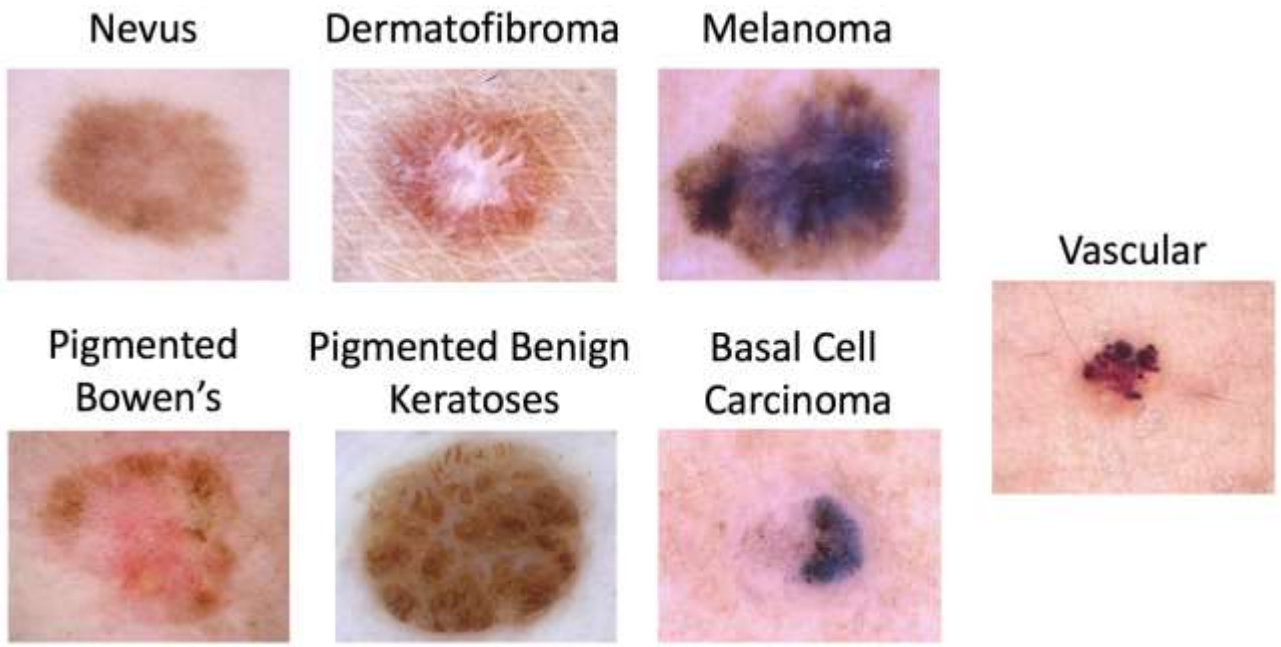
Skin cancer diagnosis through imaging classification

Pedro Leite

Ruben Pombo

The challenge



- Challenge by ISIC (International Skin Image Collaboration)
- Goal is to submit automated predictions of disease classification with dermoscopic images

The dataset

HAM10000_metadata

lesion_id	image_id	dx	dx_type	age	sex	localization	dataset
HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp	vidir_modern
HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp	vidir_modern
HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp	vidir_modern
HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp	vidir_modern

- HAM10000 from Harvard University
- Contains 10 thousand jpeg images
- Each image has 1 of the 7 diseases that we want to predict
- File with the metadata

