Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BREAST CANCER HISTOLOGY IMAGESCLASSIFICATION BASED ON HYBRID FEATURE AND XGBOOST

HAJAR SAOUD[1], PR.ABDERRAHIM GHADI[2], PR.MOHAMED GHAILANI[3]

[1,2]Data & Intelligent Systems (DIS) Team, FSTT, AbdelmalekEssaadi University, Tetouan, Morocco

[3]LabTIC Laboratory, AbdelmalekEssaadi University, Tangier, Morocco

E-mail: [1]saoudhajar1994@gmail.com, [2]aghadi@uae.ac.ma, [3]mghailani@uae.ac.ma

ABSTRACT

Breast cancer is the most frequently diagnosed cancer in women and the most fatal ones, the numbers of incidences and mortalities increase every year.The proposition of decision support system is become a primordial need for the early diagnostic of breast cancer, many systems have been proposed in the last years for the classification of breast cancer based on histology images into its two categories benign and malignantand also into 8 sub-classes using machine learning and several methods of features extraction but nuclei segmentation is an important step that should be taken in consideration because nuclei is the key to establish a diagnosis of malignancy due to the transformation that can undergo when cancer occurs. Therefore, in this paper we will try to propose an approach for breast cancer classification in early stage based on nuclei segmentation.7909 Haemotoxylin& Eosin stained histology images were used to extract significant features for breast cancer classification into its two categories: Benign and malignant also into its eight sub-classes: adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma for benign and ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma for malignant. Image segmentation was done by Multi Otsu Thresholding algorithm followed by Watershed segmentation for images elements separation and for facilitating the morphological feature extraction. Color, texture and morphological features were extracted and classification was done by XGBoost ensemble machine learning algorithm. The proposed architecture produced an accuracy of 94.67%in 400X magnification for the binary prediction of breast cancer from histopathology images. The results are competitive compared to the results of other state-of-the-art methods.

Keywords: *Breast cancer, histology Images, Multi Otsu Thresholding, Watershed Segmentation, Machine Learning, XGBoost.*

1. INTRODUCTION

The word “cancer” is a vague term which groups several diseases that can touch various organ of the body [1]. Cancer is a disease that characterized by a transformation of cells that become abnormal which multiply rapidly and can touch the nearest organs. It is one of the main causes of death in the world, almost 10 million deaths in 2020. Breast cancer comes first in terms of incidence (2,26 cases) in 2020 and fifth in term of death (685 000 cases) in 2020 [1].

The most important step in breast cancer detection is biopsy because it is the only test that can establish the diagnosis of breast cancer and the analysis of the cells extracted during the biopsy provides a set of essential and decisive information for the choice of care and treatment.

The sample extracted with the biopsy will then be examined by the microscope not only to determine the presence of cancer but also to examine cancer cells: shape, size, sensitivity to certain hormones, presence or absence of mutations, examine how cancer cells might respond to treatment and determine the origin of a secondary cancer (metastasis)[2] .this operation can be time consuming, so the proposition of decision making algorithms can help pathologist in the decision making during this operation.

The aim of this paper is to propose an approach for breast cancer classification using histology images starting with image segmentation using Multi-Otsu thresholding and watershed segmentation to extract morphologic features and fusing them with Haralicks and colors features extracted from the originals images then classification using XGBoost. Good results were

achieved with this approach compared with art of states works.

Image segmentation is an image processing operation consisting of detecting and assembling the pixels according to some criteria. We used Multi-Otsu thresholding to separate nucleus from others components of the cell and watershed segmentation to separate connected nucleus.

XGBoost is an ensemble machine learning algorithm that proved it performance in last years it is an implementation of Gradient Boosted decision trees we will use it in classification step to improve the accuracy.

The rest of this paper is structured as follow: section 2 present state of art works, section 3 give a vision about our proposed approach: materials used and methodology proposed, experimentation and results of the proposed approach in part 4 and finally conclusion.

2. RELATED WORKS

In last year's several researchers have been interested by the domain of breast cancer classification using histology images and machine learning techniques. In this section we will give a vision about same of them:

In our previous researches we tried to classify breast cancer using several databases and machine learning techniques. We started [3] by doing a comparison between supervised and unsupervised learning in the classification of breast cancer using the two databases WBCO (Wisconsin Breast Cancer Data Set Original) [4]and WBCD (Wisconsin Breast Cancer Data Set Diagnostic)[5] the results show that supervised classifiers give better accuracy than unsupervised ones because we have defined classes, then we tried [6] to do a comparison between six supervised machine learning techniques to classify breast cancer using WBCO (Wisconsin Breast Cancer Data Set Original). After that we improved the accuracy of classification using feature selection technique [7] and voting classifiers [8]. In those researches we worked with non-massive databases, In [9] we tried to test the performance of machines learning techniques on large SEER dataset also to improve their accuracy using ensemble methods . In this work we will try to classify breast cancer using histology images and methods of images segmentation.

In [10] this paper, they presented the Breakhis database for the first time then they tried to classify breast cancer into its two classes benign and malignant using 6 features extraction techniques Local Binary Patterns (LBP), Completed Local Binary Pattern (CLBP), Local Phase Quantization (LPQ), Gray Level Co-Occurrence Matrices (GLCM), PFTAS and ORB and 4 classifiers: 1-Nearest Neighbor (1-NN), Quadratic Linear Analysis (QDA), Support Vector Machines (SVM), and Random Forests of Decision Trees.

In this paper [11], authors classified breast cancer based on color and texture features and ensemble classifier the color-texture features extracted are: Normalized color space representation, Multilayer Coordinate Clusters representation, Gabor features on Gaussian color model, Gabor chromatic features, Complex Wavelet features and chromatic features, Opponent Color Local Binary pattern(OCLBP) and classification using voting classifiers.

In this work [12], they compared the performance of convolutional neural networks with various configurations for the classification of breast cancer, using Breakhis dataset histology images, into benign and malignant, and also into benign and malignant sub-classes too using SURF and DSIFT. The performance of the multi-class classification was lower when compared to the one of the binary classification due to the number of handled classes and also due to the similarities between the sub-classes.

In [13] they compared the performance of traditional machine learning and deep learning approaches for magnification dependent histopathological breast cancer image classification using Breakhis dataset. In machine learning approach the algorithms used in feature extraction are Zernike moments, Haralick, and color histogram and the machine learning that achieves the higher accuracy is Random Forest.

