A Machine Learning Approach for Ultrasound Breast Cancer Classification and Easy Accessibility through Google Colab

Chris Gomez Miles Wilderman

June 30, 2023

Abstract

Breast cancer remains a significant health concern worldwide, necessitating early detection and treatment for improved patient outcomes. This article focuses on the development of a breast cancer classification model using machine learning techniques applied to public ultrasound images. The primary objective is to assist in the early detection of breast cancer, enabling timely interventions and improved treatment efficacy. There are a few rarer types of breast cancer often looked over, which is what our code takes into account. CNN models have previously been used and shown to be the best for classifying images of this manner. We built a CNN, VGG16, VGG19, and Resnet-50 model to classify between ultrasound images of benign and malignant tumors. We also included images without any tumor as a control for our model. Our VGG 19 model achieved the highest training accuracy of 78.41% and takes new images from the web or as an upload for classification.

1 Introduction

Much has been done to research the science behind curing cancer, but still more can be done in recognizing it quickly and effectively. This can and has been done through machine learning CNN models using images of the cancer. We will further increase the performance of existing models with improved metrics and performance on different or new ultrasound images. Rosenblatt's first envisioning of a neural network applied to images in a machine learning context gave us the idea of a threshold being used in the decisions behind assessing an image (Rosenblatt, 1957). This idea was then further explored and refined, with a CNN model first being used for breast cancer classification on mammogram imaging in 1994 (Wu et al., 1994). They used 80 regions of interest for their model, which is not where CNN models are now. They also used radiographs and microcalcifications for their imaging, and the model had some computational difficulties. More recently, Jabeen K. et al. have made more improvements to these systems using pretrained models, optimization algorithms, and a probability-based serial approach for the final layer (Jabeen K. et al. 2022). This is also the latest work using ultrasound imaging as the medium for the model. However, this model and many other models will struggle with rarer types of breast cancer. If we can target these types of breast cancer in the training process, we can have a more successful model overall.

VGG19 is an extension of VGG16, featuring 19 layers. It includes 16 convolutional layers and 3 fully connected layers. The additional layers improve the representational capacity of the model, but also increase the computational complexity. VGG19 retains the same architectural principles as VGG16, with the use of 3x3 filters, max-pooling layers, and ReLU activations.

Both VGG16 and VGG19 have achieved impressive performance on various computer vision tasks, particularly in image classification. Their deep architectures enable them to capture complex visual features, making them effective at learning hierarchical representations of images. However, the main drawback of these models is their large number of parameters, which can make training and deployment computationally expensive. The ResNet (Residual Network) model is a deep convolutional neural network architecture that introduced the concept of residual learning, addressing the problem of vanishing gradients in very deep networks. It was proposed by Kaiming He et al. in their paper titled "Deep Residual Learning for Image Recognition" (He et al., 2015).

ResNet utilizes a residual block structure, where each block contains multiple convolutional layers. The key innovation of ResNet lies in the introduction of skip connections or shortcut connections that bypass some convolutional layers. These connections allow the model to learn residual map pings, focusing on the difference between the input and the desired output, which facilitates training of deeper networks.

The skip connections help in overcoming the vanishing gradient problem, where gradients diminish as they propagate backward through numerous layers. By directly propagating the input signal through the network, ResNet enables the gradient flow and facilitates the training of very deep models.

ResNet comes in various depths, such as ResNet-18, ResNet-34, ResNet-50, ResNet-101, and ResNet-152, where the number indicates the total number of layers in the network. Deeper versions of ResNet have shown improved performance on image classification and other computer vision tasks.

The ResNet architecture has had a significant impact on the field of deep learning and has become a widely adopted model for various image recognition tasks. Its effectiveness in training deep net works and its ability to achieve state-of-the-art results have made it a popular choice in the research community.

In summary, ResNet addresses the challenge of training very deep neural networks by introducing residual connections that enable the flow of gradients. These connections allow ResNet to learn residual mappings, making it easier to train deeper architectures and achieve improved performance on image recognition tasks (He et al., 2015).

1.1 Need of the System

Easy access to a system that can accurately give a prediction of the tumor as being benign or malignant is very important. Once the patient has the image, a second opinion is usually very important and can be expensive. With an easily accessible model, more people will have access to more information about their cancer and how they should go about treating it.

1.2 Applications of Proposed System

Our models can classify any input ultrasound image of breast cancer. It may be accurate for other types of cancer; however, we do not claim that it is. Our system will be applied to newly found cancer to determine the best course of action. Our system could be applied to other types of cancer if it is trained on similar images to that of which we trained our breast cancer model on. We cannot comment on the accuracy or precision of our model when tested after training on

other data because we did not test this.

