Biosc.Biotech.Res.Comm. Special Issue Vol 13 No 13 (2020) Pp-352-357

BBRC
Bioscience Biotechnology Research Communications

A Systematic Survey on Automatic Classification of Breast Cancer using Histopathology Image

Shwetha G.K[1] and K R Udaya Kumar Reddy[2]

Department of Computer Science & Engineering, NMAM Institute of Technology, Nitte
Visvesvarya Technological University, Belagavi, Karnataka State, India

ABSTRACT

In recent decades, breast cancer is the frequent cancer type in women, worldwide. The breast cancer subjects faces irreversible conditions and even death due to post treatment and diagnosis. So, automatic classification of breast cancer utilizing image techniques has great application value in the early detection of breast cancer. Due to the advance in medical field, histopathology images are regularly used in the diagnosis tool to recognize and classify the abnormality and normality cells in the images. Since, extracting non-redundant and informative features from histopathology image is a challenging task, due to heavy noise conditions, and small variant nuclei cell size. In order to highlight these issues, several models are developed by the researchers for automatic classification of breast cancer. This article investigates the existing research works performed in histopathological breast cancer classification and the problems faced by the researchers in this research area. Further, this survey article will helps the researchers to achieve significant performance in segmentation and classification of breast cancer by highlighting the problems stated in the related work section.

KEY WORDS: BREAST CANCER CLASSIFICATION, DEEP LEARNING TECHNIQUES, HISTOPATHOLOGY IMAGES, IMAGE DENOISING, MACHINE LEARNING TECHNIQUES.

INTRODUCTION

Currently, cancer is the common disease in the world, and its mortality and morbidity rate continued to increase and actively harms human health. Based on American cancer society report, breast cancer has higher mortality and diagnosis rate among women (Sudharshan et al. 2019; Feng et al. 2018). So, the early detection of breast cancer is essential to increase the survival rate. In recent times, several imaging techniques are available for early diagnosis of breast cancer like magnetic resonance imaging, diffusion tensor imaging, histopathology, computerized tomography, X-ray, etc (Xu et al. 2015). Among the available imaging techniques, histopathology images gained more attention in the researcher's communities, because of its higher resolution characteristics. In breast cancer diagnosis, histopathological image analyses consumes limited time to diagnose, but it completely relies on clinical experience and it may easily causes misdiagnosis. In order to address this issue, automated diagnosis systems are developed to enhance the detection speed and to diminish the rate of misdiagnosis (Yang et al. 2019; Wan et al. 2017; Prabu et al. 2019). In this paper, the survey on histopathological breast cancer classification is accomplished to analyse the concerns and the performances achieved by the researchers in the existing research works.

This survey paper is pre-arranged as follows; Section 2 states the overview of histopathological breast cancer classification. A few recent research papers on the topic "breast cancer classification" is surveyed in the Section 3. The objective of the research is represented in the Section 4. Conclusion of the present research study is indicated in the Section 5.

2. Overview of Histopathological Breast Cancer Classification: In recent decades, breast cancer is the 2nd

ARTICLE INFORMATION

**Corresponding Author: gk.shwetha@nitte.edu.in*

Received 9th Oct 2020 Accepted after revision 31st Dec 2020
Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRCBA

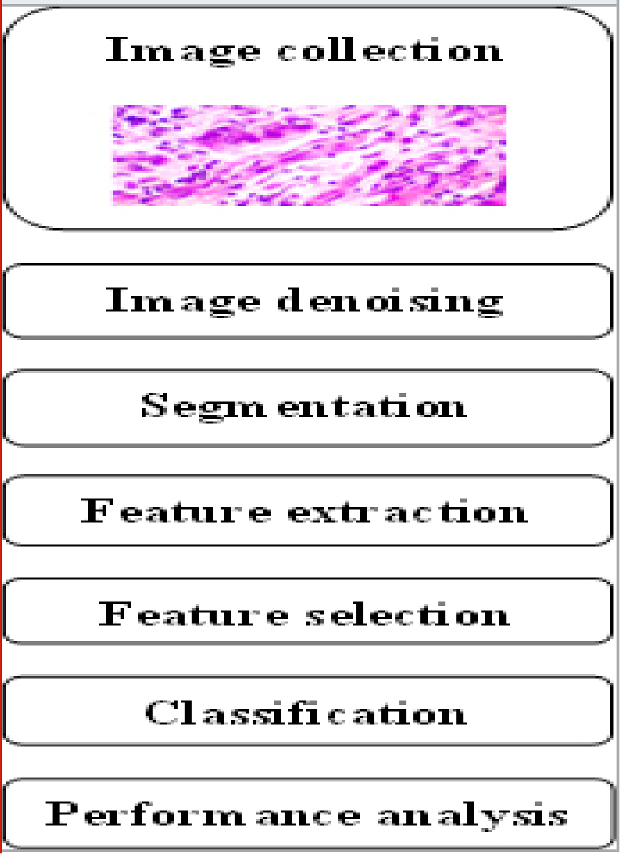
Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal

