A Review on Machine Diagnosis Methods for Breast Cancer

Guoxi Liang

Department of information & technology, Wenzhou Polytechnic, Wenzhou 325000, China.

gonseey07@qq.com

Abstract

Breast cancer is one of the common female malignant tumors. According to statistics, there are about 1.2 million new women with breast cancer each year worldwide, and about 500,000 women die of breast cancer each year. In the developed areas along the eastern coast of China, the incidence of breast cancer is relatively high. According to statistics from a hospital in Wenzhou, the average annual breast cancer surgery in the hospital is more than 1,000. Breast cancer has become one of the common cancers that trouble women. In clinical practice, some doctors have proposed that if they can integrate the patient's existing report data and disease characteristics, and realize the diagnosis and prognosis of breast cancer through AI technology, they can provide doctors with objective auxiliary diagnosis reports. The breast cancer machine diagnosis method is the key technology to realize a clinical diagnosis. This paper reviews the breast cancer machine diagnosis methods, compares the advantages and disadvantages of these methods, and finally gives a summary and prospects.

Keywords

Breast cancer, Machine diagnosis, Deep learning.

1. Introduction

Breast cancer is a common malignant tumor originating from ducts or lobular cells and is currently one of the most common female malignant tumors. According to statistics, there are about 1.2 million new women with breast cancer each year worldwide, and about 500,000 women die of breast cancer each year. In developed areas along the eastern coast of China, the incidence of breast cancer is relatively high. According to statistics from a hospital in Wenzhou, there were more than a thousand cases of breast cancer surgery last year. Breast cancer has become one of the common cancers that trouble women. How to accurately diagnose breast cancer, evaluate its prognosis and treatment responsiveness, and provide the best-individualized treatment plan has always been a clinically urgent problem to be solved. At present, in the process of diagnosis and prognosis of breast cancer, a common method is for doctors to analyze histopathological images (mainly hematoxylin-eosin, H & E and immunohistochemistry, IHC), combined with other medical data to make the final diagnosis and prognosis Evaluation. Through immunohistochemistry, the treatment of breast cancer is based on its different molecular types (4 molecular types: lumen A, lumen B, HER-2 overexpression, and basal-like) Determine the corresponding treatment strategy. Different classifications have a great influence on the surgical plan, postoperative treatment, and survival prognosis. In the classification diagnosis of breast cancer, doctors mainly make a classification based on data such as pathological reports and comprehensive diagnosis experience. For example, the doctor analyzes the histopathological image and then gives an assessment conclusion, but the morphology of the cell nucleus in the pathological image is variable and difficult to distinguish; immunohistochemical indicators affect each other and are very complex, the analysis results are time-consuming and will depend on the doctor 's experience Different from the level of knowledge, it has a strong subjectivity of doctors.

In clinical practice, some doctors have proposed that if they can integrate the existing report data and disease characteristics of patients and realize rapid analysis and prediction through AI technology, they can provide doctors with timely and effective diagnosis and prognosis during and after surgery, and even before surgery. For reference, it will make up for the lack of comprehensive objective analysis of breast cancer to a certain extent, and provide patients with a more scientific and comprehensive treatment plan to achieve better treatment results. In view of the current development of AI technology, the use of AI technology to diagnose breast cancer typing, assess its prognosis and treatment responsiveness has become a viable attempt. After the introduction of computer-aided medicine, especially the "AI + medical" method based on deep learning, it has brought new progress to the diagnosis and prognosis of breast cancer. Domestic Xu Jun [1] et al. Took breast cancer H & E pathological images as the research object, and did the focus detection, cutting, mitosis recognition and other work through deep learning methods; International Coates [2] and other scholars also studied H & E pathology based on deep learning A lot of work such as grading of nuclear morphology of images has accumulated experience for the diagnosis of breast cancer. The following is a review of breast cancer machine diagnosis methods from three perspectives.

