
CS680 Project Proposal

Shengye Chen

School of Computer Science

University of Waterloo

j57chen@uwaterloo.ca

Abstract

1 This project will be an application of computer vision techniques to solve breast
2 cancer detection problem using MRI images. I will study and analyze what has
3 been done in the previous researches that were key to the success. Next, I will
4 experiment with alternative approaches to optimize the solution and attempt to
5 produce a result close to the best performance given the constraints such as data
6 accessibility and computing resource.

7 1 Introduction

8 Breast cancer is the second most common cancer diagnosed in Canada, especially among women
9 aged over 50.[1] 1 in 8 women will be diagnosed with breast cancer in their lifetime. When the breast
10 tissue cells no longer behave normally, they may lead to breast tumours. Finding tiny tumours early
11 can prevent the disease from getting worse and reduce death rates. Using computer vision techniques
12 to screen potential breast cancer patients can significantly speed up diagnosis process. In this project
13 proposal, I will cover some interesting papers related to this topic and the preliminary steps I will
14 follow to complete this project.

15 2 Related Works

16 Many researches have been conducted to utilize computer vision to help doctors diagnosing breast
17 cancers. Based on the CDC guideline[2], there are four ways to diagnose breast cancer — breast
18 ultrasound (sonograms), diagnostic mammogram (X-ray), and breast magnetic resonance imaging
19 (MRI), and histopathological images. In this section, I will mention one paper for each medical
20 imaging type and explain why MRI is chosen for this project.

21 Histopathological images were used in [3]. Images were labeled as benign and malignant lesions
22 and each lesion had four classes. Researchers performed data augmentation on the training dataset.
23 Using ResNet, researchers achieved the best average accuracy for binary classification and multi-
24 class classification with 99.7% and 97.81% respectively.

25 In [4], the researchers adopted the Breast Ultrasound Images (BUSI) dataset to train a DarkNet-53
26 model. Data augmentation was performed to increase the diversity of the original dataset. The
27 features were extracted from the pooling layer and then the best features were selected using two
28 different optimization algorithms — the reformed binary grey wolf optimization (BGWO) and the
29 reformed differential evolution (DE) algorithm. By using feature fusion and CSVM classifier, re-
30 searchers achieved an accuracy of 99.1%.

31 Researchers in [5] used four mammography imaging datasets consisting of normal, benign, and
32 malignant pictures. They used Inception V4, ResNet-164, VGG-11, and DenseNet121 models as
33 base classifiers and ranked them with the Gompertz function. Then the decision scores of the base
34 models were adaptively combined to construct final predictions. Their model achieved a 99.32%
35 accuracy rate.

36 MRI images were used in [6] where researchers proposed a Back Propagation Boosting Recurrent
37 Widening Model (BPBRW) with a Hybrid Krill Herd African Buffalo Optimization (HKH-ABO)
38 method to diagnose breast cancer at an earlier stage. Furthermore, Wiener filter was established
39 for pre-processing the MRI noisy image content. They demonstrated their model had a 99.6%
40 accuracy rate.

41 Based on the related works above, I will choose MRI images to classify breast cancer for the follow-
42 ing reasons: compared to mammography, MRI is not affected by dangerous radiation. Additionally,
43 MRI has a higher sensitivity for detecting breast cancer than mammography and sonograms. Even
44 though the project using histopathological images achieved the best performance, it requires taking
45 tissue samples from patients, so biopsy is usually performed as the last step to affirm patients are
46 positive for cancer. Therefore, MRI is the most suitable way to quickly screen breast cancer patients
47 at an early stage.

48 **3 Project Description**

49 **3.1 Problem Definition**

50 In this project, I will apply computer vision techniques to classify grey-scale breast MRI images
51 as either positive or negative for cancer. Given the constraint of dataset availability and computing
52 power, training and testing images will be reduced-size MRI images.

53 **3.2 Project Plan**

54 After studying what has been done in related works, the preliminary steps of my project include data
55 collection, data augmentation, pre-processing, training classifiers with traditional pattern recognition
56 methods, training classifiers with CNN methods, and test result evaluation.

57 Data augmentation is a standard methodology to increase the dataset size in order to achieve a better
58 performance with CNN. It also reduces redundancy and overfitting by arranging the dataset in a
59 balanced manner. Random reflection, multiple rotations, and translations regarding horizontal or
60 vertical aspects of the images are commonly used to augment the image data.

