UDC 004.855.5
DOI: 10.20535/SRIT.2308-8893.2021.1.08

REVIEW METHODS FOR BREAST CANCER DETECTION USING ARTIFICIAL INTELLIGENCE AND DEEP LEARNING METHODS

M. NADERAN

Abstract. Nowadays, there are many related works and methods that use Neural Networks to detect the breast cancer. However, usually they do not take into account the training time and the result of False Negative (FN) while training the model. The main idea of this paper is to compare already existing methods for detecting the breast cancer using Deep Learning Algorithms. Moreover, since the breast cancer is one of the most common lethal cancers and early detection helps prevent complications, we propose a new approach and the use of the convolutional autoencoder. This proposed model has shown high performance with sensitivity, precision, and accuracy of 93,50%, 91,60% and 93% respectively.

Keywords: convolutional autoencoder, breast cancer detection, machine learning algorithms, convolutional neural networks, medical image classification.

INTRODUCTION

Breast Cancer is the second most deadly cancer mostly in women in age 20–59 worldwide. There are two classes for Breast Cancer: benign and malignant. The benign tumor cells are non–cancerous cells and grow only locally and cannot spread by invasion. Where as malignant tumor is cancerous, and the cells can grow and spread to other parts of the body. However, early detection of breast cancer could prevent cancer growth. For this reason, it is essential to train neural networks with more accurate performance and with low false negative results. According to [1], one in every eight women in the United States of America has breast cancer over their lifetime. Due to early detection and increased breast cancer awareness, there were an estimated 375,900 fewer breast cancer deaths in 2017.

REVIEW OF PREVIOUS WORKS

Authors in [2] proposed modified deep learning model where they compared two neural networks VGG and Resnet 50. In addition, they demonstrated how the number of epochs could change the performance of the model. In their experiment, the accuracy of Resnet 50 was 97% and VGG 84% with number of epoch 198 and 25, respectively. The advantage of this work is that authors worked with a whole image, which could combine two steps of preprocessing data. The first step is sliding images to recognize local patches for generating a grid of probabilistic outputs. Second step is processing to summarize the patch classifier's outputs. The limitation of this work is that authors did not include sensitivity of the model.

In [3] authors compared different methods for pre-processing data. According to the experiments, using a resample filter on input data could increase the performance of the model. Besides that, authors compared three different classifier Decision Tree (J48), NB and Sequential Minimal Optimization (SMO) on two different datasets and classifier J48 has shown the best result.

Selection of the most informative features as a classifier input is an essential task. Authors in [4] used Principle Component Analysis (PCA) method to decrease the dimensions of the feature map that will be sent to the classifier. However, there are some limitations in using PCA as a feature reduction. First, PCA is a linear relationship between features. Second, it is sensitive to the scale of the features. Third, it assumes a correlation between features and if features are not correlated, then PCA is unable to complete the principal components. To solve this problem, the Convolutional Autoencoder can be used for reducing the dimensionality of feature maps.

Dalal Bardou and other authors in [5] proposed CNN model where they compared the result of the accuracy for different magnification factors 40X, 100X, 200X and 400X. They demonstrated the result of the accuracy on 40X zoom that shows better result comparing other magnification zoom where it was 94,65%, 94,07%, 94,54% and 93,77% for 40X, 100X, 200X and 400X, respectively. Moreover, they demonstrate how pre-processing data is an essential task that must be considered while training convolutional neural networks.

Dataset BIRADS was used in experiments [6], where Support Vector Machine (SVM) method was considered for classifying breast cancer. The BIRADS dataset contain 516 benign and 445 malignant cases. During experiments authors improved the SVM model to increase false positive rate to only 13%. Moreover, SVM was compared with other methods like Decision tree, Naïve bayes, neural network and auto multilayer perceptron (MLP). Comparison result showed that sensitivity of decision tree is higher than other methods and it is 86,44%. However, the accuracy for decision tree was 78,79%. As a result, it could be concluded that the low result of accuracy does not mean a low result of sensitivity. Thus, for sensitive tasks like breast cancer detection, it is essential to consider both sensitivity and precision.

Even though SVM shows high performance [7], it cannot be used directly on images unless some pre-processing were applied on the images to extracting informative features. For example, in [8] authors used VGG as a feature extraction and applied SVM on the output of the VGG19 for classification.

