DermaVision

Team Salus

Table of Contents

Team Members

Cindy Hua

3rd Year, Toronto Metropolitan University

Nina Dang

3rd Year, York University

Mikayla Morrison

3rd Year, Toronto Metropolitan University

Mentors

Guy Morgenshtern

RBC Borealis, Machine Learning Engineer

Atishya Reddy

RBC, Software Development Engineer

The Problem

4th Leading Cause of Non Fatal Disease WorldWide

Struggle to address complexity of managing non-fatal conditions

Shortage of Licensed Dermatologists

700 Dermatologists, 38 million Canadians, overwhelming demand

Long Waitlists

Up to 1 year, Geographic distances become barriers for remote communities

Motivations

Accurate Classification

Utilize Convolutional Neural Network and Multilayer Perceptron to analyze both metadata and images.

Identify subtle patterns missed by non-specialists.

Identify diagnoses more accurately across diverse range of skin tones.

Address Shortage

Alleviate stress on medical professionals.

Improve prognosis and increase the likelihood of early intervention.

Make diagnosis more accessible to remote communities.

Project Goal

To improve access to dermatological care by developing, training, and validating a robust machine learning model that accurately diagnoses dermatological conditions using images and easily accessible patient metadata.

Focus: Malignant and Benign Skin Lesions.

Tech Stack

Python

Main programming language

Keras

Python interface for neural networks

TensorFlow

Python interface for ML models

SciKit Learn

Python library for ML tasks

OpenCV

Python library for computer vision

Other Libraries

Various Data Processing Tasks: Numpy, MatplotLib, Pandas, SciPy, Seaborn

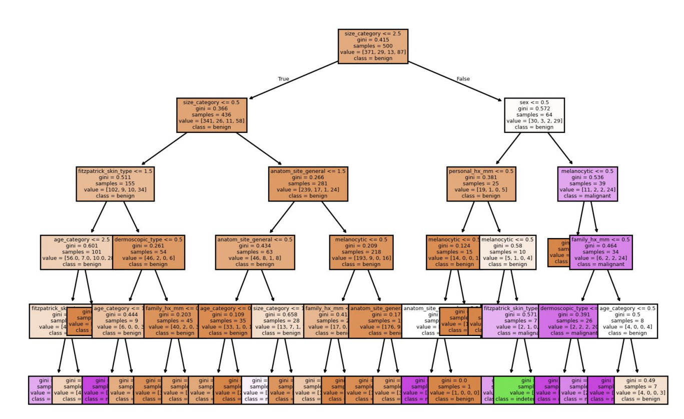
ISIC Database

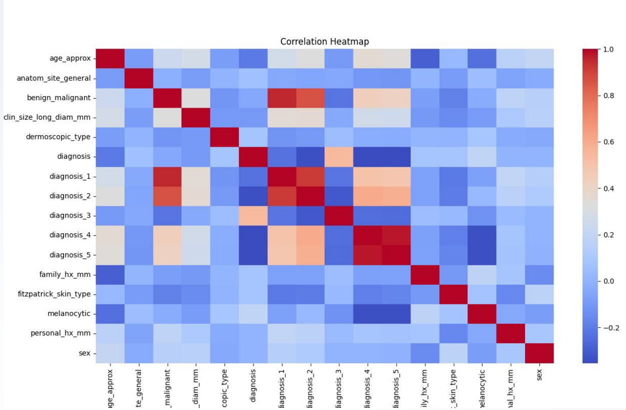
- International Skin Imaging Collaboration (ISIC)
- 503,955 Images and Metadata Records
- 30 Attributes, including:
    - Age, Sex, Anatomy Site, Size, Diagnoses, Family History, Personal History, Fitzpatrick Skin Type
- **600 metadata records required for training**
- **626 instances viable for training**
- **5 relevant features selected**

Metadata Preparation

Data Transformation

- Drop irrelevant columns
- Drop missing ‘Fitzpatrick Skin Type’
- Replace missing ‘Dermoscopic Type’
- Classify diagnoses

