A Platform for Classifying Melanoma

Wilber Eduardo Bermeo Quito

Master in Data Science

Higher Polytechnic School

University of Girona

August 7, 2023

Summery

Introduction

Motivations

- Enhance AI[1] knowledge.
- Automation as way to democratize access to research and AI solutions.
- CAD[2] system are promising path towards medical automation.

[1] Artificial Intelligence.
[2] Computer-Aided Diagnosis.

Objectives

- Gain expertise in deep learning theory and its real-world applications.
- Explore and study the optimal approach for utilizing the distribution of dermoscopy images from the dataset during the training process.
- Propose and train deep learning models using transfer learning on ISIC[3] Challenge melanoma images.
- Create an easy deployment CAD infrastructure running in Docker, with the trained models, a user-friendly web UI[4] and a HTTP API[5].

[3] Skin Imaging Collaboration.
[4] User Interface.
[5] Application Programming Interface.

Domain

Problem

Detection of Melanoma Skin Cancer

- Melanoma exhibits a high mortality rate.
- Dermoscopy procedures are utilized for melanoma detection.
- Dermoscopy images are examined by professionals to study cutaneous lesions.
- Several studies have shown that melanoma task classification using CAD systems achieve comparable or superior results to dermatologists.

Metastasis

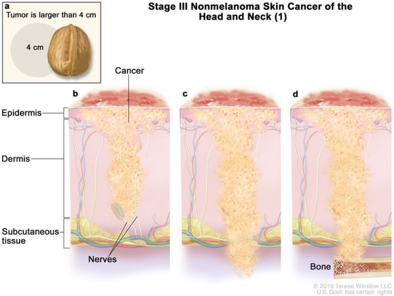

Figure 1: *Skin Cancer, Stage III. Illustration by Terese Winslow*

Solution

CAD Training and Deployment Pipeline

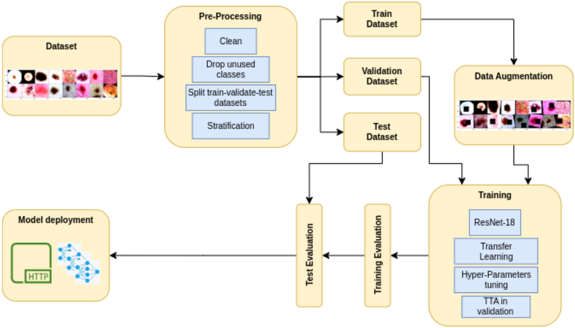

Figure 2: *CAD Infrastructure Pipeline.*

Micro-Service Architecture to Infer Images

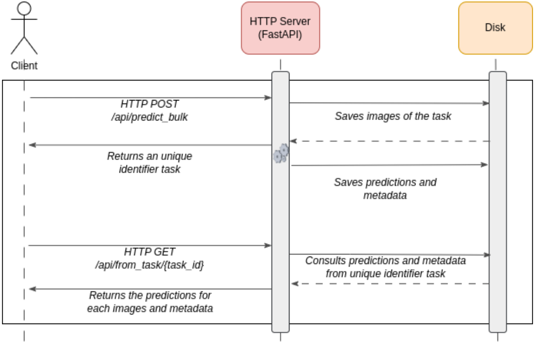

Figure 3: *Inferring Images Through the Background Task Mechanism.*

Concerns

Ethical Concern

- The solution employs "black box" models, lacking explain-ability.
- The thesis presents a CAD tool designed to aid human decision-making rather than being an autonomous decision-making system.

Regulatory Framework

- When dealing with medical images, obtaining signed consent is necessary for data publication.
- Recent research collaborations prioritize data sharing through de-identification methods to tackle these challenges.
- The thesis made use of the ISIC Archive database, which serves as a publicly accessible resource.

Data

Origin Data Decription

- The data originates from the ISIC Archive.
- It includes images from the years 2019 and 2020.
- The images are available in three different resolutions: 512x512, 768x768, and 1024x1024 pixels.
- The dataset contains more than eight distinct classes.

Used Data Decription

- Resolution selected: 512x512 pixels.
- The used dataset comprises 31,265 distinct image samples.
- Eight classes were selected to work with.
- Imbalanced dataset.

Classes Distribution in the Dataset

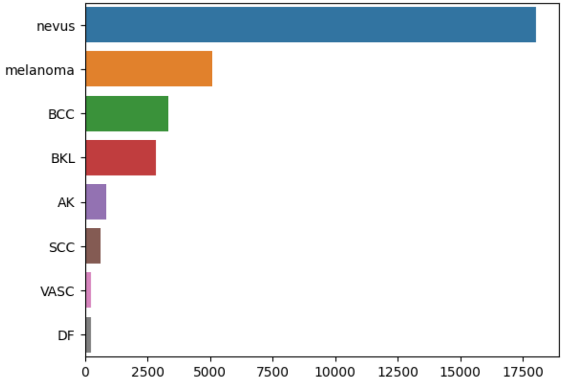

Figure 4: *Data Distribution.*

Train, Validation and Test Sets

- The dataset was stratified to ensure an equal distribution of classes in each subset.
- The training set was created using 80% of the dataset, the validation set using 10%, and the test set using the remaining 10%.