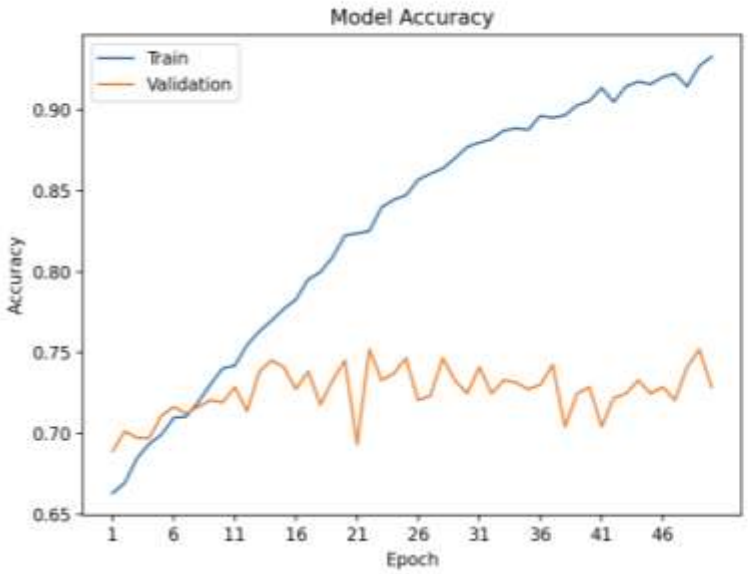
Data pre-processing

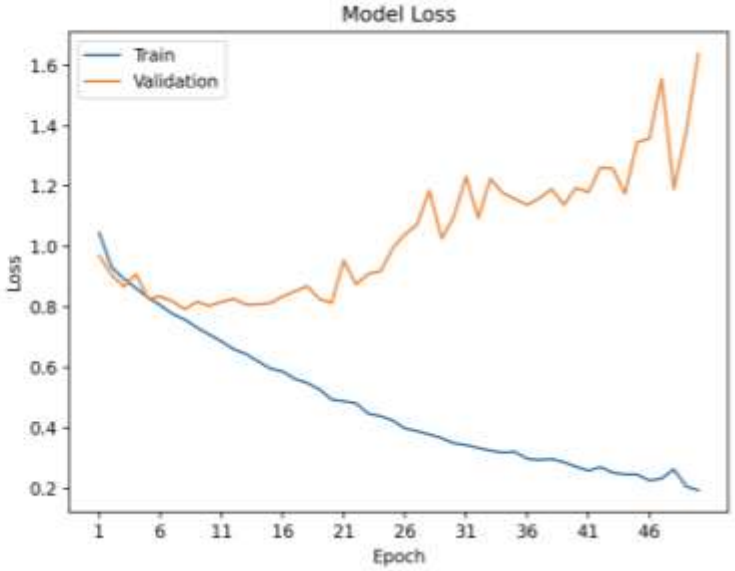
- Removed null and unknown observations
- Added to the dataset an attribute with the path to the image
- One hot-encoded the target variable
- Resize all images to 125x100

Data pre-processing

- Normalized every image
- Split the data, into:
- - Training (75%)
- - Testing (25%)
- Further split the training, into:
- - Training (90%)
- - Validation (10%)

