

BREAST CANCER GRADING OF H&E STAINED HISTOPATHOLOGY IMAGES

Vijay M. Mane and Nikhil Tagalpallewar

Department of Electronics and Telecommunication Engineering, Vishwakarma Institute of Technology, India

Abstract

Breast cancer is the common existing form of cancers amongst women. The automatic image analysis methods have an enormous potential to decrease the workload in a pathology laboratory. The grading of breast cancer histopathology images is used to find the level of breast cancer. The automatic grading of breast cancer histopathology images is a challenging task. In this paper a system for automatic detection of breast cancer grading of H&E stained histopathological images is presented. An image processing techniques such as preprocessing, segmentation, feature extraction and classification are used in this system. The segmentation of nuclei in H&E stained image is performed using color thresholding and maximum entropy thresholding. The features are computed according to Bloom Richardson grading criteria. The decision tree classifier is used to classify input image into three group i.e., low grade, intermediate grade and high grade.

Keywords:

Breast Cancer, Histopathology, H&E (Hematoxylin and Eosin), Grading, Bloom Richardson Criteria, Color thresholding, Maximum Entropy, Decision Tree

1. INTRODUCTION

The breast cancer is the most commonly observed disease among women. It is expected to over 266,120 new cases of invasive breast cancer along with 63,960 new cases of non-invasive breast cancer are to be diagnosed in American women in 2018. The classification of breast cancer depends on the grade. The grading of breast cancer histopathology images is standard practice for diagnosis and prognosis of breast cancer development [1]. The grading system shows how fast breast cancer grows and spread. The grade of breast cancer is determined by observing the appearance and behavior of cancer cell under the microscope. The modified Bloom Richardson grading system is generally used to discriminate low, intermediate and high-grade cancer. It is also known as Nottingham scoring system [1, 2]. This grading system considers 3 different features and gives score of grade in the range of 1 to 3. The breast cancer is the most widespread cancer in women's. The diagnosis and treatment of the breast cancer is done using histopathological images. The pathologists carry out evaluation of the histopathological images manually under a microscope. The accuracy of the evaluation and grading of the cancer depends on the expertise of the pathologists. This results in the variations among the pathologists. Each pathologist has to observe large number of grading cases per day. This is a very tedious and time-consuming task for the pathologists.

This paper presents an automatic system for grading of histopathological image helps the pathologists by providing second judgment and to reduce their workload. Such systems alert the pathologists to images that require more concentration and let them to focus on diagnosis and prognosis.

This paper presents a method for finding the grade of breast cancer tissue images by Bloom Richardson grading system. The

paper is organized as, section 2 presents literature survey, section 3 presents the proposed methodology, results and conclusion are discussed in section 4 and section 5 respectively.

2. RELATED WORK

The multi-resolution approach is used for grading the system by Dalle et al. [1]. For segmentation Gaussian color model are used. Low-resolution image used for detecting tubule formation region. Mitotic count and nuclear pleomorphism found on high-resolution image. Multilevel information based grading system described by Naik et al. [2], [3]. Low-level information is used by Bayesian classifier to find the pixel belong to object of interest. The High-level information used to find object boundaries using level set algorithm and gland and nuclei identification is done by template matching algorithm. The features are extracted using morphological and architectural attributes of the segmented image. These features are classified into different grades using SVM classifier. Faridi et al. [4] proposed a grading system based on multi-feature approach. The unmixing color of H&E is used in preprocessing step. The segmentation of nuclei is done by applying morphological operations and Difference of Gaussian filter on preprocessed image. The features like nuclei area, chromatin density, contour regularity and nucleoli presence are used for pleomorphism scoring. Veta et al. [5] represented different algorithm to detect mitotic activity in breast cancer histopathology images. The morphological, statistical and texture features were used. The convolutional neural network and random forest classifier is used to classify these features for finding mitosis activity detection. Ten Kate et al. presented new method for counting mitosis [6]. They used features like compact and dense objects in image which have hairy extension. The region growing algorithm is used to find mitotic nuclei. Ciresan et al. presented the mitosis detection in breast cancer histopathology images [7]. They utilized deep neural network to classify image as mitosis or non-mitosis. If the cell image is under mitosis class then they found center of mitosis. Barnes et al. presented method for identify glandular region and tubule region in breast cancer tissue [8]. They utilized random forest classifier to identify and classify tumor nuclei and true lumina. The graph-cut method applied to group of tumor nuclei and lumina to find tubule glandular regions and non-tubule glandular region. Dalle et al. proposed as algorithm to find pleomorphism in breast cancer histopathology images [9]. The candidate nuclei i.e. individual nuclei were found by applying morphology operations. The segmentation is achieved by converting Cartesian to polar space of detected epithelial nuclei. This polar space coordinate is used to find gradient of cell by Bi-quad filtering method. The nuclear pleomorphism score were calculated by Nottingham criteria. The scoring is based on size, shape and texture of segmented cell nuclei. Wan et al. proposed multi-level features based approach for grading breast cancer histopathology images [10]. Hybrid

