ClarifAI - Detecting Breast Cancer with AI

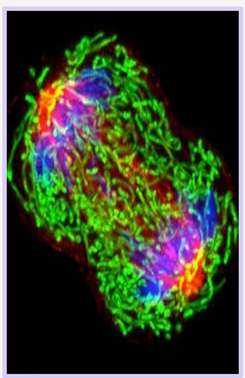
Allison Everett, Arianna Ruiz, Jordan Ghee

Research Question

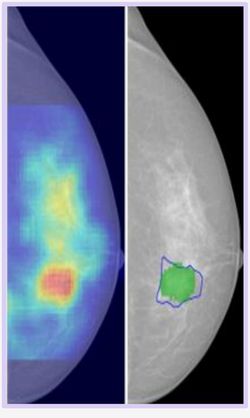
How can we develop a **machine learning** model to accurately **detect breast cancer** on scans?

Introduction

Cancer is a disease caused by the **uncontrolled growth of abnormal cells** that ignore signals to stop multiplying. These cells can spread and form tumors, disrupting healthy tissue functions. **Breast cancer**, is one of the most common forms of cancer, affecting many (Cleveland Clinic).



Advantages of AI



- AI can draw conclusions from **thousands** of images
- AI can pick up **subtle** signs of cancer that the human eye misses
- Radiologists can use AI as a **tool** to determine if more testing is required

References

Reach out to the researchers:

aeveret1@terpmail.umd.edu

amruiz@terpmail.umd.edu

Methods and Materials

```
# Setup
data_dir = "/kaggle/input/breast-cancer-screening-dbt/Breast-Cancer-Screening-DBT"  # Replace with your actual dataset path
img_height = 224
img_width = 224
batch_size = 32
seed = 42
```

Data Setup: Our model uses images from the Cancer Imaging Archive, which contains 5,060 mammograms which are separated into four categories: normal, actionable, biopsy-proven benign, and biopsy-proven cancer (TCIA).The code above shows the connection to the dataset and its categories.

```
weights_path = "/kaggle/input/efficientnetb0/efficientnetb0_notop.h5"

base_model = EfficientNetB0(include_top=False, input_shape=(img_height, img_width, 3), weights=weights_path)
base_model.trainable = False  # Freeze base model
```

Base Model: We chose EfficientNet-B0 as our base model because it has been used in a medical context in other studies, and it is well equipped to classify images (Keras). For the first training phase, the layers of the base model were frozen in order to reduce the risk of overfitting. Later on, some layers were unfrozen for fine-tuning.

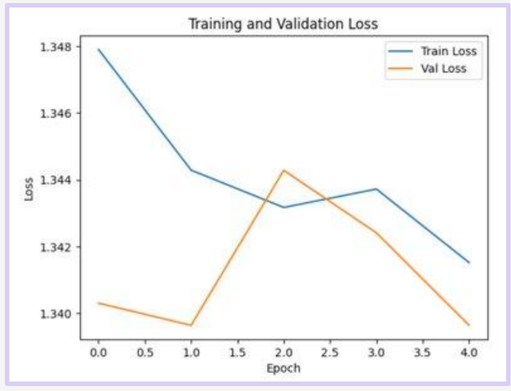
```
weights_path = "/kaggle/input/efficientnetb0/efficientnetb0_notop.h5"

base_model = EfficientNetB0(include_top=False, input_shape=(img_height, img_width, 3), weights=weights_path)
base_model.trainable = False  # Freeze base model
```

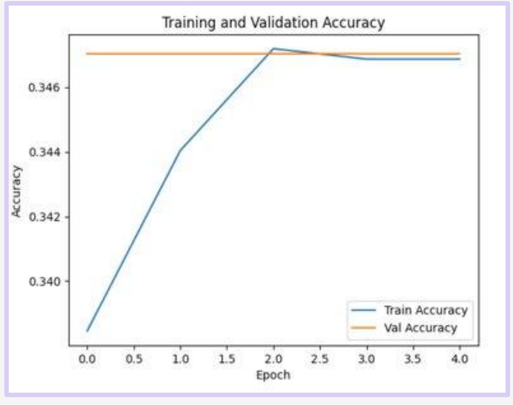
Training: As we trained our model, we noticed certain issues that we addressed in order to improve the accuracy. Some changes we made along the way include, data augmentation, sparse categorical cross entropy, unfreezing layers, and the code shown above allowed us to save our progress without restarting after each improvement.

Literature

Early detection of cancer is **critical** for improving survival rates. For example, breast cancer detected at stages 1 and 2, has a 5-year survival rate of 99%, however survival rates drop to 31% as breast cancer spreads in stages 3 and 4. The use of AI will allow for **earlier detection**, with the goal of promoting higher survival rates. Scientists are concerned about **overfitting** in AI models, where they perform well on training data but struggle with new data. Additionally, medical images must be **processed** before use in AI, and there are ongoing concerns about the **security of patient data**.

Results





- Initially, the model's accuracy was around **25%,** indicating that the predictions were random
- After refining, the accuracy increased to **37%**
- After each interaction, the model's predictions became more accurate (see Training and Validation Loss)

These results imply that our AI model can correctly detect cancer on mammograms roughly 37% of the time. This is not ideal for medical use at the moment, considering the severity of the model's use. With a larger dataset and an increase in iterations, the model's accuracy is expected to increase.