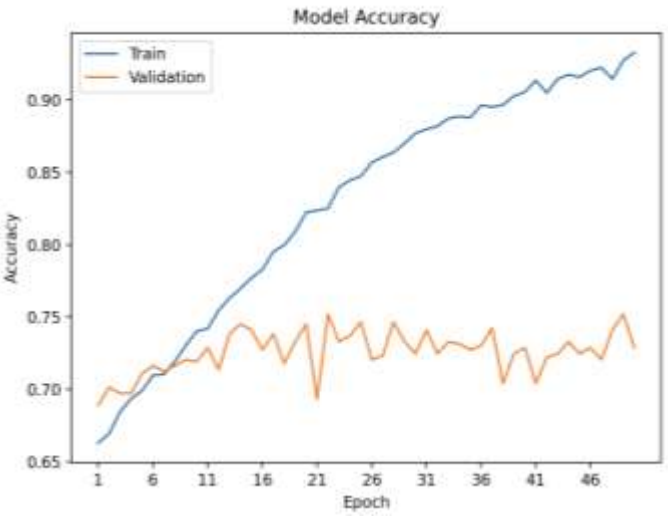
Testing

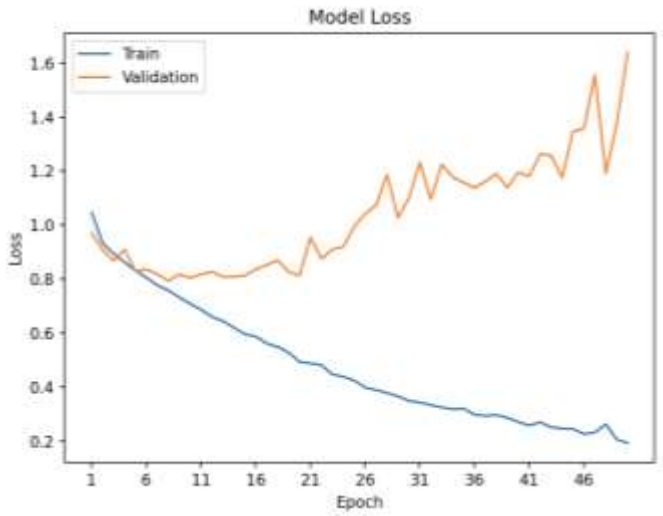


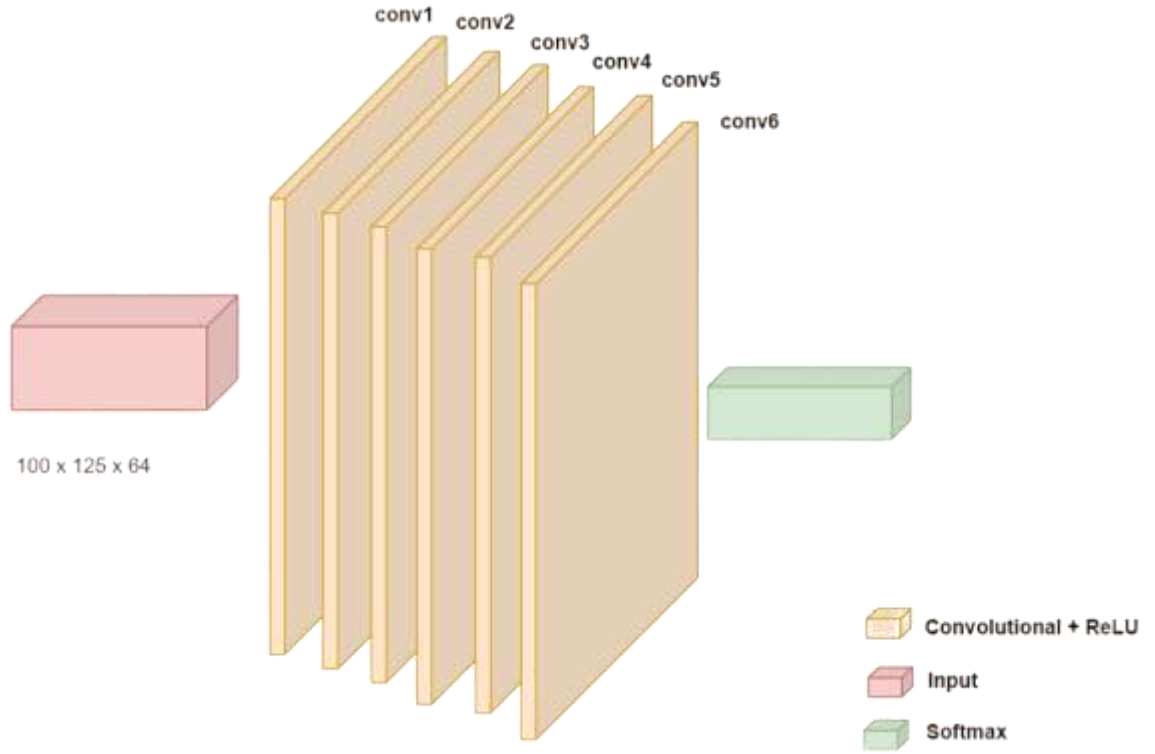


- Accuracy and Loss in testing
- Accuracy and Loss in validation

Basic Neural Network Model