active contour-based segmentation model is used for segmenting different nuclei in histopathology image. They utilized multiple SVM classifiers to find different grade of cancers.

3. METHODOLOGY

In this paper we presented a system for grading of breast cancer H&E stained histopathological images. The main objective of the proposed system is to grade the histopathology images into low, medium and high risk of breast cancer. It consists of preprocessing, nuclei detection, segmentation, feature extraction and classification of H&E stained breast cancer histopathology images to find an appropriate grade of an image. The BreakHis database is used for the experimentation [11]. The block diagram of proposed system is presented in Fig.1.

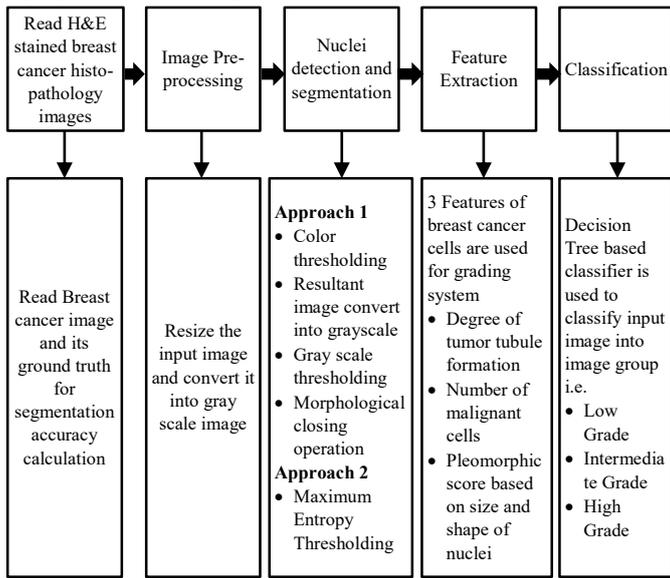


Fig.1. Block diagram of automatic breast cancer grading system

3.1 PREPROCESSING

The breast cancer cell detection and segmentation is a major step in breast cancer grading system. It consists of preprocessing step to standardize the image. The variations in image quality are observed due to many factors like natural breast cancer heterogeneity, contradictory conditions during tissue slice preparation or during image acquisition. These variations affect the nuclei segmentation accuracy. A suitable preprocessing method is required to remove such type of variation. In this proposed system. The grayscale conversion and resizing of an image is utilized in preprocessing step. The input color images are converted into grayscale image. The image is then enhanced for the contrast. The sequence of morphological operation like dilation, closing and erosion are applied.

3.2 NUCLEI DETECTION AND SEGMENTATION

The automatic detection and segmentation of nuclei is an essential step in the breast cancer grading system. In this system we used color based thresholding and maximum entropy based thresholding methods for segmenting nuclei from histopathology images. Maximum entropy thresholding is used to find the exact location of each nucleus.