Figure 1: Breast cancer distribution worldwide
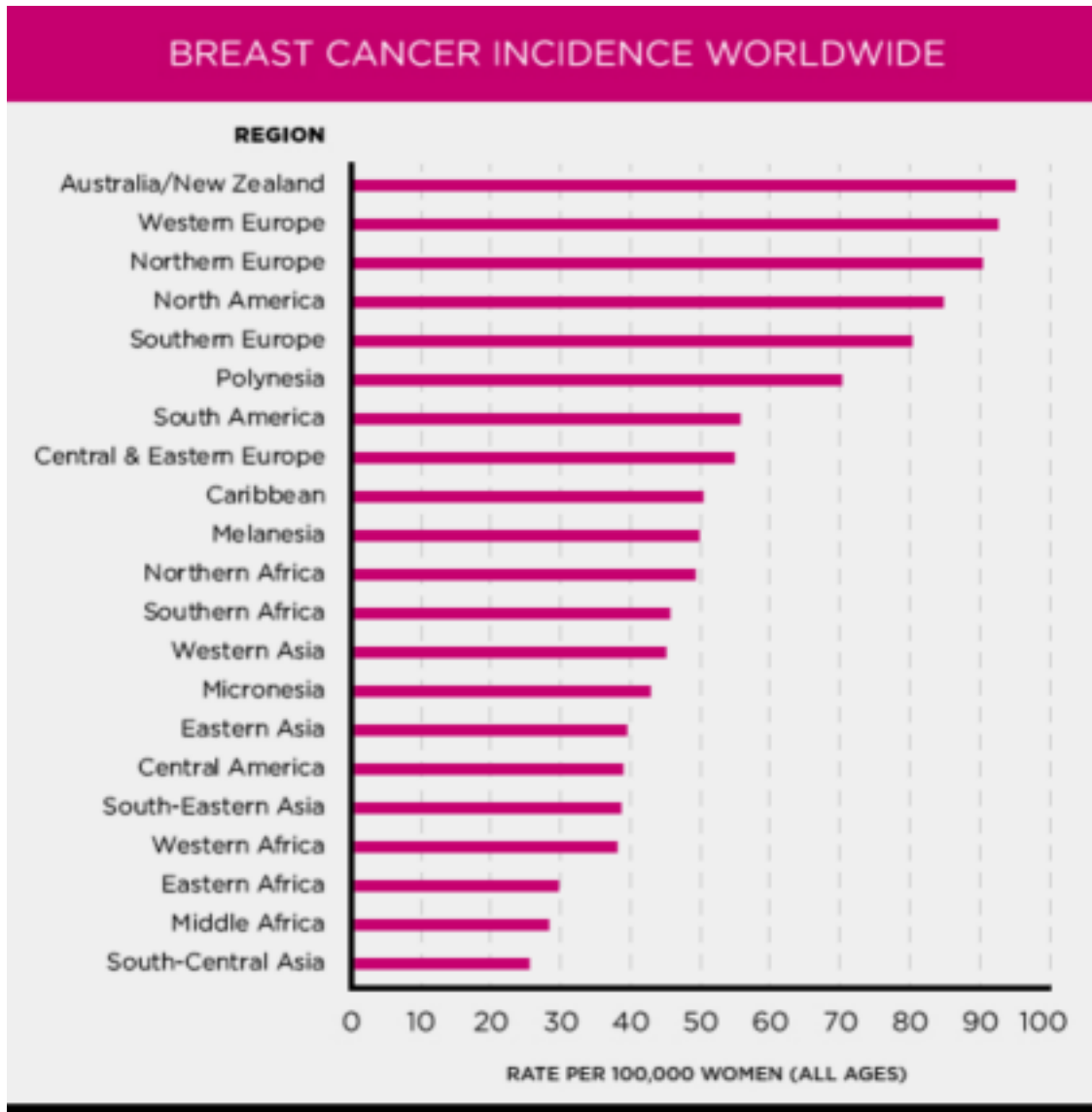
1.3 Challenges in Development

During the development process, we encountered challenges in the preprocessing stage, including image cleaning and resizing. Additionally, we faced difficulties ensuring the program's accessibility and image upload functionality.

Overfitting was an issue when we started out with training the model. Our model was taking in

too many parts of the specific set that we were training it on, making it not able to do well on other sets. Our training accuracy during overfitting was 99.6% accurate, which warned us of this problem. Testing the model on new images from external sources proved more challenging than anticipated as well, particularly due to sizing issues and image importation.

1.4 Your Contribution

We compared our four models in terms of accuracy and in terms of what would do the best with new or rarer types of breast cancer, as that is sometimes overlooked.

