# Skin Cancer Detection using Deep Learning


Kaggle Challenge: ISIC 2024 - Skin Cancer Detection with 3D-TBP

Team J: Mario El Shaer, Omar Shehata, Khaled Saleh, Aymane Aaquil

COMP433.
December 8th 2024

# Problem Statement and Project Goals

### The Problem

- Skin cancer and melanoma particularly is life-threatening
- Manual examination can be very timely and not cost-effective.
  - Also prone to Human error

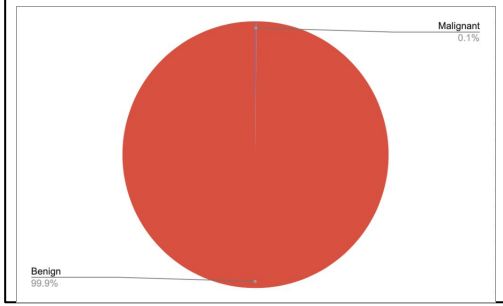
### What we want to do

- Develop a scalable and accurate skin cancer detection model.
- Minimize false negatives to prevent missed diagnoses

# Data Overview

**Data breakdown:**

400,666 benign cases Vs 393 Malignant




**Data included:**

1. JPG images
2. A CSV file with metadata including
    a. Age
    b. Sex
    c. Legion size
    d. Location

And much much more

# Methodology Overview

**Approach:**

- Data Preprocessing: Balancing, cleaning, feature selection.
- CNN Models: ResNet50 & EfficientNetB0.
- Multimodal Integration: Combining image and tabular data.
- Model Evaluation: Metrics such as precision, recall, F1-score.

# Data Processing & Feature Selection

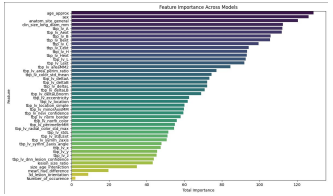
## Data Processing

- Undersampling benign cases
- Data Augmentation for malignant cases

**Purpose:** Handle class imbalance

## Feature selection

- LightGBM, XGBoost, CatBoost models to select top 20 features from the csv file.




- ResNet50 extracted the 2048 features from the images

# Model Design and Training

**Model Design:**

- ResNet50 & EfficientNetB0 were used for image feature extraction.
  - Final choice was ResNet50 as it had better results
- Fully connected neural network for the multimodal learning

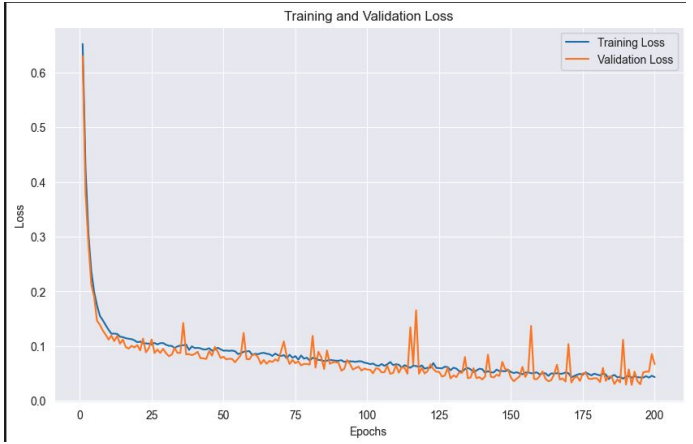
**Training Details:**

- Learning Rate: 0.00001
- Batch Size: 128
- Epochs: 200

Weighted loss function was used to prioritize malignant cases

# Result Summary

| Model | LR | BS | Epoch | Test Recall |
|---|---|---|---|---|
| Model 1 Itr.1 | 0.001 | 64 | 100 | 0.32 |
| Model 1 Itr.2 | 0.001 | 32 | 50 | 0.25 |
| Model 1 Itr.3 | 0.001 | 64 | 25 | 0.33 |
| Model 2 Itr.1 | 0.00001 | 128 | 100 | 0.46 |
| Model 2 Itr.2 | 0.00001 | 128 | 200 | 0.54 |
| Model 2 Itr.3 | 0.00001 | 64 | 200 | 0.24 |
| Model 3 Itr.1 | 0.00001 | 64 | 200 | 0.25 |
| Model 3 Itr.2 | 0.00001 | 128 | 200 | 0.54 |




| Class | Precision | Recall | F1-Score | Support |
|---|---|---|---|---|
| Negative | 0.98 | 0.96 | 0.97 | 1600 |
| Positive | 0.34 | 0.54 | 0.42 | 63 |
| Accuracy | | 0.94 | | 1663 |
| Macro Avg | 0.66 | 0.75 | 0.70 | 1663 |
| Weighted Avg | 0.96 | 0.94 | 0.95 | 1663 |

# Comparative Analysis

**Key improvements:**

- Recall: From under 40% to 54%

- This reduced the false negatives which us very crucial in medical diagnosis

Tradeoff: Precision of 0.34 was acceptable to prioritize recall

# Challenges & Limitations

**Challenges:**

- Severe class imbalance
  - < 1% Malignant cases
- Limited computing resources

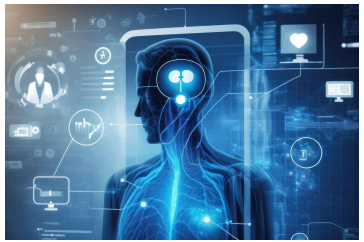
**Limitations:**

- Moderate precision for malignant cases
- Limited ability to train on larger dataset

# Conclusion & Future Work

**Conclusion:**

- Achieved 94% accuracy and 54% recall using multimodal learning
- Improved out previous models through weighted loss and feature selection



**Future Directions:**

- Ensemble models and synthetic data generation
- Exploring advanced attention-based architecture