Breast Cancer Detection from Histopathology Images: A Review

Shehras V U [1], Sindhu R [2]

[1]PG Scholar, ECE Department, NSS College of Engineering, Palakkad, Kerala, India
[2] Professor, ECE Department, NSS College of Engineering, Palakkad, Kerala, India

---***---

Abstract - *Breast cancer is the most prevalent form of cancer among women. Early detection of breast cancer will increase survival rate. Automatic image analysis methods are necessary to decrease the workload among pathologists and to improve the quality of interpretation. Nuclei detection is the first stage of identifying the cells. Division of single cell and the spreading of cancer from one part to others in the human body can also detected from histopathological images. This paper provides a review on breast cancer detection from histopathology images.*

Key Words: **Histopathology images, Cell nuclei, Metastasis, Mitosis, Convolutional Neural Network (CNN)**

1. INTRODUCTION

Cancer is a cluster of diseases involving abnormal cell growth with the potential to invade or spread to other parts of the body known as malignant tumor. A benign tumor is a tumor that does not spread around the body. Breast cancer is a cancer that forms within the cells of the breasts. Breast cancer can occur in both men and women, but it is more commonly occur in women. Breast cancer is the most common type of cancer diagnosed in women after skin cancer.

Almost all cancers can spread. The original tumor is called the primary tumor while the dispersed tumors are called metastatic tumors. Mitosis is a method where single cell divides in to two daughter cells. During mitosis, a cancerous cell makes an exact copy of it and splits into two new cells which are also cancerous. Metastasis is caused by the spread of cancer to other locations in the body. Most cancer deaths are due to metastasis.

Early detection plays a key role in cancer detection and can improve long-term survival rates. Medical imaging is a very important technique for early cancer detection and diagnosis. Manual interpretation of enormous number of medical images can be tedious and time consuming and easily causes human bias and mistakes. Therefore, Computer Assisted diagnosis (CAD) systems were introduced to assist doctors in interpreting medical images to improve their efficiency.

A biopsy is the physical examination under which a piece of sample tissue is taken out for microscopic examination. The sample is then referred to the laboratory where pathologist examines and analyses tissues under the microscope. This microscopic examination and study of biological cells and tissues are known as histopathology. Histopathology is a Greek word in which Histo means "tissues", Patho means "disease" and logo refers to "study". Therefore, histopathology means the study of tissues for the identification of diseases.

A combination of Hematoxylin and Eosin are the most commonly used stains in histopathology. Hematoxylin gets bound to Deoxyribo Nucleic Acid (DNA) and it dyes purple or blue color to the nuclei. Eosin gets bound to proteins, so it dyes pink color to other structures. These different stains will help to identify the nuclei of the cells.
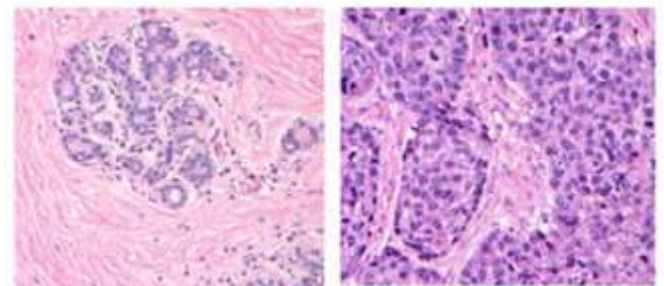

Fig-1: Hematoxylin and Eosin Images of breast cancer (a): Benign (b): Malignant

Fig-1 shows the stained histopathology images to identify benign and malignant tumor. A benign growth does not usually threaten life but it interferes with vital structures, tissues or organs. Benign growths are generally consisting of masses of cells that closely resemble the normal cells composing the tissue in which they are found. A malignant growth is consisting of cells of a typical structure and function when compared to the healthy cells surrounding them. Malignant tumor act maliciously, is capable of invading other tissues and, if untreated, usually results in death.

2. HISTOPATHOLOGICAL IMAGE ANALYSIS

There are several breast cancer detection techniques from histopathological images based on convolutional neural networks. Fast and accurate techniques where introduced to improve the quality of interpretation in the field of automatic histopathological image analysis. Automatic detection of nuclei is the most important in the identification of cells. Division of single nuclei into two new nuclei is termed as mitosis. The division of single cell in to two new daughter cells causes metastasis. The next section deals with the different breast cancer techniques.