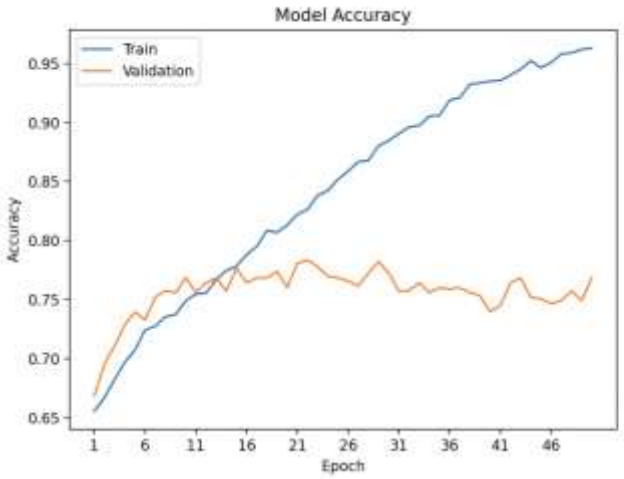


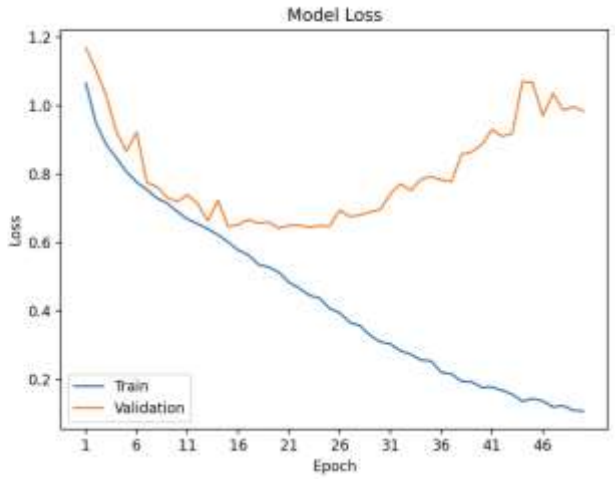


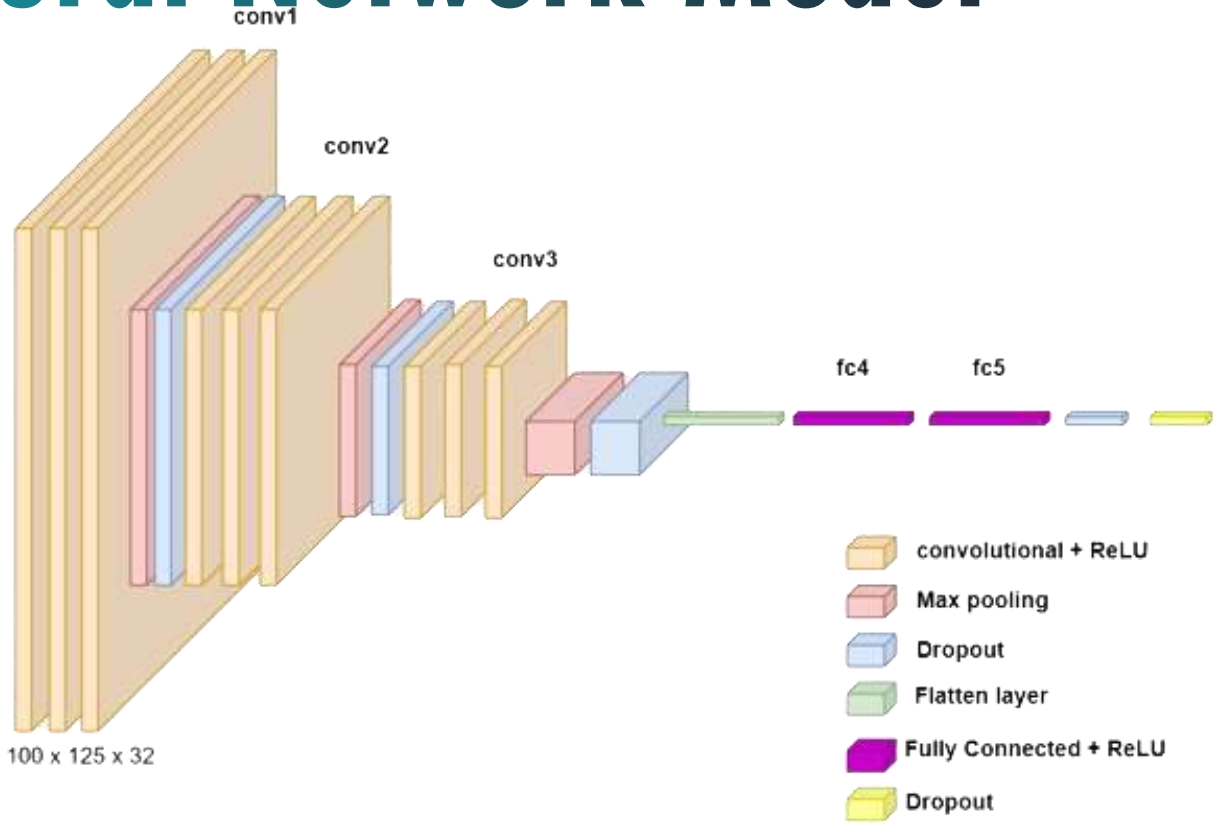


Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
15	69%	202%	72%	163%

Basic Convolutional Neural Network Model