NAAS Journal Score 2020 (4.31)
A Society of Science and Nature Publication,

Online Contents Available at: http//www.bbrc.in/
Doi: http://dx.doi.org/10.21786/bbrc/13.13/54

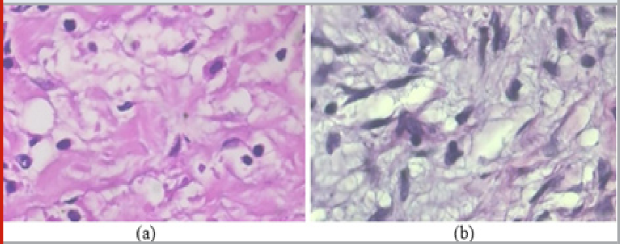
deadly cancer type in women after lung cancer. Based on the report of American cancer society, approximately 41760 people are died, due to breast cancer in the year of 2019, so it is essential to detect the breast cancer at an early stage for reducing the mortality rate. Due to the advance in medical filed, histopathological analysis is the extensively used imaging technique in breast cancer diagnosis. Usually, the automated histopathological breast cancer classification includes seven steps; image collection, image denoising, nuclei and non-nuclei segmentation, extraction of features, selection of optimal features, classification, and performance analysis. General overview of histopathological breast cancer classification is graphically indicated in figure 1.

Figure 1: Overview of histopathological breast cancer classification



1 Image collection: In present scenario, several datasets are available online for histopathological breast cancer classification such as BreaKHis, BreCaHAD, TMAD datasets, etc. In image processing application, it is necessary to choose a proper image dataset, otherwise the intent operations like segmentation, and classification cannot be achieved properly. Among the available datasets, BreaKHis is utilized in majority of the research works, where it includes 9,109 histopathological microscopic images of breast cancer, which are recorded from 82 patients under dissimilar magnifying factors; 40×, 100×, 200×, and 400×. In 9,109 microscopic images, 5429 images belongs to malignant class and residual 2480 images belongs to benign class. The sample image of BreaKHis dataset is represented in figure 2.

Figure 2: Sample image of BreaKHis dataset, a) malignant, and b) benign



2. Image denoising: After histopathology image collection, image denoising is employed to improve the visibility level of the collected images for better understanding of nuclei and non-nuclei cells. In recent times, several image denoising techniques are available for improving the quality of histology images such as filtering techniques, normalization, histogram equalization, contrast limited adaptive histogram equalization technique, etc. In histopathological breast cancer classification, image denoising phase is essential, because histopathology image is recorded using imaging instruments that may include machinery noise and impulse noise. So, image denoising techniques are preferred in histopathology based breast cancer classification to eliminate noise and to improve the visibility level of the images.

2.3 Image segmentation: Once the collected histopathological images are denoised, image segmentation is performed to segment non-nuclei and nuclei cells for early diagnosis of breast cancer. Generally, segmentation is determined as the process of dividing histopathology images into numerous regions, where the partitioned regions are homogeneous concerning to some image features like edges, blobs, corners, ridges, etc. The image segmentation phase aims in recognizing and extracting the regions which are constitute to classification. In recent scenario, several segmentation techniques are available, which are majorly classified into four types, as stated in figure 3.

Thresholding based segmentation: Multi-level Otsu thresholding, maximum entropy technique, global thresholding, etc.

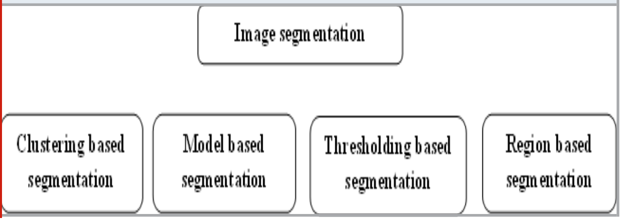
Region based segmentation: Region growing, uniform blocking, etc.

