Breast Cancer detection using Deep learning techniques

Anushka Sawant
Department of Computer Science
Central Michigan University
Mt.Pleasant,Michigan,United States
sawan1a@cmich.edu

Vamshi Jaligama
Department of Computer Science
Central Michigan University
Mt.Pleasant,Michigan,United States
jalig1v@cmich.edu

Lakshmi Sudha Aluru
Department of Computer Science
Central Michigan University
Mt.Pleasant,Michigan,United States
aluru1l@cmich.edu

***Abstract*—**

I. Introduction

Human body is made up of hundred thousands of cells. They provide structure for the body. When these cells get old or abnormal they die. Cancer develops in a human body when this process gets disrupted and also due to multiple changes in the genes. These cells start growing uncontrollably and form a mass called tumor. This cancerous tumor is again of two types malignant and benign. A benign cancerous tumor is a type of tumor which does not spread to other parts of body and a malignant tumor is which spreads to other parts of body is sometimes not curable. Carcinoma is a type of cancer that starts in cells that is made up of the skin or tissue lining organs. Breast cancer is a carcinoma type of cancer. It starts in the breast. It can be either one of the breasts. It most probably occurs in women and rarely men. Majority of the time breast cancer is not malignant. About 1 in 8 U.S. women will develop invasive breast cancer over the course of her lifetime. In 2022, an estimated 287,850 new cases of invasive breast cancer are expected to be diagnosed in women in the U.S.These occur due to genetic mutations that happen as a result of the aging process and life in general, rather than inherited mutations.[1] Many methods also available such as mammography, ultrasound, CT and MRI. Mammography is the most widely used screening method[2].Since, a lot of women are diagnosed with breast cancer our main motive is to develop a method which can efficiently detect breast cancer at an early stage. We intend to do this using Computer-aided detection using various deep learning techniques.

Screening Mammography:
The process of using low energy X- rays to examine the human breast for screening and for discovery of Breast cancer at an early stage through mass characteristics is called mammography screening. For creation of images mammograms uses doses of Ionizing Radiation [13]. It uses low energy x-rays like Mo and Rh than the rays which are using in some of the methods like Radiography, Ductography, Positron emission mammography, Magnetic resonance imaging [13] and ultrasounds [13]. But ultrasound imaging is further used in mammography for the detection of palable masses which can't be found by mammograms. There are different image screening techniques are being using for the detection of cancer like Digital breast tomosynthesis [14] which is reliable and used to provide higher diagnostic accuracy [14] and Diffuse optical mammography [5] which is used to investigate breast composition through spectral analysis. The radiologists who examined the mammography chooses the best category that describes the breast density. The breasts are almost entirely flatty and are extremely dense which makes hard to see the tumors during the mammograms [15].

Computer Aided detection for screening mammograms: Radiologists are looking for computer algorithms which not only increase detection rates of different medical conditions, but which are also cost effective and efficient. One specific way of detecting breast cancer is by using Computer-aided detection (CAD).CAD can double check,thus replacing traditional double reading by a second radiologist[4]. CAD is used for scanning digital mammograms and marking all the areas that indicate cancerous features. Radiologists generally review these marks after making their own interpretations and compare the two to reach a final assessment of the image [3]. Even-though such a technology might seem appealing, but the output obtained by using CAD can sometimes be controversial too. A large clinical trial in the United Kingdom has shown that single reading with CAD assistance has similar performance to double reading [8]. However, in the last decade multiple studies concluded that currently used CAD technologies do not improve the performance of radiologists in everyday practice in the United States [16-18]. From the above cases, we can conclude that CAD systems need to be improved before using them in everyday practice.[7].
Currently used CAD approaches are based on describing the X-ray image with meticulously designed hand-crafted features, and machine learning for classification on top of these features [5,6,7]. However, the Deep learning CNN models can solve problems which might be harder to solve using the traditional CAD systems. These models learn from complex data, perform image recognition, image enhancements and medical diagnosis. Hence, using these techniques might help radiologists get accurate results and predictions. Several studies have attempted to apply Deep Learning to analyze mammograms [9-12], but the problem is still far from being solved [7].

