

# SKIN LESION CLASSIFICATION AND DIAGNOSIS

Gladys Pao, DSI-SG-17

# PROBLEM STATEMENT

---

## 01 THE PROBLEM

- Skin cancer is the most common cancer globally
- Numerous lesions being misdiagnosed despite usage of dermoscopy
- \$673 million in overall cost

## 02 THE STAKEHOLDERS

- Dermatologists
- Hospital and skin cancer clinics in Singapore

## 03 THE PURPOSE

- Improve diagnostic rate of skin cancer
- 2 classification tasks using Convolutional Neural Networks:
  - **Specific skin lesion diagnosis**
  - **Malignant, benign, or precancerous**

# METHODOLOGY

## STEP 1:

### EXPLORATORY DATA ANALYSIS

- Analyse trends in labelled and image data



## STEP 2: IMAGE PREPROCESSING

- Standardise images (size, colour etc.)
- Data augmentation

## STEP 4: CONCLUSION

- Recommendations moving forward

## STEP 3: MODELLING AND EVALUATION

- Convolutional Neural Networks
- Evaluation metrics (Accuracy and Recall)

# THE DATASET

# DATA SOURCES

---

TOTAL NUMBER OF IMAGES

10,276

## MAIN DATASETS

- International Skin Imaging Collaboration (ISIC):
  - 2018, 2019 and 2020 datasets

## EXTERNAL DATASETS

- DermNet NZ
- Dermoskopedia
- 7-point Criteria Evaluation Database

# TARGET VARIABLES



## 7 SKIN LESIONS

- Melanocytic Nevi (NV)
- Benign keratoses (BKL)
- Vascular lesions (VASC)
- Dermatofibroma (DF)
- Melanoma (MEL)
- Basal cell carcinoma (BCC)
- Actinic Keratoses (AKIEC)

