The Investigation on Breast Cancer Prediction Technologies Based on Machine Learning Algorithms

Qinzheng Luo

Robotics and Intelligent Devices, Fuzhou University, Fuzhou, 350000, China
1809020205@stu.hrbust.edu.cn

Abstract. As one of the most common types of cancer in women in the world and early diagnosis of breast cancer is crucial to improve the survival rate and quality of life of patients worldwide. Therefore, it is of great significance to develop efficient and accurate breast cancer prediction methods for the medical field. This paper aims to explore the breast cancer prediction methods based on machine learning, especially the application of Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) in breast cancer diagnosis. In the research method, this paper introduces two kinds of breast cancer prediction methods based on machine learning. The first method is to construct a multi-layer perceptron model by using Artificial Neural Network (ANN) and optimize the network structure by using Back Propagation algorithm to realize automatic classification of mammography images. The second method uses Convolutional Neural Network (CNN), combined with a weighted Fisher algorithm and deep learning technology, to further improve the accuracy of breast cancer prediction. The research results show that the breast cancer prediction method based on machine learning has reached a high level in accuracy, sensitivity and specificity, and provides a new and effective means for early detection and diagnosis of breast cancer. The breast cancer prediction method based on machine learning proposed in this paper has important application value in the medical field, not only can improve the accuracy and efficiency of breast cancer prediction, but also can provide more timely and effective treatment suggestions for patients.

Keywords: Breast Cancer Prediction, Artificial Neural Network, Convolutional Neural Network.

1 Introduction

Breast cancer represents a significant threat to women's health worldwide. Early diagnosis is of paramount importance in improving the survival rate and reducing the mortality rate of breast cancer patients. However, traditional methods of breast cancer diagnosis often rely on doctors' experience and interpretation of medical images, which may limit the accuracy and efficiency of diagnosis to some extent. The advent of big data and artificial intelligence has led to the emergence of machine learning algorithms as a promising avenue for the construction of breast cancer prediction models [1] due to their demonstrated superior prediction in many fields [2, 3].

Y. Wang (ed.), *Proceedings of the 2024 2nd International Conference on Image, Algorithms and Artificial Intelligence (ICIAAI 2024)*, Advances in Computer Science Research 115,
https://doi.org/10.2991/978-94-6463-540-9_4

The study of breast cancer prediction models is of great significance in improving the accuracy and efficiency of breast cancer diagnosis. Firstly, accurate prediction models can assist doctors in detecting lesions in the early stages of the disease, thus enabling the implementation of timely treatment measures and improving the treatment effect. Secondly, the application of predictive models can reduce doctors' reliance on the interpretation of medical images and improve the objectivity and consistency of diagnosis. Furthermore, the analysis of predictive models enables a deeper comprehension of the pathogenesis of breast cancer, thereby providing novel insights for future treatment and prevention strategies.

Although previous studies have made some progress in the field of breast cancer prediction, there are still some problems and challenges to be overcome. On the one hand, traditional prediction methods such as Single-factor Risk Assessment Model, Logistic Regression and Decision Tree are often based on limited features and simple models, resulting in poor prediction accuracy. On the other hand, with the increasing growth and complexity of medical data, it has become a challenge to process and analyze these data effectively. Furthermore, previous studies frequently utilise a single metric to assess model performance, which is inadequate for fully reflecting the strengths and weaknesses of the model.

Previous studies have accumulated a wealth of experience and knowledge in the field of breast cancer prediction. The researchers improved the accuracy of their predictive models by constantly trying new algorithms and features. However, these studies have some limitations. First, some studies may focus too much on the predictive accuracy of the model and neglect the interpretability and stability of the model. Second, some studies may lack adequate evaluation of the generalization ability of the model, resulting in poor performance of the model in practical applications. In the field of breast cancer prediction, there are many related literatures. In order to gain insight into the core values and potential of these latest technologies, comprehensive and detailed research and analysis is essential. This not only helps us to grasp the latest developments in this field, but also provides solid support for future research directions [4].

The next part of the paper will be divided into three core modules to explore the task of breast cancer image detection in depth. Section 2 will focus on the methodology, reviewing and analyzing in detail how other researchers have used machine learning or deep learning techniques to achieve the goal of breast cancer image detection. In Section 3, the limitations and shortcomings of these methods will be discussed in depth, and the future development direction and potential research areas will be prospected. Finally, the conclusion of this paper is presented, and the future development is also discussed in Section 4.