Authors in [9] consider two essential tasks for training CNN: feature extraction, and fine tuning. During the experiment, CNN VGG-16 was used for feature extraction and ANFIS used for classification as well as SVM for compressing reason. Result of the experiment shows the combination of the CNN VGG-16 and ANFIST demonstrate high performance with precision 88,39%. Whereas precision of the architecture VGG + SVM was 87,10%.

DATASET

The BreakHis dataset was used during the experiments. The dataset includes two classes Benign and Malignant. There are 2,480 cases belonging to the benign category and 5,429 scans belonging to the malignant category. There were 5,536

training scans, 709 validation scans, and 1,581 test scans. Experiments showed even though the dataset is imbalanced, it will not affect the performance of the model.

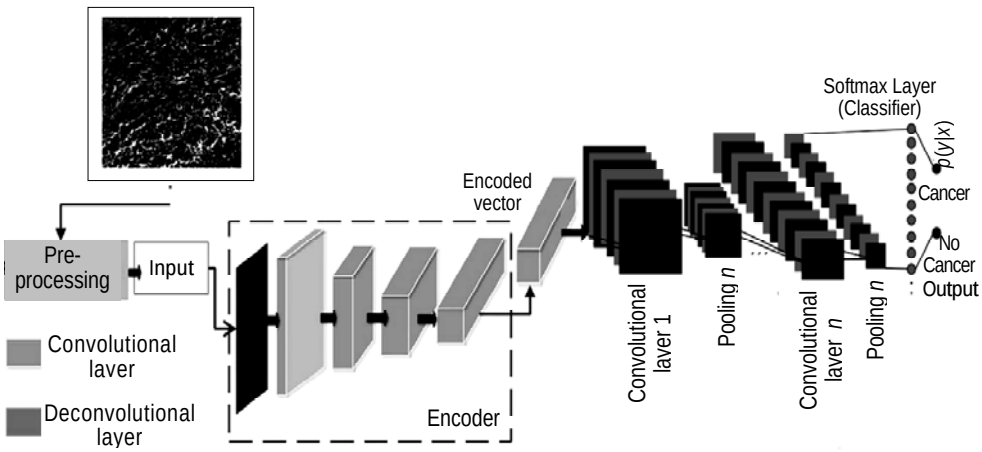
PRE-PROCESSING

There were a few steps that were considered before using the data. First, on the input data normalization was applied. Then, all data were resized using lossless compression from 700 by 460 pixels to 176 by 116 pixels. The use of lossless compression was used to reduce the input size of the data without destroying any critical information and while also preserving the aspect ratio thus maintaining the integrity of the data.

METHOD

In the current paper, a new architecture of convolutional autoencoder was proposed, where sensitivity of the model was 93,50%. Moreover, the training time was considerably lower than in the reviewed existing models.

Increasing performance of the model, required extracting the most important features. The proposed hybrid convolutional network consists of two models of convolutional autoencoder as a selection of informative features and convolutional network as a classifier. Only after achieving the most essential features, the fully connected layer was applied on the output of the convolutional layers for classifying input images as benign or malignant. The figure below demonstrates the architecture of the hybrid convolutional neural network.



Architecture of the proposed hybrid convolutional neural network

CONCLUSION

In this paper several works by various authors were reviewed. These works focused on the topic of image classification in the classification of cancer. These works tend to rely on accuracy, while sometimes neglecting the importance of sensitivity. The reviewed papers contain informative topics and ideas such as the authors in [3] use of resampling of input data as a pre-processing technique to

improve the results of the J48 network method. The analysis of accuracy in regard to different zoom levels, of the images in the dataset was proposed by authors in [5]. Authors in [8] show the use of feature extraction prior to classification using the SVM method.

According to the research and comparing related works, some of them used methods that showed high performance. However, those models are complex, and the training time is considerably high. In this paper, the proposed model showed higher performance with less training time. Faster training time is because as feature extraction, modified convolution autoencoder was used. Moreover, extraction of only informative features helped to increase the performance of the model. The sensitivity of the model has shown 93,50%.

