Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/38/20230556

Prediction and analysis of breast cancer based on transfer learning from DDSM mammography images

Shijie Han

School of Computing, University of Leeds, Leeds, LS2 9JT, the United Kingdom

1910841240@mail.sit.edu.cn

Abstract. Breast cancer is a cancer mostly found in women, causing a high morality rate. It appears as a lump in the breast initially, and as cancer cells spread, the other organs are endangered. Research shows that the earlier the breast cancer is detected, the higher rate of recovering will be. The main methods of breast cancer detection are mainly using various physical methods to probe the internal condition of the patient's breast. Breast X-ray imaging, also known as mammography, is the most widely used. It uses low-dose X-ray technology to observe the breast tissue and help doctors figure out the exact condition of the patient. With the advancements in deep learning and image recognition, the analysis of mammograms can be previously done by the computer, easing the pressure on doctors. Digital Database for Screening Mammography (DDSM) dataset is a mammography dataset specifically designed for the development of mammography detection models. It contains huge amounts of mammograms and provides the labels corresponding to them. In this study, the performances of normal deep learning model and transfer learning model on the DDSM dataset are compared. Pretrained neural network models are employed and fine-tuned on the DDSM dataset to adapt to the features of mammographic images. Experimental results demonstrate significant improvements through transfer learning, outperforming traditional methods with Convolutional Neural Network (CNN). Finally, the models are evaluated in multiple aspects, conclusions and prospects are analysed based on the results.

Keywords: Transfer Learning, Breast Cancer Detection, Mammographic Image, Convolutional Neural Network.

1. Introduction

In all cancer types in women group, breast cancer is the one with the most incidence [1]. It starts with tumors in breast cells and has the risk of expanding to the whole body [2]. Data collected in the decade shows that the incidence of breast cancer grows by nearly 230 million every year [3]. It is urgent and unavoidable to develop the capacity of breast cancer detection.

Research shows that early treatment has a lot of benefits to help breast cancer patients recover [4]. As the deep learning technology and image processing methods are getting popular, specific diagnosis tools for breast cancer detection can be designed and implemented to help doctors find the indications earlier. Among all vivo detection methods, mammography is a relatively fast and accurate method [5]. It utilizes X-ray to get one or several images of the patient's breast. Compared to Magnetic resonance imaging (MRI) and traditional ultrasound methods (excluding specially developed techniques),

DOI: 10.54254/2755-2721/38/20230556

mammography shows greater sensitivity in detecting microcalcifications [6]. Computed Tomography (CT) also uses X-ray to generate images and provides a better view and more details of the breast [7], but its high radiation dose may increase the possibility of getting cancer for patients. The widespread use of devices and the practicing of breast cancer detection programs lead to a continuous increase in the availability of mammography images for doctors to use in diagnostics. However, this also brings along difficulties when the quantity of patients is large, or the doctor is inexperienced. To help dealing with the images more quickly and accurately, machine learning and deep learning methods can be introduced to pre-classify the images and provide reference for doctors. Several models were built to deal with the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [8]. All the machine learning algorithms performed more than 90% accuracy on the test set. To solve the problem of lacking data, Salama [9] used transfer learning and data augmentation to improve the model's specificity and avoid over-fitting. They also applied a newly built U-net to provide the exact segmentation of lesions, which will increase the model's interpretability.

This research develops a conventional deep learning framework and a transfer learning approach, then evaluates the disparities in their outcomes. First, several preprocessing methods are applied to the input images. The first is denoising and windowing to minimize the effect of noise for later training. Second, the deep learning model is constructed. The model uses the convolutional neural network (CNN) to analyze the features and a maximum pooling layer to reduce dimensionality and computational complexity. The flat layer then converts the data into one dimension, followed by several dense layers, and finally outputs two categories (negative and positive). Third, pre-trained models on larger and more diverse datasets are introduced into the classification problem. By applying the pre-trained parameters obtained from the larger dataset, the ability of the model to extract general features is enhanced. Finally, several metrics were used to measure the models. It was shown that the model with transfer learning performs well with relatively high accuracy. The model will help physicians process mammograms more efficiently, thus contributing to the promotion of digital breast tomosynthesis (DBT) and breast cancer treatment.