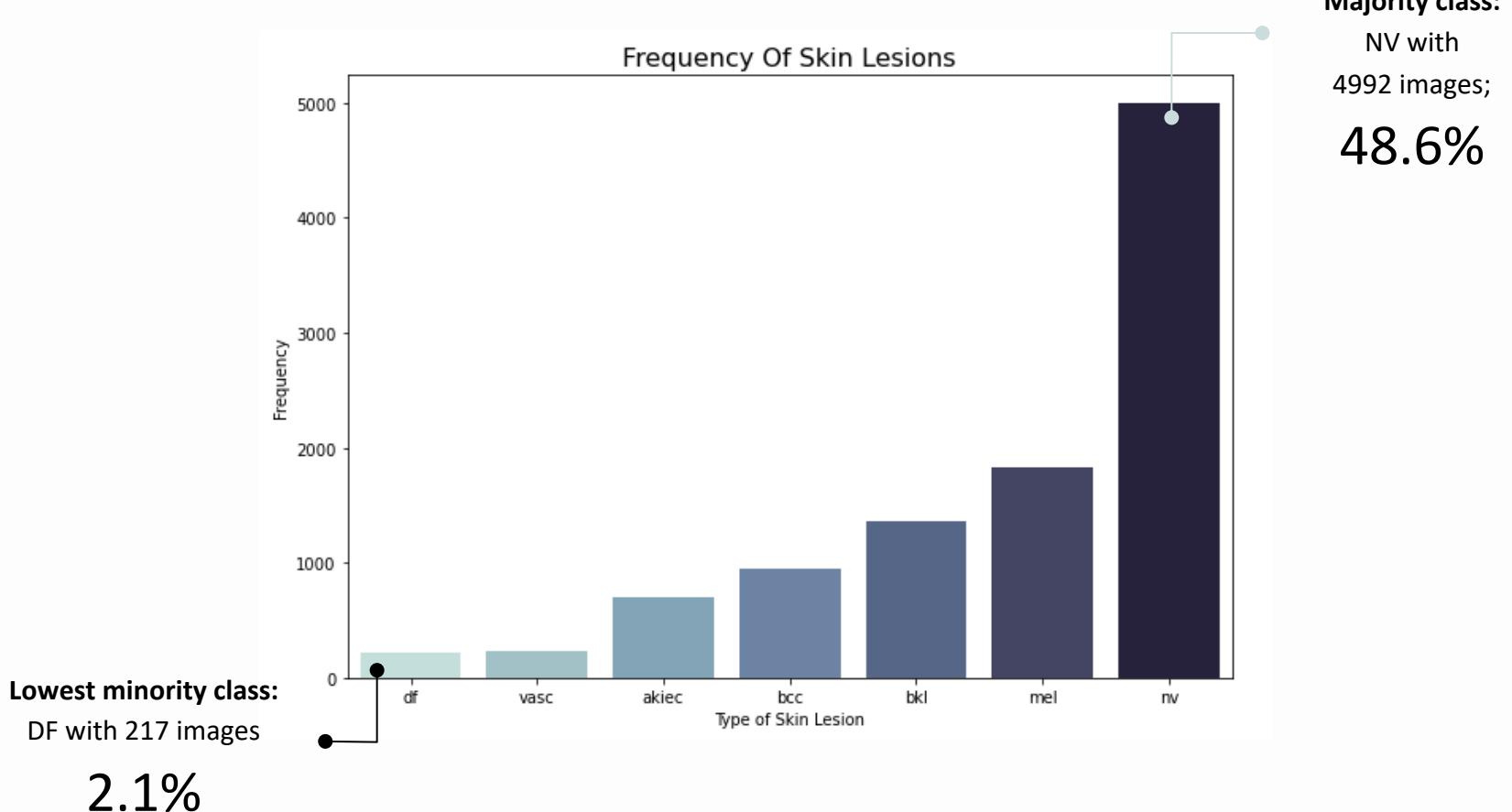


## 3 TYPES OF DIAGNOSIS

- Benign
  - NV, BKL, VASC, DF
- Malignant
  - MEL, BCC
- Precancerous
  - AKIEC

# EXPLORATORY DATA ANALYSIS

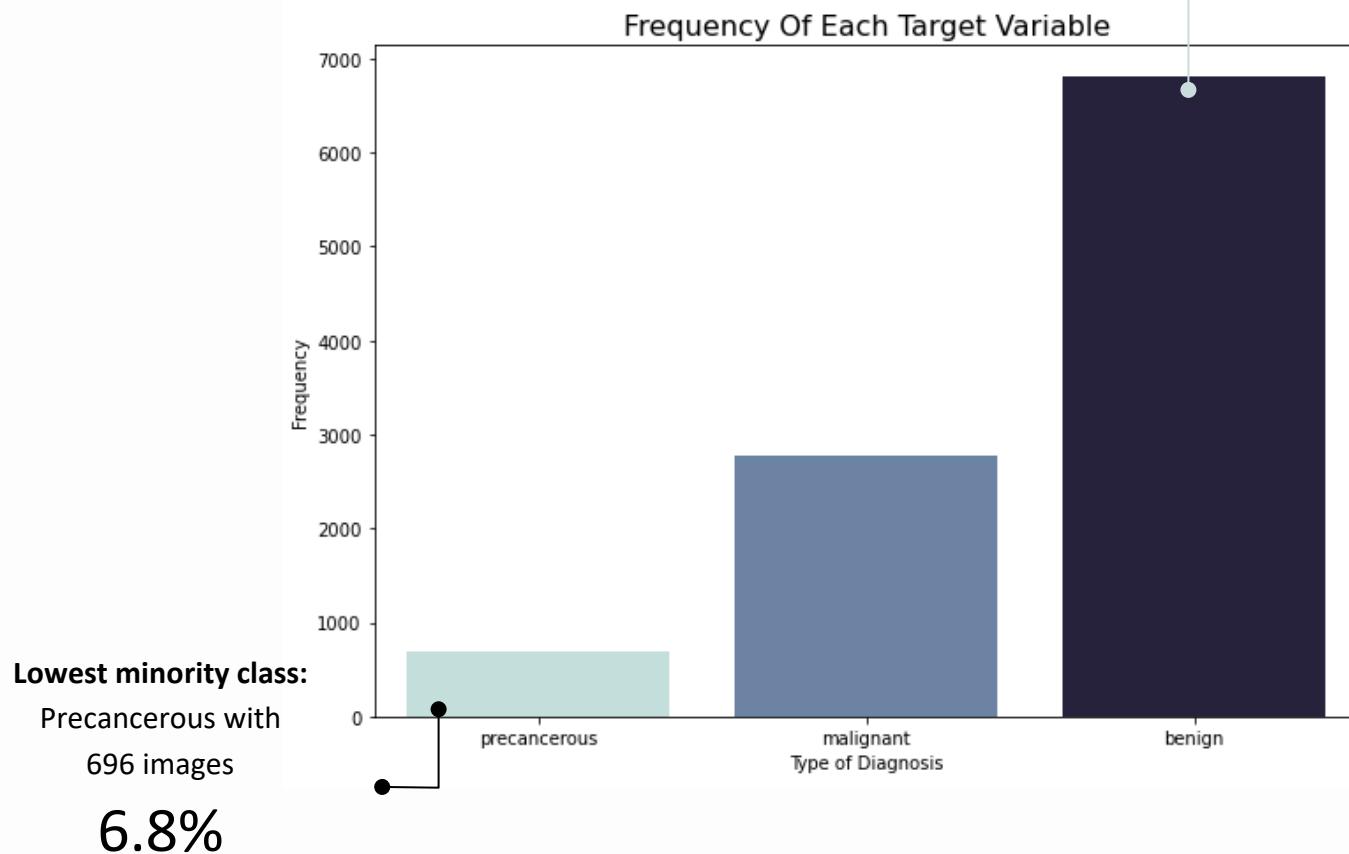