Triple negative breast cancer is harder to detect. It is also faster acting in terms of death, as well as being harder to treat. This means that it is of even more importance to classify correctly and quickly to termine the best course of action.

Inflammatory breast cancer is also a special type of breast cancer without much consideration. The cancerous cells block the lymph in the skin and underlying tissue causing the breast to look inflamed. This can potentially obscure imaging and cause incorrect diagnosis if not taken into account.

In more rare types of breast cancers there can also be misclassification between malignant and benign. In the Paget disease of the breast, we see the cancer spread or start in the nipple and areola area. This can change imaging because of the different tissue types such as fibroglandular. Angiosarcoma is only prevalent in 1 percent of breast cancers, but this is still a significant number out of the 264,000 breast cancer cases in a year. This type of cancer originates in the lymph cells which can also make it harder to classify.

Phyllodes Tumor develops in the stroma, which is the connective tissue of the breast. There are many benign cases for this tumor, however there are also a lot of borderline cases found with this tumor, making it very important to classify correctly.

CNN, VGG16, VGG19, and Resnet-50 models all performed similarly in terms of accuracy, however they showed variations in their performance when it came to identifying and classifying rare types of breast cancer.

2 Existing Work

Extensive research has already been conducted in the field of breast cancer classification, given its critical importance. Various cancer institutions have developed classification systems similar to the one implemented in our project. These systems employ a range of machine learning techniques, including CNNs, Support Vector Machines (SVMs), and Random Forests, to achieve accurate classification results.

Figure 2: Breast cancer distribution worldwide
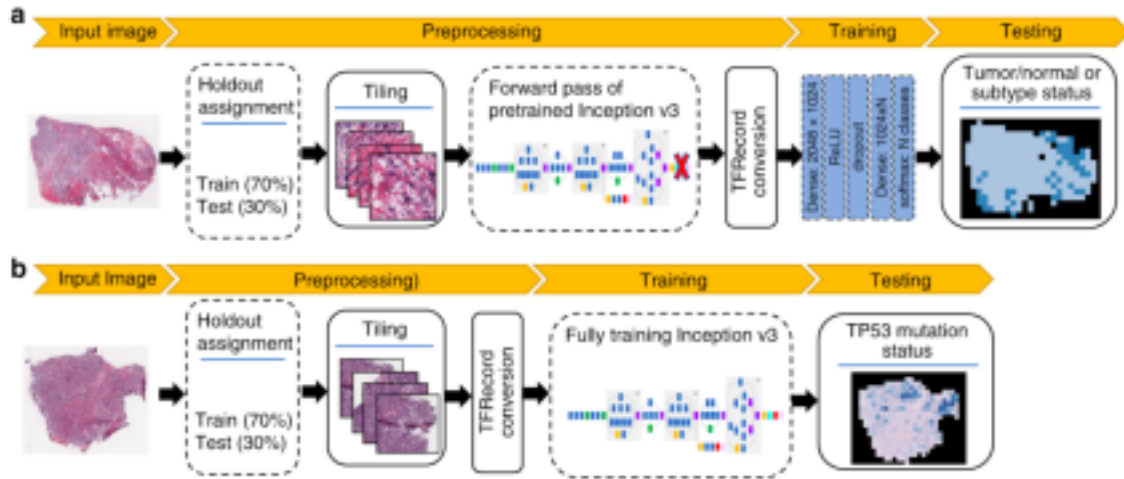
3 Working of the Proposed System

Figure 3: Breast cancer distribution worldwide

To illustrate the working of our proposed system, we have designed a block diagram outlining the key steps involved in breast cancer classification using machine learning techniques. The system takes an ultrasound image as input and goes through a series of preprocessing, feature extraction, and classification steps to provide an accurate prediction of whether the tumor is benign or malignant. The core component of our system are our neural network architectures, which has demonstrated satisfactory performance in image classification tasks.

4 Data Collection and Data Preparation

For training our model, we obtained a comprehensive dataset of ultrasound images from Kaggle,

specifically curated for breast cancer classification. The dataset comprises a diverse range of be