3.2.1 Color Based Thresholding:

The color thresholding is one of image segmentation methods whose objective is to partition an image into regions that are similar according to a set of predefined criteria. In applications where speed is an important factor, color thresholding is a fundamental approach which is more popular than other thresholding techniques [12]. The thresholds are decided by observing the color variation in H&E stained breast cancer histopathology images. The nuclei are observed in blue color. The threshold is selected according to variations in the blue color over the entire histopathology image.

3.2.2 Maximum Entropy Thresholding:

The segmentation of nuclei is carried out by auto-thresholding method. We applied binary threshold on image whose upper limit is 255 and lower limit is calculated by maximum entropy method. The maximum entropy [13] is mainly based on histogram of image. The threshold obtained using maximum entropy method is explained as below:

$$P_{ij} = \frac{n_{ij}}{N \times N}$$

where,

$N \times N$ is the Image size

n_{ij} is the Number of pixel whose grey values equal i and local average value equal j .

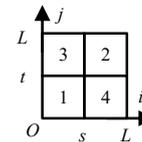


Fig.2. Two Dimensional Histogram Plane

The histogram plane is described as shown in Fig.2 where, area under 1st and 2nd block denote object and background image respectively, and 3rd and 4th block denote edges and noises. Threshold vector (t,s) where t is a threshold for pixel intensity and s is threshold for local average of pixels, should be determine. According to maximum entropy theory, the determined threshold vector should make area under 1st and 2nd block maximum information.

If the area under 1st and 2nd block has different probability distribution P_1 and P_2 determine by,

$$P_1 = \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij}$$

$$P_2 = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \tag{1}$$

The 2D discrete entropy is defined as

$$H = - \sum_i \sum_j p_{ij} \log p_{ij} \tag{2}$$

Entropy under 1st block is derived as follows:

$$H(1) = - \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} \left(\frac{p_{ij}}{P_1} \right) \log \left(\frac{p_{ij}}{P_1} \right)$$

$$H(1) = \left(\frac{-1}{P_1} \right) \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij} \log p_{ij} - p_{ij} \log P_1$$

$$H(1) = \left(\frac{1}{P_1} \right) \log P_1 \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij} - \left(\frac{1}{P_1} \right) \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij} \log p_{ij}$$

$$H(1) = \log(P_1) + \frac{H_1}{P_1} \quad (3)$$

Entropy under 2nd block can be defined as follow:

$$H(2) = \log(P_2) + \frac{H_2}{P_2} \quad (4)$$

where, H_1 and H_2 described as,

$$H_1 = - \sum_{i=0}^{s-1} \sum_{j=0}^{t-1} p_{ij} \log p_{ij} \quad (5)$$

$$H_2 = - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \log p_{ij} \quad (6)$$

Then the function of entropy is,

$$\varphi(s,t) = H(1) + H(2) \quad (7)$$

According to the maximum entropy principle, the threshold vector (s^*, t^*) should be satisfied to,

$$\varphi(s^*, t^*) = \max\{\varphi(s,t)\} \quad (8)$$

3.3 FEATURE EXTRACTION

There are three type of parameters required to grade the given H&E stained breast cancer image. The features are degree of tumor tubule formation, number of malignant cells or mitotic count and pleomorphic score.

3.3.1 Degree of Tumor Tubule Formation:

The structure which consist of Lumina (i.e. white region) surrounded by tumor cells is considered as tubule formation region [1]. To identify tubule formation region we first consider white region portion. Then morphological closing operation is applied on segmented image to remove noise in image. The sequence of morphological operations like dilation and erosion are applied to get connected nuclei. The coordinates of the entire connected nuclei region are identified by active contour method. The connected region are consider as tube, which participate in tubule formation according to it contain white Lumina region.