In this paper[14] they try to present a fine annotations for nucleus segmentation using BreakHis dataset. The texture features for breast histopathological images are studied at the region level which fails to capture variations at the pixel level. They analyze various texture features of breast histopathological images at the pixel level to segment the nucleus. The input histopathological images are preprocessed by performing contrast enhancement and color normalization to eliminate the variations in illumination and color. GLCM, filter bank and LBP texture features are extracted along with various color features in RGB and LAB color space from breast histopathological images at pixel-level. SVM and MLP classifiers are trained on

these features to classify each pixel as nucleus and background.

In this paper [15], they tried to classify breast cancer histology images into its two classes using KAZE features and a bag of features classifier. The experimental results obtained on the BreaKHis database confirm the feasibility of the proposed approach for the classification of breast cancer histopathological images at low magnification factors.

In [16]they tried to classify breast cancer into its two categories benign and malignant based on breakhis dataset, starting with image pre-processing by applying normalization technique to increases the quality of image then image segmentation using superpixel with Particle Swarm Optimizer (PSO) and Grey Wolf Optimizer (GWO) to separate nuclei from non-nuclei cells. And the features extracted are Local Direction Ternary Pattern (LDTP), perimeter, solidity, circularity, Grey Level Co-Occurrence Matrix (GLCM), eccentricity, and color autocorrelogram and finally classification by support vector machine.

A Scaling multi-instance support vector machine was proposed in[17] this paperto classify breast cancer into its two categories based on breakhis dataset.

In this study [18], they proposed an approach for prostate cancer grading using histopathology images and supervised learning methods, The segmentation process for biopsy tissue image was performed using the k-means algorithm and touching cells were separated using the watershed algorithm. Then Morphological features were extracted for prostate cancer grading and diagnosis then classification was done using support vector machine to binary classify malignant vs. benign, Grade 3 vs. Grade 4+5, and Grade 4 vs. Grade 5.

In this paper [19], a classification method based on 3-output CNN segmentation and hybrid feature extraction is proposed. In the process of segmentation, the probability growing method is used to fuse the boundary probability map and the nuclear probability map to achieve the effect of fine segmentation. For the step of feature extraction morphological, spatial and texture features were extracted and feature classification were performed using support vector machine.

In this paper [20], they proposes a CAD system to semi-automatically segment and classify H&E stained thyroid histopathology images into two classes : Normal Thyroid (NT) and Papillary Thyroid Carcinoma (PTC). Particle Swarm Optimization (PSO)-based Otsu's multilevel thresholding is used to segment the images into different binary images. The stringent area and roundness constraint eliminate artefacts and restrict the images to contain only nuclei regions.

3. MATERIAL AND METHODOLOGY

3.1 Proposed Approach

In this section we will try to present in details our proposed approach. In this paper we proposed a methodology for breast cancer classification based on histology images of Breakhis dataset. We started by image segmentation using multi-otsuthresholding and watershed segmentation to extract morphological feature from nucleus segmented and fused them with texture and color features extracted directly from original images and then we do classification by XGBoost classifier. The figure 1 gives a vision about the proposed approach.

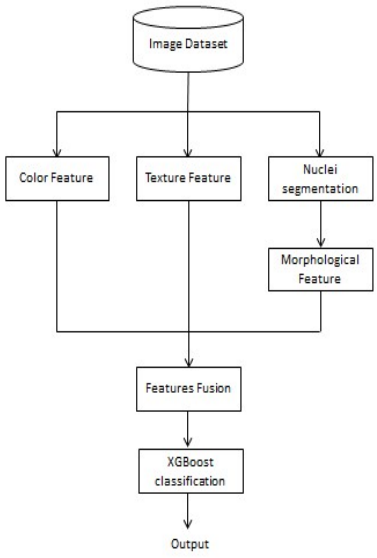


Figure 1: proposed methodology

3.2 Breast Cancer Histopathological Database

Breast Cancer Histopathological Database (BreakHis) [21]contains a total of 7909 (2,480 benign and 5,429 malignant) histopathological images collected from 82 patients (24 benign and 58 malignant) with different magnifying factors (40X, 100X, 200X, and 400X). Each class contain four sub-classes adenosis, fibroadenoma, phyllodes tumor, and tubular adenoma for benign and ductal carcinoma, lobular carcinoma, mucinous carcinoma and papillary carcinoma for malignant. Details of

BreakHis dataset are structured in tables 1, 2 and 3 from [10]. And same examples of images in 40X Magnification are given in Figure 2.

Table 1: BreakHis dataset.

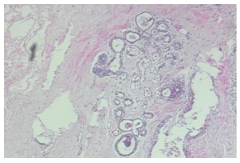
	Magnification	Benign	Malignant
40X	625	1,370	1,995
100X	644	1,437	2,081
200X	623	1,390	2,013
400X	588	1,232	1,820
Total of Images	2,480	5,429	7,909

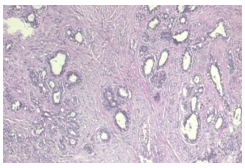
Table 2: Details of benign distribution.

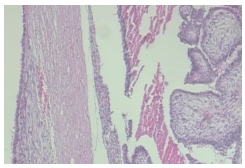
Magnification	A	F	TA	PT	Total
40X	114	253	109	149	625
100X	113	260	121	150	644
200X	111	264	108	140	623
400X	106	237	115	130	588
Total	444	1014	453	569	2480

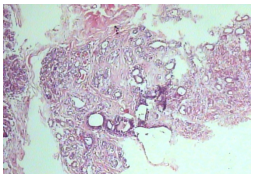
Table 3: Details of malignant distribution.

