Kaggle Competition: ISIC 2024

Skin Cancer Detection with 3D-TBP

Matthew Huang, Justin Koida, and Joenna Yaya

Context

Skin Cancer

- 3 Major Types:
    1. Basal Cell Carcinoma
    2. Squamous Cell Carcinoma,
    3. Melanoma (1)
- Caused by UV rays from the sun/tanning beds (1)
- **Skin Lesions** are areas of skin that look different from surrounding skin. They are either **benign** (non-cancerous) or **malignant** (cancerous). (2)

AI and Skin Cancer Diagnosis

- Kaggle competition via ISIC: International Skin Imaging Collaboration
- What data was provided to us?
  - .jpeg Images
  - .csv Sheets
    - target: 0 for benign, 1 malignant
    - meta-info on lesion size, symmetry, color data, body location, etc.
- **Goal**: Classify skin lesions as benign or malignant.

Our Project

- Design simple CNNs, train and evaluate them with the competition data
- Analyze competitors' submissions
- Compare different models and methods

The Data

The SLICE-3d Dataset

Over 400,000 images of skin lesions extracted from 3D Total Body Picture (TBP).

Contributed by members of the International Skin Imaging Collaboration (ISIC).

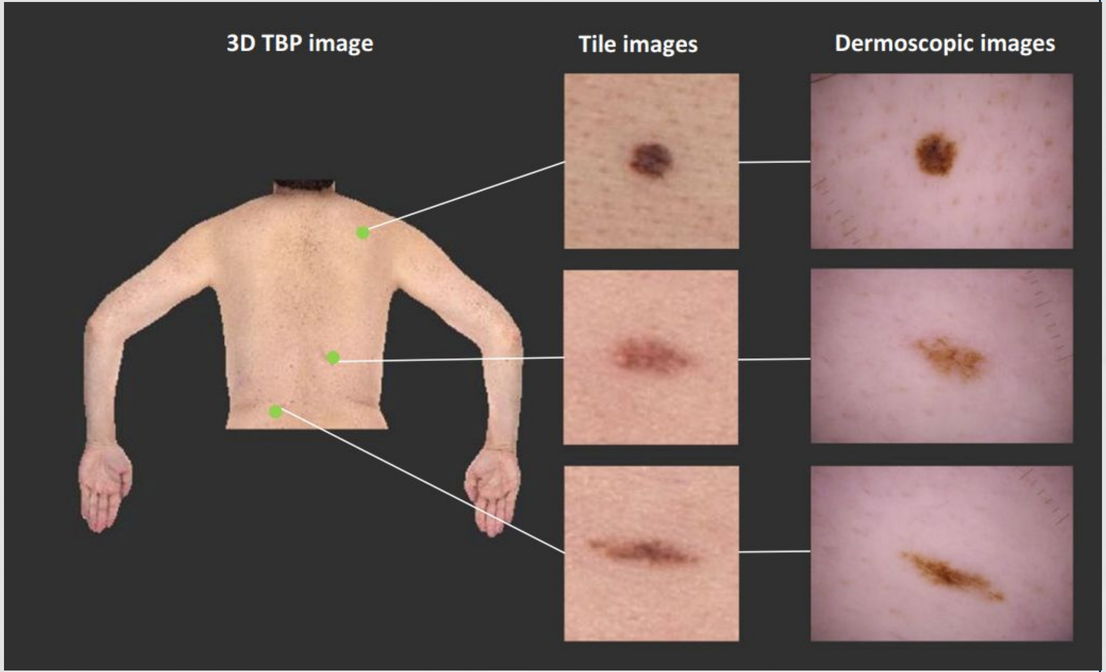
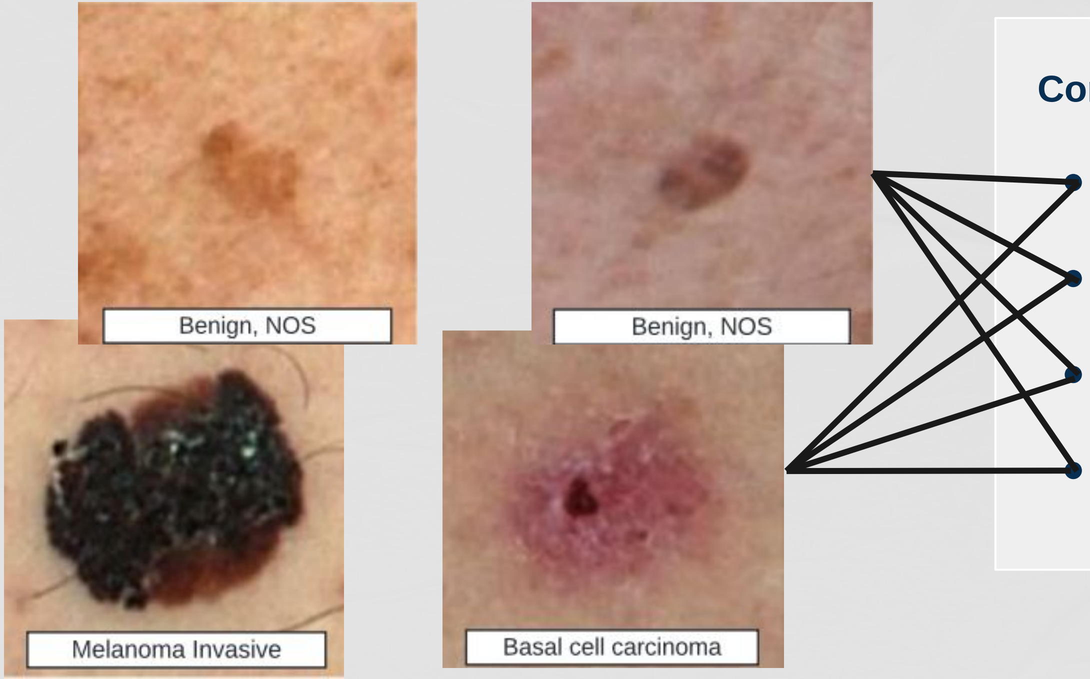