Model based segmentation: Sampling lines algorithm, active shape techniques, active appearance techniques, etc.

Clustering based segmentation: Fuzzy C means clustering, adaptively regularized fuzzy C means, K-means clustering, etc.

Feature extraction: After segmenting the nuclei and non-nuclei cells, feature extraction is accomplished by using local and global features. The feature extraction techniques maps the pixels into feature vectors in order to quantify the characteristics of segmented nuclei and non-nuclei cells.

Figure 3: Image segmentation algorithms in the medical field



Local features: Feature vectors determined from the sub division of image band are called as local features like shape, color, and texture of histopathology image. Usually, global feature descriptors works on the basis of local features.

Global features: The global feature descriptors consists of shape descriptors, local features, texture features and contour representations, which represents the image patch texture. Some of the global feature descriptors are determined as follows; histogram of oriented gradients, local ternary pattern, speeded up robust features, scale-invariant feature transform, etc.

Table 1. Performance metrics used in histopathological breast cancer detection

Segmentation related performance metrics	
Jaccard coefficient	$\frac{TP}{FP+TP+FN}$
Dice coefficient	$\frac{2TP}{2TP+FP+FN}$
Classification related performance metrics	
Accuracy	$\frac{TP+TN}{TN+TP+FN+FP}$
Precision	$\frac{TP}{FP+TP}$
Matthew's Correlation Coefficient	$\frac{TP\times TN - FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
F-score	$\frac{2TP}{FP+2TP+FN}$
Sensitivity	$\frac{TP}{FN+TP}$
Specificity	$\frac{TN}{FP+TN}$
Fowlkes-Mallows index	$\sqrt{\frac{TP}{TP+FN}\times\frac{TP}{TP+FP}}$

Feature selection: After extracting the feature vectors from the segmented nuclei and non-nuclei cells, feature selection or optimization technique is employed to select the features, which contributes to accurate or precise breast cancer classification. Few common techniques used in feature selection are listed as follows; Feature selection techniques: Ensemble feature selection algorithm, infinite algorithm, reliefF algorithm, etc. Optimization algorithms: particle swarm optimization, genetic algorithm, whale optimization algorithm, ant colony optimization, independent component analysis, principal component analysis, etc.

6 Classification: The selected optimal feature vectors are fed to a classification technique to classify the sub-stages of breast cancer. Generally, classification is determined as the mechanism of sorting objects into particular image class. Classification phase plays a crucial role in medical image applications, especially in histopathological breast cancer classification. Usually, the classification techniques are categorized into three types; semi-supervised, unsupervised, and supervised learning.

Unsupervised learning: The unsupervised learning techniques effectively identifies the unknown patterns in acquired data with-out pre-existing the labels. Some examples of unsupervised learning category are Deep Neural Network (DNN), auto encoder, Convolutional Neural Network (CNN), graph neural network, capsule network architecture, etc.

Supervised learning: Supervised learning techniques identifies the unknown patterns in acquired data by pre-existing the labels. Some examples of supervised learning category are decision tree, Support Vector Machine (SVM), Naive Bayes, etc.

Semi-supervised learning: Semi-supervised learning techniques classifies the acquired data with and without using the labels, generally it falls between the category of unsupervised and supervised learning techniques.

Some examples of semi-supervised learning category are generative methods, heuristic models, etc.

Performance measures: Performance analysis is the mechanism of acquiring, evaluating and reporting information regarding the developed model in breast cancer detection. The performance metrics utilized in this research area are listed in table 1. Where, TN is specified as true negative, TP is represented as true positive, FN is denoted as false negative, and FP is stated as false positive.

