Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method

by Riries Rulaningtyas

PAPER · OPEN ACCESS

Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method

View the article online for updates and enhancements.

Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method

Riries Rulaningtyas[1,4], Agoes Santika Hyperastuty[2], Anny Setijo Rahaju[3]

[1]*Biomedical Engineering Study Program, Department of Physics, Airlangga University, Jl. Mulyorejo, Surabaya , 60115, Indonesia*
[2]*Biomedical Engineering Master Program, Airlangga University, Jl. Mulyorejo, Surabaya , 60115, Indonesia*
[3]*Department of Pathology Anatomy, Medical Faculty, Airlangga University, Jl. Mulyorejo, Surabaya , 60115, Indonesia*

[4]E-mail: riries-r@fst.unair.ac.id

Abstract. Breast cancer is the leading type of malignant tumor which is observed in women. The effective treatment depends on its early diagnosis. The gold standard of breast cancer is histopathologic examination of cancer cells. The determination of the grading in breast cancer is determined by three factors: pleomorphic, tubular formation and cell mitosis. This paper uses pleumorfic and tubular formation pattern from breast cell histopathology images. The proposed system consists of four major steps : preprocessing, segmentation, feature extraction and classification. We use k – means clustering method for image segmentation and use Gray level Cooccurence Matrix (GLCM) for feature extraction with four features (i.e. angular second moment, contrast feature, entropy feature, and variance feature). The final step is grading classification which uses Backpropagation Neural Network. Some of important parameters will be variated in this process such as learning rate and the number of node in hidden layer. The research gives good result for the identification of breast cancer grading with 88% accuracy, 85% sensitivity, and 80% specificity.
Keywords : histopathology, breast cancer, GLCM, backpropagation.

1. Introduction

Breast cancer is one of the highest causes of human death besides cervical cancer. There were 1.7 million women that were diagnosed with this disease. Not surprisingly, breast cancer is one of the highest mortality factors, especially for women. Every year 2 out of 10,000 women in the world are estimated to have breast cancer. The high mortality rate due to breast cancer is happened in the developed and developing countries. In 2013, the American Cancer Society said there were 296,980 women and 2,240 men positive for breast cancer and 39,620 women and 410 men died. The prevalence of cancer is quite high. Based on data from the Basic Health Research of Indonesia in 2013, the prevalence of tumors in Indonesia is 1.4 per 1000 population, or about 347,000 people [1].

Many studies refer to breast cancer but almost 50% of the data used comes from mammographic images. Histopathological examination is the gold standard for examination of breast cancer grading. Histopathological examination is a microscopic examination of the organism's tissues. According to Bloom - Richardson, in determining the grading of breast cancer is determined by three factors: nuclear pleomorphism, tubular formation and mitotic activity. Of the three factors can only be seen when we see the results of histopathological examination of breast cancer. For the determination of the grading of the pathologists manually counting the number of pleomorphic, tubular and mitotic according to their respective expertise.

In this study building intelligent systems using images of breast cancer histopathology preparations which are expected to be used to help pathologists determine quickly the histopathological grading of

breast cancer. The method used is K-Means clustering for image segmentation processing, Gray Level Co-Occurance for its extraction feature and Backpropagation for intelligent system training.

In a study conducted in 2012 [2] uses image improvement with subsampling, smoothing, denoising and enhancement method. For segmentation using thresholding, edge detection, active contour, labelling and clustering. This research uses the feature extraction method based on image's intensity, there are morphology, spectral, and texture. From some of the methods used, it is found that the lack of this research is in home base data which is used that has not been standardized so that it cannot be used by others.

In 2017 [3] made a study entitled "Automated grading of breast cancer histopathology using cascaded ensembles with a combination of multi-level image features. 3 Segmentation methods using shape, length and size variables. For feature extraction using the CNN (Convolution Neural Network) method and for classification using the SVM (Support Vector Machine) method. The results obtained from the study show the level of grade comparison between I, II and III. This is seen from the level of accuracy of each grading. A high degree of accuracy means showing high grade while a small degree of accuracy shows low grade. This makes it difficult to determine the middle grade because there is no clear limit for the level of accuracy for the middle grade. From the two studies, this study tries to improve the level of accuracy by applying the image segmentation method of k-means clustering. From the results of image integration using the k-means method shows a segmentation image that has not been able to separate the image of tubular formation, mitotic cells, and nuclear pleomorphism correctly, so that feature extraction is needed where in this study the characteristics are taken from the texture of the image using the Gray Level Co-Occurrence method Matrix (GLCM) and neural network as a classification of breast cancer grading through histopathological examination of cancer cells.