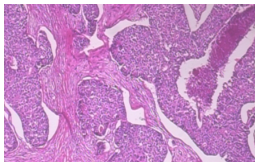
Magnification	DC	LC	MC	PC	Total
40X	864	156	205	145	1370
100X	903	170	222	142	1437
200X	896	163	196	135	1390
400X	788	137	169	138	1232
Total	3451	626	792	560	5429

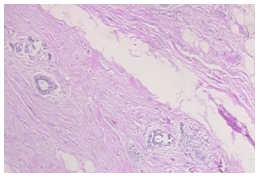

Adenosis

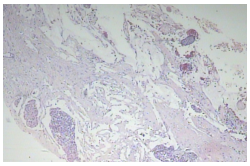

Fibroadenoma


Phyllodes tumor


Tubular adenoma


Ductal carcinoma


Lobular carcinoma


Mucinous carcinoma

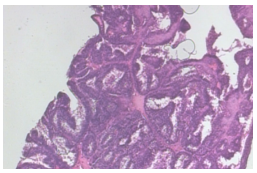
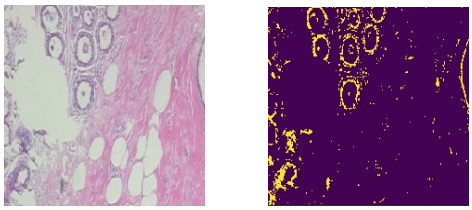

Papillary carcinoma

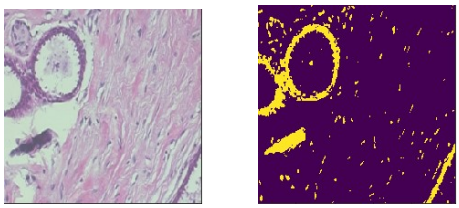
Figure 2: The four types of benign and malignant breast cancer images from the BreakHis dataset. In the 40X magnification

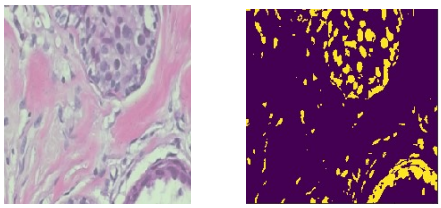
3.3 Multi-otsuthresholding

Multi-Otsu thresholding[22] is an algorithm that is used to separate the pixels of the input image transformed, into gray level image, into different classes, each one obtained according to the intensity of the gray levels within the image. The default number of classes or threshold extracted by multi-Otsu thresholding is three, but it can be determined as required.

In this paper we divided histology images into four classes to separate nuclei from the other component of histology image. Figure 3 presents the results of multi-Otsu threshold in the four magnifications.

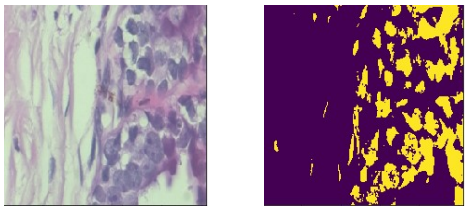

(a)40X magnification


(b)100X magnification


(c)200X magnification

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(d)400X magnification

Figure 3: Result of multi-Otsu thresholding for nuclei segmentaion

After nuclei segmentation we remark that nuclei are connect so we will use Watershed segmentation to separate connected nucleus.

3.4 Watershed segmentation

Watershed algorithm is a region-based technique used for image segmentation especially when segmenting similar connected object, it is based on the principle of drainage divide, which separates adjacent drainage basins. Thus, the objective of watershed is to find watershed lines and segmented image based on three steps: calculating distance transform, calculating local maxima points and finally labeling markers.

For 400X and 200X magnifications we suffer from the problem of over-segmentation, so we use Gaussian filter after calculating distance transform to adjust the number of local maxima found.as shown in figure 4.

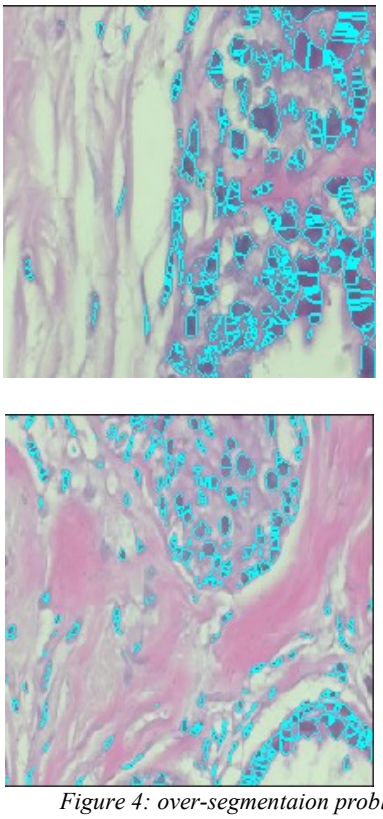
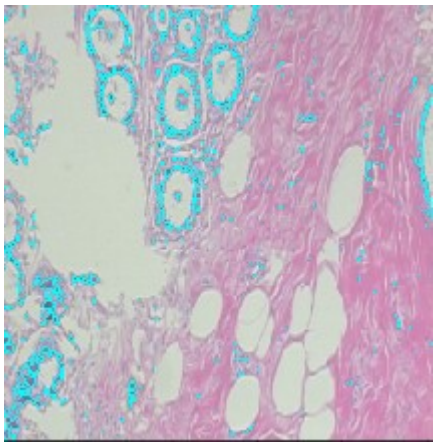
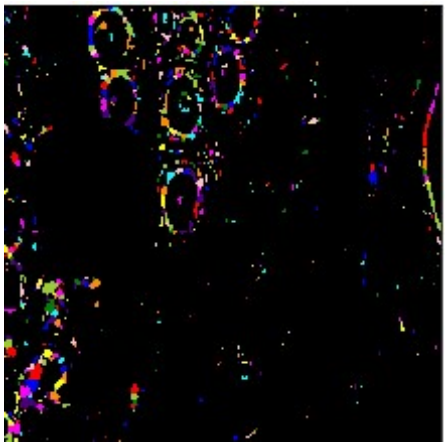


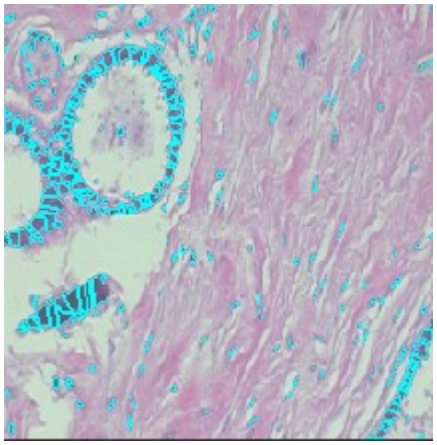
Figure 4: over-segmentaion problem

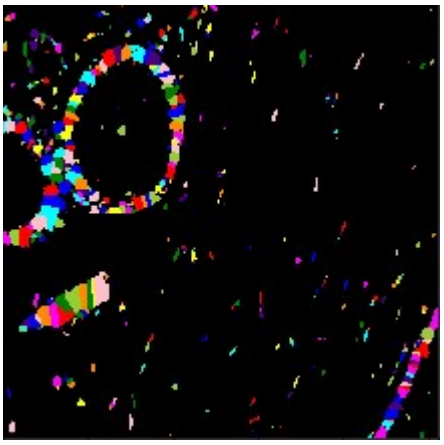
Then figure 5 present the result of watershed segmentation for every magnificastion:



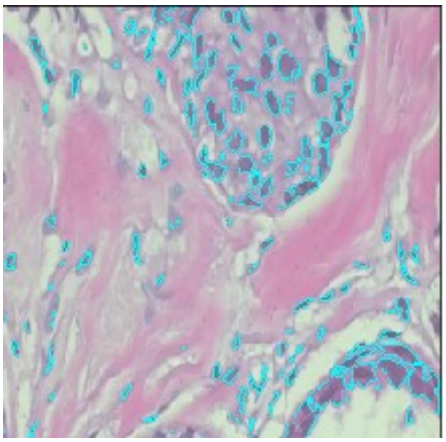


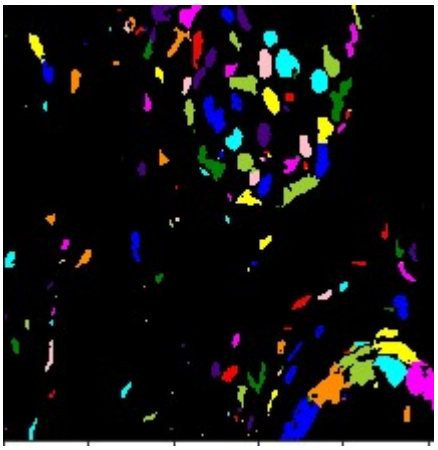
(a)40X



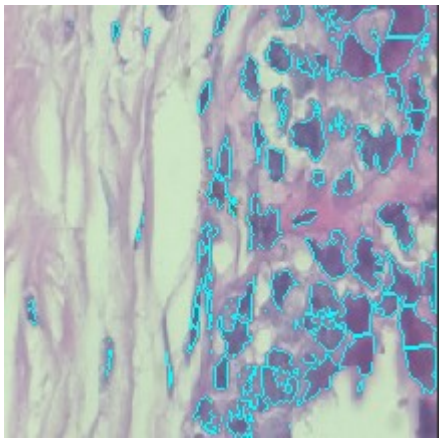


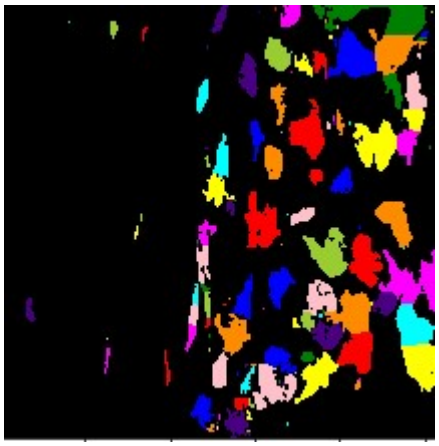
(b)100X





(c)200X





(d)400X

Figure 5:result of watershed segmentation

3.5 Feature extraction

Three feature extraction Techniques are used in this paper Haralick features to extract texture feature, Color histogram to extract color features and regionprops method proposed by Skimage library to extract morphological features after images segmentation.

3.5.1 Haralick Features

Haralicks features are one of the popular techniques of texture feature extraction that proved their performance in medical images field extracted from GLCM matrix of joint probabilities between pairs of pixels. After creating GLCM 13 Haralick features are extracted:

- Correlation
- Variance
- Inverse Difference Moment
- Sum Average
- Sum Variance
- Sum Entropy
- Entropy

•Information Measure of Correlation 2
•Difference Variance
•Difference Entropy
•Information Measure of Correlation
•Angular Second Moment
•Contrast

3.5.2 Color Histogram Features

Color is an important feature that can be extracted from cancer histology images due to the staining process. In histopathology, to differentiate between cellular components pathologists stains the nucleus and cytoplasm with different colors based on hematoxylin and eosin stain procedure. hematoxylin stains cell nuclei with purplish blue color, and eosin stains other structures with pink color. Therefore, to extract color features we used Color Histogram Feature. It represents the distribution of colors in images; here we worked in RGB space that will be divided in small color ranges, then Color Histogram Feature will represent the number of pixels that belongs to each color interval. 256 colors features were generated in this paper.

3.5.3 Morphological Features

After image segmentation, based on regionprops method, seven morphological features were extracted: number of nucleus, maximum and minimum nucleus area, portion of nucleus from the entire area of image, average of nucleus, minimum and maximum perimeter.

3.6 XGBoost classifier

In this paper we chose XGBoost in classification step because it proved its performance in last years. XGBoost is an ensemble machine learning algorithm. It is an implementation of decision tree algorithm under the Gradient Boosting framework designed for speed and performance, proposed by the researchers at the University of Washington. In XGBoost weight of variable is important because it built decision trees in sequential form. In first time the algorithm assigns weight in all the independent variables then they fed into decision tree that predict result. If the algorithm gives wrong weight it increase it and fed the variable in the second decision tree. In XGBoost the trees are built on the top of each other to correct the errors of the previous tree.

3.7 Evaluation metrics

To evaluate the performance of the proposed methodology we calculate Accuracy, Precision, F1-score and Recall based on confusion matrix.

Accuracy: Number of correct predictions divided by the total number of predictions

- Accuracy = (TP+TN)/(TP+FP+TN+FN)(1)

Precision: Number of correct predictions divided by the total number of positive predictions

- Precision = TP/(TP+FP) (2)

F1-score: Is the harmonic mean of precision and recall.

- F1-score = 2*(Precision * Recall)/(Precision + Recall) (3)

Recall: Number of correct predictions divided by the total number of Positive samples

- Recall = TP/(TP+FN) (4)

4 EXPERIMENTATION AND RESULTS

4.1 Results

In this section we will present the result obtained by our proposed approach for breast cancer classification using Breakhis dataset for binary and multi-class classification. Before the execution of our proposed method we divided the data into training and testing (70% training 30% testing) with stratified Train/Test-split.

After executing our proposed approach starting by images segmentation, features extraction and finally classification , and as shown in table 4 the higher accuracy obtained is 94.67% in 400X magnification followed by 200X magnification, then 100X and finally 40X magnification,which allows us to conclude that this approach work well in 400X magnification because the nucleus are clear and close with the microscope to study their morphology and also if magnification get smaller accuracy also get more smaller.

