“Detection and classification the breast tumors using mask R-CNN on sonograms”

Gina Cody School of Engineering and Computer Science
Concordia University
Pattern Recognition
Instructor: Adam Krzyżak
Fall 2022

Radley Carpio 40074888 Nazli Ensafi 40038607 Victor Tobar 40042766

Abstract

Cancer is the second leading cause of death in women in North America [13]. Breast cancer is the most common cancer in women on the continent. If the disease goes undiagnosed, it will rapidly spread to the regional lymph nodes [14]. The later the diagnosis, the farther in the body it travels; if it spreads to a distant part of the body, the 5-year survival rate is 29%. Therefore, doctors recommend regular screening after the age of 40 for women without a history and earlier otherwise. Standard screening methods are ultrasound and mammogram imaging proceeding by an FNA in case of detection of a suspicious mass. In an attempt to compare the efficiency of image-based vs. feature-based cancer classification, this project has reproduced the work of Jui-Ying, et al that have used a mask R-CNN on ultrasound images to achieve an accuracy of 85% for benign and malignant classification. The overall classification model produced only achieved an accuracy of 64%. Therefore, exploratory experiments were carried out and identified a potentially more reliable alternative for using sonograms for detection, segmentation, and classification. We implemented a binary normality classification model that classifies sonograms as normal or abnormal (benign and malignant). It achieved an accuracy of 94% and can be used to determine whether a biopsy is required. Finally, by testing three different binary classifiers, Gaussian Naive Bayes, and Multilayered Neural Network, an accuracy of 99% was realized on a feature-based dataset containing 30 features extracted from FNA-type biopsies.

1. Introduction

About 1 in 8 women will get breast cancer during her life, which makes the disease the most common cancer among women in North America. It is formed by an abnormal division of ducts and lobules that change the structure of the breast and produce a tumor. Factors such as nuclear to cytoplasm ratio, differentiation ability, and cell pleomorphic are used to classify tumors into two categories, benign or malignant [1]. Timely diagnosis of the disease is crucial in decreasing its mortality rate.

One of the common challenges that arise in screening is the density of the breast tissue. On a mammogram, non-dense breast tissue appears dark and transparent, whereas dense breast tissue appears as a solid white area, which causes a 30% decrease in the sensitivity of mammograms [2][3].

About half of women undergoing mammography have dense breasts. Having dense breasts increases the chance that breast cancer may go undetected by a mammogram since dense breast tissue can mask potential cancer. The most promising screening approach for dense breasts has been clinically proven to be ultrasound imaging [2][3]. Ultrasound imaging has other significant advantages, such as no ionizing radiation, real-time examination, and relative affordability, especially in comparison with MRI [4].
However, ultrasounds generally accompany FNA's (fine needle aspiration) when imaging shows an abnormal growth or area [5]. An invasive procedure, and in case of tumor heterogeneity, it only collects inconclusive information. This study is done to overcome the shortages of ultrasound/biopsy screening by objectively analyzing noninvasive ultrasound images.

Various studies have been done for breast cancer classification. Some have used features suggested by "Breast imaging reporting and data system" (BI-RADS), such as orientation, margin, shape, lesion boundary, echo pattern, and posterior acoustic feature classes in statistical models to distinguish between benign from malignant masses. In contrast, others have used artificial neural networks to evaluate characteristics such as the brightness of nodules, the number of lobules, ellipsoid shape, branch pattern, and spiculation to classify breast lesions. Li et al have compared deep learning and feature-based learning to find the effectiveness of the two and found out the deep learning methods give better results [2].

At the heart of deep learning are Neural Networks, ANNs, whose name and structure mimic how biological neurons work with one another. Convolutional Neural Networks are a variation of ANNs that are primarily used to solve challenging image-driven pattern recognition tasks [6]. They have been proven to be an excellent tool for problems dealing with collars and borders. Region-based Convolutional Neural Networks, R-CNNs are specially designed for object detection. Because it takes a significant amount of time to train the network, as you would have to classify 2000 region proposals per image, Fast R-CNNs have emerged [7]. Instead of feeding the region proposals to the CNN, Fast R-CNNs feed the input image to the CNN to generate a convolutional feature map [8]. From the map, they identify the region of proposals and wrap them into squares. Using a RoI pooling layer, they are then reshaped into a fixed size to be fed into a fully connected layer. From the RoI feature vector, a softmax layer is used to predict the class of the proposed region and the offset values for the bounding box. This works faster because

2000 region proposals don’t have to be fed to the convolutional neural network every time. The convolution operation is done only once per image, and a feature map is generated from it.