Tubule formation score is calculated as follows. The total numbers of nuclei participating in tube-like structure formation is calculated. The total number of nuclei are available in input image are calculated. The percentage of degree of tumor tubule formation is calculated by following formula,

$$\text{Degree of tumor tubule formation } (D) = (\text{Total number of nuclei participate in tube like structure formation}) / (\text{Total number of nuclei available in entire image}) \quad (9)$$

According to degree of tubule tumor formation the score is calculated as below,

$$S_1 = \begin{cases} 1 & \text{if } D > 75\% \\ 2 & \text{if } 10\% < D < 75\% \\ 3 & \text{if } D < 10\% \end{cases} \quad (10)$$

3.3.2 Mitotic Count:

The manual detection and counting of malignant cells is used by pathologist to find the grade. The automatic mitotic detection and counting process provides potential solution to relieve pathologist workload. The Mitosis appearance is identified by their special features. The mitotic cells appear in compact, highly dense object and usually have hairy extension at boundary. The shape and size of mitotic cell is primarily depend on internal state of the cell, phase of division when the mitosis was capture, the cutting angle and preparation factors. They may appear in elliptical, circular or rod shape. In some cases we cannot differentiate mitosis nuclei due to more and tightly packed nearby cells.

In this system, region growing algorithm is used to detect the mitotic cells in an image. The properties of mitotic cells like compactness and high dark pixel density are used for region growing algorithm. The darkest pixel in the image is selected as a starting point or 'seed' of the region growing algorithm. Then aggregate the pixels surrounding by it showing same property as seed point. Connected pixels are considered in region only if the predefined conditions are satisfied. The score for mitotic count is decided according to Bloom Richardson criteria.

3.3.3 Pleomorphic Score:

Pleomorphic score is calculated by using the features of nuclei like shape, size and texture. The size feature is decided according to mean and standard deviation of the population of the segmented nuclei. The shape feature is decided according to roundness of segmented cell nuclei. The roundness is calculated according to following equation,

$$R = 4\pi A^2 / P \quad (11)$$

where,

R is the roundness of an object

A is the area of an object

P is the perimeter of an object

The roundness is calculated for each and every nucleus. Then it is compared with elliptical object having major axis equal to maximum distance of center of nuclei to counterpoint of that nuclei and minor axis is taken as 80% of major axis. The difference of these two roundness is considered as feature factor for finding the shape of nuclei. The texture feature is calculated according to grey level of the nuclei. The score for pleomorphism is calculated using these features and Bloom Richardson grading system.

3.4 CLASSIFICATION

The extracted features as explained above are utilized for the classification and grading of cancer tissue image by using decision tree classifier. This type of classifier is used when the input data can be divided into specific number of classes. It poses a series of carefully set questions about the attribute of the test record. Each time it will receive an answer until it got conclusion to reach particular class. The final score and grade of system is calculated using decision tree classifier based on Bloom Richardson criteria as shown in Fig.3.

4. RESULTS AND DISCUSSION

The presented method is implemented using Python 3.6.3 and OpenCV 3.3.0 in Windows 10 environment on BreakHis database. The H&E stained breast cancer histopathology image which consists of a large number of cancerous nuclei is shown in Fig.4. The white spots in the image denote the lumen area of tissue, pink part indicates the cytoplasm and blue part indicates nuclei which is our region of interest. Fig.5 and Fig.6 show the result of color thresholding and maximum entropy thresholding techniques respectively. The result of segmentation techniques for ten patient's samples is shown in Table.1. According to the results, the accuracy of segmentation algorithm mostly depends on the quality of input sample images. In most of these cases, the accuracy of the color thresholding method is more than the maximum entropy method. The color thresholding segmentation result is used for further processing.

Fig.7 represents the tube-like structure appearance in the input image according to the algorithm explained in section 3.3.1. This type of structure is considered for the degree of tumor tubule formation. Fig.8 represents the total number of mitotic cells present in the image. To find the pleomorphic score, segmentation of single nuclei which does not overlap with other nuclei is required. Such types of nuclei are shown in Fig.9. The final score and the results of intermediate steps are shown in Fig.10 collectively.