2. Material and Method

This type of research is analytical descriptive where in this study observes the image of breast biopsy result, all of which had been done by anatomical pathologists. All data stored in the Anatomy Pathology laboratory in the form of slides. The slides will be examined 10 times glance at different places. After that the doctor will add and take the average calculation results to determine the histopathological grading. This research uses 150 slides of Histopathology Anatomy (HPA) breast cancer biopsy with the grading score is determined by Anatomical Pathology experts. The data are consisted of 50 slides of grading 1, 50 slides of grading 2, and 50 slides of grading 3. The slide's retrieval techniques were random data retrieval from hospital, either the slide's results from tissue retrieval biopsy.

This research uses CLAHE (contrast-limited adaptive histogram equalization) method for contrast enhancement, k – means clustering for image segmentation. The result of image segmentation still has a lack of performance. It is failed to separate the tubular and nuclei images. To improve it, we use GLCM method for feature extraction to obtain the texture of segmented images, then the texture values could become the input for grading classification that is done by backpropagation neural network.

2.1. Nottingham Grading System (NGS) / Bloom-Richardson

At present, NGS is a scoring system recommended by World Health Organization (WHO), American Joint Committee on Cancer (AJCC), European Union (EU), and the Royal College of Pathologists (UK RC Path), and international consensus states that the NGS system is considered a 'gold standard' for grading breast cancer [4]. This modification has increased the ability to determine grading by anatomical pathologists   [4].

Grading Breast Cancer

Tubule Formation

Majority of tumor (>75%)   1
Moderate Degree (10-75%) 2
Little or none(<10%)   3

Grading Breast Cancer

Mitotic Count:

0-9 Mitoses/10hpf   1
10-19 Mitoses/10hpf   2
20 or> Mitosis/10hpf   3

Grading Breast Cancer

Nuclear Pleomorphism

Small Reguler Uniform Cell   1
Moderate Nuclear Size & Variation   2
Marked Nuclear Variation   3

Grading Breast Cancer

Combined Histologi Grade

Low Grade (I)   3-5
Intermediate Grade (II)   6-7
High Grade (III)   8-9

Figure 1. Scoring Criteria Based on NGS [4]

2.2. Contrast-Limited Adaptive Histogram Equalization (CLAHE)

The slides of HPA breast cancer will observed by microscope. The slide is captured by using digital camera microscope. The digital images from the digital camera have low qualities because of the environment brightness influences. We try to implement CLAHE method to improve the contrast of HPA digital slide images. The first step, we change the color channel of images RGB into CIE Lab channel. Then we use one individually color channel Luminance from the CIE Lab channel to be analyzed. The distribution of single channel can be adapted by using exponential distribution as [5].

$$g = g_{min} - \left(\frac{1}{\alpha}\right) * \ln[1 - P(f)] \tag{1}$$

Where :
g_{min} = minimum pixel value
α = the clip parameter,
P(f) = Cumulative Probability Distribution

The histogram of output region is matched with the histogram that is specified by the distribution type using Rayleigh distribution [5].

$$y = P\big(f(x\backslash b)\big) = \int_0^x \frac{x}{b^2} e^{\left(\frac{-x^2}{2b^2}\right)} \tag{2}$$

2.3. k – Means Clustering

K-means is one of the non-hierarchical data clustering methods that try to partition existing data into one or more clusters or groups. This method partitions data into clusters or groups so that the data which has the same characteristics are grouped into the same cluster and data that have different characteristics are grouped into other groups. The algorithm for performing k-means clustering is as follows:

Step 1 : Determine the desired number of clusters (k)
Step 2 : Determine the value of the initial centroids randomly
Step 3 : Calculate the distance of each data to each centroid [6] [7]

$$d(X,Y) = \sum_j \left(X_j - Y_j\right)^2 \tag{3}$$

Where : X_j = the data, Y_j = the centroids
Step 4 : Group the data into clusters based on the closest distance (minimum) to a cluster [6][7]

$$\emptyset : \{1, \dots, n\} \rightarrow \{1, \dots, k\} \tag{4}$$

Where : n = the number of data, k = the number of clusters

$$\mu_j(\emptyset) = argmin_{\mu \in \mathcal{R}^d} \sum_i d(X_i, \mu)\mathrm{I}(\emptyset(i) = j) \tag{5}$$

Step 5 : Recalculate the value of the centroids by calculating the mean value of the data from each cluster

Step 6 : Perform steps 3-5 until the value of the centroids no longer changes.