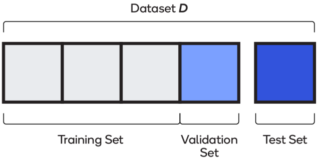

Figure 5: *Holdout Set Scheme. Illustration by Qualcomm*

Data Augmentation

The train dataset (Figure 6),

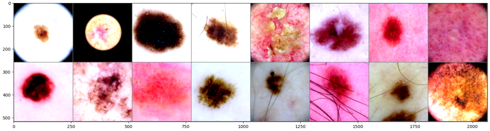

Figure 6: *Random Sample of Images.*

Is mapped into an augmented train dataset (Figure 7).

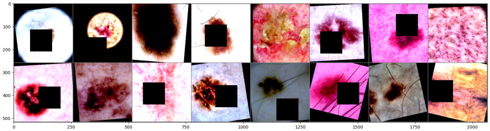

Figure 7: *Augmented Random Sample of Images.*

Modeling

General Modeling Information

- Eight trained models with different ML thecniques.
- Used the ResNet18 pre-trained weights.
- SGD as optimizer.
- Cross-entropy as loss function.
- Model training performance were evaluated with AUC metric.

	M0	M1	M2	M3	M4	M5	M6	M7
Model Architecture	R18M	R18M	R18M	R18M	R18DM	R18DM	R18DM	R18DM
Epochs	20	20	20	20	40	40	40	40
Batch Size	400	400	400	400	1024	1024	1024	1024
Scheduler		SLR	CALR	CAWR		SLR	CALR	CAWR
Data Augmentation	No	No	No	No	Yes	Yes	Yes	Yes
Dropout Regularization	No	No	No	No	Yes	Yes	Yes	Yes
GPU	TT4	TT4	TT4	TT4	NA100	NA100	NA100	NA100
Training Time	1h 45m	1h 22m	1h 43m	1h 38m	1d 7h 30m	1d 7h 4m	1d 7h 1m	1d 12h 55m

Table 1: *Training Information For Each Model. Empty spaces represent non-use of that feature.*

	Train AUC	Val AUC	Train Recall	Val Recall	Train Acc	Val Acc
M0	0.952	0.903	0.756	0.676	0.835	0.778
M1 ⋆	0.947	0.900	0.695	0.633	0.829	0.779
M2 ∗	0.933	0.895	0.658	0.609	0.808	0.765
M3 •	0.935	0.896	0.663	0.605	0.811	0.767
M4	0.886	0.877	0.478	0.475	0.757	0.750
M5 ⋆	0.867	0.861	0.423	0.403	0.728	0.717
M6 ∗	0.874	0.868	0.451	0.440	0.738	0.728
M7 •	0.877	0.849	0.470	0.432	0.742	0.732
Mean	94.175%	89.850%	69.300%	63.075%	82.075%	77.225%
SD	0.921%	0.370%	4.509%	3.260%	1.327%	0.727%
Mean	87.600%	86.875%	45.550%	44.400%	74.125%	73.175%
SD	0.787%	0.655%	2.445%	3.084%	1.204%	1.372%

Table 2: *Train & Validaton Metrics.*

M3 vs. M7

(c) AUC Curves

(a) Loss Curves

(b) Accuracy Curves

Figure 8: *M3 vs. M7. Combined AUC, Loss and Accuracy Curves.*

Testing

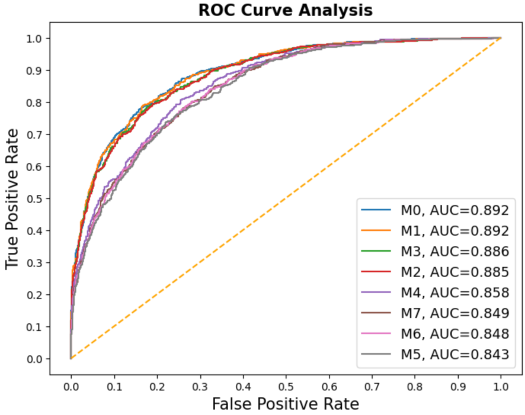

Figure 9: *ROC-AUC Results in Test Dataset.*

Workflow & Results

Workflow Methodology

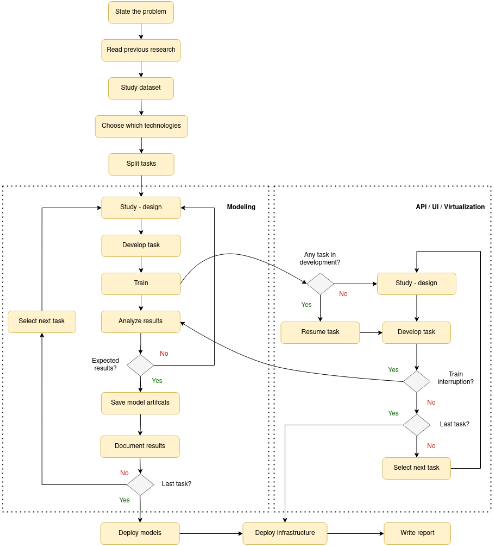

Figure 10: *Activity Diagram Describing the Workflow Methodology.*

Results