The result for the degree of tumor tubule formation and the number of mitotic cells for ten samples of image are shown in Table.2. The score for each feature, total score, and final grades for ten sample images are shown in Table.3.

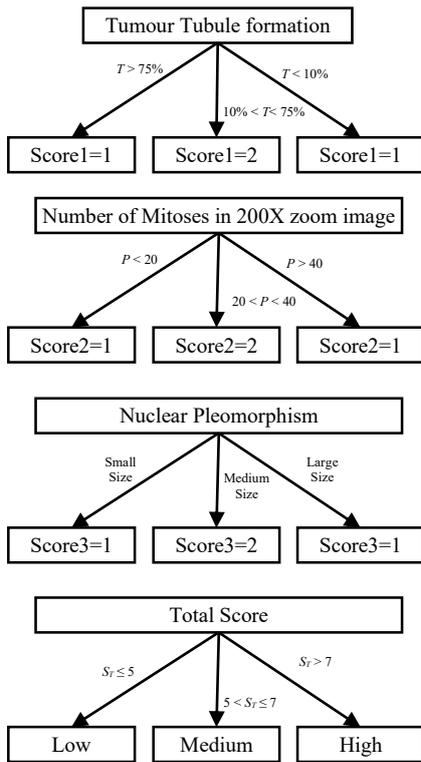


Fig.3. Decision tree classifier

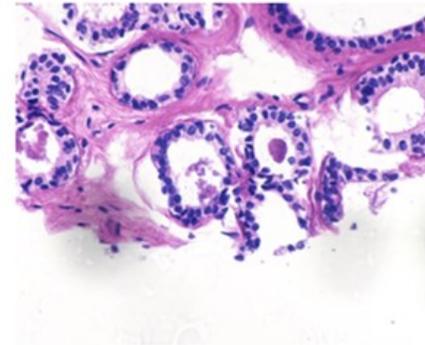


Fig.4. Input H&E stained histopathology image

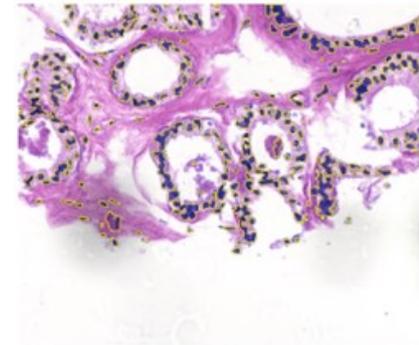


Fig.5. Segmentation result using color thresholding

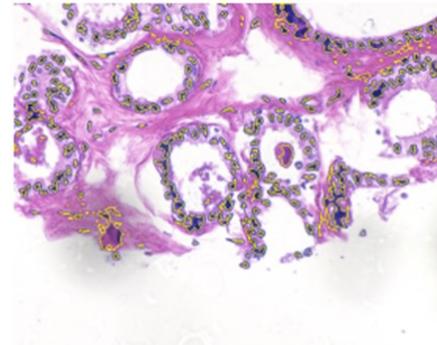


Fig.6. Segmentation result using Maximum entropy thresholding

Table.1. Segmentation algorithm comparison

Patient ID	Color Thresholding (%)	Maximum Entropy Thresholding (%)
ytma10_010704_benign1	32.9	63.31
ytma10_010704_malignant1	27.86	88.03
ytma12_010804_benign1	84.66	70.5
ytma12_010804_malignant1	57.22	23.41
ytma23_022103_benign1	94.59	71.16
ytma23_022103_malignant1	85.53	85.69
ytma49_042003_benign1	85.88	42.67
ytma49_042003_malignant1	27.57	3.08
ytma49_042203_benign1	98.19	73.41
ytma49_042203_malignant1	90.19	61.13