2. Review of research Methods

Back in 2013, MIT Technology Review magazine ranked deep learning as the top ten breakthrough technology of the year. As the most promising deep learning method in the field of machine learning, deep learning has been widely used in the fields of image, speech, and text recognition. In the field of biomedicine, deep learning technology has also been widely used, such as bioinformatics, medical image segmentation, lesion detection and classification, tumor diagnosis, etc. Smart medicine based on deep learning has brought new changes to the entire medical industry, as well as new advances in the diagnosis and prognosis of breast cancer. Xu Jun [1] et al. Took breast cancer H & E pathological images as the research object, and conducted deep detection methods for lesion detection, cutting, and mitosis recognition; scholars such as Coates [2] also studied the cell nucleus of H & E pathological images based on deep learning A lot of work such as morphological classification has accumulated experience for the diagnosis of breast cancer. There are many machine diagnosis methods for breast cancer, and the main types are summarized below.

2.1 Convolutional Neural Networks Based Methods

Convolutional Neural Networks (CNN) is a representative structure of deep learning models and a current research hotspot in deep learning. It was proposed by LeCun et al. [3] in 1989. It is a feed-forward artificial neural network with a multi-layer network structure, usually including an input layer, convolution layer, activation function, pooling layer, and fully connected layer. CNN has powerful feature extraction capabilities, which can extract higher-level features. In 2014, Hinton [4] and Suen [5] and others used CNN technology to achieve good results on many different pattern recognition problems. The research by Hafemann et al. [6] in 2016 showed that on micro and macro pictures, the characterization capabilities of the texture descriptors extracted by CNN surpassed the traditional texture descriptors. The effect is very obvious. Jiao et al. [8] developed a CNN-based CAD (Computer-Aided Diagnosis) system in 2016 to classify breast cancer lumps. They first used CNN to extract high-level and medium-level features, and then combined to train the model, and combined the CNN automatically extracted intensity information with depth features, and achieved good results. In 2017, several algorithms on breast diagnosis were proposed one after another. Al-Masni et al. [9] proposed a CAD system based on regional deep learning technology in 2017, using ROI (Region Of Interest) as the perceptive object, and carried out on the mammography data set with ROI information. Training, and directly optimize the detection performance. This method can also detect multiple targets simultaneously. Therefore, the CAD system they proposed can complete feature extraction and detect and classify breast masses on CNN, which is a fast and accurate target detector. Bayramoglu et al. [10] proposed an independent magnification-based breast cancer pathological image classifier based on deep learning in 2017. The accuracy rate is higher than that of traditional classifiers trained separately for different magnifications. Spaniel et al. [11] et al. Proposed a breast cancer pathological image classification algorithm based on DeCAF features in 2017. This algorithm uses pre-training to initialize model parameters and ultimately obtains a high classification accuracy. Li et al. [12] adopted a full convolution design in 2017 to develop an end-to-end training algorithm for breast cancer diagnosis of complete breast X-ray images. The algorithm completely uses CNN, so that you can input images of any size. It only needs to annotate the lesions in the first stage of training. After the training model recognizes the local patches, the weights of the complete image classification network can be initialized, and then the model can be migrated to a full-image classifier, which can be annotated without ROI. End-to-end training, which greatly reduces the dependence on lesion annotation. Compared with the previous method, this design is simple, and the performance is more superior. Liu Juan et al. [21] used a convolutional neural network model based on AlexNet, assisted transfer learning and data enhancement methods to divide the image into breast ductal carcinoma in situ, breast invasive ductal carcinoma, breast fibroadenoma and breast hyperplasia Four categories, and did a classification study and so on.