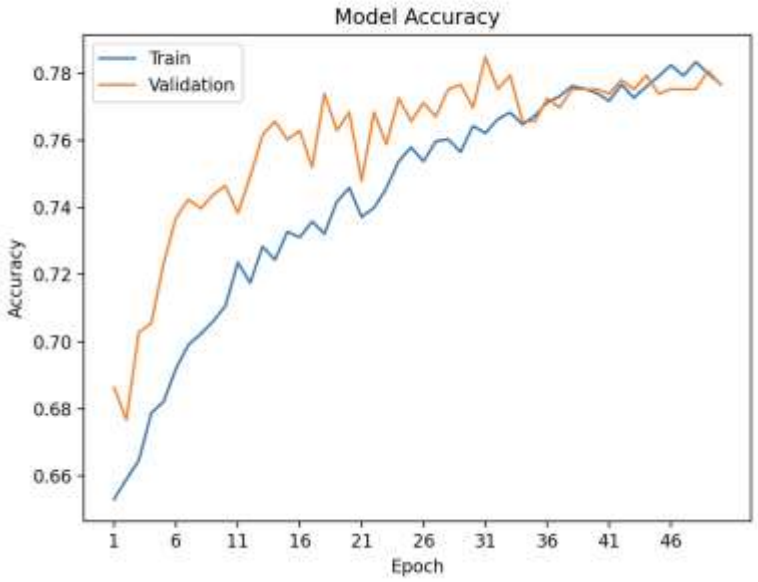


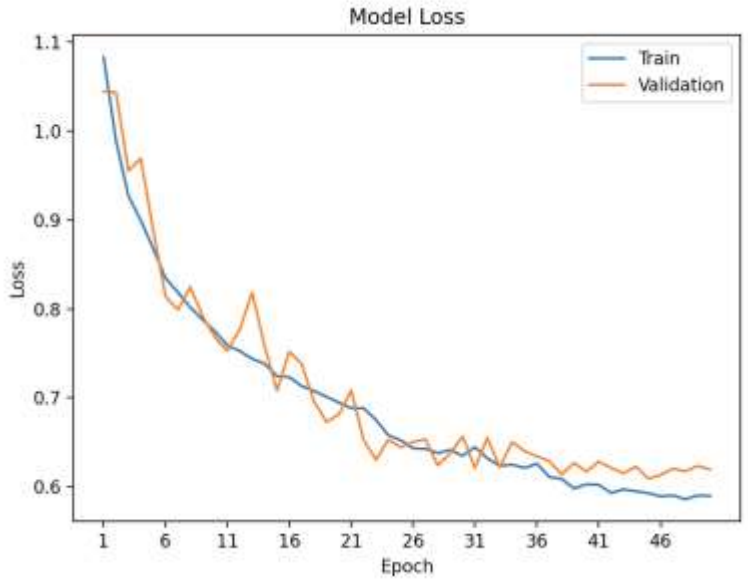




Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
58	75%	106%	76%	98%

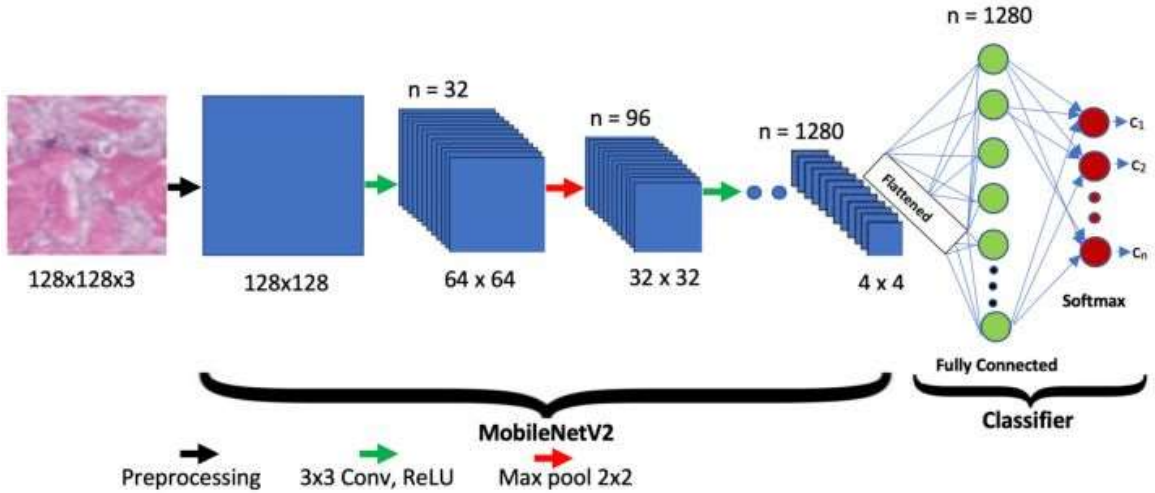
Basic Convolutional Neural Network Model 2.0