Calculation of the distance between centroids and data can be done using the euclidean distance, cityblock, cosine, correlation, and hamming equations.

2.4. Gray Level Co-Occurrence Matrix (GLCM)

The GLCM matrix is calculated from the pixel value in pairs and has a certain intensity value with d is the distance between two pixels (x1, y1) and (x2, y2), θ is defined as the angle between the two, then the GLCM matrix is the spatial distribution of $P_{d,\theta}(i,j)$. Many papers write various versions of GLCM calculations, especially regarding the angle problem used. The following is an illustration that illustrates the angle direction with a distance of 1 pixel and there are 4 types of angles used by the features in GLCM [10]:

1. Angular Second Moment Feature (ASM Feature), measure pixel uniformity in the images with

$$fe_1 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j)^2 \tag{6}$$

2. Contrast Feature
   To measure the gray various level between referenced pixels and its neighbour.

$$fe_2 = \sum_{n=0}^{N_g-1} |i-j|^2 * \left\{ \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j) \right\} \tag{7}$$

3. Entropy Feature
   To measure the irregularity of the gray distribution

$$fe_3 = - \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} P_{d,\theta}(i,j) * \log(P_{d,\theta}(i,j)) \tag{8}$$

4. Variance Feature
   To measure the spread ability of the pixel values at around the combination of referenced pixels and the neighbourhood pixels.

$$fe_4 = \sum_{i=0}^{N_g-1} \sum_{j=0}^{N_g-1} (i-\mu)^2 P_{d,\theta}(i,j) \tag{9}$$

2.5. Neural Network

In Figure 2, the example of backpropagation neural network configuration that consists of 3 neuron units in the input layer, x1, x2, and x3; 1 hidden layer with 2 neurons, namely z1 and z2; and 1 unit in the output layer, y. The weight that connects x1, x2, and x3 with the first neuron in the hidden layer, is v11, v21, and v31; bl1 and bl2 are the bias weights leading to the first and second layers of the hidden layer. The weight that connects z1 and z2 with the neurons in the output layer, is w1 and w2. The bias weight of b2 connects the hidden layer with the output layer. The activation function used between the input layer and the hidden layer, and between the hidden layer with the output layer is the logsig activation function.
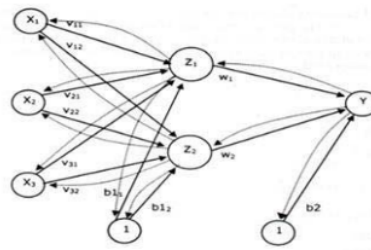

Figure 2. The Architecture of Backpropagation Neural Network [7]

Each ouput units (Yk, k=1,2,3,...,m) improve the bias and their weights (j = 0,1,2,...,p) :

$$Wjk(new) = Wjk(old) + \Delta wjk \tag{10}$$

$$b2k(new) = b2k(old) + \Delta b2k \tag{11}$$

Each hidden units (Zj, j=1,2,3,...,p) improve the bias and their weights (i=0,1,2,...,n) :