Jui-Ying, et al have used a mask R-CNN approach, which is based on Faster R-CNN and has the advantage of automatic image segmentation. It Defines the bounding box of tumors and draws a contour of the tumor area before lesion classification between benign and malignant. They developed a model that detects, segments, and classifies breast lesions with ultrasound images. Their model was trained and tested using a total of 307 images, 178 classified as benign and 129 classified as malignant. More specifically, 80% of the images were used for training, and 20% were for testing. Ultimately, their model achieved an accuracy of 85% [2]. The target points of this study are the biopsy results of the patients whose ultrasound images are the input to the model.

2. Material and Methodology

2.1. Material

Similarly to the original paper, we have used ultrasound images as an input to the model, see Figure 1. The data includes breast ultrasound images among 600 women between 25 and 75 years old collected in 2018. The dataset consists of 780 sonograms, in PNG format, with an average image size of 500*500 pixels. Additionally, each image is accompanied by the tumor contour image in PNG format, which will serve as the ground truth for the mask R-CNN [9]. The images are categorized into three classes, normal, benign, and malignant. The dataset was originally used in an article by Al-Dhabyani et al in 2020.

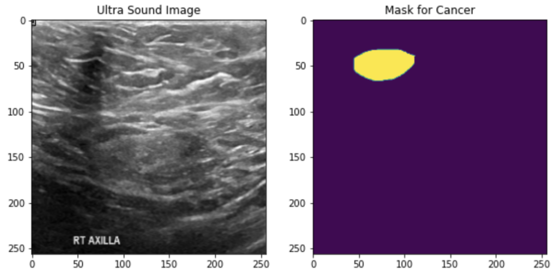
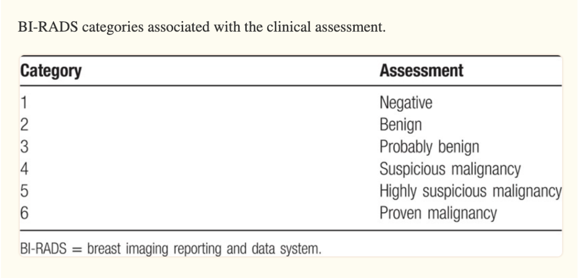


Figure 1. Sample ultrasound image and the contour of the mass, used in this project and originally used by Al-Dhabyani et al in 2020.

In the original paper by Jui-Ying, et al, The tumors in the collected ultrasound images were delineated and contoured by an experienced radiologist, then classified into 6 BI-RADS categories, see Table 1, by a physician. If the mass was in category 3, the medical team assessed whether or not to proceed with a biopsy. Alternatively, in the case of category 4 or higher, biopsies were mostly suggested. The results of tumor contours and biopsies were then used as the ground truth for Mask R-CNN network training [2].



BI-RADS categories associated with the clinical assessment.

Category	Assessment
1	Negative
2	Benign
3	Probably benign
4	Suspicious malignancy
5	Highly suspicious malignancy
6	Proven malignancy

BI-RADS = breast imaging reporting and data system.

Table 1. BI-RADS tumour assessment

2.2. Mask R-CNN Methodology

Mask R-CNN is state-of-the-art in terms of image segmentation and instance segmentation [10]. Mask R-CNN was developed on top of Faster R-CNN, a Region-Based Convolutional Neural Network explained above.

To comprehend Mask R-CNNs, we need first to understand image segmentation. Image segmentation or pixel-level classification is the task of clustering parts of an image that belong to the same object class.The clusters or segments then locate the objects and their boundaries, such as lines and curves [10].