Table 4: result of binary classification.

Magnification	Accuracy	Precision	F1-score	Recall
40X	91.19%	91.12 %	91.09 %	91.19 %
100X	92.57 %	92.52 %	92.49 %	92.57 %
200X	94.43%	94.41 %	94.38 %	94.43 %
400X	94.67%	94.65 %	94.64 %	94.67 %

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Confusion matrix and ROC Curve also arepresented to further evaluate the performance of the proposed approach.The figure 6,7,8 and 9 present confusion matrix and ROC Curve of the four magnification.

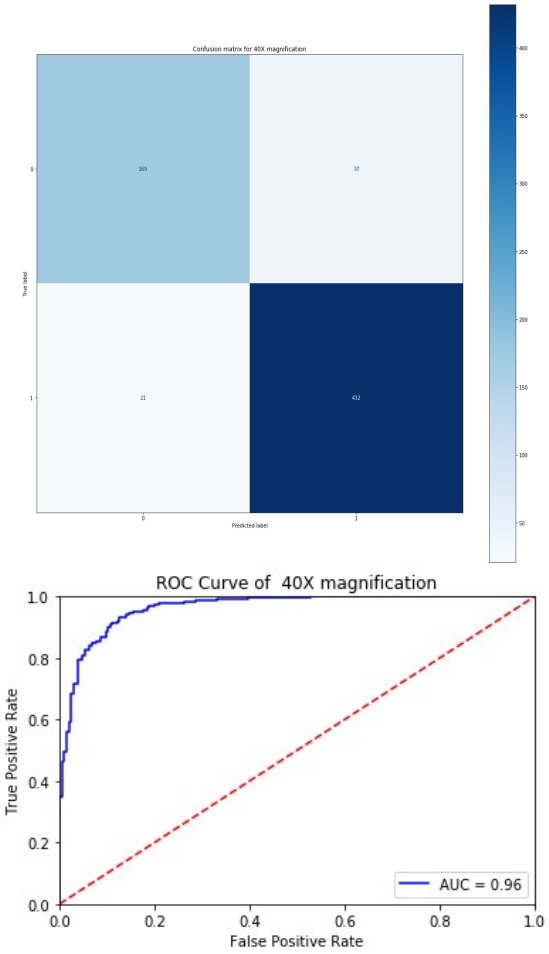
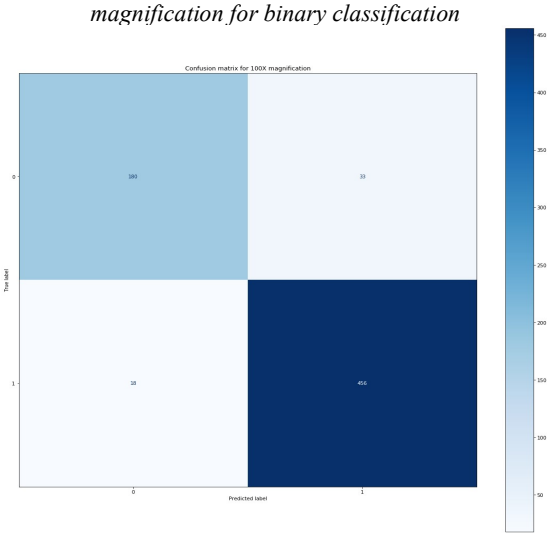


Figure6 :Confuxion matrix and Roc Curve of 40X magnification for binary classification



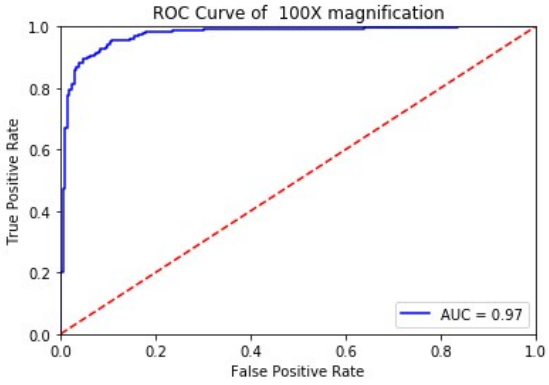
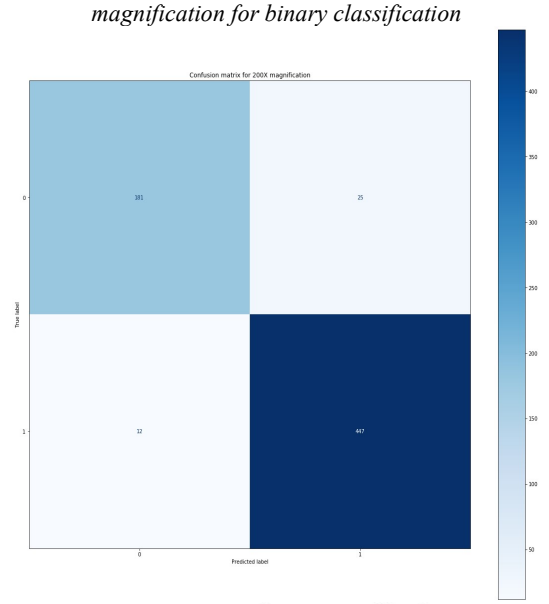


Figure7 : Confuxion matrix and Ruc Curve of 100X magnification for binary classification



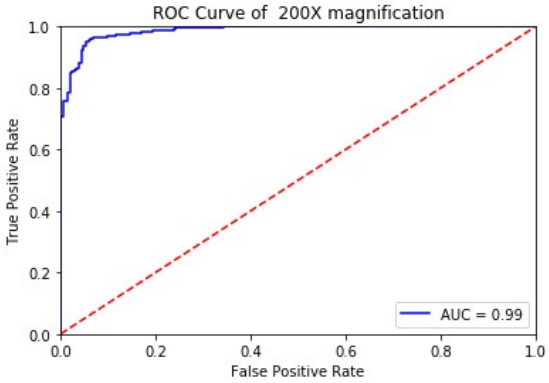


Figure8 :Confuxion matrix and Roc Curve of 200X magnification for binary classification

Journal of Theoretical and Applied Information Technology
15th January 2023. Vol.101. No 1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