2. Methodology

2.1. Dataset description

The mammographic images were sourced from the Digital Database for Screening Mammography (DDSM) Mammography dataset [10]. It consists of numerous tf-records files, each containing 2D mammographic scans representing various diagnoses. With a collection of 55,890 images, each image in the dataset is categorized into five groups, including one for normal case, two for benign cases and two for malignant cases [10]. For ease of binary classification, the labels can be grouped: 0 for negative and 1 for positive outcomes. Figure 1 showcases a positive and a negative sample from the DDSM Mammography collection.

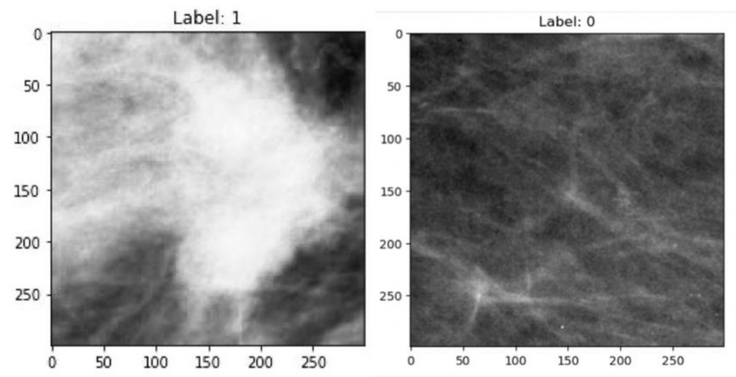

Figure 1. An image with '0' label and an image of '1' label in DDSM Mammography.

2.2. Proposed approach

In this project, the author builds two models: one is a traditional CNN used as a base model, and the other is a pre-trained model using an architecture known as ResNet50 (a deep residual network known for its outstanding performance in various image classification tasks). In base model, several convolutional layers are employed, with multiple convolutional kernels sliding over the mammograms to capture the features. Max-pooling is utilized between pairs of convolutional layers to diminish the spatial size. A flatten layer is applied then to convert the two-dimensional images into a one-dimensional sequence. Then several Fully Connected Layers (FCL) furtherly process the features and produce an output as the model's prediction finally. In this project, the output shows the chance that the image has signs of breast cancer. A value near 1 means it's likely positive for cancer, and a value near 0 means it's likely not. To build the pretrained model, the base model's convolutional layers are replaced by the ones in the pretrained figuration. The new parameters are sourced from a ResNet-50 model that was pretrained on ImageNet (a vast visual database), and these parameters are kept static to retain the capability of feature extraction. Following that, the upper parts of the fully connected layers, along with some convolutional layers, are adjusted to optimize the model's performance. Finally, the model's output size is adjusted to suit the binary classification problem in DDSM.

2.2.1. Data Analysis and Filtering. After extracting the dataset, the label distribution in DDSM is plotted in Figure 2. As shown in Figure 2, positive examples only take up 13% of the total data, with the negative examples significantly outnumbering them. This imbalanced distribution might be due to the dataset's intent to cater to multi-class classification. Among the negative samples, distinctions like normal, benign calcifications, and benign tumors further add to their count. However, such distribution can pose difficulties for binary classification tasks. To address this, the author writes a script to filter the label files, selecting an equal number of two types of labels for training. The distribution of the final dataset is shown in Figure 3.

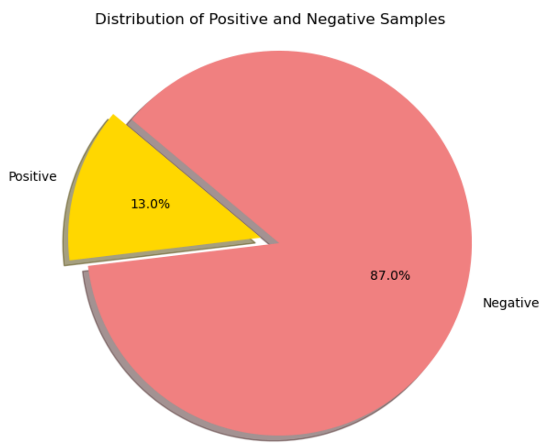

Figure 2. Distribution of Labels in DDSM Mammography.

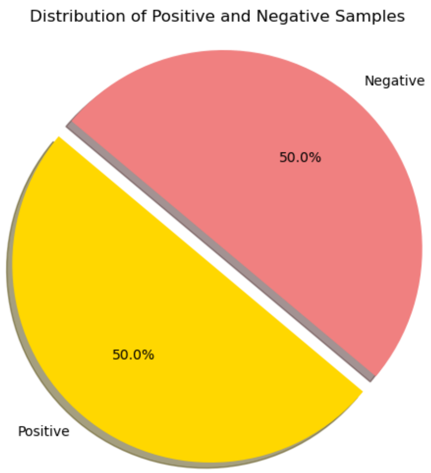

Figure 3. Distribution of Balanced Labels in DDSM Mammography.