$$vij(new) = vij(old) + \Delta vij \tag{12}$$

$$b1j(new) = b1j(old) + \Delta b1j \tag{13}$$

3. Result and Discussion
The result of image enhancement in histopathology breast cancer slides is shown in Figure 3. From these figures, we can distinguish the better performance between before and after image's contrast enhancement. After enhancement, the images appear clearer than original digital images. This results could help the clinicians and diagnosing software when observing the image become more accurate.
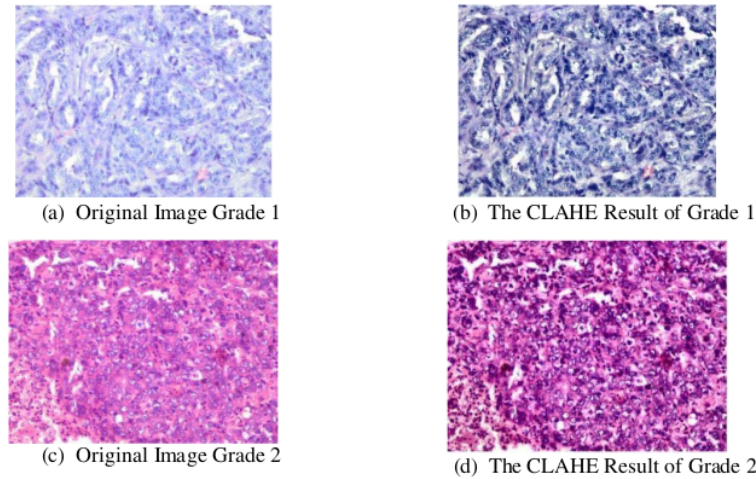
(a) Original Image Grade 1

(b) The CLAHE Result of Grade 1

(c) Original Image Grade 2

(d) The CLAHE Result of Grade 2

Figure 3. The result of CLAHE Method

The result of image segmentation using $k-$ means clustering is shown in Figure 4. This method can separate the pleomorphism, mitotic, and tubule formation. The lack of this method is still give unclear for these different patterns, then it could be make the difficulty of grading classification. This research uses the feature extraction method based on texture GLCM to find the accurate feature from the segmented results.
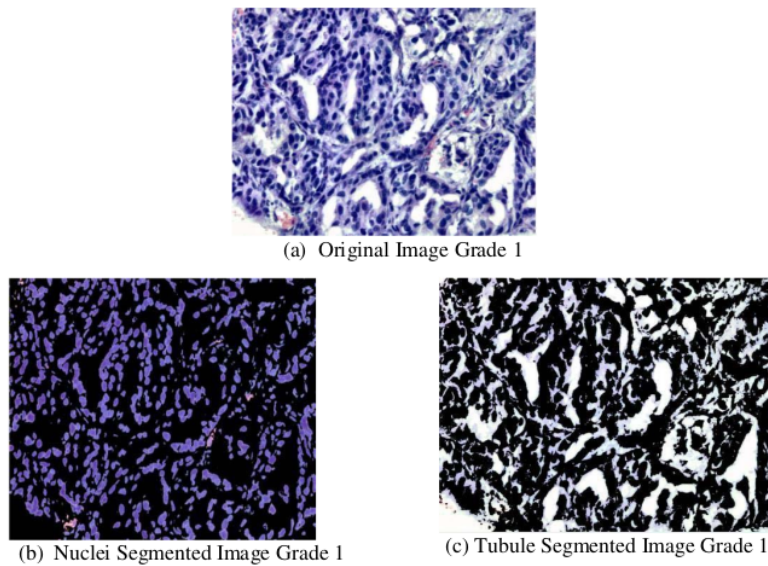