2.1 Breast Cancer Detection Techniques

M. M. Dundar *et.al* [1] developed a prototype system for classifying breast microscopic tissues to distinguish between Usual Ductal Hyperplasia (UDH) and actionable subtypes Atypical Ductal Hyperplasia (ADH), and Ductal Carcinoma In Situ (DCIS). The introduced system automatically evaluates digitized slides of tissues for certain cytological criteria and classifies the tissues based on the features derived. The developed system has great potential in improving diagnostic accuracy and reproducibility.

H. Su *et.al* [2] proposed a method to apply a fast scanning deep convolutional network (fCNN) to pixel wise region segmentation. The aim of fCNN was to removes the redundant computations in the original CNN without affecting its performance. This method is robust, efficient and scalable based on fast scanning deep neural network.

A method based on the extraction of image patches for training CNN and from the combination of these patches final classification was done by F.A. Spanhol *et.al* [3]. This method aims to allow using high resolution images from BreaKHis as input to existing CNN in order to avoid adaptations of the model that led to a more complex and computationally costly architecture. When compared to previously reported results obtained by other machine learning models trained with handcrafted textural descriptors, a better performed CNN is introduced.

T. Wan *et.al* [4] presented a novel image analysis based method for automatically identifying different breast cancer grades such as low, intermediate, and high in digitized histopathology. To segment cell nuclei in histopathology images, an improved hybrid active contour model based segmentation method was introduced. This method provides an accuracy of 0.92 for low Vs high, 0.77 for low Vs intermediate, 0.76 for intermediate Vs high.

In 2018, a method for automatic classification of breast cancer using neural network were introduced by S. Kaymak *et.al* [5]. The proposed methodology represents the images using discrete Haar Wavelets which provides better image content representation and then in putting them in to neural networks. Classification of the images was achieved using Back Propagation Neural Network (BPNN) and Radial Basis Neural Network (RBFN).

B. Gecer *et.al* [6] proposed a system that classifies whole slide images of breast biopsies in to five diagnostic categories. A saliency detector by using a pipeline of four sequential fully convolutional networks was used for multi scale processing. For diagnosis, patch based multi class convolutional network was used. Finally, a fusion of saliency detector and the fully convolutionalized classifier network for pixel wise labelling of the whole slide.

In 2019, C. Kaushal *et.al* [7] introduced a Computer Assisted Diagnosis (CAD) system to diagnose and detect breast cancer from histopathology images which will helps pathologists to detect the cancer earlier. CAD system will help to reduce the time taken to diagnose the cancer as well as computational complexity. Since it requires large amount of data for training, data augmentation with the support of various deep learning techniques had led to provide reliable and accurate results.

R. Yan *et.al* [8] proposed a new hybrid model combining both convolutional and recurrent deep neural networks for breast cancer image classification which integrates the advantages of the both convolutional and recurrent deep neural networks. In this model, both the short and long term spatial correlations were preserved between the patches. This new hybrid model provides an average accuracy of 91.3%.

When compared to manual detection system, Computer Aided diagnosis system will provide better results. In deep learning, this is generally done by using convolutional neural network for extracting features and these features are classified using fully connected network. S. Dabeer *et.al* [9] trained a convolutional neural network and provided a prediction accuracy of up to 99.86%.

X. Li *et.al* [10] studied a practical, generalizable, and self-interpretable solution to pathology image based cancer diagnosis in 2019. The proposed method learns discriminative patterns in weak-supervised manner from Whole slide imaging (WSI) labels and explains its diagnosis results via inferring locations of abnormalities in a histopathology image.

Nuclei detection is the important stage of identifying the cells from histopathological images. Accurate detection and segmentation of nuclei can be done using convolutional networks and auto encoder from high resolution images of breast cancer.

2.2 Nuclei Detection

Y.Al-Kofahi *et.al* [11] presented a robust and accurate method for the segmentation of nuclei. The image foreground is automatically extracted using a graph-cuts-based binarization. A novel method combining Multiscale Laplacian of Gaussian filtering constrained by distance map based adaptive scale selection were used to detect nuclear seed points. An initial segmentation which is refined using a second graph cuts based algorithm incorporating the method of alpha expansions were performed using these seed points.

