To clarify the experiments, this section presents the parameters used in descriptors and classifiers algorithms. We extracted LBP features using non-rotation-invariant uniform patterns variant, $radius = 1$ and 8 neighbours. PFTAS was implemented using mahotas function [Coelho 2013]. GridSearch was applied to find the best parameters for SVM. A exhaustively search for c, $gamma$ and $kernel$, parameters of the classifier, is performed for this function. Best values were employed in each experiment. Euclidean Distance was calculated in KNN. Variations of k, from 1 to 9, were analyzed for KNN, where the best results were obtained with $k = 1$. MLP and Decision Trees were implemented with defaults parameters in scikit-learn library [Pedregosa et al. 2011]. These methods remain to be more explored.

In the next section, we present the results obtained using the proposed approach.

3. Experimental Results

Firstly, the results in image level are shown. Table 1 presents the accuracy resulted of leave-one-patient-out validation on training patches (those selected by a pathologist).

Table 1. Accuracy on image level - training patches (in %)

| | (0/1+), 2+, 3+ and NOISE | | | | 0, 1+, 2+, 3+ and NOISE | | | |
	SVM	KNN	MLP	Tree	SVM	KNN	MLP	Tree
HSV	88.44	87.75	**90.20**	84.09	82.67	80.60	**86.51**	82.52
HSV_MS	88.62	87.80	**89.70**	85.87	82.77	81.07	**85.22**	83.14
HSV_RGB	88.36	86.16	**89.40**	85.18	82.55	76.44	**85.71**	81.76
LBP	**58.37**	49.83	57.67	49.82	**50.80**	41.83	49.37	39.87
PFTAS	**79.87**	68.99	76.03	69.64	**69.77**	60.27	68.44	59.53

Analyzing our results, the texture descriptors employed did not discriminate the evaluated patches correctly. We obtained satisfactory results in both approaches, with four and five classes. Then, only color descriptors were used in patient level classification. Color descriptors with SVM, KNN ($k = 1$), MLP and Decision Tree were used in image level, to distinguish patches and create probabilities to scoring HER-2. By using these probabilities to predict HER-2 score, SVM, KNN, MLP and Decision Tree were experimented in patient level.

Regarding the performance at patient level, Table 2 shows an overall increase in accuracy when classifying only with three classes, probably related to the similarity among 0 and 1+ classes. Also, SVM is a promising classifier to scoring HER-2 based on probabilities created by any color descriptor which classified patches using KNN classifier(HSV+KNN, HSV_MS+KNN, HSV_RGB+KNN).

Table 2. Accuracy on patient level - HER-2 scoring (in %)

| | (0/1+), 2+ and 3+ | | | | 0, 1+, 2+, 3+ | | | |
	SVM	KNN	MLP	Tree	SVM	KNN	MLP	Tree
HSV+SVM	86.27	**90.20**	84.31	88.24	61.54	**65.38**	50.00	61.54
HSV+KNN	**94.12**	86.27	92.16	92.16	**86.27**	70.59	78.43	72.55
HSV+MLP	**86.27**	82.35	86.27	76.47	**67.31**	65.38	51.92	50.00
HSV+Tree	**78.43**	60.78	70.59	78.43	**53.85**	30.77	46.16	48.08
HSV_MS+SVM	**86.27**	86.27	86.27	76.47	57.69	59.62	53.85	**65.38**
HSV_MS+KNN	**94.12**	86.27	92.16	88.24	**86.27**	70.59	74.51	74.51
HSV_MS+MLP	86.27	84.31	**88.24**	84.31	**73.08**	51.92	57.69	67.69
HSV_MS+Tree	68.63	62.75	**72.55**	70.59	**59.62**	59.62	50.00	46.16
HSV_RGB+SVM	**86.27**	80.39	86.27	86.27	55.77	55.77	50.00	**57.69**
HSV_RGB+KNN	**94.12**	82.35	86.27	84.31	**84.31**	70.59	74.83	72.55
HSV_RGB+MLP	**88.24**	82.35	86.27	76.47	**73.08**	55.77	63.46	67.31
HSV_RGB+Tree	**84.31**	66.67	82.35	62.75	**61.54**	40.38	50.00	46.15