Fig. 1

Image Data

Meta-Data



Corresponding Features:

- Patient age
- Lesion area
- Body Location
- Etc.

Fig. 2a-c

Pre-processing for Meta-Data

- Filling Missing Values
- Dropping detrimental features
- Normalizing Quantitative Data
- One Hot Encoding Categorical Data

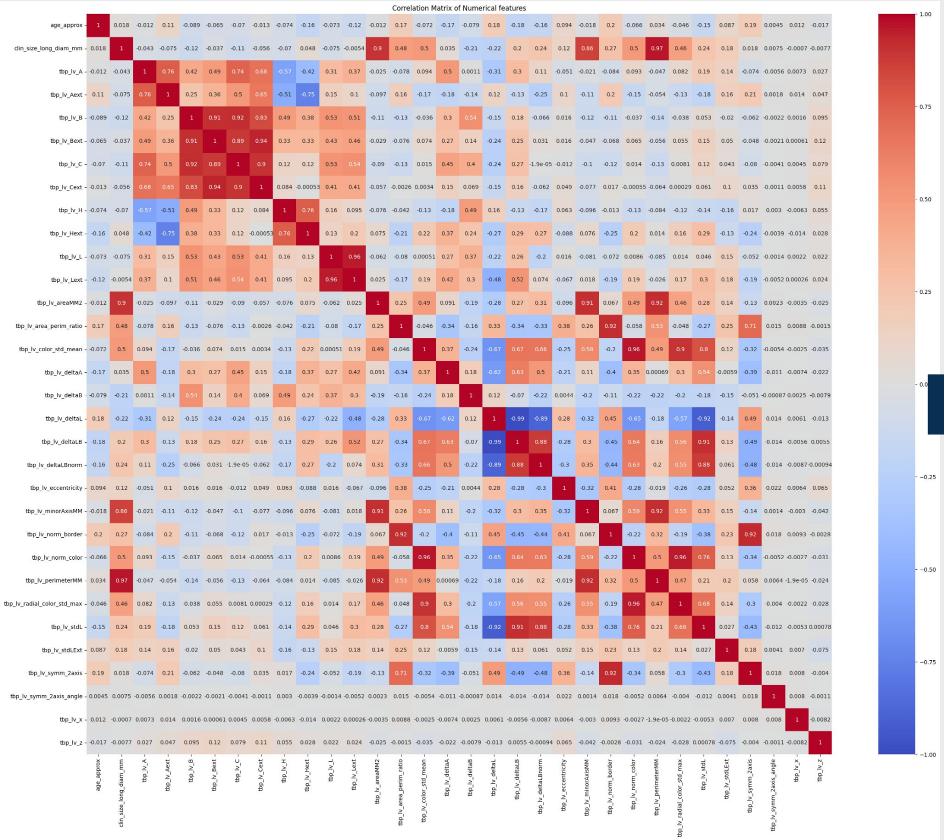


Fig. 3a

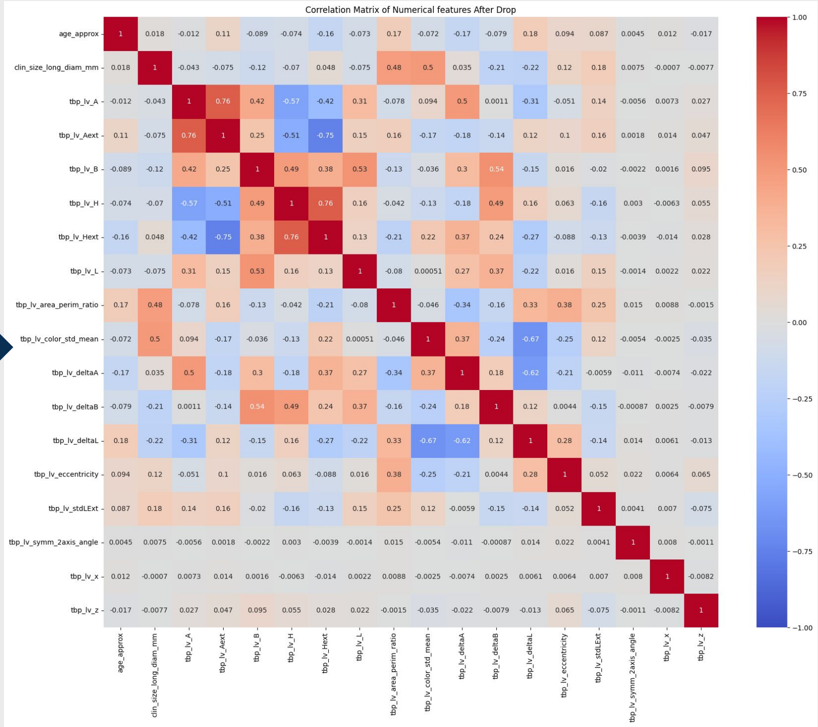


Fig. 3b

Reducing Collinearity

Predicting from Meta-Data

Logistic Regression

- Created Logistic Regression class in PyTorch
- Experimented with different loss functions
- Implemented class balancing techniques