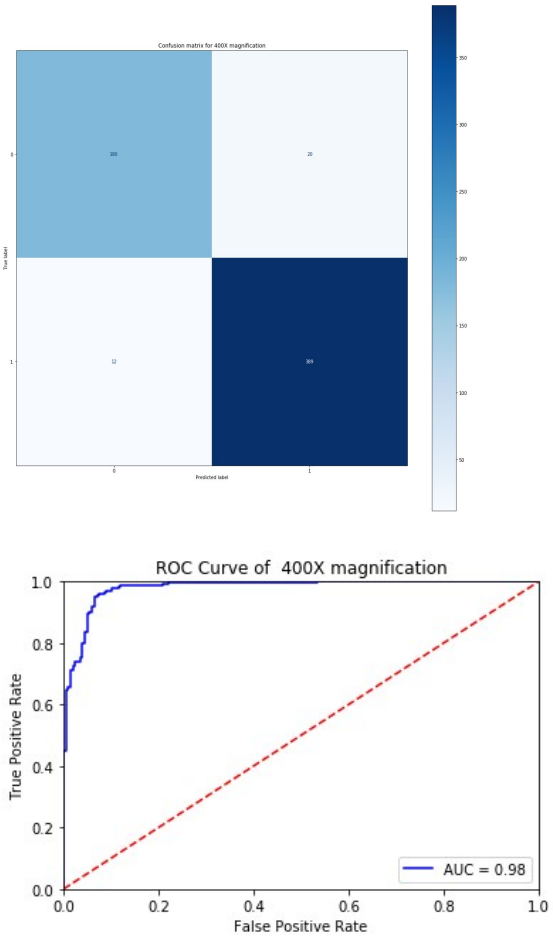


Figure9 :Confuxion matrix and Roc Curve of 400X magnification for binary classification

Also we applied our approach in eight sub-classes of breakhis dataset table 5 presents the obtained results, in multi-class classification higher accuracy is obtained also by 400X magnification which mean that is also more suitable also for multi-class classification. Followed by 200X magnification, then 100X and finally 40X magnification.

Table 5: result of sub-class classification

Magnification	Accuracy	Precision	F1-score	Recall
40X	75.72%	76.05%	74.89 %	75.72 %
100X	77.87%	78.19%	77.39 %	77.87 %
200X	77.89 %	77.81%	77.32 %	77.89 %
400X	78.86%	78.08%	77.88 %	78.86 %

Figures bellow present the confusion matrix and ROC Curve of the four magnification of eight class classification.

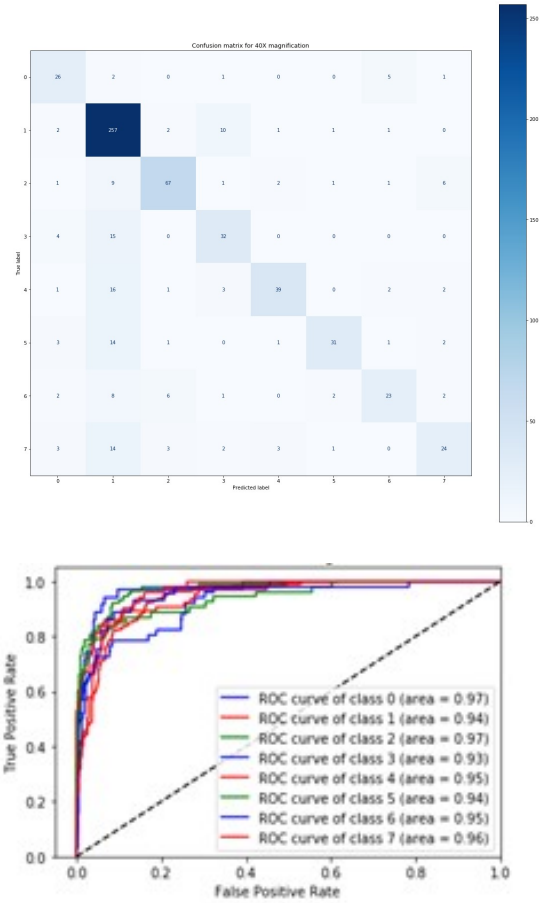


Figure10 :Confuxion matrix and Roc Curve of 40X magnification for Multi-class classification

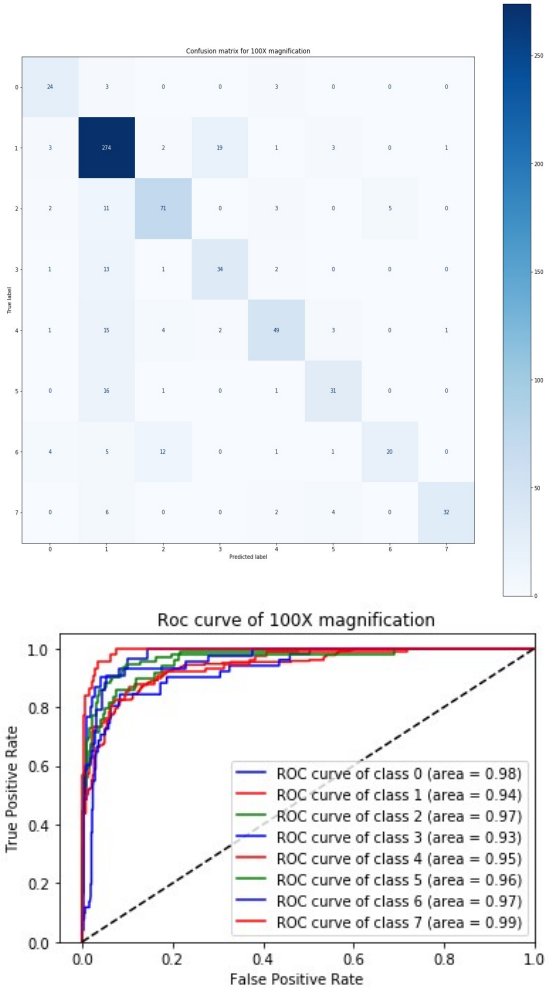
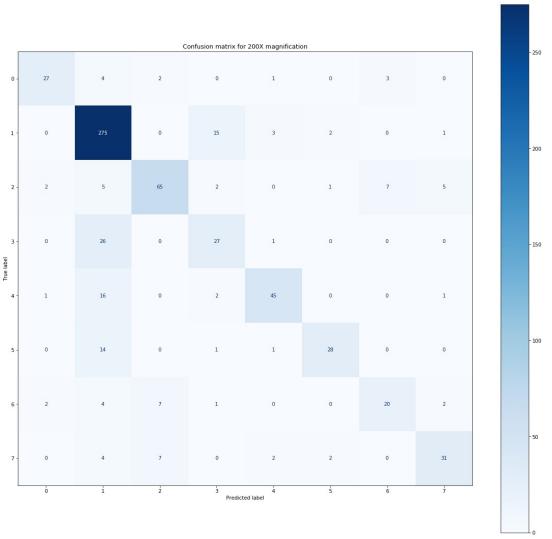