II. METHODS

A. Dataset

This dataset consists of images from the DDSM [1] and CBIS-DDSM [3] datasets. The images have been pre-processed and converted to 299x299 images by extracting the ROIs. The data is stored as tfrecords files for TensorFlow. The following dataset contains a total of 55,890 training examples. Out of these examples, 14 (percent) are positive and remaining 86(percent) are negative and divided into 5 different tfrecords files. There is one discrepancy in the data though, which is that the data was separated into training and test as per the division in the CBIS-DDSM dataset and the test files have also been divided equally into test and validation data. However, the split between test and validation data was done incorrectly, that resulted in the test NumPy files containing only masses and validation files having only the calcifications. But this discrepancy can be solved by combining both so that we can have balanced and complete test data.

Data Availability

The CBIS-DDSM dataset was obtained from: https://wiki.cancerimagingarchive.net/display/Public/CBIS-DDSM

This CBIS-DDSM (Curated Breast Imaging Subset of DDSM) is an updated and standardized version of the Digital Database for Screening Mammography (DDSM) . The CBIS-DDSM collection includes a subset of the DDSM data selected and curated by a trained mammographer. The DDSM dataset is available on: http://marathon.csee.usf.edu/Mammography/Database.html.

Pre-processing the Data The dataset consists of negative images from the DDSM dataset and positive images from the CBIS-DDSM dataset. The data was pre-processed to convert it into 299x299 images. The negative (DDSM) images were tiled into 598x598 tiles, which were then resized to 299x299. The positive (CBIS-DDSM) images had their ROIs extracted using the masks with a small amount of padding to provide context. Each ROI was then randomly cropped three times into 598x598 images, with random flips and rotations, and then the images were resized down to 299x299. The images are labeled with two labels:

labelnormal - 0 for negative and 1 for positive

label - full multi-class labels, 0 is negative, 1 is benign calcification, 2 is benign mass, 3 is malignant calcification, 4 is malignant mass.

B. Methodology

R-CNN stands for Regional Based Convolutional Neural Network. It is a machine learning model used for object detection. We use R-CNN to detect object in a given image. It combines the rectangular region proposals with convolutional neural network features. It is basically a two stage algorithm. In the first stage the image is divided into regions where there is a probability to find a object. In the second stage it classifies the object in the region accordingly. The Faster R-CNN detector adds a region proposal network (RPN) to generate region proposals directly in the network instead of using an external algorithm like Edge Boxes. The RPN uses Anchor Boxes for Object Detection. Generating region proposals in the network is faster and better tuned to your data The Faster R-CNN model builds on the Fast R-CNN model. A region proposal network is added to produce the region proposals instead of getting the proposals from an external algorithm [19].Selective search method is used in Faster R-CNN along with RPN.