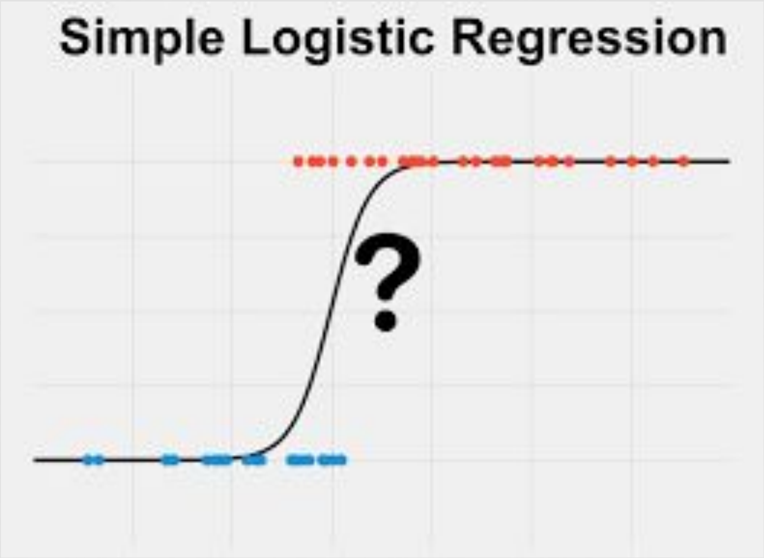


Fig. 4

Random Forest Classifier and XGBoost

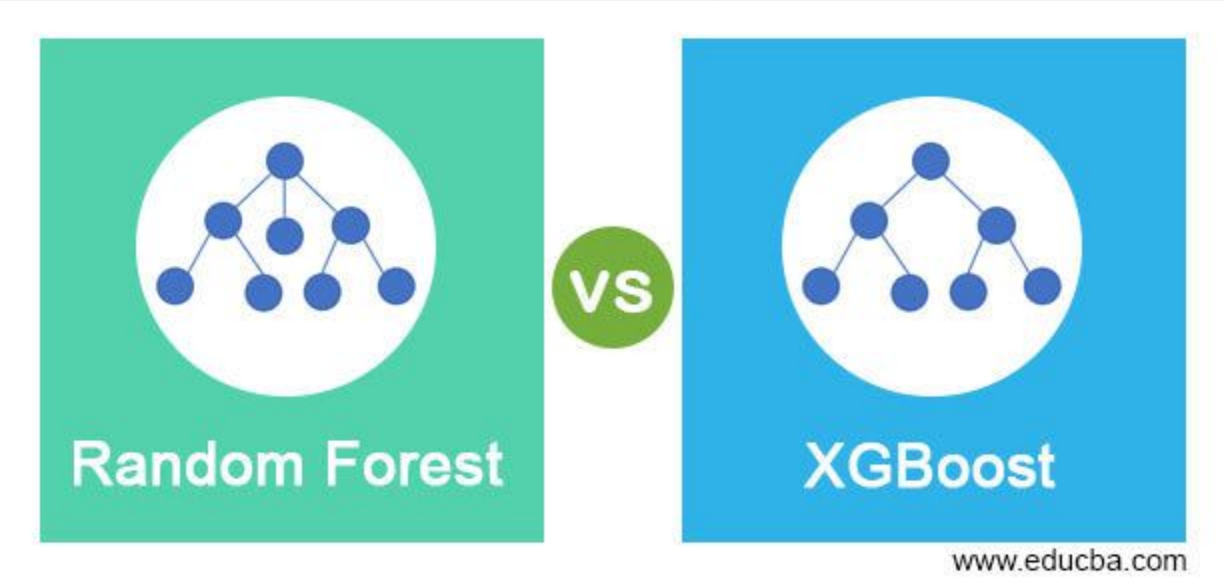


Fig. 5

Predicting from Image Data

SimpleCNN()

Image → Convolution → Pool → Convolution → Pool → FC → (**Benign**, Malignant)