Table 2. Overview of the related works

Methodology	Dataset	Performance measure
Deep feature fusion model based on CpasNet and CNN (Prabu et al. 2019)	BreaKHis	Deep feature fusion model achieved mean classification accuracy of 92.71%, 94.52%, 94.03%, and 93.54% on 40×, 100×, 200×, and 400×.
Deep neural network (ResNet-18) (Wang et al. 2021)	BreaKHis	Achieved 98.42% of accuracy in binary classification, and 92.03% of accuracy in eight class classification.
Hybrid model based on CNN and RNN (Togaçar et al. 2020)	Real time	Achieved classification accuracy of 91.30% in four class classification.
BreastNet model (Budak et al. 2019)	BreaKHis	Obtained best binary classification result of 98.80% and four class classification result of 98.51%.
Bi-LSTM and FCN	BreaKHis	Obtained average classification accuracy of 95.69%, 93.61%, 96.32%, and 94.29% on 40×, 100×, 200×, and 400×.
DNN architecture (Xie et al. 2019)	BreaKHis	Mean classification accuracy of 97.98% in benign class, and 88.37% in malignant class.
Inception-ResNet-V2 and Inception V3 networks (Xie et al. 2019)	BreaKHis	Mean accuracy of 99.27%, 98.97%, 98.90%, and 98.74% on 40×, 100×, 200×, and 400×.
VGG 16, and AlexNet with SVM (Saha et al. 2018)	BreaKHis	Accuracy of 90.96%, 90.58%, 91.37%, and 91.30% on 40×, 100×, 200×, and 400×.
Deep learning model with handcrafted features (Man et al. 2020)	BreaKHis	Achieved 92% of precision, 88% of recall and 90% of f-score.
DenseNet 121-AnoGAN model (Man et al. 2020)	BreaKHis	Mean classification accuracy of 99.13%, 96.39%, 86.38%, and 85.20% on 40×, 100×, 200×, and 400×.

Related works: (Wang et al. 2021) implemented a deep feature fusion model for breast cancer detection using histopathology images, where the developed model includes the benefits of CpasNet and CNN. At first, a dual channel structure was developed for convolutional and capsule feature extraction, and then combine spatial and semantic features to get discrimination information. Additionally, routing coefficients were optimized for modifying the loss function and then embed the routing mechanism into optimization process. In this literature study, the developed model performance was validated on a benchmark BreaKHis dataset. In the experimental phase, the developed model achieved mean classification accuracy of 92.71%, 94.52%, 94.03%, and 93.54% on BreaKHis dataset (40×, 100×, 200×, and 400×), respectively. However, CNN requires enough training samples, orelse it leads to overfitting problem and reduces the generalization ability.

(Yan et al. 2020) developed a transfer learning based methodology for automatic diagnosis of breast cancer utilizing histopathological images. In this literature, deep neural network (ResNet-18) was developed for refining the network on histopathology images that avoids overfitting issue and speed up the training process. Hence, the experimental outcomes on BreaKHis dataset showed the effectiveness of the developed model. In binary class classification, developed model achieved 98.42% of accuracy, 98.75% of precision, 99.01% of recall, 98.88% of f-measure, and 96.19% of MCC. Similarly in eight class classification, the developed model achieved 92.03% of accuracy, 91.39% of precision, 90.28% of recall, 90.77% of f-measure, and 89.38% of MCC. However, the developed model was computationally expensive and occurred vanishing gradient concern, due to short-cut connections in ResNet-18 network.

(Togaçar et al. 2020) developed a new hybrid model for automatic breast cancer detection, where the developed model combines the benefits of CNN and Recurrent Neural Network (RNN). Hence, the developed model effectively preserves the long and short term spatial correlations between the patches based on the richer multi-level feature representation of the histopathology images. In the experimental phase, developed hybrid model achieved average classification accuracy of 91.30% in four class classification on an on-line available dataset. The undertaken dataset consists of 3771 histopathology images related to breast cancer. However, the unbalanced distribution and resource shortage were considered as the major issues in this literature.

(Budak et al. 2019) used BreastNet model to perform automatic breast cancer detection by utilizing histopathology images. Initially, the collected data was processed by the augmentation techniques to change the features of each image like shift, brightness, flip and rotation. Next, select the important key regions in the images using attention module, and then hyper-column technique was applied to obtain precise and stable classification. The developed BreastNet model comprises of residual, pooling, dense and convolutional blocks to achieve better classification performance. In the experimental section, developed model achieved best binary classification result of 98.80% and four class classification result of 98.51% on BreaKHis dataset (40×, 100×, 200×, and 400×). Here, the developed BreastNet model was deeper (maximum number of layers), so it includes the issues like exploding/vanishing gradients.