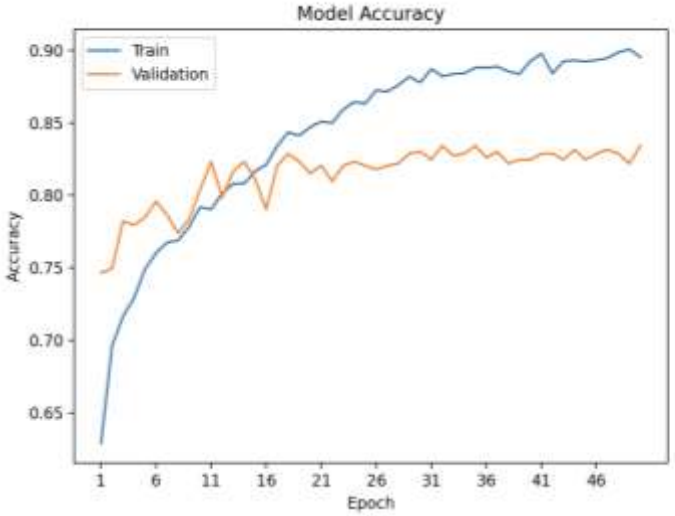


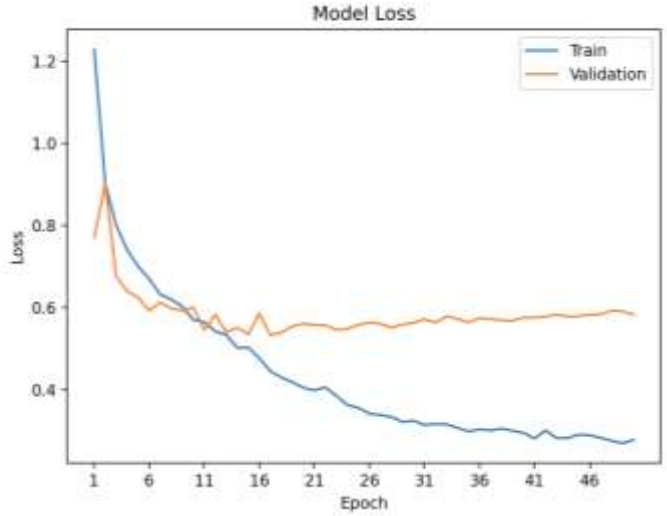


Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
58	75%	65%	77%	61%

Mobile Net

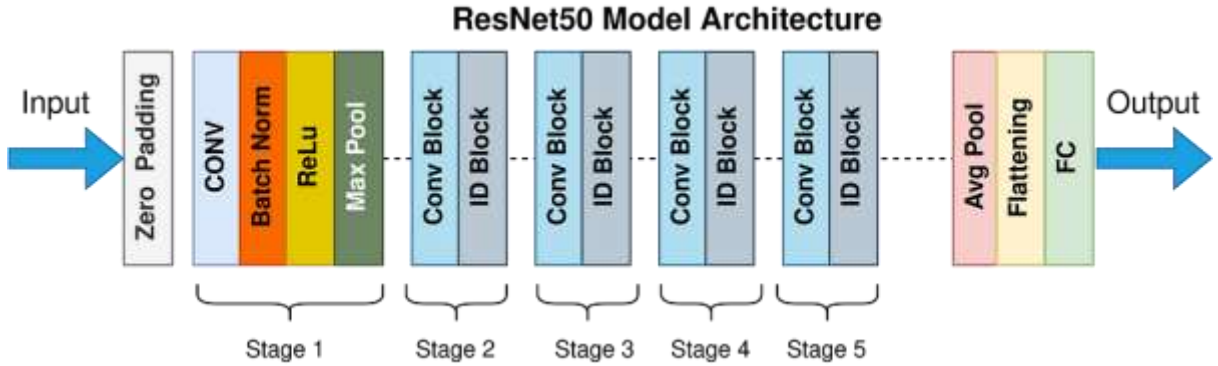


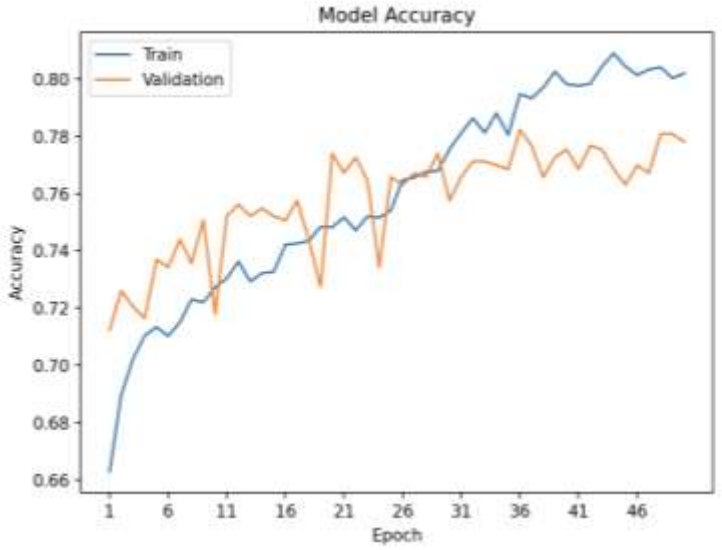


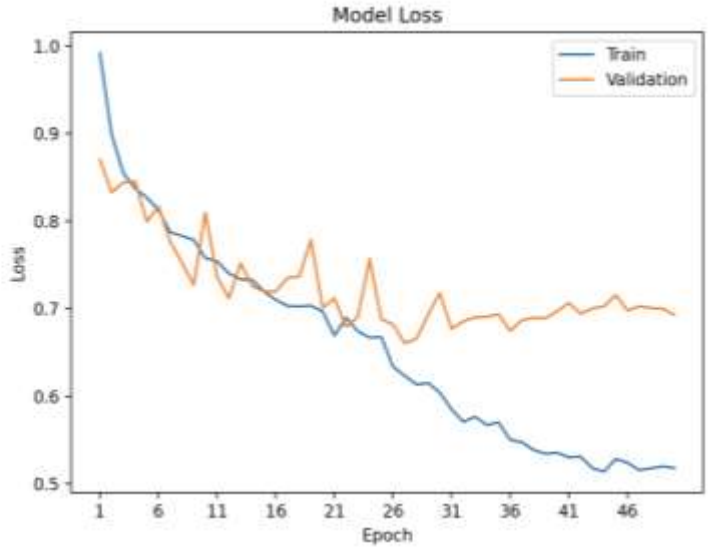


Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
54	81%	60%	83%	58%