128x128 2x1

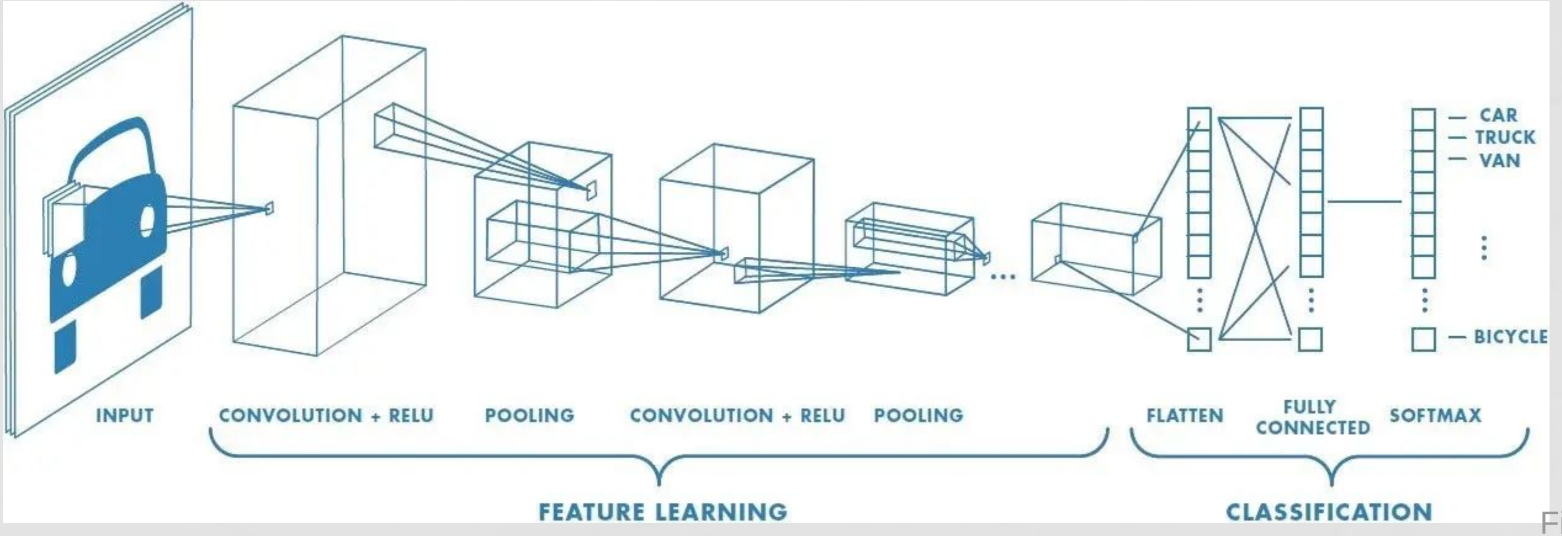


Fig. 6

BiggerCNN()

12 layers, 17 million parameters

Additions

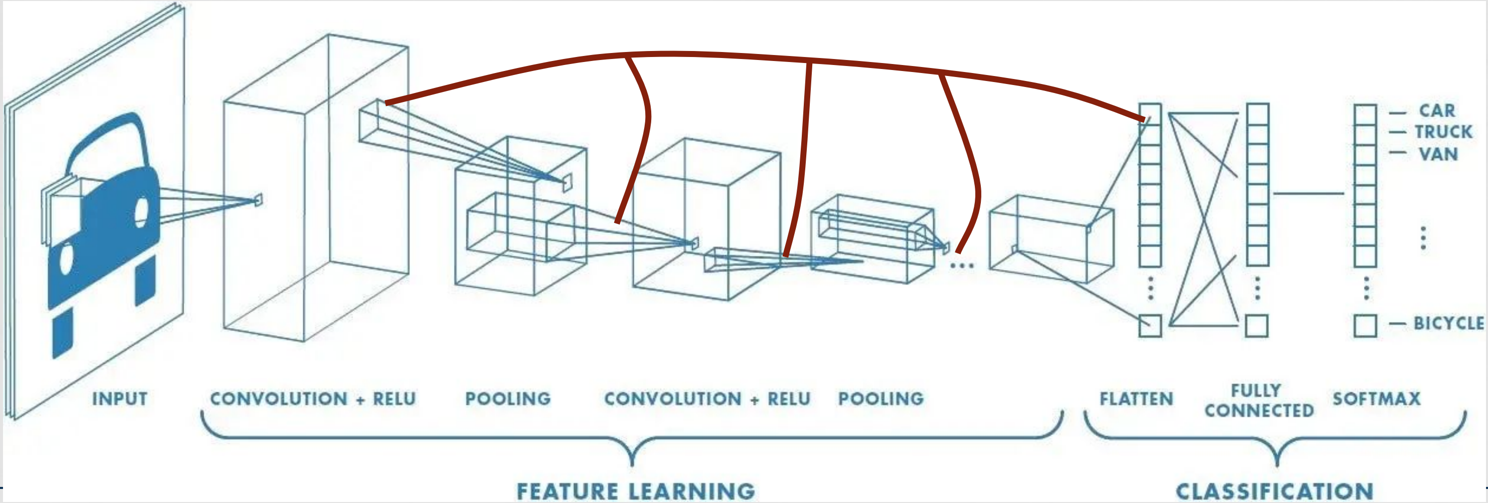
- Batchnorm
- Dropout layers

Resnet18()