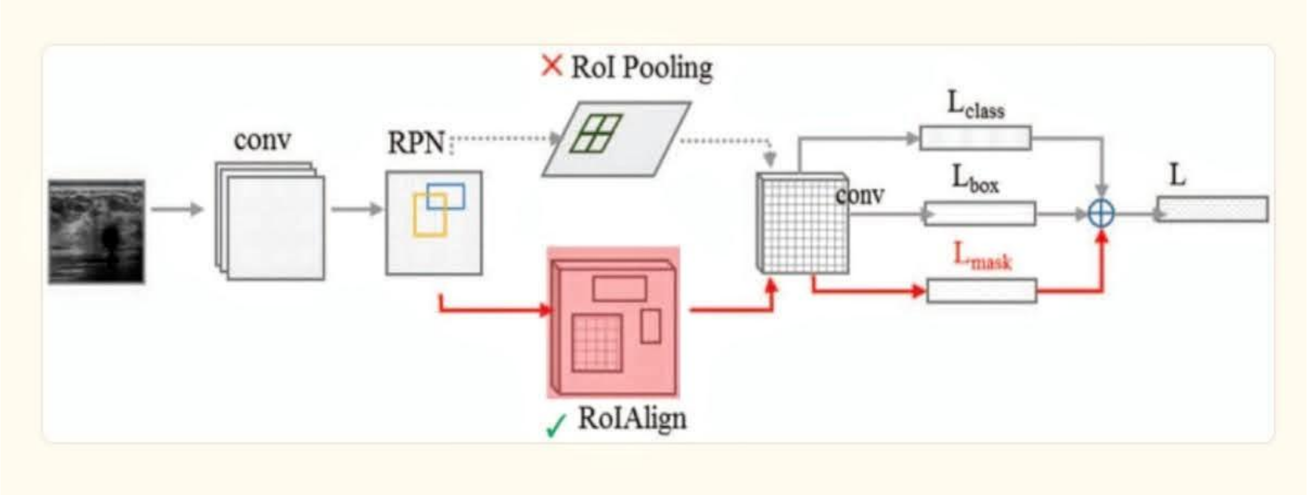


Figure 2. The network architecture of Mast R-CNN.
R-CNN = regions with convolutional neural network,
RoI = region of interest, RoIAlign = region of interest alignment,
RoIPool = region of interest pooling.

Mask R-CNNs use region proposal networks (RPN) to extract features and tighten the bounding boxes. As a feature extraction method, they use RoIPool. Then max pooling is used to quantify each RoI region and solve the problem of sizes of RoI features at different scales. Unfortunately, this process causes misplacement in the original image RoI and extraction features because of the loss of spatial information. Mask R-CNN tackles this issue by replacing RoI pooling, used by Faster R-CNN, with ROI alignment or RoIAlign. Finally, it uses the mask branch to mark the results of RoIAlign for the object area. See Figure 2 for an overview of the architecture [2][10].

The loss function of Mask R-CNNs has one extra element compared to the loss function of R-CNNs, which is the loss of the mask. See formulas 1, 2, and 3 for the details of the loss functions and Table 2 for the notation [2].

$$L = L_{class} + L_{box} + L_{mask}$$

Formula 1. Loss function of Mask R-CNNs

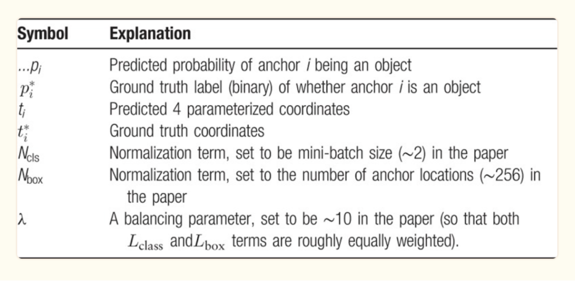
$$L_{\text{class}} + L_{\text{box}} = \frac{1}{N_{\text{cls}}} \sum_i L_{\text{cls}}(p_i, p_i^*) + \frac{1}{N_{\text{box}}} \sum_i p_i^* L_1^{\text{smooth}}(t_i - t_i^*)$$

$$L_{\text{cls}}(\{p_i, p_i^*\}) = -p_i^* log p_i^* - (1 - p_i^*) log(1 - p_i^*)$$

Formula 2. Loss function of R-CNNs

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \le i,j \le m} \left[y_{ij} log \circ y_{ij}^k + \left(1 - y_{ij}\right) log\left(1 - \circ y_{ij}^k\right)\right]$$

Formula 3.. Loss function associated with the Mask, average binary cross-entropy loss



Symbol	Explanation
$...p_i$	Predicted probability of anchor *i* being an object
p_i^*	Ground truth label (binary) of whether anchor *i* is an object
t_i	Predicted 4 parameterized coordinates
t_i^*	Ground truth coordinates
N_{cls}	Normalization term, set to be mini-batch size (~2) in the paper
N_{box}	Normalization term, set to the number of anchor locations (~256) in the paper
λ	A balancing parameter, set to be ~10 in the paper (so that both L_{class} andL_{box} terms are roughly equally weighted).