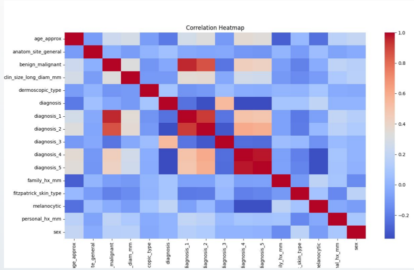


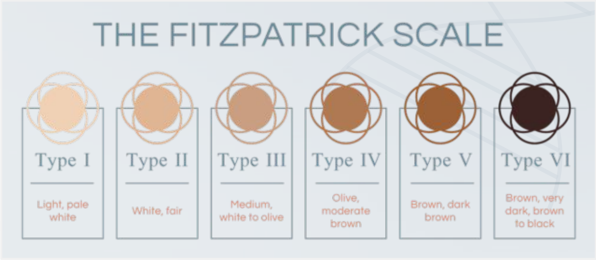


Features Selected

- 5 relevant features selected
  - 'anatom_site_general'
  - 'fitzpatrick_skin_type'
  - 'sex'
  - 'age_approx'
  - 'clin_size_long_diam_mm'

Challenges





Lack of Correlation

- Minimal strong positive correlations
- More complex model required for accurate diagnosis based on metadata

Lack of Diversity

- Fitzpatrick skin type recorded for 626 of 503,000+ records
- 6 categories total, 4 categories in dataset
  - Minimal coverage of medium tones
  - No coverage of deeper tones

"Under-representation in datasets results in lower performance in deep learning models for underrepresented skin types...This issue can potentially lead to the exclusion of certain groups of people by AI-based models." - Skin Type Diversity in Skin Lesion Datasets: A Review

Image Preparation

Image Resizing

- Resize all images to 244 x 244 pixels using the equation:

$$I_{resize} = resize(I_{original}, (W_{new}, H_{new}))$$

Image Segmentation

- Attempted both traditional thresholding and texture-based segmentation using Gabor kernels

Image Normalization

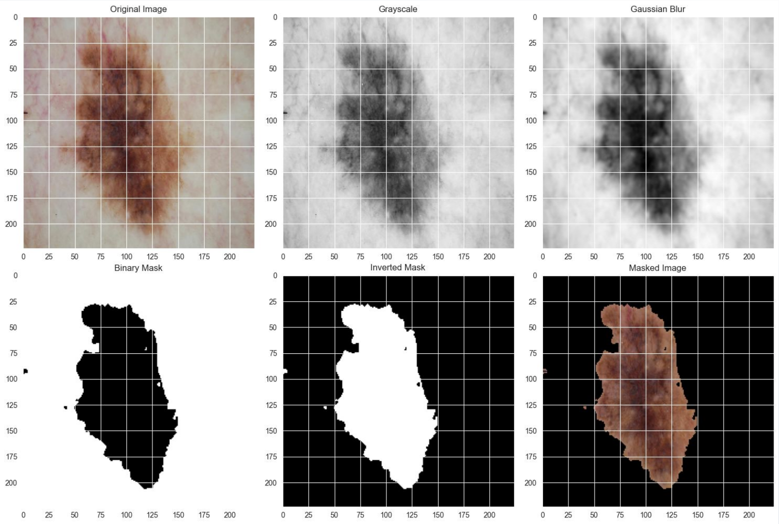
- Converted all pixel values to have a range of -1 to 1 using the equation:

$$I_{normalized}(x, y) = \frac{I_{original}(x, y)}{127.5} - 1$$

Image Augmentation

- Flipped and rotated images to give more diversity and more data for the neural network.