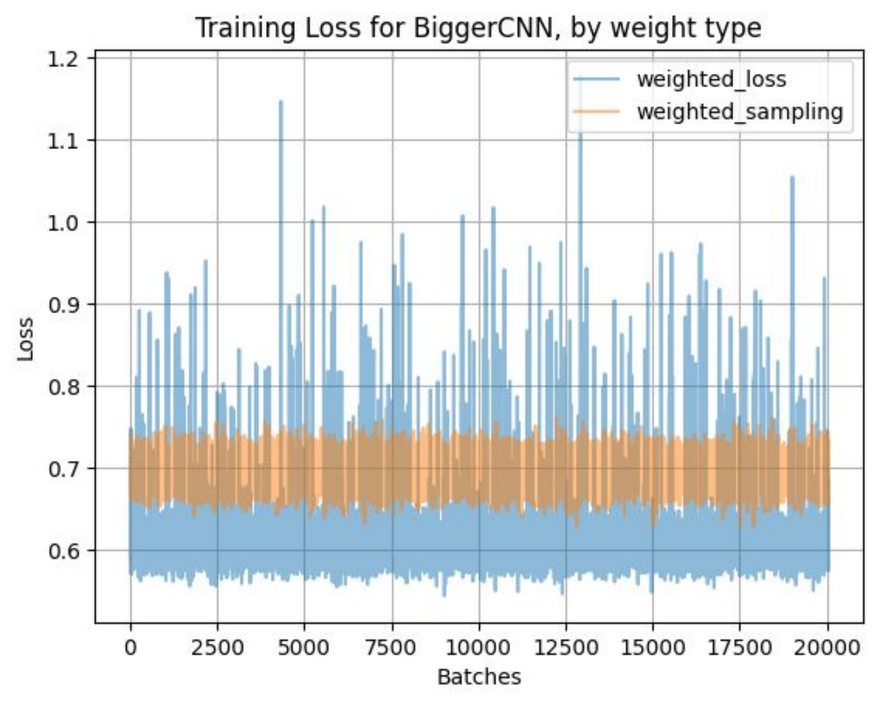
18 layers, 11 million parameters

Additions

- Batchnorm
- Dropout Layers
- **Residual Connections**



Problem: Data Imbalance

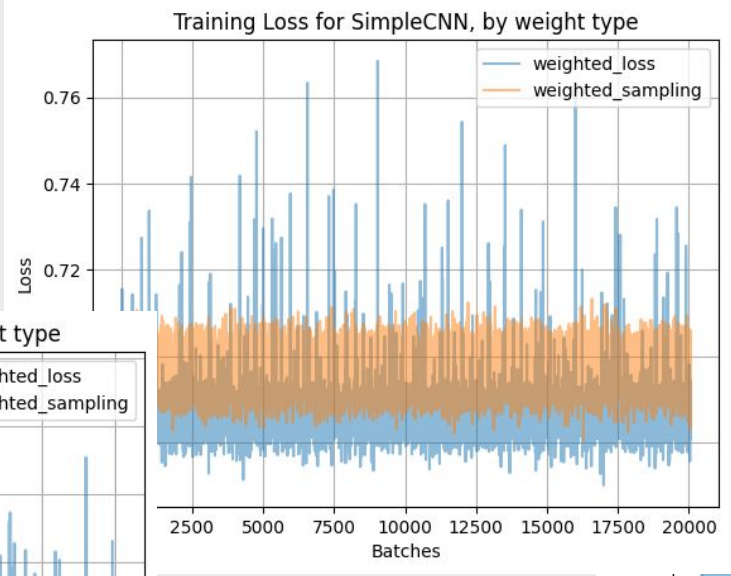


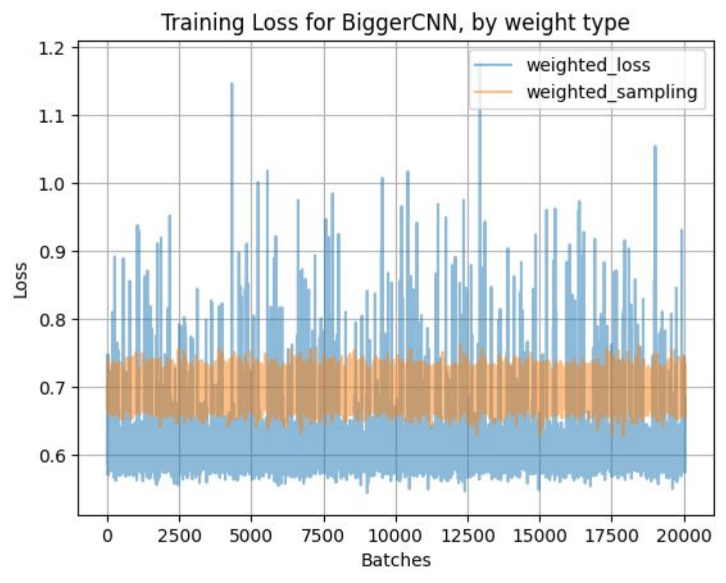
Weighted Sampling is much more stable

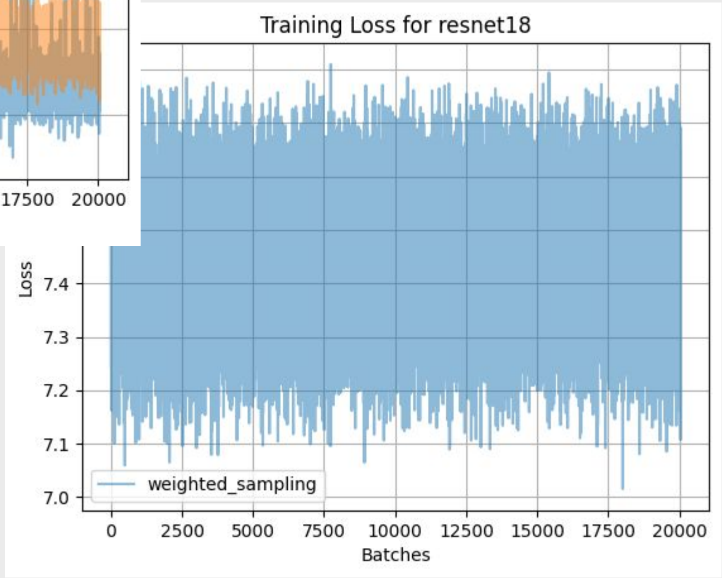
Benign:Malignant ratio is ~**1000:1**

To prevent models from only predicting Benign, during training we tried **weighing**

- the **loss**: increase penalty for missing Malignant
- the **sampling**: oversample Malignant, undersample Benign







Problem: No Downward Trend

Results:
Not Great

Results

	Logistic Regression	Random Forest	XGBoost	SimpleCNN	BiggerCNN	Resnet18
Unweighted	P: 0% R: 0%	P: 0% R: 0%		P: 0% R: 0%		
Weighted Loss	P: 0.25% R: 89%	P: 18% R: 3%	P: 2% R: 64%	P: 0% R: 0%	P: 0% R: 1%	
Weighted Sampling				P: 0.25% R: 68%	P: 3.9% R: 11%	P: 8% R: 2.67%

P: Precision of **Malignant** Class, the proportion of predicted malignants which are actually malignant.

R: Recall of the **Malignant** Class, the proportion of true malignant cases found.

Thanks!

References