Figure11 : Confuxion matrix and Ruc Curve of 100X magnification for Multi-class classification



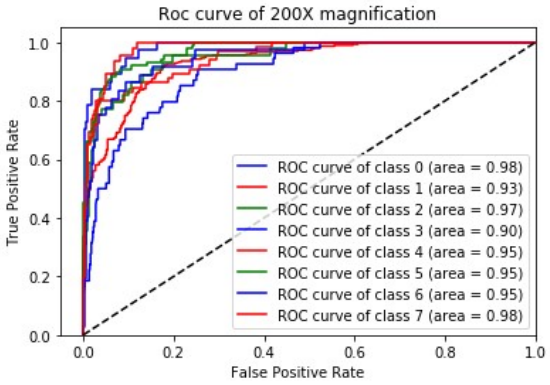
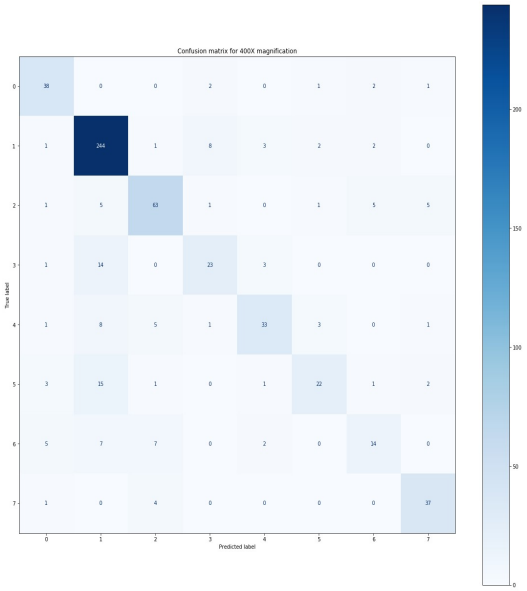


Figure12 : Confuxion matrix and Ruc Curve of 200X magnification for Multi-class classification



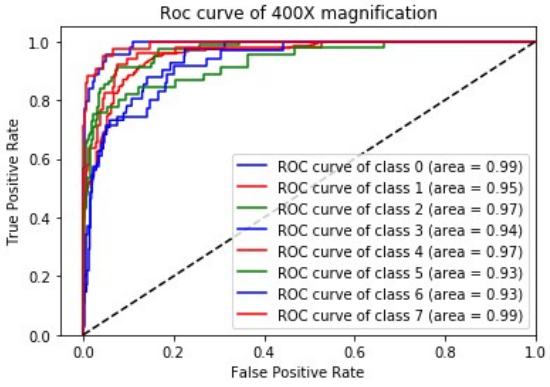


Figure13 : Confuxion matrix and Ruc Curve of 400X magnification for Multi-class classification

4.2 Discussion

In this paper we tried to propose an approach for breast cancer classification into its two

categories: benign and malignant and also into 8 sub-classes based on nucleus segmentation and XGBoostclassification. First, we started by nuclei segmentation following two steps: Multi-otsuthresholding for nuclei segmentation from non-nuclei elements then watershed segmentation to separate connected nucleus then feature extraction: texture and color feature were extracted directly from the original images and morphology features from the segmented images. And finally classification with XGBoost.Our approach was examined within Breakhis dataset. And as we know the classification of cancer and nuclei segmentation are a challenging tasks but in this paper we succeeded in the proposition of an approach that had classify breast cancer with higher results using machine learning approach.

To evaluate our approach we chooseas evaluation metric: Accuracy, Precision, Recall and F1-score that are calculated based on the confusion matrix and also the ROC curve to further assess the performance of the proposed approach in different magnifications.

For binary classification the accuracy achieved is between 91.19%and 94.67%and between 75.72% and 78.86% for multi-class. The raison of the low result in multi-class classification is the higher number of classes and also similarity between tabular adenoma and fibroadenoma as shown in the confusion matrix of multi-class classification and also the difficulty of classifying fibroadenoma in general.

4.3 Comparison with art of state works

In this section comparison with art of state works will be done, in [12] a comparison between convolutional neural network andmachine learning is done for binary and multi-class breast cancer classification in one of the configuration they combine SURF with CNN. Then in [13] they also compared machine learning approach with deep learning approach, for machine learning approach the feature extracted were Zernike Moments Features, Haralick Features and Color Histogram Features and the classifiers used were Random Forest, Support vector machine, Adaboost, Multilayer perceptron and K-nearest neighbor finally in[17],they proposed Scaling multi-instance support vector machine based on PFTAS and multi-instance SVM. Our research gives an encouraging results comparing with art of stat works using machine learning approach for binary classification as presented in table 6.

Magnification	Work	Features	classifiers	Accuracy
40X	[12]	SURF	CNN	85.45%
	[13]	Zernike Moments Features, Haralick Features and Color Histogram Features	Random Forest	87.69%
	[17]	PFTAS	multi-instance SVM	85.3%
	Our work	Texture features, color features and morphological features	XGBoost	91.19%
100X	[12]	SURF	CNN	79.77%
	[17]	PFTAS	multi-instance SVM	88.3%
	Our work	Texture features, color features and morphological features	XGBoost	92.57 %
200X	[12]	SURF	CNN	78.97%
	[13]	Zernike Moments Features, Haralick Features and Color Histogram Features	Random Forest	89.82%
	[17]	PFTAS	multi-instance SVM	89.8%
	Our work	Texture features, color features and morphological features	XGBoost	94.43%
400X	[14]	Filter Bank	Multilayer perceptron	86%
	[13]	Zernike Moments Features, Haralick Features and Color Histogram Features	Random Forest	85.65%
	[17]	PFTAS	multi-instance SVM	86.8%
	Our work	Texture features, color features and morphological features	XGBoost	94.67%

5 CONCLUSION

To conclude, as we know the diagnostic of breast cancer is challenging task and time consuming for pathologist. So in this paper we tried to propose an approach for breast cancer classification based on Histology images to help pathologist in the diagnosis of breast cancer. A combination of colors, texture and morphological feature was done to increase the accuracy of classification. Color features were extracted based on color histogram feature extraction technique, texture feature based on haralick features extraction technique and morphological features based on image segmentation using multi-otsuThresholding for region of interest segmentation to segment nuclei from others component of cell followed by watershed segmentation to overlapping cell nucleus. The higher accuracy obtained is 94.67% in 400Xmagnification for binary classification followed by 94.67% in 200X, 92.57 % in 100X and 91.19% in 40X and for multi-classification the higher accuracy is 78.86% in 40X magnification.