2.2 Weakly Supervised Based Methods

In 2016, Quellec et al. [7] proposed a weakly supervised computer-aided detection and diagnosis system for mammography, which can be diagnosed only by using labels at the overall image level. The method adaptively divides the mammary gland into multiple regions, then extracts the detected lesion features from each region and merges them, and then classifies the mammary gland X-ray image as normal or abnormal. Choukroun et al. [13] proposed a computer-aided detection and diagnosis system for weakly supervised learning in 2017, which solved the detection of abnormal results of mammography by a new deep learning framework built on the MIL (Multiple Instance Learning) paradigm And classification issues. The characteristic of this method is that MIL can be used to automatically discover discriminative examples in mammograms. The results of this system are comparable to supervised methods trained on fully annotated data sets. Zhu et al. [14] also proposed an end-to-end trained deep MIL neural network for classifying breast X-ray images without ROI annotation in 2017. The goal is to predict whether the entire breast X-ray image contains malignant masses. The study also used CNN to efficiently obtain the features of all examples. Al-Antari et al. [15] proposed a CAD system for breast cancer diagnosis based on DBN (Deep Belief Network) in 2017. The system includes automatic detection of lumps, ROI extraction, feature extraction, and DBN classification modules, whose goal is to identify Normal, benign, and malignant breast tissue.

2.3 Generative Adversarial Network Based Methods

Generative Adversarial Network (GAN) [16] was proposed by Goodfellow et al. In 2014 and is called by LeCun as an exciting network, which has achieved good results in extracting fitting data distribution. The GAN method separately trains a generator and a discriminator, and the discriminator is used to discriminate whether the data comes from real training data or data generated by the generator. Game training is used to make the generator generate data that matches the distribution of real training data. Based on GAN, Mirza et al. [17] proposed a conditional generation adversarial network in 2014 to generate images of specified tags, provide conditions for expanding the sample size, and increase the sample data to prevent the model from overfitting and improve the recognition accuracy of the model. Radford et al. [18] proposed a deep convolutional generation adversarial network in 2015, in which GAN was used for image feature extraction to achieve semi-supervised image recognition. Denton et al. [19] used the Laplacian pyramid in 2015 to introduce the neural network structure of the rolling machine into the Laplacian pyramid method, using GAN to train and generate images at each layer, and generate high-quality images layer by layer, And finally get high-resolution effective images that can be used to expand the training data. Singh et al. [20] used Conditional Generative Adversarial and Convolutional Networks (cGAN) for the segmentation of breast masses in 2018. Experiments conducted on dozens of malignant tumors extracted from public DDSM datasets and their internal private datasets confirmed that the cGAN structure is very suitable for accurately depicting quality regions, especially when training data is limited, indicating that the generation of confrontation networks For lesion tracking.

3. Conclusion

Despite so much fruitful work, compared with the complexity of breast cancer treatment and prognosis, these research work are still far from clinical application, mainly as follows: ① Most of the existing research is based on a few The database has a limited and single amount of data, and the etiology of different countries and regions often varies greatly. Therefore, it is important to establish a local breast cancer data set to serve this region; Most of the medical data are X-rays and pathological images, and there are few multimodal data. Therefore, research on multimodal data based methods has become one of the problems to be solved; Inferiority comparison is dominant, and there are few clinical trials. We expect to carry out more in-depth research in the following areas to promote the application of breast cancer machine diagnosis.

First, the pathological image is extremely complex, and it is difficult to form a standardized image for downstream research due to the shooting by different instruments. Therefore, the denoising and standardization of pathological images will be an important part of the follow-up research. This work is of great significance for the generalization of auxiliary diagnosis and needs further research.

Second, the average accuracy of the mentioned methods has much room for improvement. It is a work worth looking forward to continue to study and improve the diagnostic methods and further improve the average accuracy of the same method prediction, which needs further in-depth study.

The third is that during the experiment, the researchers found that different data of the same disease type have great differences, which poses great challenges to the application of the system. It is necessary to study a wider range of data and cover various differences in order to make the proposed diagnostic method more convincing when it is applied to clinical practice.

Acknowledgements

This work was supported in part by the Basic Research Project of Wenzhou, China, under Grant G20190020, in part by the Service and Technology Innovation Project of Wenzhou Science and Technology Association under Grant 2019KXCX-RH07, Wenzhou, China, and in part by the Key Scientific Research Project of Wenzhou Polytechnic under Grant WZY2019004, Wenzhou, China.

References

[1] Jun Xu, Lei Xiang, Qingshan Liu. Stacked Sparse Autoencoder (SSAE) for Nuclei Detection on Breast Cancer Histopathology Images. IEEE TRANSACTIONS ON MEDICAL IMAGING (2016), VOL. 35, NO. 1, pp.119-129.