(1) The Skin Cancer Foundation (2024, February 1). *Defining Skin Cancer*. Skin Cancer Foundation. https://www.skincancer.org/blog/defining-skin-cancer/

(2) The Skin Cancer Foundation. (2024 July). *Skin Cancer 101*. Skin Cancer Foundation. https://www.skincancer.org/skin-cancer-information/

(3) Nicholas Kurtansky, Veronica Rotemberg, Maura Gillis, Kivanc Kose, Walter Reade, Ashley Chow. (2024). *ISIC 2024 - Skin Cancer Detection with 3D-TBP.* Kaggle. https://kaggle.com/competitions/isic-2024-challenge

(Fig. 1) Kurtansky, et all. (2024). The SLICE-3D dataset: 400,000 skin lesion image crops extracted from 3D TBP for skin cancer detection. Scientific Data. 11. 10.1038/s41597-024-03743-w.

(Fig. 4) Simple Logistic Regression [Image]. (n.d.). In *Statstest.com*. Retrieved from https://www.statstest.com/simple-logistic-regression/

(Fig. 5) Random Forest vs XGBoost [Image]. (2021, August). In *EDUCBA*. Retrieved from https://cdn.educba.com/academy/wp-content/uploads/2021/08/Random-Forest-vs-XGBoost-1.jpg

(Fig. 6) A Guide to Convolutional Neural Networks — the ELI5 way | Saturn Cloud Blog. (2023, November 20). https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/