Although SVM and MLP had better results in the image level evaluation, the

feature vector generated from classes' occurrences by the KNN achieved an overall higher accuracy. A likely motive is that the two classifiers did not adapt to patches outside the ones selected by the pathologist, which were more class homogeneous. Since patches can be heterogeneous, meaning that each patch has certain cells and membrane which can be classified in different HER-2 scores, it is difficult to correctly represent them. KNN seems to better classify these peculiarities. Due to this heterogeneity other probability descriptors created by SVM, MLP and Tree were not discriminative for HER-2 score.

The worst results in patient level were resulted from patches classified by Decision Tree with any descriptor (HSV+Tree, HSV_MS+Tree, HSV_RGB+Tree). As it is observed in Table 1, this classifier did not perform well in image level. Despite the fact the accuracy percentage is not much lower than others, the results presented in this step only involve patches analyzed by the pathologist. In a WSI, more difficult patches can be present, with more heterogeneous classes and thus, this result appears to be a consequence of the image level classification.

Tables 3 shows a confusion matrix of the best results obtained (three classes classification). In our approach we consider 0/1+ as negative, 2+ as limit and 3+ as positive.

Table 3. SVM Confusion Matrix

	HSV+KNN			HSV_MS+KNN			HSV_RGB+KNN		
	0/1+	2+	3+	0/1+	2+	3+	0/1+	2+	3+
0/1+	24	0	0	23	1	0	23	1	0
2+	2	11	1	1	12	1	1	12	1
3+	0	0	13	0	0	13	0	0	13

In CADs (Computer Aided Decision) focused on cancer treatment decision, it is important to evaluate specificity and sensitivity. The three confusion matrices presented 100% sensitivity and specificity. It means patients that should receive treatment with trastuzumab will receive it and patients that do not need to be treated with trastuzumab, will not be. These metrics are about negative(0/1+) and positive classes(3+). Mistakes between 2+ and other are very common, thus a FISH test is required to confirm HER-2 positivity in 2+ slides. Despite some 2+ confusions, our results are still very promising and might assist pathologists as a second opinion.

Our method avoids segmentation and do not need manual intervention, different of several works reviewed. In Table 4 we compared our method with others described before. HER2NET proposed by [Saha and Chakraborty 2018] had a better accuracy than ours. Although both works have used the same dataset, partitions for training and test were different. Also, HER2NET depends on manual intervention for ROI selection and includes a segmentation step.

4. Conclusion

The purpose of this study is to provide a technique able to scoring HER-2 in histopathological slides. Our results show that the proposed approach using classical machine learning techniques and color descriptors is very promising. Since we only used simple features and also without combining them, results may be improved by a more broadly study

Table 4. Comparison with related works

	Manual Intervention	Segmentation	Remarks
[Brügmann et al. 2012]	Yes	Yes	92.3% agreement
[Masmoudi et al. 2009]	Yes	Yes	83% agreement
[Hall et al. 2008]	Yes	Yes	87% AUC
[Joshi et al. 2007]	Yes	Yes	84% correlation
[Vandenberghe et al. 2017]	No	Yes	83% concordance
[Saha and Chakraborty 2018]	Yes	Yes	98.33% accuracy
Proposed work	**No**	**No**	**94.12% accuracy**

of descriptors and combination of them. Also, we have a limitation about the number of samples. Studies in other datasets and with a greater volume of samples may lead to improvements and show a more reliable result.

As described in literature review, most classical approaches include segmentation, which is known to introduce errors in next steps. Their concordance was around 85%, being increased by using deep learning techniques. Nonetheless, our approach achieved more than 90% accuracy, avoiding explicit segmentation and extraction of structure properties such as cell nuclei, membrane, size and shape of these. Besides, it is fully automated and can easily works in simple desktop computers. Thus, findings presented in this study support the idea of cheap techniques to help in pathologists routine.