P.Wang *et.al* [12] proposed an automatic quantitative image analysis technique. To enhance the image quality, top-bottom Hat transform were applied in nuclei segmentation. To obtain the region of interest, Wavelet decomposition and Multi scale region grown (WDMR) were used. To split overlapped cells, a Double Strategy Splitting Model (DSS)

containing adaptive mathematical morphology and Curvature Scale Space (CSS) corner detection method were applied for better accuracy and robustness.

J. Xu *et.al* [13] in 2015 introduced a Stacked Sparse Auto Encoder (SSAE) based algorithm which is an instance of deep learning strategy for efficient nuclei detection on high resolution histopathological images of breast cancer. The SSAE learns high level features from pixel intensities in order to identify distinguishing features of cell nuclei.

For the efficient and accurate detection of cell nuclei, H. Xu *et.al* [14] presented an automatic technique based on generalized Laplacian of Gaussian (gLoG) filter in 2016. They proposed computationally efficient nuclei seed detection algorithm based on directionals gLoG kernels and mean-shift clustering to remove false seeds in the image background.

In 2017, L. Hou *et.al* [15] proposed a sparse Convolutional Auto Encoder (CAE) that uses visual characteristics of nuclei for simultaneous unsupervised nucleus detection and feature extraction in histopathology tissue images. CAE detects and encodes nuclei in image patches into sparse feature maps that encode both the location and appearance of nuclei.

In 2018, X. Li *et.al* [16] proposed a staged identification framework for cell nuclei in colon cancer histopathology images. A cascade residual fusion block was presented to enhance the detection performance during the decoding process. A multicropping module was designed for effectively capturing contextual feature contents around the centre of a nucleus for reducing the impact of uncertainty.

P. Naylor *et.al* [17] described a new method to automatically segment nuclei from Haematoxylin and Eosin (H&E) stained histopathology images with fully convolutional networks. They addressed the problem of segmenting touching nuclei by formulating the segmentation problem as a regression task of the distance map. They showed that the fully convolutional networks are well suited for the task of nuclei segmentation.

The spreading of cancer cells is mainly due to the division of single cell to two daughter cells. The cancer causing cells will replace with the normal cells and led to the growth of abnormal cells out of control.

2.3 Mitosis Detection

In 2016, A. Albayrak *et.al* [18] proposed a deep learning based feature extraction method by Convolutional Neural Network (CNN) for automated mitosis detection for cancer diagnosis and grading by histopathological images. After Pre processing, cell structures are found by combined clustering based segmentation and blob analysis. CNN is a prominent deep learning method on image processing tasks which were used for extracting discriminative features. They used a robust kernel based classifier, support vector

machine (SVM) for final classification of mitotic and non-mitotic cells.

H. Chen *et.al* [19] introduced a fast and accurate method to detect mitosis by designing a novel deep cascaded convolutional neural network composed of two components. They proposed a coarse retrieval model to identify and locate the candidates of mitosis while preserving a high sensitivity. A fine discrimination model utilizing knowledge transferred from cross-domain is developed to further single out mitoses from hard mimics.

In S. Albarqouni *et.al* [20] proposed a new concept for learning from crowds that handle data aggregation directly as part of the learning process of CNN via additional crowd sourcing layer AggNet. An aggregation layer is introduced to aggregate the prediction results with annotation results from multiple participations.

In 2017, M. Saha *et.al* [21] presented a supervised model to detect mitosis signature for breast histopathology WSI (Whole Slide Images) images. This model was designed using deep learning architecture with handcrafted features. The proposed architecture had an improved 92% precision, 88% recall and 90% F-score.

B.K. Sabeena *et.al* [22] transformed a pre-trained Convolutional Neural Network by coupling random forest classifier with the initial fully connected layers to extract discriminant features from nuclei patches and to precisely prognosticate the class label of cell nuclei. The designed framework gives higher classification accuracy by carefully fine tune the pre trained model and pre processing the extraction features.

For weakly supervised breast cancer diagnosis, C. Li *et.al* [23] proposed a deep learning scheme with novel loss function in 2019. This method utilizes a deep segmentation network to produce segmentation map. Then a filtering operation was applied to produce the detection results.

The spread of cancer cells from the place where they first formed to another part of the body after the division of nuclei. In metastasis, cancer cells break away from the original (primary) tumor, travel through the blood or lymph system, and form a new tumor in other organs or tissues of the body.

2.4 Metastasis Detection

M. Valkonen *et.al* [24] described a machine learning approach for detection of cancerous tissue from scanned whole slide images. This method was based on feature engineering and supervisor leaning with random forest model. Several local descriptors related to image texture, spatial structure, and distribution of nuclei was the features extracted from the whole slide images.