Image Segmentation - Process



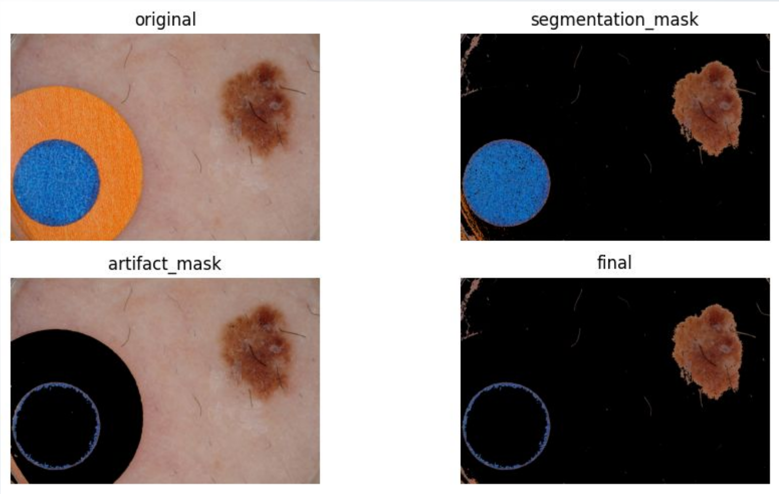
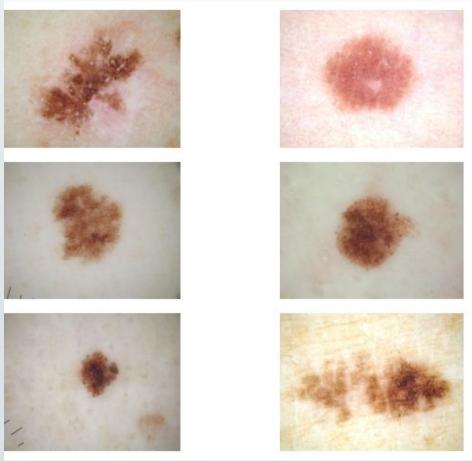
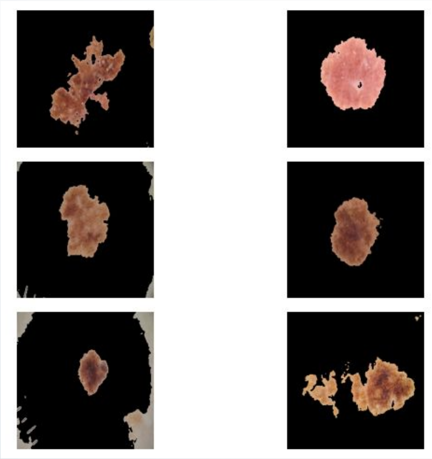