(Feng 2018) developed Bi-directional Long Short Term Memory (Bi-LSTM) and Fully Convolutional Network (FCN) for breast cancer detection. In this literature study, the FCN was utilized as an encoder to extract the features, which were fed into Bi-LSTM in order to ensure that histopathology images were utilized as direct input. Experimental result demonstrated that the developed model achieved significant performance in breast cancer detection, and the developed model achieved mean classification accuracy of 95.69%, 93.6%, 96.32% and 94.29% on BreaKHis dataset (40×, 100×, 200×, and 400×). However, Bi-LSTM network suffers from vanishing gradient issue that hampers the learning of histopathology images.

(Xie et al. 2019) developed a DNN architecture to perform cell nuclei recognition. Related to conventional learning techniques, the developed architecture achieved better performance in heavy noise conditions. In the experimental phase, the developed novel DNN architecture achieved mean classification accuracy of 97.98% in benign class, and 88.37% in malignant class. However, large datasets were needed for data training in DNN model, which was one of the major concerns faced by the researchers in histopathological breast cancer classification.

(Deniz et al. 2018) developed Inception-ResNet-V2 and Inception V3 networks to solve multi-class and binary class problems in breast cancer detection by utilizing transfer learning techniques. In this literature, the sub-classes were balanced with ductal carcinoma by tilting the histopathology images left, right, up and down, and rotate the images in 90° and 180° clockwise directions in order to overcome imbalance issue in the breast cancer classification.

Simulation results showed that the developed architectures achieved classification accuracy of 99.27%, 98.97%, 98.90%, and 98.74% on BreaKHis dataset (40×, 100×, 200×, and 400×). Major two issues in Inception-ResNet-V2 and Inception V3 networks were sensitive to noise and computationally expensive. (Deniz et al. 2018) used deep and transfer learning feature extraction techniques like Visual Geometric Group (VGG) 16, and AlexNet for extracting features from the images. The obtained features were fed to SVM for classifying the histopathology images as benign and malignant classes. The developed model achieved mean accuracy of 90.96%, 90.58%, 91.37% and 91.3% on BreaKHis dataset (40×, 100×, 200×, and 400×). The SVM was a binary classification technique which supports only binary class classification, which was inadequate to multi-class classification.

(Man et al. 2020) designed a deep learning architecture with handcrafted features for histopathology breast cancer detection. The developed deep learning architecture contain two fully connected layers, four rectified linear units, four max pooling layers, and five convolution layers. Additionally, the handcrafted features comprises of intensity, morphological, and textural features in order to avoid overfitting problem in the developed architecture. In the experimental phase, the developed architecture almost achieved 92% of precision, 88% of recall and 90% of f-score on BreaKHis dataset. In this study, the developed architecture was trained on 40×, 100×, 200×, and 400× magnification images, where the developed architecture was not worked properly on higher magnification images.

(Man et al. 2020) developed a new approach; denseNet 121-AnoGenerative Adversarial Networks (GAN) for automatic detection of breast cancer. DenseNet 121-AnoGAN model solves the mislabeled patch issue to improve the classification performance. The developed denseNet 121-AnoGAN model achieved classification accuracy of 99.13%, 96.39%, 86.38%, and 85.20% on BreaKHis dataset (40×, 100×, 200×, and 400×), respectively. As a future enhancement, the novel methods are needed to be developed for solving data imbalance problem in the medical domain. The overview of the related works is given in table 2.

Objective of the research: Objectives to achieve better performance in breast cancer classification using histopathology image are listed below,

- To develop a new denoising technique to enhance the visibility level of the histopathology images. It is essential to select the proper denoising technique, or else the intent operations gets affected.
- To develop a hybrid clustering/thresholding method to enhance segmentation performance in the complex or overlapped background region, also it improves the computational speed.
- To develop a hybrid feature extraction technique (combination of descriptor level features) for extracting features from the histopathology images, where the extracted feature vectors are less exposed to overfitting problem and provides better accuracy for the classification.
- To develop a new ensemble feature selection/ improved optimization algorithm to diminish the dimensions of the extracted feature vectors for better classification.
- To develop an effective unsupervised classification technique to classify the sub-types of breast cancer for early diagnosis.
- To develop an effective deep learning based classification technique in order to deal with data imbalance and uncertainty estimation problem.