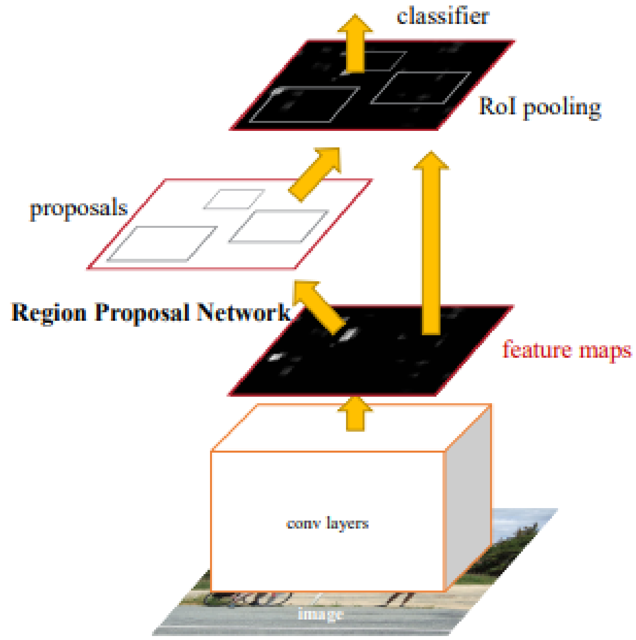

Fig. 1. Diagram of Faster R-CNN model for CAD in mammography

It is mainly divided into the following parts in structure: The Convolution layer which can extract the feature map of the entire picture. Next is the Region Proposal Network (RPN) which is the core part of Faster R-CNN, RPN uses the Selective Search method which is used in Fast R-CNN algorithm to create a region proposal. Using this RPN to obtain a region proposal can be more quickly and efficiently use the CNN network. RPN generates anchors while generating a region proposal. The judgement function determines whether the anchors are foreground or background, and then adjusts the anchors through border regression to obtain an accurate region proposal. RoI pooling: It can deal with the problem that different sizes of feature maps input to the network with fully connected layer. The fixed size is obtained by up-sampling. The classification layer is responsible for judging which class an object belongs to and the regression layer adjusts the positions of regions of interest (RoIs) to obtain the final object detection result [20]. The model proposed has a VGG16 network, which is 16 layer deep CNN[21]. The last layer detects whether the image has malignant or benign lesions. The model gives a box for each lesion that it detects and gives is score which tells how confident the model is about the lesion detected. If there is a case of multiple images of the breast of the same patient then we take average scored of each image. This model has used joint optimization which means using optimized object detection part and classifier part at the same time, during training. Vertical and horizontal flipping is also applied to the dataset. We will be using Google collab as our development environment.

III. Prepare Your Paper Before Styling

A. Abbreviations and Acronyms

B. Units

C. Equations

D. LaTeX-Specific Advice

E. Some Common Mistakes

F. Authors and Affiliations

G. Identify the Headings

H. Figures and Tables

Acknowledgment

References

[1] https://www.breastcancer.org/facts-statistics

[2] N. Khuriwal and N. Mishra, "Breast Cancer Diagnosis Using Deep Learning Algorithm," 2018 International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), 2018, pp. 98-103, doi: 10.1109/ICACCCN.2018.8748777.

[3] Keen JD, Keen JM, Keen JE. Utilization of Computer-Aided Detection for Digital Screening Mammography in the United States, 2008 to 2016. J Am Coll Radiol. 2018 Dec;15(1 Pt A):44–48. doi: 10.1016/j.jacr.2017.08.033. [PubMed] [CrossRef] [Google Scholar] [Ref list]

[4] Paquerault S, Hardy PT, Wersto N, Chen J, Smith RC. Investigation of optimal use of computer-aided detection systems: the role of the "machine" in decision making process. Acad Radiol. 2010 Sep;17(9):1112–21. doi: 10.1016/j.acra.2010.04.010. [PubMed] [CrossRef] [Google Scholar] [Ref list]

[5] Christoyianni, I., Koutras, A., Dermatas, E. Kokkinakis, G. Computer aided diagnosis of breast cancer in digitized mammograms. Comput. medical imaging graphics 26, 309–319 (2002).

[6] Hologic. Understanding ImageCheckerR CAD 10.0 User Guide – MAN-03682 Rev 002 (2017).

[7] Ribli, D., Horváth, A., Unger, Z. et al. Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep 8, 4165 (2018). https://doi.org/10.1038/s41598-018-22437-z

[8] Gilbert, F. J. et al. Single reading with computer-aided detection for screening mammography. New England Journal of Medicine 359, 1675–1684 (2008).

[9] Kooi, T. et al. Large scale deep learning for computer aided detection of mammographic lesions. Med. image analysis 35, 303–312 (2017).

[10] Becker, A. S. et al. Deep learning in mammography: Diagnostic accuracy of a multipurpose image analysis software in the detection of breast cancer. Investig. Radiol. (2017).

[11] Dhungel, N., Carneiro, G. Bradley, A. P. Fully automated classification of mammograms using deep residual neural networks. In Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on, 310–314 (IEEE, 2017) .

[12] Lotter, W., Sorensen, G. Cox, D. A multi-scale cnn and curriculum learning strategy for mammogram classification. In Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, 169–177 (Springer, 2017).

[13] https://en.wikipedia.org/wiki/Mammography

[14] https://en.wikipedia.org/wiki/Breast$_imaging$

[15] https://www.cancer.gov/types/breast/mammograms-fact-sheet

[16] Lehman, C. D. et al. Diagnostic accuracy of digital screening mammography with and without computer-aided detection. JAMA internal medicine 175, 1828–1837 (2015).

[17] Fenton, J. J. et al. Influence of computer-aided detection on performance of screening mammography. New England Journal of Medicine 356, 1399–1409 (2007).

[18] Fenton, J. J. et al. Effectiveness of computer-aided detection in community mammography practice. Journal of the National Cancer institute 103, 1152–1161 (2011).

[19] ttps://www.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.htmlmw$_25d18973-df6c-48ef-aaa9-31a4ec9e6705$

[20] Du, Lixuan Zhang, Rongyu Wang, Xiaotian. (2020). Overview of two-stage object detection algorithms. Journal of Physics: Conference Series. 1544. 012033. 10.1088/1742-6596/1544/1/012033.

[21] Simonyan, K. Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).