Image Segmentation - Results

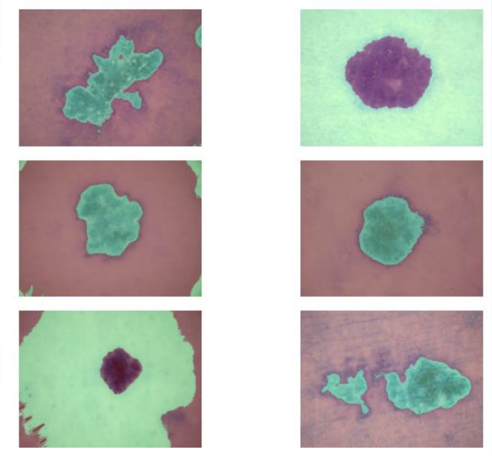
Original images



Using Otsu Thresholding

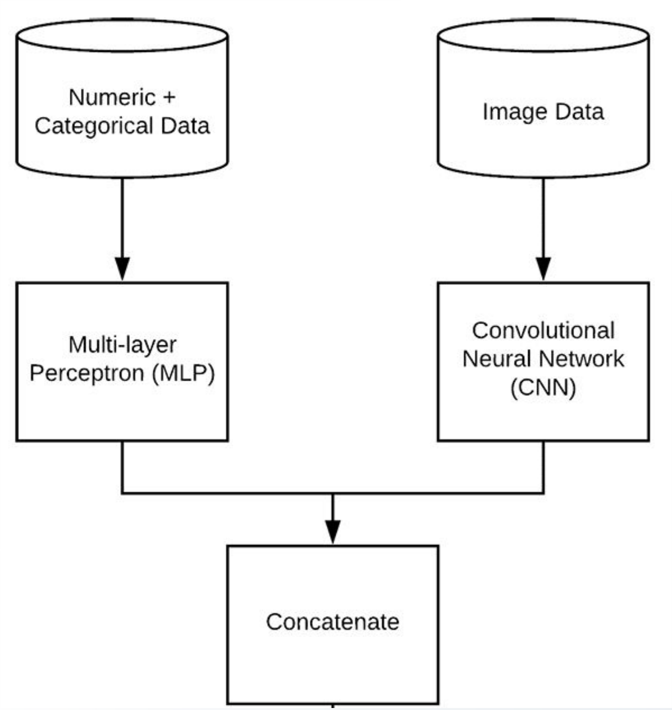


Using Gabor Kernels



Mixed-Input Model

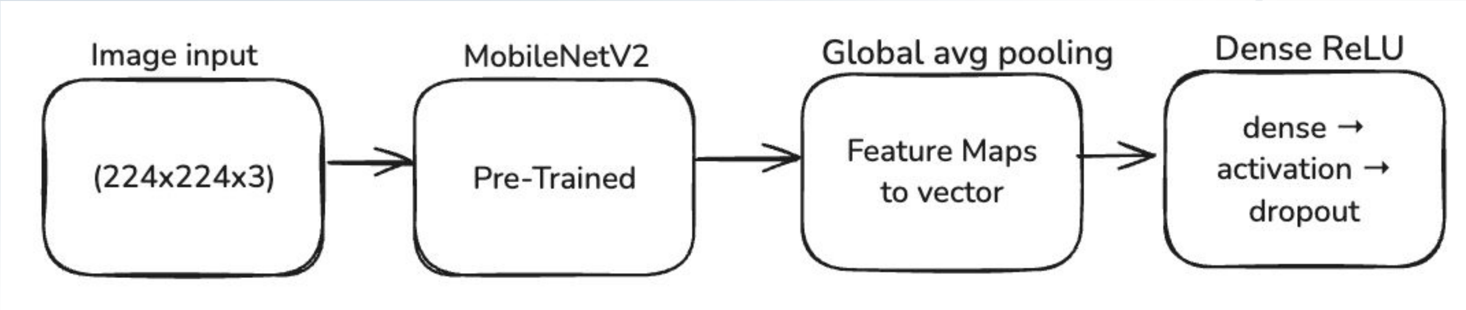
- Considers both the image data and patient metadata
- 2 branches processes different data types separately
- Concatenates features before final classification



Convolution Neural Network (CNN) - image branch

- CNN: neural network designed for image analysis
- Extracts features through convolutional filters

- MobileNetV2 pre-trained on ImageNet as base model
- Global average pooling reduces spatial dimensions
- Dense layer with dropout to prevent overfitting

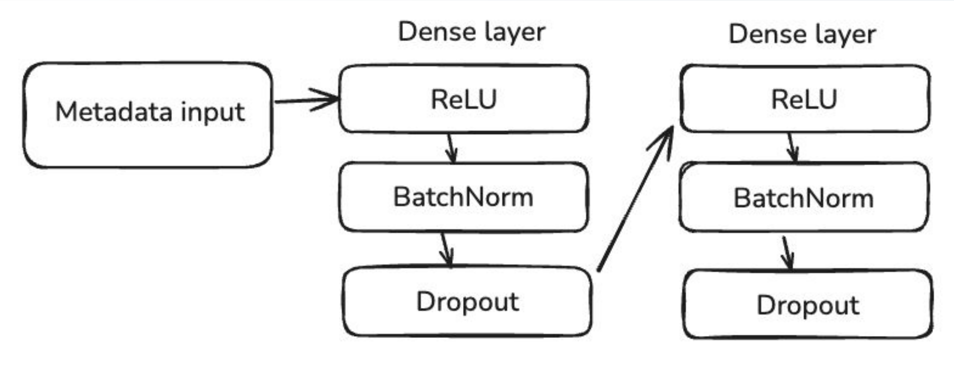


Multilayer Perceptron (MLP) - metadata branch

- MLP: Feedforward fully connected neural network for learning relationships between features
- Two-layer network with batch normalization and ReLU activation

Using features selected:

- anatom_site_general
- fitzpatrick_skin_type
- sex
- age_approx
- clin_size_long_diam_mm (max diameter of the lesion (mm))

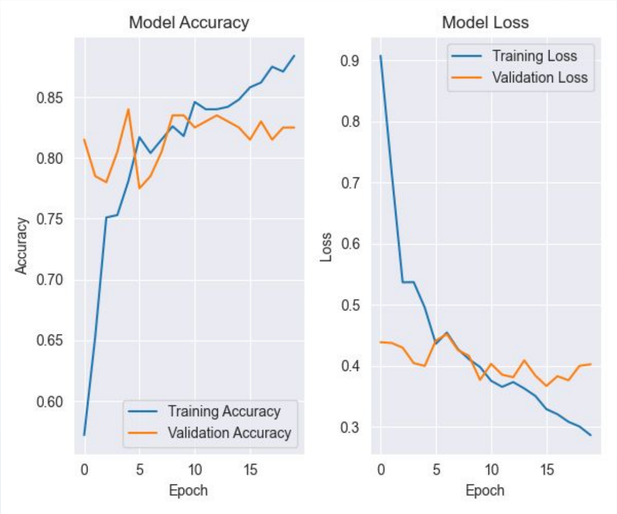


Metrics (20 epochs)