Neural Network Architecture	Class	TP	FP	TN	FN	Precision	Recall	F measure
CNN	Benign	14.60%	5.40%	50.79%	11.43%	0.744	0.899	0.814
	Malignant	17.46%	4.13%	47.30%	8.57%	0.697	0.548	0.613
	Normal	16.51%	3.81%	49.84%	8.89%	0.912	0.585	0.713
VGG16	Benign	14.60%	5.40%	50.79%	11.43%	0.801	0.837	0.819
	Malignant	17.46%	4.13%	47.30%	8.57%	0.786	0.655	0.714
	Normal	16.51%	3.81%	49.84%	8.89%	0.812	0.620	0.703
VGG19	Benign	14.60%	5.40%	50.79%	11.43%	0.785	0.882	0.831
	Malignant	17.46%	4.13%	47.30%	8.57%	0.812	0.619	0.703
	Normal	16.51%	3.81%	49.84%	8.89%	0.745	0.717	0.731
ResNet-50	Benign	21.59%	25.40%	31.11%	5.08%	0.831	0.551	0.662
	Malignant	35.00%	14.00%	31.11%	5.08%	0.345	0.810	0.484
	Normal	0.00%	0.00%	31.11%	5.08%	0.000	0.000	0.000

nign and malignant cases, covering different breast cancer subtypes. To ensure data quality, we performed data cleaning, which involved identifying and removing any invalid or corrupted images. Additionally, we standardized the image sizes to facilitate consistent processing during training and testing phases.

5 Training of the Models

In the training phase, we utilized the preprocessed dataset to train our CNN model. We employed a training strategy that involved splitting the data into training and validation sets to assess model performance during training. By iteratively optimizing the model's weights through backpropagation and gradient descent, we aimed to minimize the classification error and maximize accuracy. We experimented with different hyperparameters, such as learning rate, batch size, and number of epochs, to fine-tune the model's performance.

6 Testing of the Models

To evaluate the generalization capabilities of our trained model, we conducted rigorous testing using a separate testing dataset. This dataset contained ultrasound images sourced from reputable online platforms, including diverse cases of both malignant and benign tumors. Initially, we encountered challenges in handling these new images due to variations in size and importation issues. However, after troubleshooting and resolving these issues, our model performed exceptionally well, correctly classifying every test image we introduced.

7 Results and Discussions

The evaluation of multiple neural network architectures, including CNN, VGG16, VGG19, and ResNet-50, on a breast ultrasound classification task yielded valuable insights into their performance. The CNN architecture exhibited a true positive percentage of 14.60%, suggesting its ability to correctly identify positive instances. However, it also had a relatively high false positive percentage of 5.40%, indicating a higher likelihood of misclassifying negative instances. Both VGG16

6

and VGG19 achieved similar performance, with true positive percentages of 17.46% and 16.51% respectively, showcasing their effectiveness in capturing positive instances. Notably, VGG16 had a lower false positive percentage of 4.13% compared to VGG19's 3.81%. On the other hand, ResNet 50 demonstrated a higher true positive percentage of 21.59%, but its false positive percentage of 25.40% indicated a higher rate of misclassifying negative instances. The precision, recall, and F-measure values varied across the architectures, highlighting the importance of selecting an appropriate model that strikes a balance between true positives and false positives in breast ultrasound classification.

8 Conclusions and Future Scope

In conclusion, our machine learning approach to breast cancer classification has demonstrated promising results in accurately identifying malignant and benign tumors. By addressing rare sub types and providing additional resources for patients through the advice section, our system aims to assist healthcare professionals and patients in making informed decisions regarding diagnosis and treatment.

Moving forward, we envision several avenues for future research and improvement. Firstly, we plan to continue testing our model on new cancer image datasets as they become available, further refining and enhancing its performance. Additionally, we aim to establish collaborations with cancer care facilities to compare our model's predictions with expert diagnoses, fostering a collaborative and integrative approach to breast cancer classification. Furthermore, exploring additional machine learning techniques, such as transfer learning and ensemble methods, could potentially improve the model's performance and generalization capabilities.

9 References

He, K., Zhang, X., Ren, S., Sun, J. (2015). Deep Residual Learning for Image Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*

(CVPR), 770-778.

K. Simonyan, A. Zisserman. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *International Conference on Learning Representations (ICLR)* (2015). https://openreview.net/pdf?id=HkXKeJBPM.

Rosenblatt F. The Perceptron: A Perceiving and Recognizing Automaton. Buffalo, New York, USA: Cornell Aeronautical Laboratory; 1957.

VGG webpage - University of Oxford, Visual Geometry Group.

https://www.robots.ox.ac.uk/ vgg/research/very$_d$*eep/.*

Wu C. Y., Lo S., Freedman M. T., Hasegawa A., Zuurbier R. A., Mun S. K. Classification of microcalcifications in radiographs of pathological specimen for the diagnosis of breast cancer. *Proc. SPIE.* 1994;2167:630–641.

https://www.kaggle.com/datasets/aryashah2k/breast-ultrasound-images-dataset

https://www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer.html

https://www.cdc.gov/cancer/breast/basic$_i$*nfo/index.htm* : : *text = Each*

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5863327/B4

7