Application of Convolutional Neural Network in Accurate Breast Cancer Identification

Zhiyang Li*, Haoyu Guo, Yande Xiang, Wentao Kuang, Lu Chen

Tianjin University of Science and Technology, 300222， China;

*E-mail: * a17730209318@163.com*

www.tust.edu.cn

Abstract

As one of the most common cancers in women, breast cancer has the highest incidence in the world. Nearly 600,000 people die from breast cancer each year, and early detection is essential for breast cancer treatment. In recent years, the rapid development of artificial intelligence has provided unprecedented ideas for the precise diagnosis and treatment of breast cancer. In this paper, the practical application of artificial intelligence convolutional neural network in breast cancer recognition is studied, which greatly improves the detection speed and saves a lot of time for doctors to further judge the condition.

Keywords: breast cancer, artificial intelligence, convolutional neural network, recognition

1. Introduction

In 2021, there will be 2.5 million breast cancer patients in China. The overall survival rate of breast cancer in the country in the past five years is less than 50%. The incidence of breast cancer in urban areas has reached 34.3 cases per 100,000 people, twice that of rural areas. Globally, women lose more disability-adjusted life years to breast cancer than any other type of cancer. Women in every country in the world develop breast cancer at any age after puberty, but the incidence is increasing later in life. Invasive ductal carcinoma is one of the most common types of breast cancer, and 80% of breast cancers are invasive ductal carcinoma. Currently, the diagnosis of invasive ductal carcinoma often requires a combination of tests, including a physical examination and some imaging tests. This process is not only time-consuming, but also extremely cumbersome. For some places where medical facilities are backward, the conditions for testing are not even available.

As a new technology developed by today's science and technology, artificial intelligence has made remarkable progress in the application of imaging and pathology[1]. This paper studies the practical application of artificial intelligence convolutional neural network in breast cancer recognition, which can achieve the effect of batch detection. Using AI can improve the efficiency of breast cancer diagnosis, reduce the workload of doctors, and get better feedback in clinical medicine.

The rest of this article is organized as follows. The second part introduces the source of the dataset, and the third part introduces the convolutional neural network and explains the architecture of the convolutional neural network used in this article. The fourth part is the effect of model detection, which shows that the recognition effect of the model is better through the confusion matrix and learning curve. The fifth part is a summary of the full text.

2. BreakHis Dataset

Spanhol et al. published the BreakHis (breast cancer histopathological database) dataset in 2016. The dataset contains 7909 pathological images of breast histopathology from 82 patients. The dataset currently

contains four different types of benign breast tumors histologically: adenopathy (A), fibroadenoma (F), lobular tumors (PT), and tubular adenomas (TA); and four malignancies (breast cancer): carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC) and carcinoma (PC). The sample was derived from a breast tissue biopsy section, stained with hematoxylin and eosin (HE) and labeled by a pathologist.

This allows the BreakHis dataset to not only advance the research of benign and malignant binary classification algorithms, but also promote the research of pathological classification algorithms with more significant clinical significance. The dataset BreaKHis is divided into two broad categories: benign tumors and malignant tumors. Histological benign is a term that refers to lesions that do not meet any criteria for malignancy, e.g., obvious cellular atypia, mitosis, basement membrane disruption, metastasis, etc. As a rule, benign tumors are relatively "innocent", grow slowly and remain localized. Malignancy is synonymous with cancer: lesions can invade and destroy adjacent structures (local infiltrative) and spread to distant places (metastasis) leading to death. The samples present in the current dataset are collected by the SOB method, also known as partial mastectomy or excisional biopsy. Compared to any needle biopsy method, this type of procedure allows the removal of larger tissue samples and general anesthesia in the hospital. This dataset can provide strong support for the smooth progress of the study. A partial picture of the dataset is shown in Fig.1.

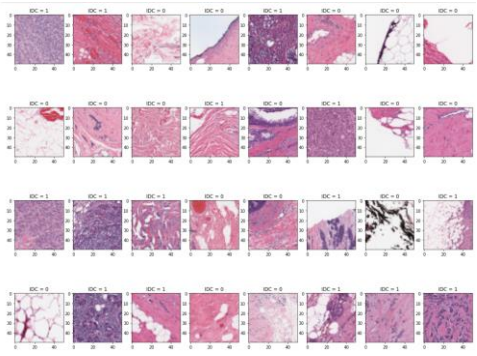

Fig.1 A partial picture of the dataset

3. Convolutional Neural Networks

Convolutional neural networks are variants of multilayer perceptrons (MLPs) developed by biologists Huebor and Wiesel in earlier research on the feline visual cortex, where cells have a complex structure that is very sensitive to subregions of visual input space called receptive fields. Convolutional neural network is a multi-layer supervised learning neural network, and the convolutional layer and pool sampling layer of the hidden layer are the core modules to realize the feature extraction function of convolutional neural network[2].

The network model uses the gradient descent method to minimize the loss function, and adjusts the weight parameters in the network layer by layer, and improves the accuracy of the network through frequent iterative training[3]. Compared with the original neural network, the convolutional network structure can better adapt to the structure of the image, and perform feature extraction and classification at the same time, so that feature extraction can help feature classification. And weight sharing can reduce the training parameters of the network, making the neural network structure simpler and more adaptable[4]. The figure below shows the architecture of a convolutional neural network in breast cancer. The neural network architecture used in the model is shown in Fig.2.