In 2018, P. Grover *et.al* [25] proposed an algorithm for automated detection of breast cancer metastasis in whole

slide images. To improve the detection accuracy and overall time to localize tumorous regions, proposed algorithm leverages the capability of advanced image processing and machine learning.

K. Fukuta *et.al* [26] developed a system for the automatic detection of breast cancer metastasis in whole slide images of histopathological lymph node sections. The system focuses on extracting feature and spatial information. They created tumor probability heat maps and perform post-processing to make patient level predictions.

In 2018, H. Lin *et.al* [27] presented a framework by leveraging fully convolutional networks that overcomes the major speed bottleneck in whole-slide image analysis for efficient inference. For ensuring accurate detection on both micro- and macro-metastases, framework reconstructs dense heat maps. This method achieved superior performance compared to other method on Camelyon 2016 Grand Challenge dataset.

In order to increase the survival of patients by early diagnosis of breast cancer, L. Tapak *et.al* [28] compared the performance of six machine learning techniques, two traditional methods for the prediction of breast cancer and metastasis. Results shows that the average specificity of all techniques was greater than 94% and when compared to other techniques, Support Vector Machine (SVM) and Linear Discriminant Analysis (LDA) have greater sensitivity of 73%.

3. CONCLUSIONS

The analysis of histopathological images remains the most widely used method for breast cancer diagnosis. The use of computer-assisted analysis of histopathological images is a promising way to improve the analytical and predictive capabilities. The computer aided diagnosis of breast cancer from Histopathology will reduce the complexity and increase the accuracy.

To classify breast cancer efficiently, different neural networks are used. Auto encoder based algorithms are proposed to identify the distinguishing features of cell nuclei. Deep learning based feature extractions for mitosis detection are performed using Convolutional Neural Networks. There are various machine learning approach for the detection of cancerous tissue from scanned Whole Slide Images.

REFERENCES

[1] M. M. Dundar et al., "Computerized classification of intraductal breast lesions using histopathological images," IEEE Trans. Biomed. Eng., vol. 58, no. 7, pp. 1977–1984, 2011.

[2] H. Su, F. Liu, Y. Xie, F. Xing, S. Meyyappan, and L. Yang, "Region segmentation in histopathological breast cancer images using deep convolutional neural network," Proc. - Int. Symp. Biomed. Imaging, vol. 2015-July, pp. 55–58, 2015.

[3] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast Cancer Histopathological Image Classification using Convolutional Neural Networks," Int. Jt. Conf. Neural Networks (IJCNN 2016), 2016.

[4] 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.

[5] S. Kaymak, A. Helwan, and D. Uzun, "Breast cancer image classification using artificial neural networks," Procedia Comput. Sci., vol. 120, pp. 126–131, 2018.

[6] B. Gecer, S. Aksoy, E. Mercan, L. G. Shapiro, D. L. Weaver, and J. G. Elmore, "Detection and Classification of Cancer in Whole Slide Breast Histopathology Images Using Deep Convolutional Networks," Pattern Recognit., 2018.

[7] C. Kaushal, S. Bhat, D. Koundal, and A. Singla, "Recent Trends in Computer Assisted Diagnosis (CAD) System for Breast Cancer Diagnosis Using Histopathological Images," IRBM, vol. 1, pp. 1–17, 2019.

[8] R. Yan et al., "Breast cancer histopathological image classification using a hybrid deep neural network," Methods, no. February, pp. 1–9, 2019.

[9] S. Dabeer, M. M. Khan, and S. Islam, "Cancer diagnosis in histopathological image: CNN based approach," Informatics Med. Unlocked, p. 100231, 2019.

[10] X. Li, M. Radulovic, K. Kanjer, and K. N. Plataniotis, "Discriminative Pattern Mining for Breast Cancer Histopathology Image Classification via Fully Convolutional Autoencoder," IEEE Access, vol. 7, no. c, pp. 36433–36445, 2019.

[11] Y. Al-Kofahi, W. Lassoued, W. Lee, and B. Roysam, "Improved automatic detection and segmentation of cell nuclei in histopathology images," IEEE Trans. Biomed. Eng., vol. 57, no. 4, pp. 841–852, 2010.