(a) Original Image Grade 1

(b) Nuclei Segmented Image Grade 1

(c) Tubule Segmented Image Grade 1

Figure 4. The Result of Image Segmentation for Grade 1

(a) Original Image Grade 3

(b) Nuclei Segmented Image Grade 3
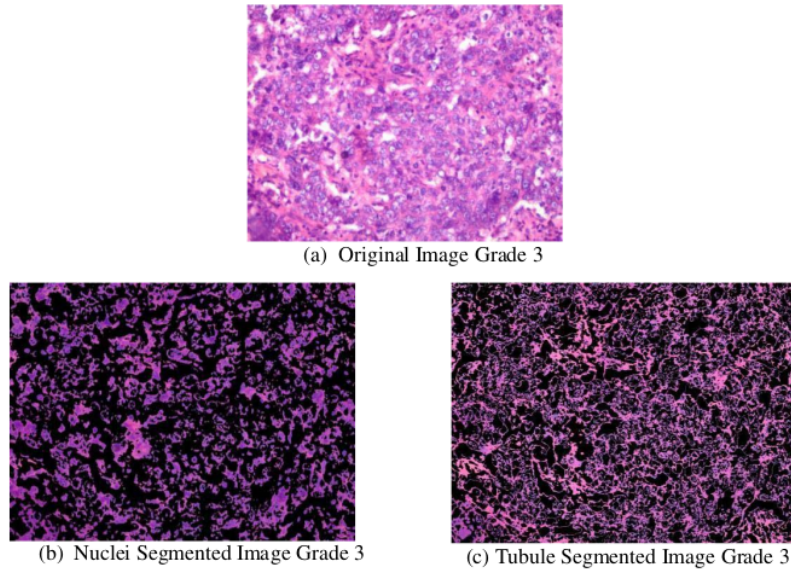

(c) Tubule Segmented Image Grade 3

Figure 5. The Result of Image Segmentation for Grade 3

Feature extraction is done on training data as well as testing to find out the characteristics that describe image diagnosis. This research uses the second-order statistical method, Gray Level Co- occurrence Matrices (GLCM) to determine the four texture features of the image, namely energy, entropy, contrast and entrophy. GLCM is used to determine the probability of pixel pairs neighboring to a certain degree of gray, distance, and angle. The GLCM research was carried out in four angular directions : 0°, 45°, 90°, and 135° with a distance of 1 pixel. [9] recommends only using the average value of texture features from all four angles if you want to avoid the dependence on neighboring pixels.

Backpropagation training in this research has aim to obtain a weight that is able to produce the calculation of the output as close as possible to the target so as to produce high accuracy. The initial weights are chosen randomly with a range of values from -1 to 1. The training phase is carried out by variations in backpropagation parameters, namely the maximum epoch, hidden layer, and learning rate. The results of backpropagation training are presented in Figure 5. The training results are expressed in terms of percentage accuracy. Training accuracy also illustrates the ability of backpropagation to recognize a given pattern. For input variations with the highest accuracy obtained from training with MSE value 0.0068. The backpropagation testing has aim to find out the backpropagation capabilities that have been built and trained to recognize new patterns. The backpropagation neural network can classify the histopathology grading of breast cancer with 88% accuracy, 85% sensitivity, and 80% specificity.

Conclusion

The results of image segmentation in breast cancer histopathology images are able to separate the histopathology grading pattern. The highest training accuracy is obtained using a variety of parameter

values generated from GLCM and the training data process in backpropagation neural network is 1 for the learning rate, 10 neurons in the hidden layer, and 10000 maximum epochs, with 88% accuracy, 85% sensitivity, and 80% specificity.

It is necessary to develop a method that is carried out by developing a more accurate feature extraction method by applying the feature selection method, so that it is expected to increase the accuracy value of the classification process.

References

[1] Kementerian Kesehatan RI, *"Riset Kesehatan Dasar 2013- Kanker Payudara"*, Badan Penelitian dan Pengembangan Kesehatan Kementerian Kesehatan RI, (2013).

[2] Lei He, L. Rodney Long, S. Antani, G., R. Thoma*, "Histology image analysis for carcinoma detection and grading"*, NCBI, *(2012).*

[3] T., Wana, J., Cao, J., Chen, Z., Qin, *"Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features"*, Neorocomputing, 229, 34-44, (2017).

[4] Histological Grading And Prognosis In Breast Cancer A Study Of 1409 Cases Of Which 359 Have Been Followed For 15 Years,H. J. G. Bloom And W. W. Richardson, From The Meyerstein Institute Of Radiotherapy And The Bland-Sutton Institute Of Pathology Of The Middlesex Hospital, London, W,1957.

[5] Singh, B.B, Patel, S., 2017, *"Medical Image Enhancement using CLAHE Enhencement and Wavelet Fusion* ", International Journal of Computer Applications, Volume 167 – No.5.

[6] J.,Wu, 2012, *Advances in K-Means Clustering-Data Mining Thinking*, Springer Theses, Recognizing Outstanding Ph.D Research, London.

[7] R., Rulaningtyas, A.B. Suksmono, T.L.R, Mengko, P. Saptawati, 2016 *"Nearest Patch Matching for Color Image Segmentation Supporting Neural Network Classification in Pulmonary Tuberculosis Identification"*, AIP Conference Proceeding.

[8] L., Fauset, 1994, *Fundamental of Neural Network.* Prentice Hall, Inc, United Stated.