CONCLUSION

This survey paper addressed several machine learning techniques, deep learning techniques, feature extraction techniques, feature selection techniques and classifiers applied in histopathology based breast cancer classification. Additionally, advantages and limitations of existing techniques are investigated on the topic “breast cancer classification”. By using validation measures, several concerns pertaining to sensitivity, accuracy, specificity are also addressed in this article. Though, this survey is helpful for the researchers to understand denoising methods, segmentation methods, feature extraction techniques, feature selection techniques and classifiers used in histopathological breast cancer detection to achieve better performance in segmentation and classification phases.

REFERENCES

E. Deniz, A. Sengür, Z. Kadiroglu, Y. Guo, V. Bajaj, and Ü. Budak, “Transfer learning based histopathologic image classification for breast cancer detection”, Health information science and systems, vol.6, no.1, pp.18, 2018.

H. Yang, J.Y. Kim, H. Kim, and S.P. Adhikari, “Guided soft attention network for classification of breast cancer histopathology images”, IEEE transactions on medical imaging, vol.39, no.5, pp.1306-1315, 2019.

J. Xie, R. Liu, J. Luttrell IV, and C. Zhang, “Deep learning based analysis of histopathological images of breast cancer”, Frontiers in genetics, vol.10, pp.80, 2019.

J. Xu, L. Xiang, Q. Liu, H. Gilmore, J. Wu, J. Tang, and A. Madabhushi, “Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images”, IEEE transactions on medical imaging, vol.35, no.1, pp.119-130, 2015.

M. Saha, C. Chakraborty, and D. Racoceanu, “Efficient deep learning model for mitosis detection using breast histopathology images”, Computerized Medical Imaging and Graphics, vol.64, pp.29-40, 2018.

M. Togaçar, K.B. Özkurt, B. Ergen, and Z. Cömert, “BreastNet: A novel convolutional neural network model through histopathological images for the diagnosis of breast cancer”, Physica A: Statistical Mechanics and its Applications, vol.545, pp.123592, 2020.

P. Wang, J. Wang, Y. Li, P. Li, L. Li, and M. Jiang, “Automatic classification of breast cancer histopathological images based on deep feature fusion and enhanced routing”, Biomedical Signal Processing and Control, vol.65, pp.102341, 2021.

P.J. Sudharshan, C. Petitjean, F. Spanhol, L.E. Oliveira, L. Heutte, and P. Honeine, “Multiple instance learning for histopathological breast cancer image classification”, Expert Systems with Applications, vol.117, pp.103-111, 2019.

Prabu, S., Balamurugan, V. and Vengatesan, K., 2019. Design of cognitive image filters for suppression of noise level in medical images. Measurement, 141, pp.296-301.

R. Man, P. Yang, and B. Xu, “Classification of Breast Cancer Histopathological Images Using Discriminative Patches Screened by Generative Adversarial Networks”, IEEE Access, vol.8, pp.155362-155377, 2020.

R. Yan, F. Ren, Z. Wang, L. Wang, T. Zhang, Y. Liu, X. Rao, C. Zheng, and F. Zhang, “Breast cancer histopathological image classification using a hybrid deep neural network”, Methods, vol.173, pp.52-60, 2020.

S. Boumaraf, X. Liu, Z. Zheng, X. Ma, and C. Ferkous, “A new transfer learning based approach to magnification dependent and independent classification of breast cancer in histopathological images”, Biomedical Signal Processing and Control, vol.63, pp.102192, 2021.

T. Wan, J. Cao, J. Chen, and Z. Qin, “Automated grading of breast cancer histopathology using cascaded ensemble with combination of multi-level image features”, Neurocomputing, vol.229, pp.34-44, 2017.

Ü. Budak, Z. Cömert, Z.N. Rashid, A. Sengür, and M. Çıbuk, “Computer-aided diagnosis system combining FCN and Bi-LSTM model for efficient breast cancer detection from histopathological images”, Applied Soft Computing, vol.85, pp.105765, 2019.

Y. Feng, L. Zhang, and J. Mo, “Deep manifold preserving autoencoder for classifying breast cancer histopathological images”, IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2018.

Y. Feng, L. Zhang, and Z. Yi, “Breast cancer cell nuclei classification in histopathology images using deep neural networks”, International journal of computer assisted radiology and surgery, vol.13, no.2, pp.179-191, 2018.