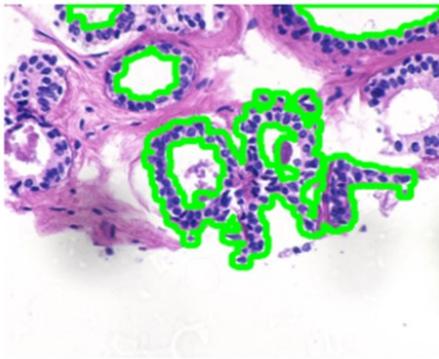


Fig.7. Tube-like structures in input images

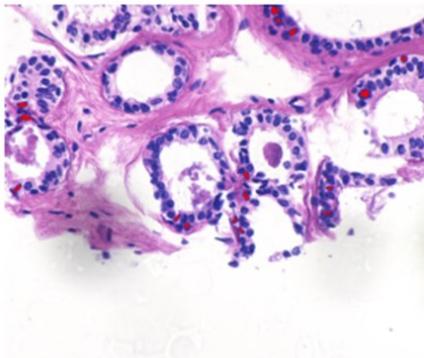


Fig.8. Mitotic cell consideration for mitotic count

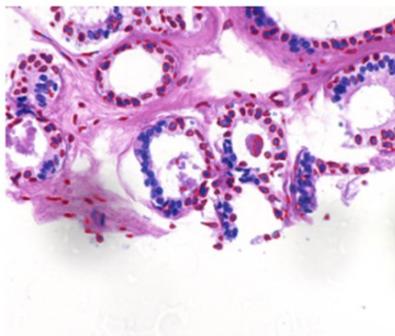


Fig.9. Nuclei consider for Pleomorphism Score

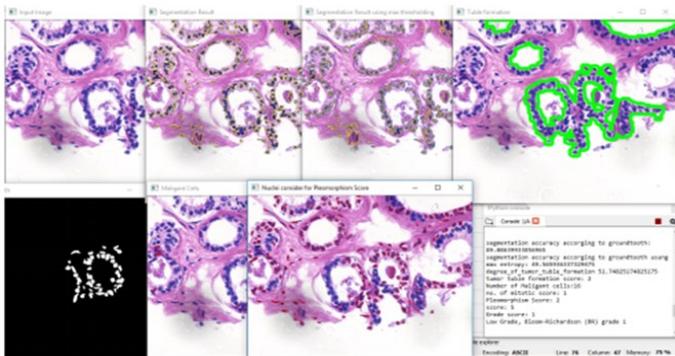


Fig.10. Final result of algorithm

Table.2. Degree of tubule tumor formation and number of malignant cells are available in image

Patient ID	Degree of tubule tumor formation (%)	No. of Malignant cells in 200×Zoom
ytma10_010704_benign1	0.29	11
ytma10_010704_malignant1	0.16	20
ytma12_010804_benign1	23.73	20
ytma12_010804_malignant1	35.64	3
ytma23_022103_benign1	11.95	36
ytma23_022103_malignant1	2.11	60
ytma49_042003_benign1	81.36	83
ytma49_042003_malignant1	0.68	8
ytma49_042203_benign1	13.43	60
ytma49_042203_malignant1	1.52	33

Table.3. Grading result on histopathology images

Patient ID	Tumor tubule formation score	No of mitotic count score	Pleomorphic Score	Total Score	Grade Score	Remark
ytma10_010704_benign1	3	1	2	6	2	Medium Grade
ytma10_010704_malignant1	3	2	1	6	2	Medium Grade
ytma12_010804_benign1	2	2	2	6	2	Medium Grade
ytma12_010804_malignant1	2	1	2	5	1	Low Grade
ytma23_022103_benign1	2	2	2	6	2	Medium Grade
ytma23_022103_malignant1	3	3	2	8	3	High Grade
ytma49_042003_benign1	1	3	2	6	2	Medium Grade
ytma49_042003_malignant1	3	1	2	6	2	Medium Grade
ytma49_042203_benign1	2	3	2	7	2	Medium Grade
ytma49_042203_malignant1	3	2	2	7	2	Medium Grade

