The Classification of Breast Masses Using Machine Learning

Ankur Raghavan, Sleiman R. Ghorayeb, PhD*

*Bioengineering, School of Engineering and Applied Science, Hofstra University

Abstract

This research project used machine learning to analyze ultrasound images of breast masses and to classify them into either Fibroadenoma (FA) or Invasive Ductal Carcinoma (IDC).

Engineering Goals

- To develop a machine learning algorithm that will classify breast ultrasound images with an accuracy greater than 90%
- To develop an Ultraportable system with an App that allows for the easy use of this algorithm by health care professionals.

Introduction

Classification of breast has been a longstanding area of interest in the field of oncology and medical imaging. Traditionally, pathologists have relied on histopathological examination of tissue samples obtained through biopsies to classify breast tumors into distinct categories. This process involves examining cellular and tissue characteristics, such as cell morphology, organization, and the presence of invasive features, to differentiate between malignant and benign tumors. The traditional approach, while effective, is labor-intensive, time-consuming, and subject to inter-observer variability based on the expertise of the pathologist.

In addition to histopathological analysis, medical imaging has played a pivotal role in aiding tumor classification. Mammography, ultrasound, and magnetic resonance imaging (MRI) are commonly used imaging modalities in breast tumor assessment. These imaging techniques provide valuable information regarding tumor size, shape, margins, and other features that contribute to the classification process. However, human interpretation of these images is subjective and can vary based on the experience and expertise of the radiologist.

In recent years, there has been a paradigm shift toward integrating machine learning (ML) into the classification process. ML algorithms can analyze vast amounts of imaging and clinical data to discern patterns and features that may not be easily identifiable through traditional methods. This has led to the development of automated classification systems that can distinguish between invasive ductal carcinoma (IDC) and fibroadenoma (FA) with a high degree of accuracy. By harnessing the power of ML, we aim to enhance the efficiency and precision of breast tumor classification, ultimately improving patient outcomes.

Methodology

Setup:

- Downloaded Jupyter Notebook
- Downloaded TensorFlow
- Downloaded Dataset, provided by Ultrasonic Research Lab

Data Pre-Processing:

- Split into Training and Testing Datasets (70%/30%)
- Converted grayscale images to 2D arrays of doubles

Neural Network:

- Created and trained the Sequential Machine Learning Model
- Tuned parameters
  - Number and size of layers
  - Optimizer
- Converted to a .tflite file for the iOS App

Testing:

- Tested the model by using the testing dataset.
- Generated a graph showing the accuracy increase as the model trained

App Development:

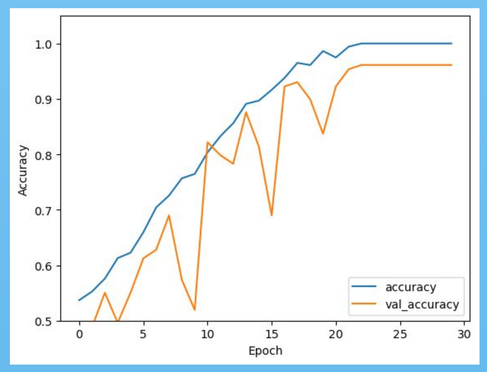
- Downloaded the tensorflow script and model onto XCode
- Added buttons to upload a photo.
- Implemented the tensorflow model which classifies the uploaded images
- Added functionality to zoom in on images once clicked.

Results

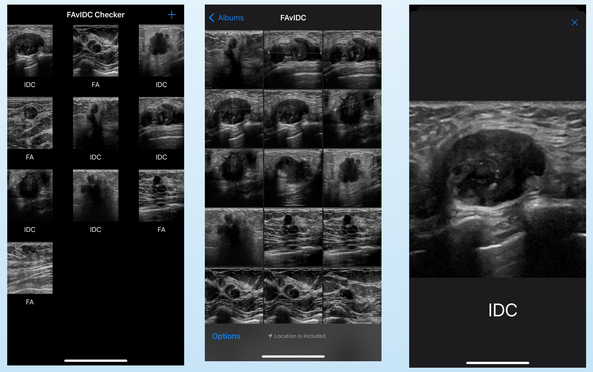
All of the Engineering Goals were met.

The model achieved a maximum accuracy of 96%.

An iOS App was created which implements the machine learning algorithms.



Discussion



The developed machine learning model exhibited a remarkable classification accuracy of 96% in distinguishing between invasive ductal carcinoma (IDC) and fibroadenoma (FA). This high accuracy demonstrates the model's robustness and effectiveness in automated breast tumor classification. To further enhance its utility and accessibility for healthcare professionals, an intuitive and user-friendly application was designed. This application provides a streamlined interface, granting healthcare professionals easy access to the classification model. Through this application, medical practitioners can conveniently upload patient data, such as medical images and relevant clinical information, and receive real-time classification results. This technological solution not only optimizes the diagnostic process but also facilitates timely decision-making, ultimately contributing to improved point-of-care, patient care, and outcomes.

Conclusions / Significance

In summary, this research represents a significant breakthrough in the field of breast tumor identification, specifically in distinguishing between invasive ductal carcinoma (IDC) and fibroadenoma (FA). The machine learning model achieved an impressive accuracy rate of 96%, showcasing the potential of artificial intelligence to enhance diagnostic abilities in medicine. By integrating this model into a user-friendly application, healthcare professionals may now have a fast and reliable tool for tumor classification. This not only simplifies the diagnostic process but also optimizes treatment decisions, ultimately leading to improved patient outcomes. The successful application of this technology emphasizes the importance of innovative solutions in addressing critical medical challenges and points to ongoing advancements at the intersection of machine learning and healthcare.

Citations

Ghorayeb SR, et. al, “A Potential Artificial Intelligence Protocol for Texture Analysis of Breast Masses in Ultrasound Images,” 2020 BMES Annual Meeting, San Diego, CA, Oct 14–17, 2020.

https://www.cancer.gov/types/breast

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6754844/

https://www.kaggle.com/

https://www.tensorflow.org/lite/guide/ios