[9] S., Kusumadewi, *"Membangun Jaringan Saraf Tiruan (menggunakan MATLAB dan Excel Link)".* Yogyakarta: Graha Ilmu, (2004).

[10] Y., Purwanto, R., Soelaiman, Irwanto, *"Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra"*, Jurnal Teknik ITS, (2012).

[11] A., Eleyan, H., Demierl, *"Co-occurrence Matrix and its Statistical Features as a new Approach For Face Recognition"*, Turkish Journal of Electrical Engineering and Computer Sciences, 19(1) : 97-107, January, (2011).

[12] K. D. Kunze, G. Haroske, V. Dimmer, W. Meyer and F. Theissig, *"Grading and Prognosis of Invasive Ductal Mammary Carcinoma by Nuclear Image Analysis in Tissue Sections"*, *Institute ofPathological Anatomy, Medical Academy "Carl Gustav* Carus'; *Dresden, GDR,* (1989).

[13] F. Albregtsen, *"Statistical Texture Measures Computed from Gray Level Coocurrence Matrices"*, Image Processing Laboratory, Department of Informatics, University of Oslo, (2008).

Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method

7	www.fda.gov Internet Source	1%
8	Yansu Hu, Zhijun Wang, Xiaoguang Fan, Junjie Li, Ang Gao. "Material microstructures analyzed by using gray level Co-occurrence matrices", Chinese Physics B, 2017 Publication	1%
9	IFMBE Proceedings, 2009. Publication	1%
10	e-journal.unair.ac.id Internet Source	1%
11	www.tib.eu Internet Source	1%
12	www.sciencepub.net Internet Source	1%
13	www.scribd.com Internet Source	1%
14	Lian Liu, Xiukun Yang, Yao Liu, Xiaojun Jing, Qingbo Ji. "A characteristic absorption peak interval method based on subspace partition for FTIR microscopic imaging classification", Measurement Science and Technology, 2018 Publication	1%
15	www.hindawi.com Internet Source	<1%

16 "Sensors and Image Processing", Springer Science and Business Media LLC, 2018
Publication

<1 %

17 T. Sathwik, R. Yasaswini, Roshini Venkatesh, A. Gopal. "Classification of selected medicinal plant leaves using texture analysis", 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), 2013
Publication

<1 %

18 theses.lib.polyu.edu.hk
Internet Source

<1 %

19 diva-portal.org
Internet Source

<1 %

20 unsworks.unsw.edu.au
Internet Source

<1 %

21 zone.biblio.laurentian.ca
Internet Source

<1 %

22 George Livanos, Michalis Zervakis, Nikos Pasadakis, Marouso Karelioti, George Giakos. "Deconvolution of petroleum mixtures using mid-FTIR analysis and non-negative matrix factorization", Measurement Science and Technology, 2016
Publication

<1 %

23   Teuku Tahlil, Richard J. Woodman, John Coveney, Paul R. Ward. "Six-months follow-up of a cluster randomized trial of school-based smoking prevention education programs in Aceh, Indonesia", BMC Public Health, 2015
Publication   <1 %

24   A.N. Zulkeflee, W.N.J. Wan Yussof, R. Umar, N. Ahmad, F.S. Mohamad, M. Man, E.A. Awalludin. "Detection of a new crescent moon using the Maximally Stable Extremal Regions (MSER) technique", Astronomy and Computing, 2022
Publication   <1 %

25   aip.scitation.org
Internet Source   <1 %

26   hdl.handle.net
Internet Source   <1 %

27   repository.stikim.ac.id
Internet Source   <1 %

28   Elbetel Taye Zewdie, Abel Worku Tessema, Gizeaddis Lamesgin Simegn. "Classification of breast cancer types, sub-types and grade from histopathological images using deep learning technique", Health and Technology, 2021
Publication   <1 %

29 Asha Das, Madhu S. Nair, S. David Peter. "Computer-Aided Histopathological Image Analysis Techniques for Automated Nuclear Atypia Scoring of Breast Cancer: a Review", Journal of Digital Imaging, 2020
Publication

<1 %

Exclude quotes	On	Exclude matches	Off
Exclude bibliography	On		

Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method

FINAL GRADE

/100

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9