Table 2. Notations in the loss functions

3. Implementation and Experiments

3.1. Implementation

The images from each class of sonogram were divided into an 80% training and 20% testing split. Splitting each image class equivalently ensured the same ratio of each sample type in both sets. In addition, the sonograms' detection, segmentation, and classification were divided into two different models, one responsible for the detection and segmentation and the second responsible for taking the generated segmentations to apply the classification.

Regarding segmentation, the first model follows the structure of a U-Net model, a type of convolutional neural network used for image segmentation tasks. The U-Net architecture is known for its efficiency and effectiveness in segmenting images into different regions or classes, see Figure. 3.

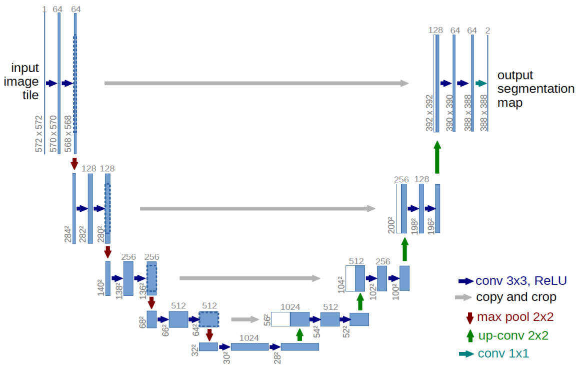


Figure. 3 U-Net model architecture.

.

It consists of a series of convolutional blocks followed by max-pooling layers, dropout layers, and up-sampling layers and is compiled with the Adam optimizer and binary cross-entropy loss function [12].

A convolutional block consists of two convolutional layers, each followed by an optional batch normalization layer and an activation layer. This is a standard building block in many convolutional neural network architectures, allowing the network to learn more complex and abstract features from the input data. Furthermore, batch normalization and ReLU activation can help improve the performance and stability of the network.

Max pooling is a technique used in image processing and computer vision to down sample an image. This is typically done by dividing the image into a grid and then, for each grid cell, taking the maximum value of all the pixels in that cell. This reduces the size of the image and makes the image's features more robust to small changes in the position of the objects in the image. This can be useful for tasks such as object recognition, where the goal is to identify the objects in an image regardless of their position within the image.

Up-sampling is a technique used in image processing to increase the size of an image. This is typically done by inserting zeros between the original image's values and then applying a low-pass filter to smooth out the resulting image. This can be useful for tasks such as image segmentation, where the goal is to produce a detailed output image of the same dimensions as the input image.

Regarding the classification of the generated segmentations, the model consists of a sequence of layers, including convolutional, max pooling, and dense layers. It is compiled with the Adam optimizer and binary cross-entropy loss function and designed to output probabilities for each class via the Softmax activation function on the final dense layer.

3.2 Experiments

3.2.1 Using Sonogram Images

Training the segmentation model with a batch size of 10 and 100 epochs achieved a validation accuracy of 92%. In a subsequent attempt, the model with a smaller batch size of 5 and 200 epochs achieved a validation accuracy of 94%. The lower batch size value removed some unwanted generalizations and demonstrated that additional tweaking of the hyperparameters could yield improved results.

However, our multiclass classification model that distinguishes between normal, benign, and malignant yielded a 64% accuracy.

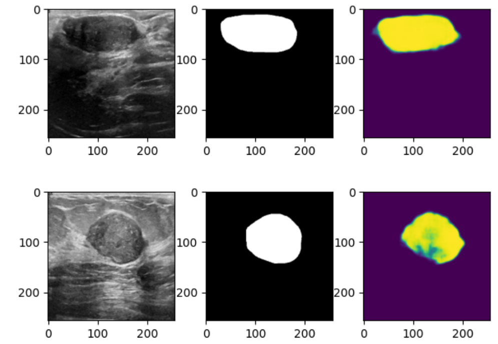


Figure 4. From the left, ultrasound image, mask, segmentation of a benign tumor

Figure 4 is an example of good segmentation of the sonograms produced by the first model. However, Figure 5 shows two inaccurate segmentations of tumors on similarly looking sonograms. The above segmentation is incomplete and only captures the top of the tumor. In contrast, the sonogram produces over-segmentation by capturing the tumor itself along with the dark area beneath it.