No image segmentation, max val_accuracy


```
Test Loss: 0.3859
Test Accuracy: 0.8500
```

image segmentation, max recall



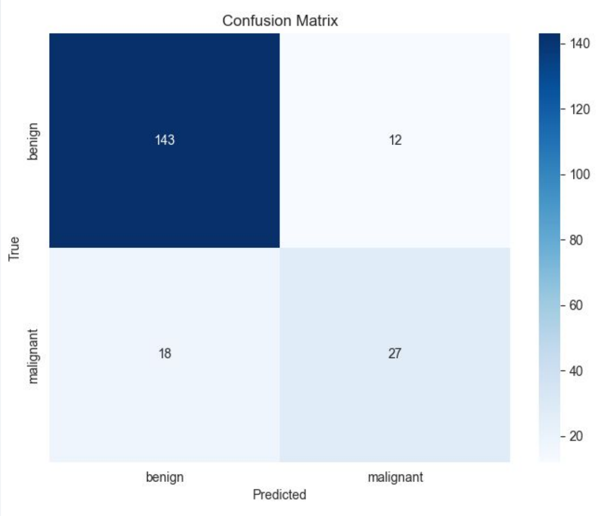

```
Test Loss: 0.3771
Test Accuracy: 0.8350
```

image segmentation, max val_accuracy


```
Test Loss: 0.3731
Test Accuracy: 0.8400
```

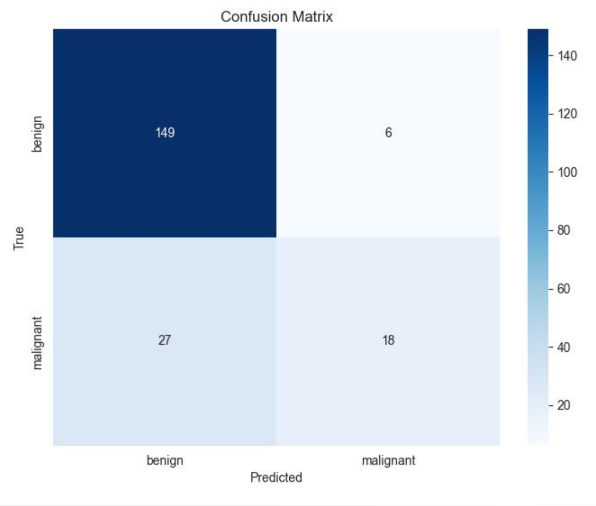
Metrics

No image segmentation, max val_accuracy



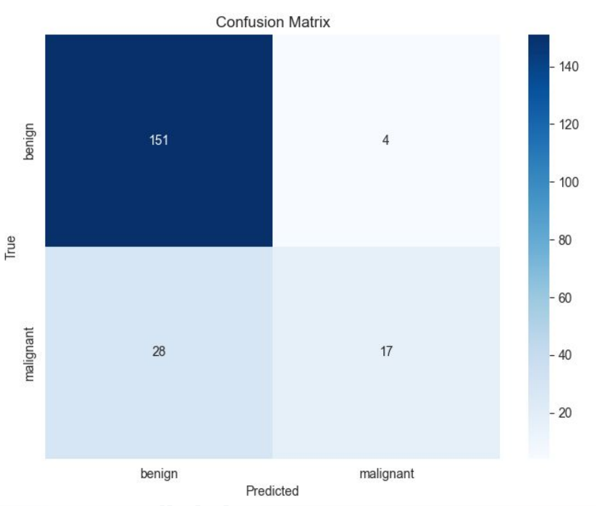
	precision	recall	f1-score
benign	0.89	0.92	0.91
malignant	0.69	0.60	0.64
accuracy			0.85

image segmentation, max recall



	precision	recall	f1-score
benign	0.85	0.96	0.90
malignant	0.75	0.40	0.52
accuracy			0.83

image segmentation, max val_accuracy



	precision	recall	f1-score
benign	0.84	0.97	0.90
malignant	0.81	0.38	0.52
accuracy			0.84