Furthermore, we propose to compare classical machine learning and deep learning techniques, and also to employ images obtained in different clinical conditions.

Acknowledgement

References

[Aktan et al. 2016] Aktan, P. E., Hatipoğlu, G., and Arica, N. (2016). Risk classification for breast cancer diagnosis using her2 testing. In *2016 24th Signal Processing and Communication Application Conference (SIU)*, pages 2133–2136.

[Breiman 1984] Breiman, L. (1984). *Classification and Regression Trees*. Routledge.

[Brügmann et al. 2012] Brügmann, A., Eld, M., Lelkaitis, G., Nielsen, S., Grunkin, M., Hansen, J. D., Foged, N. T., and Vyberg, M. (2012). Digital image analysis of membrane connectivity is a robust measure of her2 immunostains. *Breast Cancer Research and Treatment*, 132(1):41–49.

[Coelho 2013] Coelho, L. P. (2013). Mahotas: Open source software for scriptable computer vision. *Journal of Open Research Software*, 1(3):131 – 155.

[Coelho et al. 2010] Coelho, L. P., Ahmed, A., an Joshua Kangas, A. A., Sheikh, A.-S., Xing, E. P., Cohen, W. W., and Murphy, R. F. (2010). Structured literature image finder: extracting information from text and images in biomedical literature. *Web Semantics: Science, Services and Agents on the World Wide Web*, 8(2-3).

[Dasarathy 1991] Dasarathy, B. V. (1991). *Nearest Neighbor (NN) Norms NN pattern Classification Techniques*.

[DeSantis et al. 2014] DeSantis, C., Ma, J., Bryan, L., and Jemal, A. (2014). Breast cancer statistics, 2013. *CA: a cancer journal for clinicians*, 64(1):52–62.

[Dobson et al. 2010] Dobson, L., Conway, C., Hanley, A., Johnson, A., Costello, S., O'Grady, A., Connolly, Y., Magee, H., O'Shea, D., Jeffers, M., and Kay, E. (2010). Image analysis as an adjunct to manual her-2 immunohistochemical review: a diagnostic tool to standardize interpretation. *Histopathology*, 57(1):27–38.

[Gaur et al. 2016] Gaur, U., Kourakis, M., Newman-Smith, E., Smith, W., and Manjunath, B. (2016). Membrane segmentation via active learning with deep networks. In *Image Processing (ICIP), 2016 IEEE International Conference on*, pages 1943–1947. IEEE.

[Goode et al. 2013] Goode, A., Gilbert, B., Harkes, J., Jukic, D., and Satyanarayanan, M. (2013). Openslide: A vendor-neutral software foundation for digital pathology. *Journal of pathology informatics*, 4.

[Hall et al. 2008] Hall, B. H., Ianosi-Irimie, M., Javidian, P., Chen, W., Ganesan, S., and Foran, D. J. (2008). 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*, 8(1):11.

[INCA 2014] INCA (2014). Estimativa 2014 – incidência de câncer no brasil. http://www.inca.gov.br/wcm/dncc/2013/apresentacao-estimativa-2014.pdf. Visited em 08/11/2015.

[INCA 2018] INCA (2018). Estimativa 2018 – incidência de câncer no brasil. http://www.inca.gov.br/estimativa/2018/index.asp. Visited em 03/05/2018.

[Jacques et al. 2015] Jacques, F., Isabelle, S., Rajesh, D., Sultan, E., Colin, M., Marise, R., Maxwell, P. D., David, F., and Freddie, B. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in globocan 2012. *International Journal of Cancer*, 136(5):E359–E386.

[Jeung et al. 2012] Jeung, J., Patel, R. A., Vila, L., Wakefield, D. S., and Liu, C. (2012). Quantitation of her2/neu expression in primary gastroesophageal adenocarcinomas using conventional light microscopy and quantitative image analysis. *Archives of pathology & laboratory medicine*, 136 6:610–7.