This approach worked well in binary classification but the result rest low in multi-class classification in future work we will try to propose a strong approach for multi-class classification of breast cancer histology images using deep learning feature that can also improve more the accuracy of binary classification.

REFERENCES:

[1]'Cancer'. https://www.who.int/fr/news-room/fact-sheets/detail/cancer (accessed Oct. 09, 2022).

[2]'La biopsie | Fondation contre le Cancer'. https://www.cancer.be/le-cancer/jeunes-et-cancer/les-examens/la-biopsie (accessed Oct. 10, 2022).

[3]H. Saoud, A. Ghadi, and M. Ghailani, 'Proposed approach for breast cancer diagnosis using machine learning', in Proceedings of the 4th International Conference on Smart City Applications, Casablanca Morocco, Oct. 2019, pp. 1–5. doi: 10.1145/3368756.3369089.

[4]'UCI Machine Learning Repository: Breast Cancer Wisconsin (Original) Data Set'. https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(original) (accessed Mar. 26, 2018).

[5]'UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set'. https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic) (accessed May 04, 2022).

[6]H. Saoud, A. Ghadi, M. Ghailani, and B. A. Abdelhakim, 'Application of Data Mining Classification Algorithms for Breast Cancer Diagnosis', in Proceedings of the 3rd International Conference on Smart City Applications - SCA '18, Tetouan, Morocco, 2018, pp. 1–7. doi: 10.1145/3286606.3286861.

[7] H. Saoud, A. Ghadi, M. Ghailani, and B. AnouarAbdelhakim, 'Using Feature Selection Techniques to Improve the Accuracy of Breast Cancer Classification: Special Issue on Data and Security Engineering', 2019, pp. 307–315. doi: 10.1007/978-3-030-11196-0_28.

[8]H. Saoud, A. Ghadi, and M. Ghailani, 'Hybrid Method for Breast Cancer Diagnosis Using Voting Technique and Three Classifiers', in Innovations in Smart Cities Applications Edition 3, Cham, 2020, pp. 470–482. doi: 10.1007/978-3-030-37629-1_34.

[9]H. Saoud, A. Ghadi, and M. Ghailani, 'BREAST CANCER DIAGNOSIS USING MACHINE LEARNING AND ENSEMBLE METHODS ON LARGE SEER DATABASE', .Vol., no. 3, p. 11, 2021.

[10]F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, 'A Dataset for Breast Cancer Histopathological Image Classification', IEEE Trans. Biomed. Eng., vol. 63, no. 7, pp. 1455–1462, Jul. 2016, doi: 10.1109/TBME.2015.2496264.

[11]V. Gupta and A. Bhavsar, 'Breast Cancer Histopathological Image Classification: Is Magnification Important?', in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, Jul. 2017, pp. 769–776. doi: 10.1109/CVPRW.2017.107.

[12]D. Bardou, K. Zhang, and S. M. Ahmad, 'Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks', IEEE Access, vol. 6, pp. 24680–24693, 2018, doi: 10.1109/ACCESS.2018.2831280.

[13]S. Boumaraf et al., 'Conventional Machine Learning versus Deep Learning for Magnification Dependent Histopathological Breast Cancer Image Classification: A Comparative Study with Visual Explanation', Diagnostics, vol. 11, no. 3, p. 528, Mar. 2021, doi: 10.3390/diagnostics11030528.

[14]R. Rashmi, K. Prasad, C. B. K. Udupa, and V. Shwetha, 'A Comparative Evaluation of Texture Features for Semantic Segmentation of Breast Histopathological Images', IEEE

Access, vol. 8, pp. 64331–64346, 2020, doi: 10.1109/ACCESS.2020.2984522.

[15]D. Sanchez-Morillo, J. González, M. García-Rojo, and J. Ortega, 'Classification of Breast Cancer Histopathological Images Using KAZE Features', in Bioinformatics and Biomedical Engineering, vol. 10814, I. Rojas and F. Ortuño, Eds. Cham: Springer International Publishing, 2018, pp. 276–286. doi: 10.1007/978-3-319-78759-6_26.

[16]R. Saturi and P. Chand Parvataneni, 'Histopathology Breast Cancer Detection and Classification using Optimized Superpixel Clustering Algorithm and Support Vector Machine', J. Inst. Eng. India Ser. B, vol. 103, no. 5, pp. 1589–1603, Oct. 2022, doi: 10.1007/s40031-022-00745-3.

[17]H. Seo, L. Brand, L. S. Barco, and H. Wang, 'Scaling multi-instance support vector machine to breast cancer detection on the BreaKHis dataset', Bioinformatics, vol. 38, no. Supplement_1, pp. i92–i100, Jun. 2022, doi: 10.1093/bioinformatics/btac267.

[18]Bhattacharjee et al., 'Quantitative Analysis of Benign and Malignant Tumors in Histopathology: Predicting Prostate Cancer Grading Using SVM', Appl. Sci., vol. 9, no. 15, p. 2969, Jul. 2019, doi: 10.3390/app9152969.

[19]C. Yu, H. Chen, Y. Li, Y. Peng, J. Li, and F. Yang, 'Breast cancer classification in pathological images based on hybrid features', Multimed.Tools Appl., vol. 78, no. 15, pp. 21325–21345, Aug. 2019, doi: 10.1007/s11042-019-7468-9.

[20]J. Angel Arul Jothi and V. Mary Anita Rajam, 'Effective segmentation and classification of thyroid histopathology images', Appl. Soft Comput., vol. 46, pp. 652–664, Sep. 2016, doi: 10.1016/j.asoc.2016.02.030.

[21]'Breast Cancer Histopathological Database (BreakHis) - LaboratórioVisãoRobótica e Imagem'. https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/ (accessed Nov. 26, 2022).

[22]'Multi-Otsu Thresholding — skimage v0.19.2 docs'. https://scikit-image.org/docs/stable/auto_examples/segmentation/plot_multiotsu.html#id2 (accessed Oct. 11, 2022).