Next Steps

Refine and Extend Cancer Detection Model

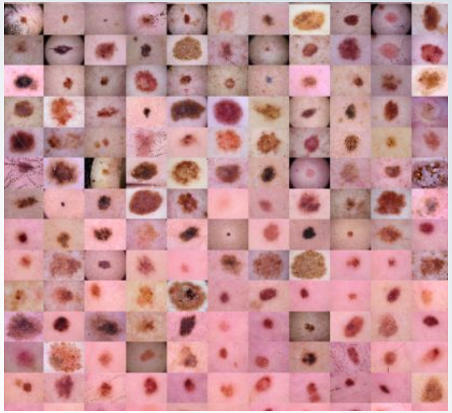
- Improve f1-score, precision, and recall metrics
- Identify specific types of melanoma and benign conditions

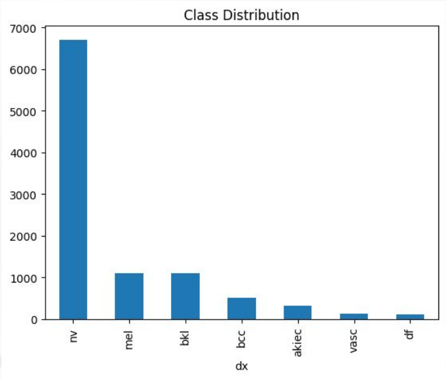
Increase Transparency

- From an AI governance perspective, we want to increase transparency before deployment
- Doctors, patients, and companies will want to know how the model predicts before release

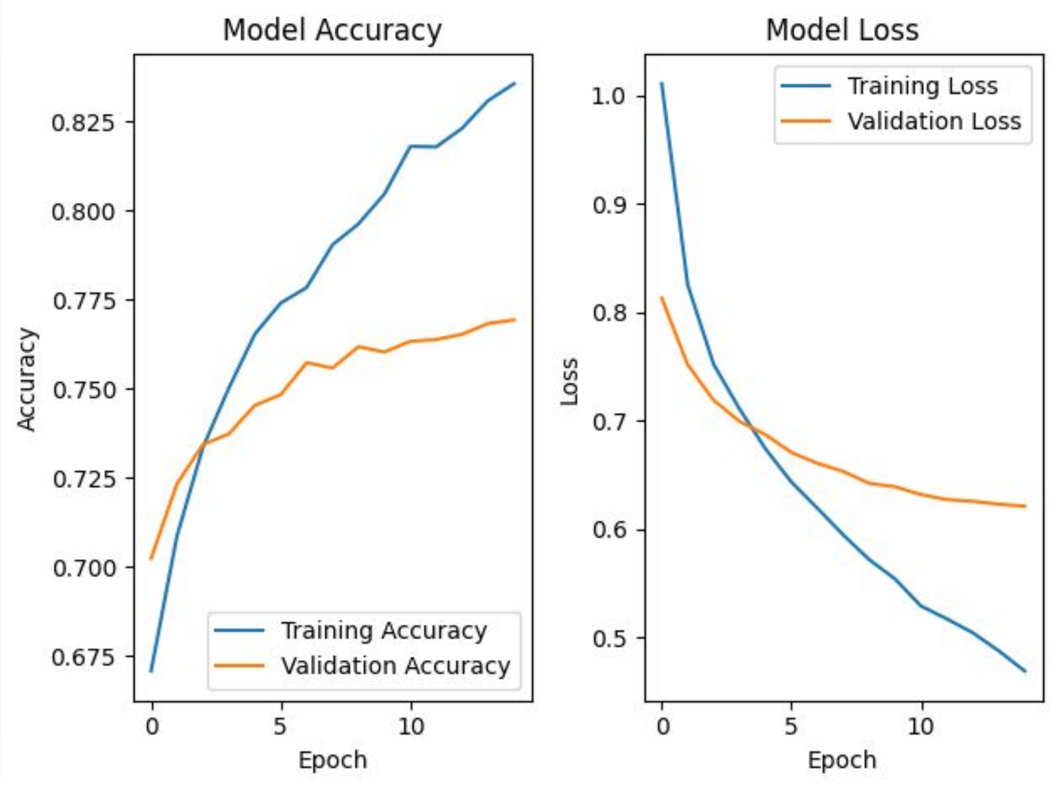
HAM10000 Dataset

- ISIC 2018 Challenge, HAM10000 Dataset
- International Skin Imaging Collaboration (ISIC)
- 10,015 Images and Metadata Records
- Used **only** for image branch