61 Pre-processing step is commonly adopted to remove noise and enhance contrast. In order to enhance
62 the qualities of breast cancer images, median filter and Gaussian filter are two options to remove
63 noise. The main reason to use these filters are their abilities to maintain the original image features,
64 like edges and corners. Images are often required to be resized into a uniform size in order to fit
65 pre-trained deep neural networks.

66 I plan to experiment with different methods to train classifiers, including but not limited to naïve
67 Bayes and SVM for traditional approach (proved to be feasible in [7]), and ResNet for CNN ap-
68 proach (proved to be feasible in [3][5]).

69 **3.3 Expected Outcome**

70 I will attempt to produce a result that is as close to the highest accuracy as possible. Due to my lack
71 of medical expertise, I will not be able to explain why images are classified as positive or negative.
72 This is the common drawback with neural networks regardless of their creator or application. My
73 learnings from this project is the most important outcome.

74 **4 Dataset**

75 As I mentioned before, MRI images used in this project will be grey-scale and in reduced size. The
76 dataset from Kaggle[8] consists of well-labeled and balanced MRI images, with 740 healthy and
77 740 sick images. Each image has a dimension of 540×250 pixels. This dataset does not have
78 annotated cancer region and this could increase the difficulty for machine learning models to find
79 which area is the most useful to detect breast cancer. High-resolution and annotated medical images
80 are difficult to access and often require professional software to process. Moreover, it will cost large
81 amount of resource to train models on high-quality images. Therefore, the Kaggle dataset should
82 serve as a good starting point for this project.

References

- [1] Darren R. Brenner, Hannah K. Weir, Alain A. Demers, Larry F. Ellison, Cheryl Louzado, Amanda Shaw, Donna Turner, Ryan R. Woods, and Leah M. Smith. “Projected estimates of cancer in Canada in 2020”. *Canadian Medical Association Journal*, vol. 192, no. 9 (2020) (cit. on p. 1).
- [2] Centers for Disease Control and Prevention. *How Is Breast Cancer Diagnosed?* URL: https://www.cdc.gov/cancer/breast/basic_info/diagnosis.htm (visited on 10/25/2022) (cit. on p. 1).
- [3] Hanan Aljuaid, Nazik Alturki, Najah Alsubaie, Lucia Cavallaro, and Antonio Liotta. “Computer-aided diagnosis for breast cancer classification using deep neural networks and transfer learning”. *Computer Methods and Programs in Biomedicine*, vol. 223 (2022), p. 106951 (cit. on pp. 1, 2).
- [4] Kiran Jabeen, Muhammad Attique Khan, Majed Alhaisoni, Usman Tariq, Yu-Dong Zhang, Ameer Hamza, Artūras Mickus, and Robertas Damaševičius. “Breast Cancer Classification from Ultrasound Images Using Probability-Based Optimal Deep Learning Feature Fusion”. *Sensors*, vol. 22, no. 3 (2022) (cit. on p. 1).
- [5] Ayman Altameem, Chandrakanta Mahanty, Ramesh Chandra Poonia, Abdul Khader Jilani Saudagar, and Raghvendra Kumar. “Breast Cancer Detection in Mammography Images Using Deep Convolutional Neural Networks and Fuzzy Ensemble Modeling Techniques”. *Diagnostics*, vol. 12, no. 8 (2022) (cit. on pp. 1, 2).
- [6] Kranti Kumar Dewangan, Deepak Kumar Dewangan, Satya Prakash Sahu, and Rekhram Janghel. “Breast cancer diagnosis in an early stage using novel deep learning with hybrid optimization technique”. *Multimedia Tools and Applications*, vol. 81, no. 10 (Apr. 2022), pp. 13935–13960 (cit. on p. 2).
- [7] Maha M. Alshammari, Afnan Almuhanha, and Jamal Alhiyafi. “Mammography Image-Based Diagnosis of Breast Cancer Using Machine Learning: A Pilot Study”. *Sensors*, vol. 22, no. 1 (2022) (cit. on p. 2).
- [8] Uzair Khan. *Breast Cancer Patients MRI’s Dataset*. URL: <https://www.kaggle.com/datasets/uzairkhan45/breast-cancer-patients-mris> (visited on 10/25/2022) (cit. on p. 2).