2 Method

2.1 Introduction of Machine Learning Workflow

Understanding the workflow of machine learning in breast cancer image detection tasks is critical when exploring its application. The machine learning workflow consists of

several key steps that together build a complete framework for data analysis and modeling. First of all, data preparation will be carried out, which involves collecting, cleaning and organizing the relevant breast image dataset. This is followed by data preprocessing, which is designed to improve data quality and may include operations such as image scaling, normalization, noise removal, etc. to ensure consistency and accuracy of model inputs.

After the data preprocessing is completed, the model building stage is entered. At this stage, researchers choose the appropriate machine learning algorithm according to the task requirements, and construct the model framework accordingly. The training process of the model is to optimize the parameters of the model by using the marked data set, so that the model can accurately identify the abnormal regions in the breast image.

After training, the model needs to be tested to evaluate its performance. In the testing phase, independent and untrained data sets are usually used to evaluate the performance of the model on unknown data by calculating the accuracy rate, recall rate, F1 score and other indicators of the model. In addition, the generalization ability of the model can be further verified by cross-validation and other strategies.

The whole machine learning workflow is shown in Figure 1, which clearly shows all links from data preparation to model testing, providing researchers with complete guidance and reference.

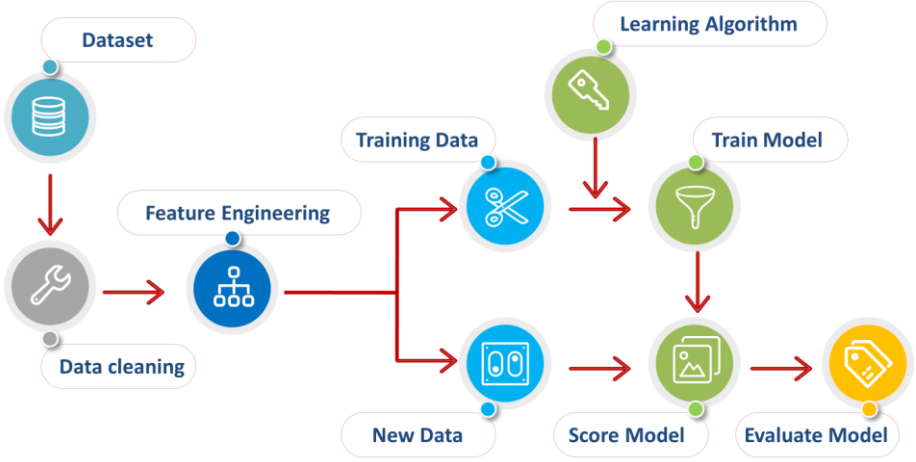

Figure 1. The workflow of machine learning algorithms.

2.2 ANN-based Breast Cancer Detection

Back Propagation Neural Network. Yuan chose to use Back Propagation (BP) neural network to deal with the complexity and nonlinear characteristics of breast cancer pathology data, utilizing its self-learning and adaptive capabilities [6].

This study took the Medical College of Wisconsin breast cancer dataset, which contains multiple associated characteristic variables and diagnoses. Standardization and

normalization pretreatment were performed to ensure data validity. In the neural network design, this study determines the number of layers and neurons according to the characteristics of the dataset, ensuring that the input layer matches the number of feature variables, and the output layer corresponds to the number of classifications. The optimal structure of hidden layer is determined by experiment and empirical formula, and S-type activation function is selected to capture complex relations.

During the model training phase, this study adjusts network weights and biases using a backpropagation algorithm to minimize prediction errors. Set appropriate learning rate and iteration times, and adopt early stop method, regularization and other strategies to avoid overfitting. After training, the model performance was evaluated comprehensively using indicators such as classification accuracy, confusion matrix and Receiver Operating Characteristic (ROC) curve, and adjusts the network structure and hyperparameters to optimize the performance according to the evaluation results.

Finally, this study explained the trained model in depth and applied it to new breast cancer cases to assist doctors in diagnosis and prediction. The application of this model improves the accuracy and efficiency of diagnosis and provides strong support for breast cancer diagnosis and treatment.

Artificial Neural Network. Li et al. used artificial neural network technology to diagnose and predict breast cancer. They began by collecting extensive data on demographics, menstrual fertility history, family and personal disease histories through questionnaires. To ensure the accuracy of the model, the researchers used the regression method of Logistic regression model to filter out the variables with the most predictive value and optimize the input layer [7].

In the model construction, they chose feedforward neural network with L-BFGS algorithm. Feedforward neural network can learn to recognize complex data patterns, while L-BFGS algorithm optimizes model parameters efficiently and improves prediction accuracy. During the study, the data were randomly divided into a training set for model training and a test set to evaluate model performance. Finally, a model with good performance for breast cancer diagnosis and prediction is obtained after several iterations.