Results for CNN Branch

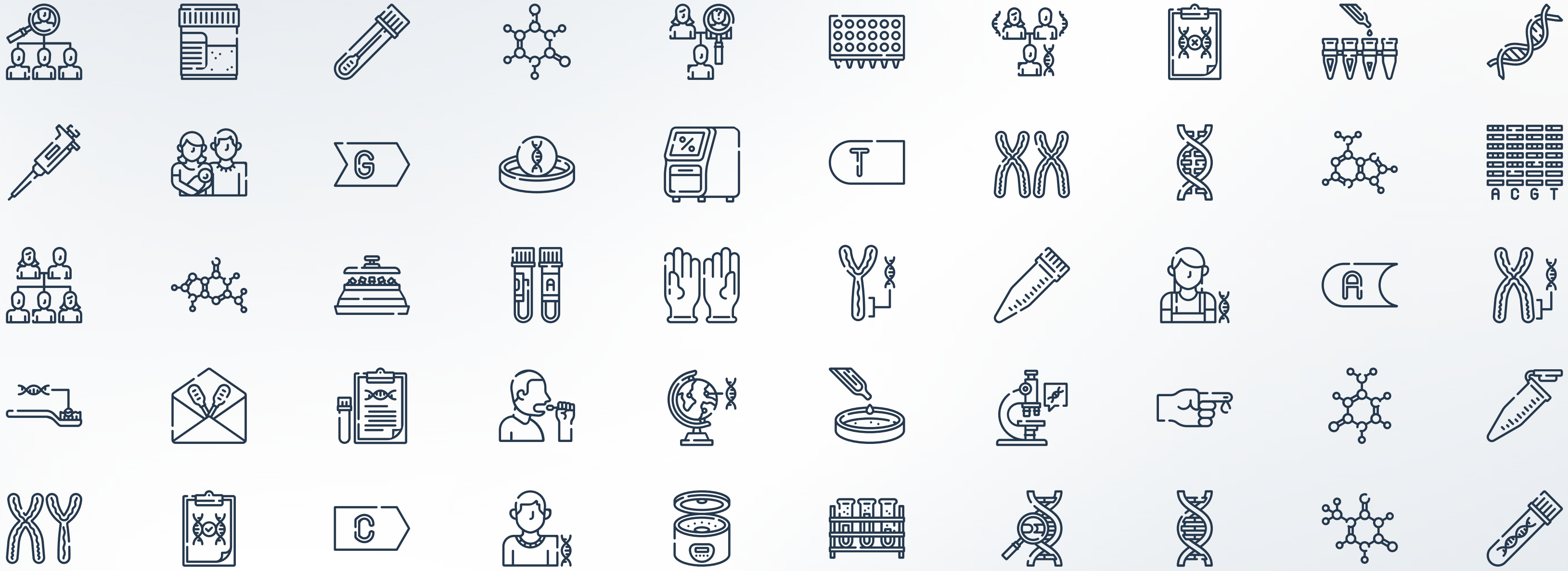


- For training the model to classify 7 types of skin lesions, we had an accuracy of 77.2%
- We aim to continue to investigates results with multiple images as well.

Thank You

GitHub Repo:

Icon pack



...and our sets of editable icons

You can **resize** these icons without losing quality.
You can **change the stroke and fill color**; just select the icon and click on the **paint bucket/pen**.
In Google Slides, you can also use **Flaticon’s extension**, allowing you to customize and add even more icons.

Educational Icons

Medical Icons

Business Icons

Teamwork Icons

Help & Support Icons

Avatar Icons

Creative Process Icons

Performing Arts Icons

Nature Icons

SEO & Marketing Icons