Res Net

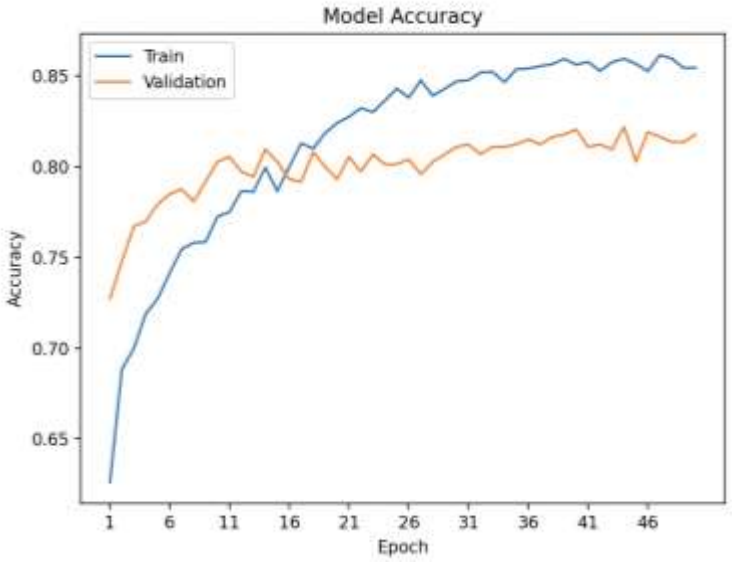


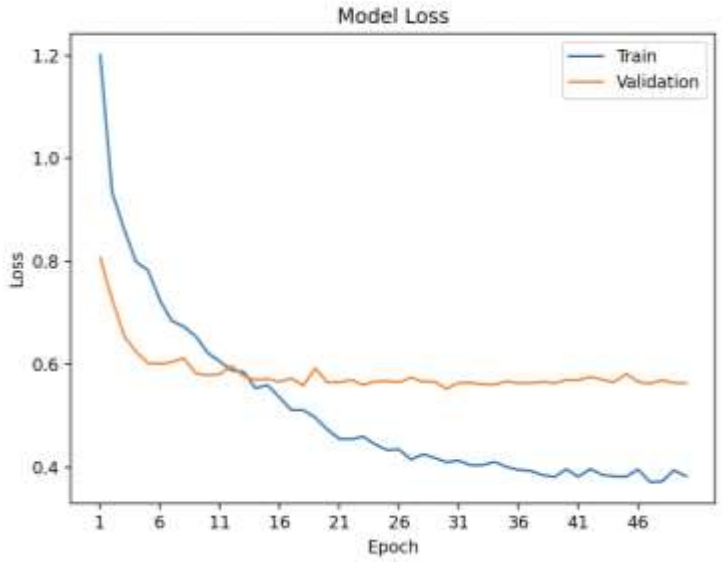


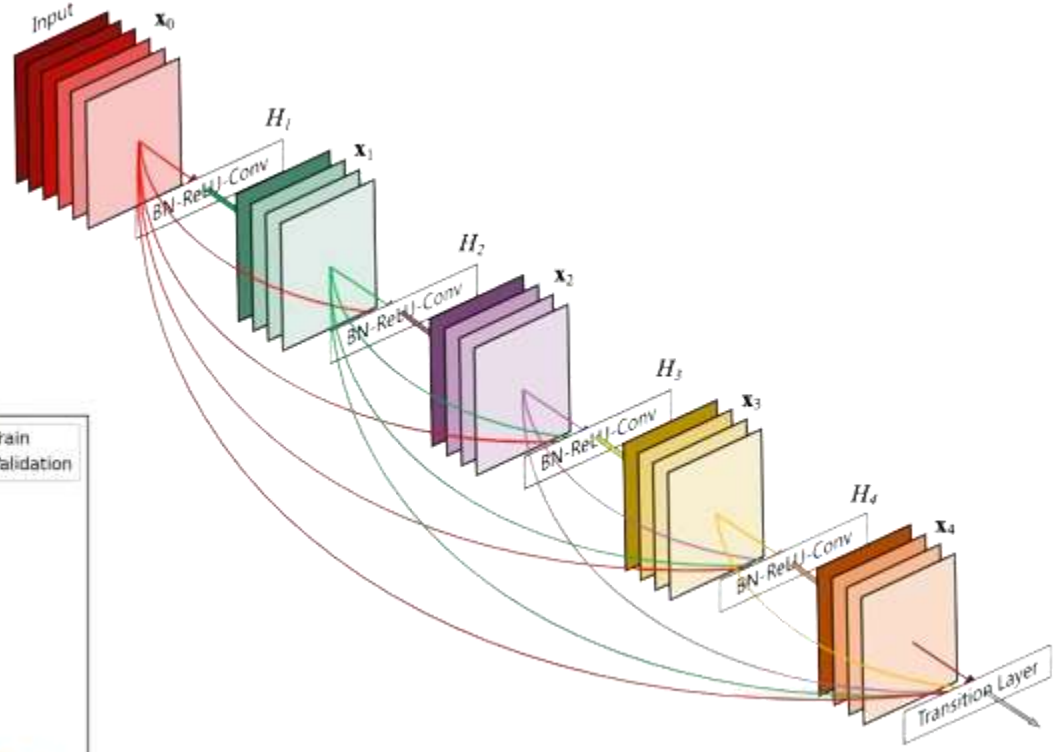


Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
217	75%	78%	77%	69%

Dense Net

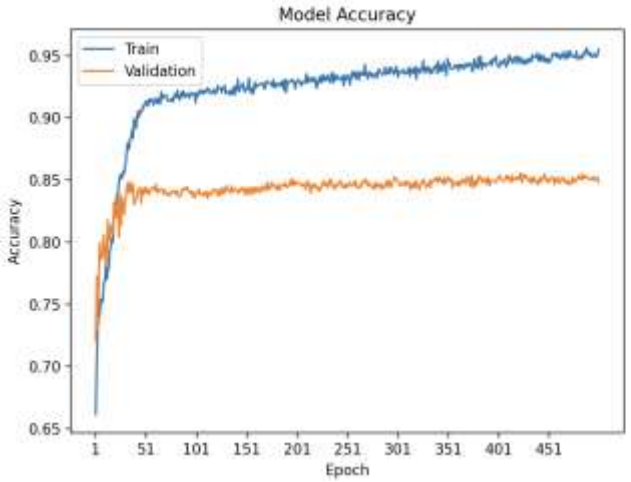


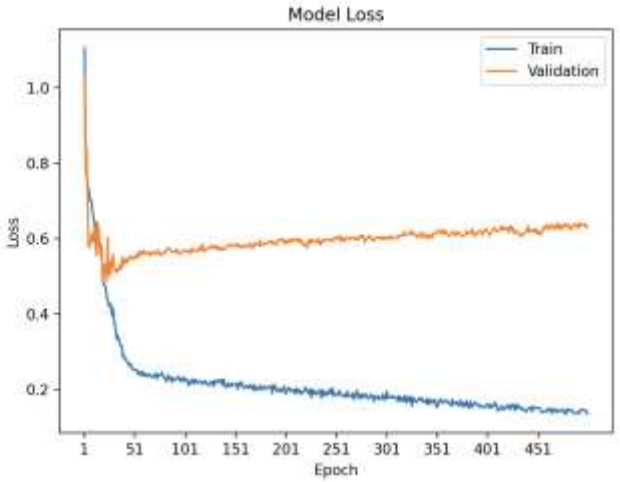




Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
135	82%	57%	81%	56%

Mobile Net 2.0





- Trained more layers
- Bigger dropout
- Batch Normalization
- Added regularization to the output layer
- Smaller learning rate
- Bigger data augmentation

Time (minutes)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
1092	83%	68%	84%	62%

Epoch	Time (seconds)	Accuracy (Test)	Loss (Test)	Accuracy (Validation)	Loss (Validation)
500º	115	95	13	85	63