In addition, the researchers also carried out variable contribution analysis, revealing the degree of influence of different variables on the prediction results, which not only explained the prediction mechanism of the model, but also provided valuable reference for subsequent research. This research not only demonstrates the potential of artificial neural networks in the field of medical diagnosis, but also provides new methods and ideas for early detection and treatment of breast cancer.

2.3 CNN-based Breast Cancer Detection

Weighted Fisher Convolutional Neural Network. As one of the most common kinds of malignancies in women, breast cancer poses a serious threat to women's health. With the rapid increase of medical image data and the progress of computer image classification technology, medical image classification based on artificial intelligence has

shown great potential in scientific research and clinical practice. Hao et al. proposed a new weighted Fisher convolutional neural network model to improve the accuracy and efficiency of breast cancer detection [8].

The weighted Fisher criterion is applied to LetNet-5 convolutional neural network, and the classification ability of the model is enhanced by optimizing the distance between homogeneous samples and the distance between heterogeneous samples. Experiments show that the model performs well on both public data sets and real hospital data, effectively improves the convergence time and recognition rate of the convolutional neural network, and especially shows better effect compared with the improved AlexNet.

This research not only provides a new idea for the auxiliary diagnosis of breast cancer, but also provides a valuable reference for the development of medical image classification. The application of weighted Fisher criterion in breast cancer detection indicates that artificial intelligence has broad application prospect and clinical value in medical field.

Deep Learning Convolutional Neural Network. Li has conducted in-depth research on the diagnosis and prediction of breast cancer using deep learning convolutional neural network models. First of all, she obtained rich case information and pathological characteristics through statistical analysis of breast cancer data sets published in real clinical practice, which provided a solid foundation for subsequent model construction [9].

In terms of method, the author chooses convolutional neural network, which is an excellent model in image recognition. Based on the characteristics of breast cancer data, she designed and trained a convolutional neural network model that automatically extracted key features from image data and based on them diagnosed and predicted breast cancer.

In the experiment, the author selects the current epidemic breast cancer data set, and uses the trained convolutional neural network model to model and analyze. The experimental results show that the model can accurately and effectively predict malignant breast cancer patients. Compared with the traditional support vector machine model, the accuracy, specificity, sensitivity and area under the curve of the proposed model are significantly improved.

This result not only demonstrates the effectiveness and potential of deep learning convolutional neural networks in breast cancer diagnosis, but also provides medical personnel with a new, more intelligent and efficient diagnostic tool. By reducing the diagnostic burden on healthcare professionals, the model is expected to improve early detection and treatment efficiency of breast cancer, resulting in better treatment outcomes and quality of life for breast cancer patients.

Convolutional Neural Network. Zhao Qian et al. used the convolutional neural network as the main tool to diagnose and predict breast cancer. Because of its powerful feature extraction and pattern recognition capabilities, CNN has shown great potential in the field of medical image processing, especially in mammographic image analysis [10].

First, the researchers collected a large number of mammographic images to construct a dataset that included images of normal, benign, and malignant breast disease. This is the basis for CNN to train and learn.

Then, the researchers designed a multi-layer CNN model. The model typically includes an input layer, a combination of a plurality of convolution layer and a pooled layer, a fully connected layer and an output layer. The input layer receives the mammogram image directly, while the convolution layer is responsible for extracting local features in the image, such as edges, textures, etc. The pooling layer down-samples the feature map output by the convolution layer to reduce the amount of computation and enhance the robustness of the model. The fully connected layer integrates the extracted features and maps them to the final classification result.

In the training process, the researchers used the back-propagation algorithm and gradient descent method to optimize the parameters of the model, so that the model can more accurately identify the disease characteristics in the mammogram image. At the same time, in order to prevent overfitting, the researchers also adopted data enhancement, Dropout and other techniques.

The trained CNN model can automatically analyze and diagnose the new mammogram, and predict whether it is normal, benign or malignant breast disease. Compared with other traditional algorithms, the researchers found that CNN has higher accuracy and efficiency in breast cancer diagnosis and prediction.

3 Discussion

Although neural networks have made remarkable achievements in medical image processing and disease prediction due to their powerful learning and analytical capabilities, this field still faces many limitations and challenges, that need to be solved further.

The first is the data constraints. In breast cancer diagnosis and prediction, high quality labeled data is the basis of training neural network model. However, the acquisition and annotation of medical data often face many difficulties, such as the complexity of data collection, the professionalism of annotation and the size of data sets. These factors limit the training effect and generalization ability of neural network model, and make the model have some limitations in practical application.

Then, the "black box" nature of neural networks is also a concern. When neural network models make diagnosis and prediction, their internal decision-making process is often difficult to explain, which makes it challenging for doctors to fully trust the judgment results of the model. In the medical world, accuracy is important, but interpretability is also critical. Physicians need to understand why the model makes such judg-

ments in order to better understand the condition and develop treatment options. Therefore, improving the interpretability of neural network model is one of the important directions of future research.