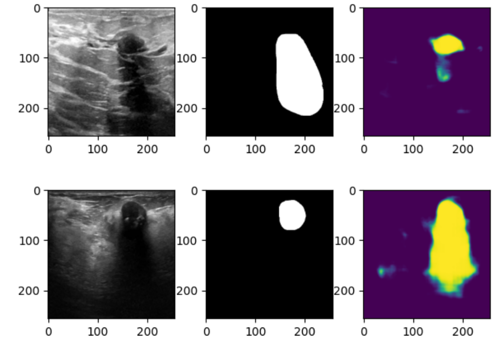


Figure 5. From the left, ultrasound image, mask, segmentation of a benign tumor

Malignant tumors are sometimes overly under-segmented by the model. For example, see an accurate segmentation on top and an utterly inaccurate one on the bottom in Figure 6 below.

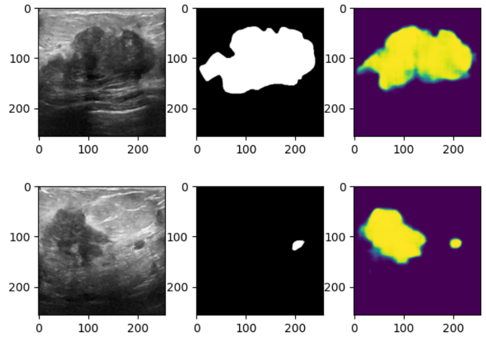


Figure 6. From the left, ultrasound image, mask, segmentation of a malignant tumor

In the case of normal sonograms, Figure 7 illustrates the difference between a good and a bad segmentation. The above image shows that for a correct segmentation for a normal sonogram, only faint specs from the dark spots of the image should be picked up. Whereas for a bad segmentation found on the image below, the model identifies a round dark spot as a tumor by over-segmentation.

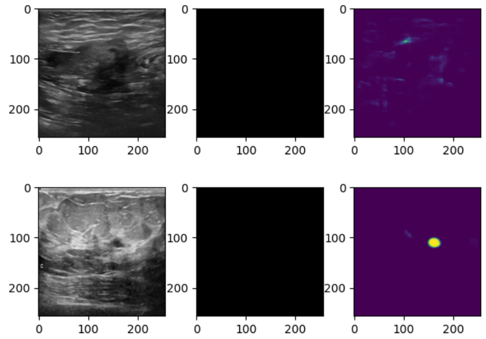


Figure 7. From the left, ultrasound image, mask, segmentation of a normal tumor

The noise in the sonogram images is visible in Figures 4 to 7, which explains how it can be difficult for the machine to discern the outline of the tumors and cause overall low accuracy.

In an attempt to clarify if a biopsy should be carried out, we implemented a binary classification model that distinguishes between normal vs. benign or malignant. The new model yielded 93% accuracy, which could be improved as the size of the dataset increases.

3.2.1 Using FNA Results

We further strived for an alternative computer-aided diagnostic system by implementing three different binary classifiers, Gaussian Naive Bayes, Multilayered Neural Network, and K-Means, and a dataset containing FNA results.

We have used the 'Wisconsin Breast Cancer Diagnostic Dataset'. Features are computed from a digitized image of a fine needle aspirate of a breast mass. They describe ten real-valued characteristics of the cell nuclei present in the image. These features include cell nuclei's radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension. Additionally, each of these ten features is subdivided into three sub-categories, mean, standard error, and worse, resulting in a total of 30 features. Features are recorded with four significant digits and no missing attribute values. Class distribution is 357 benign and 212 malignant [13].

With the knowledge that the F1 score is a better measure when generally comparing classifiers and the fact that for cancer diagnostic tasks, recall and false negative rates are more informative than accuracy, we have decided to use accuracy, merely because the original paper has used this measure to report the results of their work.

We have produced an accuracy of 99% using the Wisconsin dataset and our most reliable classifier, the multilayered neural network.
The results from the classification based on FNA driven dataset show that, despite the invasive nature of the procedure, it produces more accurate results.

4. Remarks

We acknowledge that the results of this comparison would be far more reliable if the datasets that we have used and that of the paper were collected from the very same patients and contained a larger number of samples. And That the comparison of their performances just based on their accuracy is not perfectly appropriate. In the case of cancer diagnostic tasks, recall and false negative rates are more informative than accuracy. The decision to use accuracy has been merely because the original paper has used this measure to report the results of their work.