5. CONCLUSION

The first step towards prognosis and diagnosis of breast cancer is to find grade or stage of breast cancer. A system for automatic grading of breast cancer H&E stained histopathology images is presented in this paper. This system is useful to reduce the work load of pathologist and make attention of pathologist toward critical cases. To implement this system various image processing techniques are utilized. The preprocessing techniques such as gray level conversion and resizing of images are applied to increase the image quality and improve the speed of execution. To detect the nucleus which is our region of interest color thresholding and

maximum entropy thresholding segmentation algorithm is used. These thresholding algorithms give segmentation accuracy up to 70%. The parameters are derived from Bloom Richardson criteria. The results are computed in the form of degree of tubule tumor formation, mitotic count and pleomorphic score. These parameters are applied to decision tree classifier to find the final grade of input image in the form of low, medium and high.

REFERENCES

- [1] J.R. Dalle, W.K. Leow, D. Racoceanu, A.E. Tutac and T.C. Putti, "Automatic Breast Cancer Grading of Histopathological Images", *Proceedings of Annual Conference on IEEE Engineering in Medicine and Biology Society*, pp. 3052-3055, 2008.
- [2] S. Naik, S. Doyle, S. Agner, A. Madabhushi, M. Feldman and J. Tomaszewski, "Automated Gland and Nuclei Segmentation for Grading of Prostate and Breast Cancer Histopathology", *Proceedings of 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 284-287, 2008.
- [3] Scott Doyle et al., "Automated Grading of Breast Cancer Histopathology using Spectral Clustering with Textural and Architectural Image Features", *Proceedings of 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pp. 114-117, 2008.
- [4] Pegah Faridi et al., "Cancerous Nuclei Detection and Scoring in Breast Cancer Histopathological Images", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1-7, 2016.
- [5] Mitko Veta et al., "Assessment of Algorithms for Mitosis Detection in Breast Cancer Histopathology Images", *Medical Image Analysis*, Vol. 20, No. 1, pp. 237-248, 2015.
- [6] T.K. Ten Kate et al., "Method for Counting Mitoses by Image Processing in Feulgen Stained Breast Cancer Sections", *Cytometry*, Vol. 14, No. 3, pp. 241-250, 1993.
- [7] Dan C. Ciresan et al., "Mitosis Detection in Breast Cancer Histology Images with Deep Neural Networks", *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 1-8, 2013.
- [8] K. Nguyen, "Automatic Glandular and Tubule Detection in Histological Grading of Breast Cancer", *Proceedings of the SPIE*, Vol. 9420, pp. 1-7, 2015.
- [9] Jean Romain Dalle et al., "Nuclear Pleomorphism Scoring by Selective Cell Nuclei Detection", Available at: <https://www.comp.nus.edu.sg/~leowwk/papers/wacv2009-npscore.pdf>.
- [10] T. Wan et al., "Automated Grading of Breast Cancer Histopathology using Cascaded Ensemble with Combination of Multi-Level Image Features", *Neurocomputing*, Vol. 229, pp. 34-44, 2017.
- [11] Fabio A. Spanhol et al., "A Dataset for Breast Cancer Histopathological Image Classification", *IEEE Transactions on Biomedical Engineering*, Vol. 63, No. 7, pp. 1455-1462, 2016.
- [12] Rafikha Aliana A. Raof, "Comparison of Colour Thresholding Method using RGB and HSI Information for Ziehl-Neelsen Sputum Slide Images", *Proceedings of 10th International Conference on Information Sciences Signal Processing and their Applications*, pp. 1-8, 2010.
- [13] K. Meethongjan and D. Mohamad, "Maximum Entropy-Based Thresholding Algorithm for Face Image Segmentation", Available at: <https://pdfs.semanticscholar.org/9081/8e0ab85b6a5f03cd28bc5c23c90a0971c36e.pdf>.