2.2.2. Preprocessing. For the downloaded data, it is initially read into array format. Subsequently, image preprocessing operations are performed, including windowing, resizing, and image enhancement. Windowing is a commonly used technique in medical image processing. Its purpose is to enhance the visibility of images, making certain image details clearer. Since medical images may contain a wide range of grayscale values, directly displaying these images could result in unclear visualization of many details. Windowing is mainly a choice of a particular span of grayscale values, designated as the "window" for visualization. This range is then expanded to encompass the entire grayscale spectrum of the display device. Grayscale values beyond this window are clipped and set to the maximum or minimum display values. Resize refers to the process of changing the dimensions of an image, commonly used in medical image processing to adjust the size of images for various purposes such as analysis, visualization, or integration into a specific system. In this project, the size of mammograms is 299x299 initially and converted to 75x75 for later processing. This helps reduce image dimensions to speed up the training process. The result of resizing is shown in Figure 4. Finally, image augmentation is employed to increase data diversity. The methods used mainly involve applying techniques such as rotation, scaling, and other transformations to enhance slices. These methods contribute to enhancing data diversity.

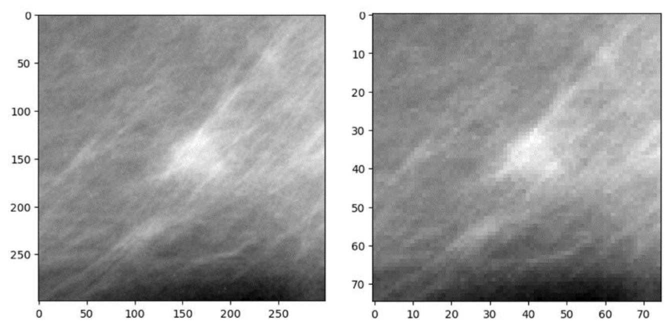

Figure 4. Comparison of an image in 299x299 and 75x75 size.

2.2.3. Training. In the training phase, both models are trained on preprocessed data. Various parameter combinations are experimented with to achieve the best model configuration. These combinations include the number and configuration of both convolutional and fully connected layers, learning rates, proportion of dropouts, early stopping policies, and cross-validation. For the basic model, the testing process begin with determining the optimal number of convolutional layers and kernel sizes. This is followed by investigating the best number of layers and dimensions for the fully connected layers. For the pretrained model, fewer experiments with convolutional layers and similar experiments of the rest settings are performed.

For both models, determining the optimal learning rate is achieved primarily using the learning rate decay algorithm provided by the Adam optimizer, which controls the learning rate by monitoring its changing rate. Similarly, an early stopping policy is applied, which monitors the changes in the learning rate. Training is terminated when the change in the learning rate falls below a certain threshold, aiming to prevent overfitting. Lastly, dropout is also applied by randomly omitting a certain number of neurons in various layers of the model. This also helps prevent overfitting.

3. Result and discussion

The primary metrics for evaluating both models include accuracy and loss. Other assessment tools consist of confusion matrices, associated calculations, and Receiver Operating Characteristic (ROC) curves. Accuracy is the most straightforward metric for assessing a model. It is the ratio of accurate forecasts to the overall predictions made by the model. During the training process, the accuracy changes of both models are illustrated in Figure 5. It can be observed that the pretrained model converges to a higher accuracy, while the basic model achieves only a 70% accuracy on the validation set.

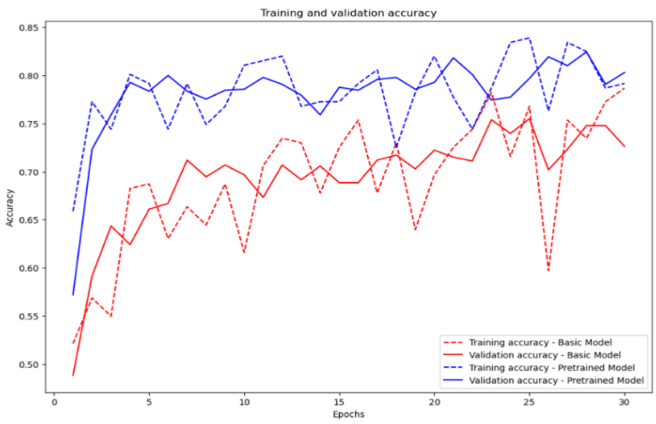

Figure 5. Accuracy Curves of the Models.