REFERENCES

1. *Breast Cancer Facts*. Available: https://www.nationalbreastcancer.org/breast-cancer-facts
2. L. Shen et al., "Deep Learning to Improve Breast Cancer Detection on Screening Mammography", *Sci Rep*, 9, 12495, 2019. Available: https://doi.org/10.1038/s41598-019-48995-4
3. S.A. Mohammed, S. Darrab, S.A. Noaman, and G. Saake, "Analysis of Breast Cancer Detection Using Different Machine Learning Techniques", in *Data Mining and Big Data. DMBD 2020. Communications in Computer and Information Science,* vol. 1234, 2020. Springer, Singapore. Available: https://doi.org/10.1007/978-981-15-7205-0_10
4. Habib Dhahri, Eslam Al Maghayreh, Awais Mahmood, Wail Elkilani, and Mohammed Faisal Nagi, "Automated Breast Cancer Diagnosis Based on Machine Learning Algorithms", *Journal of Healthcare Engineering*, vol. 2019, article ID 4253641, 11 pages, 2019. Available: https://doi.org/10.1155/2019/4253641.
5. Dalal Bardou, Kun Zhangi and Sayed Mohammad Ahmad, "Classification of Breast Cancer Based on Histology Images Using Convolutional Neural Networks", *2018 IEEE*, vol. 6, 2018.
6. Sandeep Chaurasia, Prasun Chakrabarti, and Neha Chourasia, "Prediction of Breast Cancer Biopsy Outcomes – An Approach using Machine Leaning Perspectives", *International Journal of Computer Applications*, vol. 100, no.9, 2014.
7. Md. Milon Islam, Hasib Iqbal, Md. Rezwanul Haque, and Md. Kamrul Hasan, "Prediction of Breast Cancer Using Support Vector Machine and K-Nearest Neighbors", *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*, 21–23 Dec 2017, Dhaka, Bangladesh.
8. Md. Kamrul Hasan and Tajwar Abrar Aleef, "Automatic Mass Detection in Breast Using Deep Convolutional Neural Network and SVM Classifier", *arXiv preprint arXiv*:1907.04424v1 [cs.CV], 9 Jul 2019.
9. Yuriy P. Zaychenko, Kostiantyn A. Zdor, and Galib Hamidov, "Brain tumor diagnostics with application of hybrid fuzzy convolutional neural networks", *System Research and Information Technologies*, no. 1, 2020. Available: https://doi.org/10.20535/SRIT.2308-8893.2020.1.06.

Received 26.01.2021

INFORMATION ON THE ARTICLE

Maryam Naderan, Educational and Scientific Complex "Institute for Applied System Analysis" of the National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine, e-mail: ma.naderan@gmail.com

ОГЛЯД МЕТОДІВ ВИЯВЛЕННЯ РАКУ МОЛОЧНОЇ ЗАЛОЗИ З ВИКОРИСТАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ І МЕТОДІВ ПОГЛИБЛЕНОГО НАВЧАННЯ / М. Надеран

Анотація. Існує безліч схожих праць і методів, у яких використовуються нейронні мережі для виявлення раку молочної залози. Утім зазвичай у них не враховуються час навчання і результати помилкового негативного FN під час навчання моделі. Основною ідеєю цієї роботи є порівняння вже наявних методів виявлення раку молочної залози з використанням глибоких алгоритмів навчання. Оскільки рак молочної залози є одним з найбільш поширених смертельних видів раку, а його раннє виявлення запобігає ускладненням, запропоновано новий підхід і використання згорткового автоенкодера. Запропонована модель показала високу продуктивність із чутливістю (recall), точністю (precision) і точністю (accuracy) 93,50%, 91,60% і 93% відповідно.

Ключові слова: згортковий автоенкодер, розпізнавання рака молочної залози, алгоритми машинного навчання, згорткові нейронні мережі, класифікація медичних зображень.

ОБЗОР МЕТОДОВ ВЫЯВЛЕНИЯ РАКА МОЛОЧНОЙ ЖЕЛЕЗЫ С ИСПОЛЬЗОВАНИЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА И МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ / М. Надеран

Аннотация. Существует множество похожих работ и методов, в которых используются нейронные сети для обнаружения рака молочной железы. Однако обычно в них не учитываются время обучения и результаты ложного отрицательного FN при обучении модели. Основной идеей этой работы является сравнение уже существующих методов обнаружения рака молочного железы с использованием глубоких алгоритмов обучения. Поскольку рак молочной железы является одним из наиболее распространенных смертельных видов рака, а его раннее обнаружение предотвращает осложнения, предложен новый подход и использование сверточного автоэнкодера. Предложенная модель показала высокую производительность с чувствительностью (recall), точностью (precision) и точностью (accuracy) 93,50%, 91,60% и 93% соответственно.

Ключевые слова: сверточный автоэнкодер, распознавание рака молочной железы, алгоритмы машинного обучения, сверточные нейронные сети, классификация медицинских изображений.