[Joshi et al. 2007] Joshi, A. S., Sharangpani, G. M., Porter, K., Keyhani, S., Morrison, C., Basu, A. S., Gholap, G. A., Gholap, A. S., and Barsky, S. H. (2007). Semi-automated imaging system to quantitate her-2/neu membrane receptor immunoreactivity in human breast cancer. *Cytometry Part A*, 71A(5):273–285.

[Kumar et al. 2013] Kumar, V., Abbas, A. K., and Aster, J. C. (2013). *Robbins Basic Pathology*. Elsevier Health Sciences.

[LeCun et al. 1998] LeCun, Y., Bottou, L., Orr, G. B., and Müller, K.-R. (1998). Efficient backprop. In *Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop*, pages 9–50. Springer-Verlag.

[Masmoudi et al. 2009] Masmoudi, H., Hewitt, S. M., Petrick, N., Myers, K. J., and Gavrielides, M. A. (2009). Automated quantitative assessment of her-2/neu immuno-

histochemical expression in breast cancer. *IEEE Transactions on Medical Imaging*, 28(6):916–925.

[Ojala et al. 1996] Ojala, T., MattiPietikäinen, and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distribution. *Pattern Recognition*, 1(29):51–59.

[Pedregosa et al. 2011] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

[Qaiser et al. 2018] Qaiser, T., Mukherjee, A., PB, C. R., Munugoti, S. D., Tallam, V., Pitkäaho, T., Lehtimäki, T., Naughton, T., Berseth, M., Pedraza, A., Mukundan, R., Smith, M., Bhalerao, A., Rodner, E., Simon, M., Denzler, J., Huang, C.-H., Bueno, G., Snead, D., Ellis, I. O., Ilyas, M., and Rajpoot, N. (2018). Her2 challenge contest: a detailed assessment of automated her2 scoring algorithms in whole slide images of breast cancer tissues. *Histopathology*, 72(2):227–238.

[Saha and Chakraborty 2018] Saha, M. and Chakraborty, C. (2018). Her2net: A deep framework for semantic segmentation and classification of cell membranes and nuclei in breast cancer evaluation. *IEEE Transactions on Image Processing*, 7149(c):1–1.

[Skaland et al. 2008] Skaland, I., Øvestad, I. T., Janssen, E., Klos, J., Kjellevold, K. H., Helliesen, T., and Baak, J. P. A. (2008). Comparing subjective and digital image analysis her2/neu expression scores with conventional and modified fish scores in breast cancer. *Journal of Clinical Pathology*, 61(1):68–71.

[Slamon et al. 2001] Slamon, D. J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M., Baselga, J., and Norton, L. (2001). Use of chemotherapy plus a monoclonal antibody against her2 for metastatic breast cancer that overexpresses her2. *New England Journal of Medicine*, 344(11):783–792.

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

[Vandenberghe et al. 2017] Vandenberghe, M. E., Scott, M. L. J. S., Scorer, S. W., Magnus Söderberg, D. B., and Barker, C. (2017). Relevance of deep learning to facilitate the diagnosis of her2 status in breast cancer. *Scientific Reports*, 7(45938).

[Vapnik and Cortes 1995] Vapnik, V. and Cortes, C. (1995). Support-vector networks. *Machine Learning*, 20:273–297.

[Viale et al. 2016] Viale, G., Paterson, J., Bloch, M., Csathy, G., Allen, D., Dell'Orto, P., Kjærsgaard, G., Levy, Y. Y., and Jørgensen, J. T. (2016). Assessment of her2 amplification status in breast cancer using a new automated her2 iqfish pharmdx™ (dako omnis) assay. *Pathology - Research and Practice*, 212(8):735 – 742.

[Yaziji et al. 2004] Yaziji, H., Goldstein, L. C., Barry, T. S., Werling, R., Hwang, H., Ellis, G. K., Gralow, J. R., Livingston, R. B., and Gown, A. M. (2004). HER-2 testing in breast cancer using parallel tissue-based methods. *Jama*, 291(16):1972–1977.