[12] P. Wang, X. Hu, Y. Li, Q. Liu, and X. Zhu, "Automatic cell nuclei segmentation and classification of breast cancer histopathology images," Signal Processing, vol. 122, pp. 1–13, 2016.

[13] J. Xu et al., "Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images," IEEE Trans. Med. Imaging, vol. 35, no. 1, pp. 119–130, 2016.

[14] H. Xu, C. Lu, R. Berendt, N. Jha, and M. Mandal, "Automatic Nuclei Detection Based on Generalized Laplacian of Gaussian Filters," IEEE J. Biomed. Heal. Informatics, vol. 21, no. 3, pp. 826–837, 2017.

[15] L. Hou et al., "Sparse autoencoder for unsupervised nucleus detection and representation in histopathology images," Pattern Recognit., vol. 86, pp. 188–200, 2019.

[16] X. Li, W. Li, S. Member, R. Tao, and S. Member, "Staged Detection – Identification Framework for Cell Nuclei in Histopathology Images," IEEE Trans. Instrum. Meas., vol. PP, pp. 1–11, 2019.

[17] P. Naylor, M. Laé F. Reyal, and T. Walter, "Segmentation of Nuclei in Histopathology Images by Deep Regression of the Distance Map," IEEE Trans. Med. Imaging, vol. 38, no. 2, pp. 448–459, 2019.

[18] A. Albayrak and G. Bilgin, "Mitosis detection using convolutional neural network based features," CINTI 2016 - 17th IEEE Int. Symp. Comput. Intell. Informatics Proc., pp. 335–340, 2017.

[19] H. Chen, Q. Dou, X. Wang, J. Qin, and P. A. Heng, "Mitosis detection in breast cancer histology images via deep cascaded networks," 30th AAAI Conf. Artif. Intell. AAAI 2016, pp. 1160–1166, 2016.

[20] S. Albarqouni, C. Baur, F. Achilles, V. Belagiannis, S. Demirci, and N. Navab, "AggNet: Deep Learning From Crowds for Mitosis Detection in Breast Cancer Histology Images," IEEE Trans. Med. Imaging, vol. 35, no. 5, pp. 1313–1321, 2016.

[21] M. Saha, C. Chakraborty, and D. Racoceanu, "Efficient deep learning model for mitosis detection using breast histopathology images," Comput. Med. Imaging Graph., vol. 64, pp. 29–40, 2018.

[22] K. Sabeena Beevi, M. S. Nair, and G. R. Bindu, "Automatic mitosis detection in breast histopathology images using Convolutional Neural Network based deep transfer learning," Biocybern. Biomed. Eng., vol. 39, no. 1, pp. 214–223, 2019.

[23] C. Li, X. Wang, W. Liu, L. J. Latecki, B. Wang, and J. Huang, "Weakly supervised mitosis detection in breast histopathology images using concentric loss," Med. Image Anal., vol. 53, pp. 165–178, 2019.

[24] M. Valkonen, K. Kartasalo, K. Liimatainen, M. Nykter, L. Latonen, and P. Ruusuvuori, "Metastasis detection from whole slide images using local features and random forests," Cytom. Part A, vol. 91, no. 6, pp. 555–565, 2017.

[25] P. Grover, "Automated Detection of Breast Cancer Metastases in Whole Slide Images," 2018 First Int. Conf. Secur. Cyber Comput. Commun., pp. 1–6, 2018.

[26] K. Fukuta, "Identifying metastatic breast cancer using deep texture representation," The Univercity of Tokyo , Tokyo Medical and Dental University Camelyon17.Grand-Challenge.Org.

[27] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P. A. Heng, "Fast ScanNet: Fast and Dense Analysis of Multi-Gigapixel Whole-Slide Images for Cancer Metastasis Detection," IEEE Trans. Med. Imaging, vol. 38, no. 8, pp. 1948–1958, 2019.

[28] L. Tapak, N. Shirmohammadi-Khorram, P. Amini, B. Alafchi, O. Hamidi, and J. Poorolajal, "Prediction of survival and metastasis in breast cancer patients using machine learning classifiers," Clin. Epidemiol. Glob. Heal., vol. 7, no. 3, pp. 293–299, 2019.

BIOGRAPHIES

Shehras V U, Currently pursuing M.Tech in Communication Engineering from NSS College of Engineering, Palakkad, Kerala, India

Dr. Sindhu R, Professor, ECE Dept, NSS College of Engineering, Palakkad, Kerala, India