5. Conclusion

Pattern recognition, which takes up a large area in the domain of artificial intelligence, has reached new heights in clinical cancer research in recent years. In this project, we have explored how AI could assist in cancer diagnosis and prognosis. The work consists of two main parts. The first part, based on a paper by Jui-Ying et al, focuses on the segmentation and classification of 780 ultrasound images from women between the ages of 25 and 75 collected in 2018, using a methodology based on the Faster Region-Based Convolutional Neural Network, called Mask R-CNN. The data was then segmented into three categories: normal, malignant, and benign. Unfortunately, the multi-class classification yielded only 64% accuracy.
Furthermore, to clarify if a biopsy should be carried out, we further implemented a binary classification model that distinguishes between normal vs. benign or malignant, which produced a promising accuracy of 93%. The second part uses a feature-based approach utilizing the Wisconsin breast cancer dataset. The dataset contains 569 instances of 30 features that are computed from a digitized image of a fine needle aspirate of a breast mass. The dataset has been fed to three different binary classifiers, and the best result has been found to be that of our multilayered neural network, 99% accuracy, with 100 epochs and a batch size of 10. While image-based classification offers an automatic and non-invasive methodology, the results of classification based on FNA have proven to be more promising by far. Although with the speed that research in image processing advances and as datasets grow larger, the image-based approaches become more and more reliable and applicable.

References

[1]*cancer-stats*.(n.d.).https://cancer.ca/en/cancer-information/cancer-types/breast/statistics

[2]Jui-Ying, Kuan-Yung, Ken Ying-Kai, Po-Hsin, Geoffrey, & Tzung-Chi. (2019). Detection and classification of breast tumours using mask R-CNN on sonograms.*Medicine*. https://doi.org/10.1097/MD.0000000000015200

[3] *Dense breast tissue: What it means to have dense breasts*. (2022, February 25). MayoClinic.https://www.mayoclinic.org/tests-procedures/mammogram/in-depth/dense-breast-tissue/art-20123968

[4]*tests-procedures*.(n.d.).https://cancer.ca/en/cancer-information/cancer-types/breast/statistics

[5]*FNA*.(n.d.).https://cancer.ca/en/treatments/tests-and-procedures/fine-needle-aspiration-fna

[6]Education, I. C. (2021, August 3). *Neural Networks*.https://www.ibm.com/cloud/learn/neural-networks

[7]O'Shea, & Nash. (2015). An Introduction to Convolutional Neural Networks. *Computer Science(Neural and Evolutionary Computing)*. https://doi.org/10.48550/arXiv.1511.08458

[8]Gandhi, R. (2018, December 3). *R-CNN, Fast R-CNN, Faster R-CNN, YOLO — Object Detection Algorithms*.Medium.https://towardsdatascience.com/r-cnn-fast-r-cnn-faster-r-cnn-yolo-object-detection-algorithms-36d53571365e

[9]Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, *28*, 104863. https://doi.org/10.1016/j.dib.2019.104863

[10]Odemakinde, E. (2022, July 11). *Everything about Mask R-CNN: ABeginner's Guide*.viso.ai.https://viso.ai/deep-learning/mask-r-cnn/

[11]Klingler, N. (2022, August 2). *Image Segmentation with Deep Learning (Beginner Guide)*. viso.ai. https://viso.ai/deep-learning/image-segmentation-using-deep-learning/

[12]Ronneberger, O., Fischer, P., & Brox, T. (2015, May). U-Net: Convolutional networks for biomedical image segmentation. arXiv:1505.04597v1 [cs.CV]. Retrieved from https://arxiv.org/abs/1505.04597v1

[13]*UCI Machine Learning Repository: Breast Cancer Wisconsin (Diagnostic) Data Set*. (n.d.). https://archive.ics.uci.edu/ml/datasets/breast+cancer+wisconsin+(diagnostic)

[14]*Leading Causes of Death-Females-All races/origins*. (2021, June 21). Centers for Disease Control and Prevention. https://www.cdc.gov/women/lcod/2017/all-races-origins/index.htm

[15]*Breast Cancer - Statistics*. (2022, May 24). Cancer.Net. https://www.cancer.net/cancer-types/breast-cancer/statistics