Then, both models are tested. Confusion matrices are calculated and shown in Figure 6. According to the calculations, the accuracy of the basic model is 75.71%, while the accuracy of the pretrained model is 83.33%. This indicates that the pretrained model performs better in terms of overall predictive ability. Likewise, calculate the recall values for both models. In the context of tumor detection, recall quantifies the possibility of missing diagnoses. Based on the calculations, the recall value for the basic model is 57.84%, while for the pretrained model is 80.39%. This indicates that the pretrained model has a lower indicator for missed detections (false negatives). In conclusion, the pretrained model outperforms the basic model in terms of training stability, final accuracy, and recall.

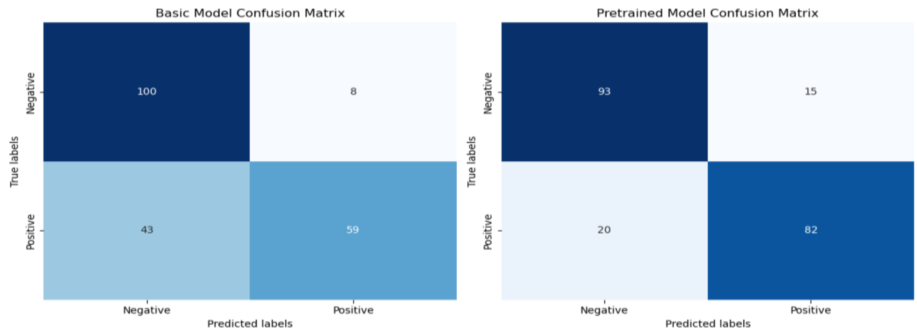

Figure 6. Confusion Matrixes of the Models.

4. Conclusion

This research mainly centers on implementing transfer learning to the DDSM mammography dataset. A CNN model with pretrained parameters is built to classify the mammograms. The model employs CNN layers to extract features from images and utilizes the weights learned from big datasets. Subsequently, these spatial features are passed to flatten layers, which can convert the data into one-dimensional format. Finally, the output layer gives the prediction. Various measures are evaluated to gauge the performance and efficacy of the pretrained model. The results show that the pretrained model has much better performance than the traditional CNN model. In the future, the interpretability and robustness of the model will be considered as the next phase of research. This study will furtherly apply more latest applications of deep learning tools in mammogram classification, aiming to provide more effective tools for doctors.

References

[1] World Health Organization (WHO) 2021 Breast Cancer https://www.who.int/news-room/fact-sheets/detail/breast-cancer
[2] Sharma G N Dave R Sanadya J Sharma P and Sharma K K 2010 Various types and management of breast cancer: An overview Journal of Advanced Pharmaceutical Technology and Research
[3] Lei S Zheng R Zhang S et al 2021 Global patterns of breast cancer incidence and mortality: A population‐based cancer registry data analysis from 2000 to 2020 Cancer Communications 41(11): pp 1183-1194
[4] Ginsburg O Yip C H Brooks A et al 2020 Breast cancer early detection: A phased approach to implementation Cancer 126: pp 2379-2393
[5] Jaglan P Dass R Duhan M 2019 Breast cancer detection techniques: issues and challenges Journal of The Institution of Engineers (India): Series B 100(4): pp 379-386
[6] Tot T Gere M Hofmeyer S et al 2021 The clinical value of detecting microcalcifications on a mammogram Seminars in cancer biology Academic Press 72: pp 165-174
[7] Boone J M Lindfors K K 2006 Breast CT: potential for breast cancer screening and diagnosis pp 351-356
[8] Agarap A F M 2018 On breast cancer detection: an application of machine learning algorithms on the wisconsin diagnostic dataset Proceedings of the 2nd international conference on machine learning and soft computing pp 5-9
[9] Salama W M Aly M H 2021 Deep learning in mammography images segmentation and classification: Automated CNN approach Alexandria Engineering Journal 60(5): pp 4701-4709
[10] Eric A Scuccimarra 2018 DDSM Mammography https://www.kaggle.com/datasets/skooch/ddsm-mammography