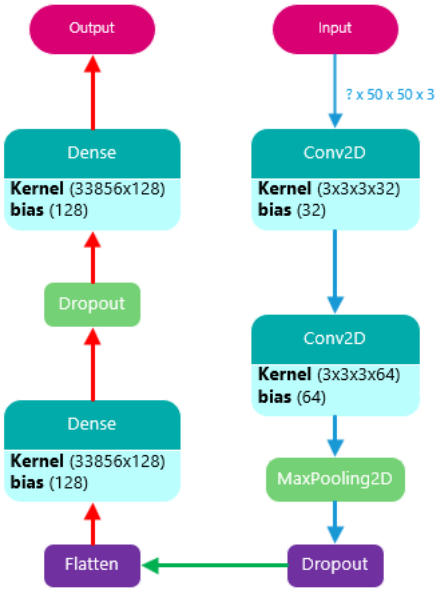

Fig.2 Neural network architecture

4. The Effect of Detection

The confusion matrix, also known as the error matrix, is a standard format for representing accuracy evaluation, represented in the form of a matrix of n rows and n columns. The specific evaluation indicators include overall accuracy, cartographic accuracy, user accuracy, etc., which reflect the accuracy of image classification from different aspects. In artificial intelligence, confusion matrices are visualization tools, especially for supervised learning, and in unsupervised learning, they are generally called matching matrices. The confusion matrix after detection is shown in Fig.3.

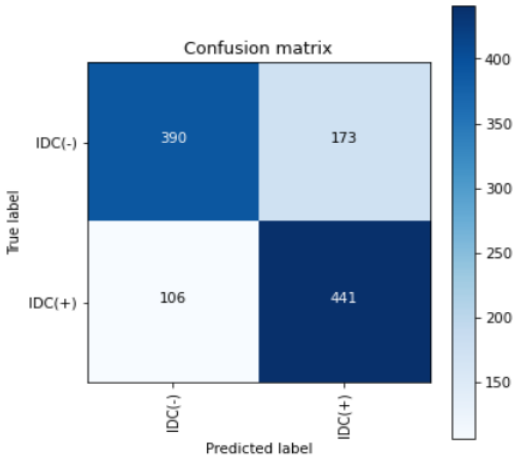

Fig.3 Detect confusion matrix

The confusion matrix shows that the overall recognition accuracy reaches 75%, and the IDC(-) recognition effect can reach 79%. From the graph, we can see that the learning curve shows that the validation set score is always higher than the training set score. The learning curve after the test is shown in Fig.4.

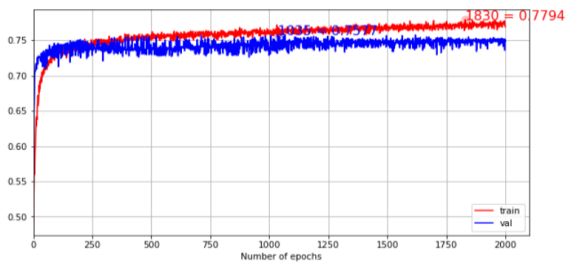

Fig.4 learning curve

The learning curve shows that the training curve and cross-validation curve of different shapes will not have too much overfitting, and both the confusion matrix and the learning curve indicate that the model does not have high bias and has high reference value. In summary, the recognition effect of the model is better.

5. Conclusion

Machine learning technology is an important branch of artificial intelligence, and it is proposed that machine learning-based technology is expected to play a greater role in cancer screening. The times are constantly developing, and human society, which is entering the intelligent era, is facing new challenges and opportunities. This paper uses convolutional neural networks for breast cancer recognition, which has a good recognition effect and can help doctors carry out preliminary screening, thereby greatly saving time.

References

1. Wang J, Liu Q, Xie H, et al. Boosted efficientnet: Detection of lymph node metastases in breast cancer using convolutional neural networks. *Cancers*, 2021, 13(4): 661.
2. Masud M, Eldin Rashed A E, Hossain M S. Convolutional neural network-based models for diagnosis of breast cancer. *Neural Computing and Applications*, 2020: 1-12.
3. Zarbega T S A, Gültepe Y. Semantic segmentation of cell nuclei in breast cancer using convolutional neural network. *International Journal of Computer Applications*, 2020, 176(22): 1-8.
4. Ahmed S R A, UÇAN O N, Duru A D, et al. Breast cancer detection and image evaluation using augmented deep convolutional neural networks. *Aurum journal of engineering systems and architecture*, 2018, 2(2): 121-129.

Authors Introduction

Mr. Zhiyang Li

He is a second-year undergraduate student at Tianjin University of Science and Technology. His research interests are computer vision and machine learning.

Mr. Haoyu Guo

He is currently studying in the School of Biology of Tianjin University of Science and Technology and has achieved certain results in related professional competitions.

Mr. Yande Xiang

He is currently pursuing his undergraduate degree at the School of Electronic Information and Automation, Tianjin University of Science and Technology.

Mr. Wentao Kuang

He is currently pursuing his bachelor's degree at the School of Mechanical Engineering, Tianjin University of Science and Technology.

Ms. Lu Chen

She is currently studying for an undergraduate degree at Tianjin University of Science and Technology.