Privacy protection is another issue that cannot be ignored. Medical data relates to the patient's personal privacy and life and health, and its security is very important. When using neural networks for breast cancer diagnosis and prediction, researchers must ensure that patient privacy is not compromised or abused. This requires researchers to take strict security measures in data processing, model training and result output to protect the privacy of patients.

In addition, the applicability of the neural network is also a problem to be considered. Although neural networks do well in certain types of breast cancer diagnosis and prediction tasks, they may not be suitable for all situations. Different types of breast cancer, different stages of the disease, and differences between different patients may affect the performance of neural networks. Therefore, researchers need to select appropriate neural network models according to specific conditions, and combine with doctors 'experience and judgment to improve the accuracy and reliability of breast cancer diagnosis and prediction.

Faced with these challenges and limitations, researchers can take a series of measures to improve the effectiveness of neural networks in breast cancer diagnosis and prediction. First, more advanced neural network architectures and algorithms, such as Recurrent Neural Networks, etc., can be explored to improve the performance of the model. Second, the fusion analysis of multi-modal data can be introduced, such as combining various image data such as mammography, Magnetic Resonance Imaging, ultrasound, clinical history, gene data and other information to provide richer input features for the model, so as to improve the accuracy of diagnosis and prediction. In addition, transfer learning and generated countermeasure network can be used to solve the problems of insufficient data annotation and limited generalization ability of models. Finally, it is required to strengthen the interpretability research of the model, help doctors understand the decision-making process of the model and improve their trust in the model by means of visualization and feature importance analysis [10].

4 Conclusion

This paper investigates the application of artificial intelligence algorithms in breast cancer diagnosis and prediction, especially methods based on Artificial Neural Network and Convolutional Neural Network and discusses their potential application value and challenges.

In breast cancer detection, this paper introduces the principles of Back Propagation Neural Network and Artificial Neural Network. BPNN reduces prediction error by adjusting parameters and achieves effective classification. ANN simulates human brain neurons to construct models with strong learning and generalization capabilities to improve detection accuracy. In addition, this paper also discusses the application of Weighted Fisher Convolutional Neural Network, Deep Learning Convolutional Neural Networks and traditional Convolutional Neural Network in breast cancer prediction.

These methods extract data features through convolution operation and combine hierarchical learning with deep network structure to improve detection efficiency and accuracy, overcoming the limitations of traditional methods.

Artificial intelligence algorithms also face many challenges in breast cancer detection, such as data imbalance, noise interference and model generalization ability. Looking forward, more studies are required to explore the application of machine learning, especially deep learning in breast cancer detection, to provide more powerful support for clinical diagnosis and treatment.

References

1. Ji, G., Tang, H., Bao, T., et al.: Research Progress in the Risk Prediction Models of Breast Cancer. Medical Review 24(19), 3782-3786+3791 (2018).
2. Amiri, Z., Heidari, A., Navimipour, N.J., Unal, M., Mousavi, A.: Adventures in data analysis: A systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. Multimedia Tools and Applications 83(8), 22909-22973 (2024).
3. Qiu, Y., Chen, H., Dong, X., Lin, Z., Liao, I.Y., Tistarelli, M., Jin, Z.: Ifvit: Interpretable fixed-length representation for fingerprint matching via vision transformer. arXiv preprint arXiv:2404.08237 (2024).
4. Zhou, L., Wang, Y., Tian, L., et al.: Effect Evaluation of Different Machine Algorithms in Breast Cancer Risk Prediction Model. Traditional Chinese Medicine Information 40(08), 23-28. (2023).
5. Juejin. Implementing Data Preprocessing with Python - Juejin (juejin.cn) (2024).
6. Yuan, X.: Analysis and Prediction of Pathological Diagnosis of Breast Cancer Based on BP Neural Network. Dalian University of Technology (2020).
7. Li, X., Mo, M., Wu, F., et al.: Artificial neural network models based on questionnaire survey for prediction of breast cancer risk among Chinese women in Shanghai. Oncology 38(09), 883-893 (2018).
8. Hao, K., Hu, L., Ding, X.: A Model Based on Weighted Fisher Let-Net－5 Criterion Convolutional Neural Network for Breast Cancer Detection. Computer Technology and Development (2022).
9. Li, M.: Application of deep learning technology in breast cancer diagnosis. China Medical Engineering 29(01), 1-3 (2021).
10. Zhao, J.: Detection algorithm of breast disease based on convolution neural network. Microelectronics and Computers, (7), 48-53 (2019).