[2] Coates AS. Evolution of the St. Gallen Consensus process for the optimal treatment of women with breast cancer. The Breast 24 Supp 1, PG 0.1. 2015.

[3] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L.D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Computation, vol. 1, no. 4, pp. 541–551, 1989.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems 25, 2012, pp.1097–1105.

[5] X. X. Niu and C. Y. Suen, "A novel hybrid cnn–svm classifier for recognizing handwritten digits," Pattern Recognition, vol. 45, no. 13181325, 2012.

[6] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, "Forest species recognition using deep convolutional neural networks," in International Conference on Pattern Recognition, 2014, pp. 1103–1107.

[7] Quellec G., Lamard M., Cozic M.,et al. Multiple-instance learning for anomaly detection in digital mammography.IEEE Trans Med Imaging, 2016, 35:1604-1614.

[8] Jiao Z., Gao X., Wang Y., et al.A deep feature based framework for breast masses classification. Neurocomputing, 2016,197:221-231.

[9] Al-Masni MA, Al-Antari MA, Park JM,et al. Detection and classifi-cation of the breast abnormalities in digital mammograms via re-gional Convolutional Neural Network. Seogwipo:IEEE,2017.

[10] Bayramoglu N., Kannala J., Heikkilä J. Deep learning for magnification independent breast cancer histopathology image classification International Conference on Pattern Recognition. IEEE, 2017:2440-2445.

[11] Spanhol F. A., Oliveira L. S., Cavalin P. R., et al. Deep features for breast cancer histopathological image classification. IEEE International Conference on Systems, Man and Cybernetics.IEEE, 2017:1868-1873.

[12] Li S. End-to-end training for whole image breast cancer diagnosis using an all convolutional design [J/OL].arXiv,2017：ar Xiv:1711.05775

[13] Choukroun Y., Bakalo R., Benari R.,et al.Mammogram classification and abnormality detection from nonlocal labels using deep multiple instance neural network. Eurographics Workshop Vis Comput Biomed, 2017,17:11-19.

[14] Zhu W., Lou Q., Vang YS, et al. Deep multi-instance networks with sparse label assignment for whole mammogram classification. Cham:Springer International Publishing,2017.

[15] Al-Antari MA, Al-Masni MA, Park SU, et al. An Automatic com-puter-aided diagnosis system for breast cancer in digital mammo-grams via deep belief network. J Med Biol Eng,2018，38:443-456.

[16] I. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, SherjilOzair, Aaron Courville, and YoshuaBengio. Generative adversarial nets. In Advances in Neural Information Processing Systems, pages 2672–2680, 2014.

[17] M. Mirza and S. Osindero. Conditional generative adversarial nets. CoRR, abs/1411.1784, 2014.

[18] Radford A, Metz L, Chintala S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. Computer Science, 2015.

[19] Denton, Emily L, Chintala, Soumith, Szlam, Arthur, and Fergus, Rob. Deep generative image models using a laplacian pyramid of adversarial networks. In Cortes, C., Lawrence, N.D., Lee, D.D., Sugiyama, M., Garnett, R., and Garnett, R. (eds.), Advances in Neural Informa-tion Processing Systems 28, pp. 1486–1494. Curran As-sociates, Inc., 2015.

[20] Singh V.K. et al. (2018) Conditional Generative Adversarial and Convolutional Networks for X-ray Breast Mass Segmentation and Shape Classification. In: Frangi A., Schnabel J., Davatzikos C., Alberola-López C., Fichtinger G. (eds) Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. MICCAI 2018. Lecture Notes in Computer Science, vol 11071. Springer, Cham.

[21] Zheng Qunhua1Duan Huifang,Shen Yao1Liu Juan,Yuan Jingping. BＲEAST CANCEＲ HISTOLOGICAL IMAGE CLASSIFICATION BASED ON CONVOLUTIONAL NEUＲAL NETWOＲK AND TＲANSFEＲ LEAＲNING. Computer Applications and Software. Vol.35, No.7. Jul, 2018.