# CLASS IMBALANCE



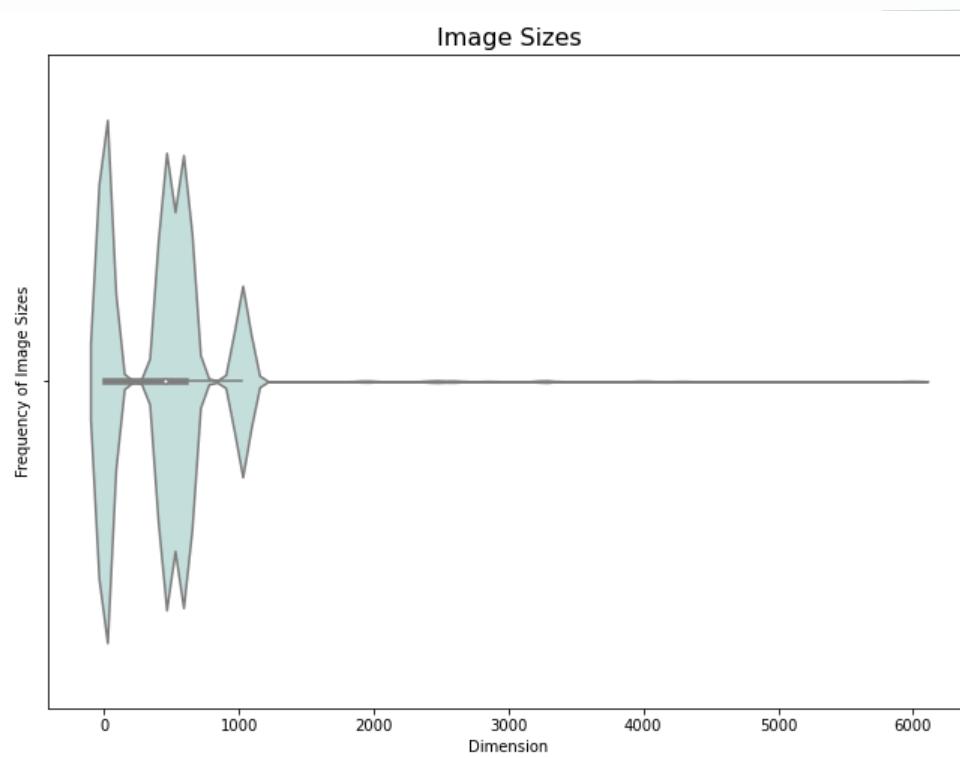
# CLASS IMBALANCE

Majority class:  
Benign with 6801 images

66.2%

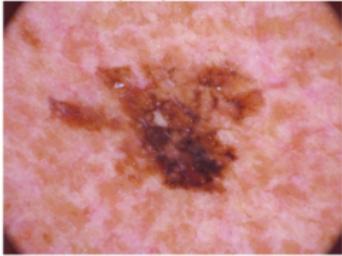


# IMAGE SIZES

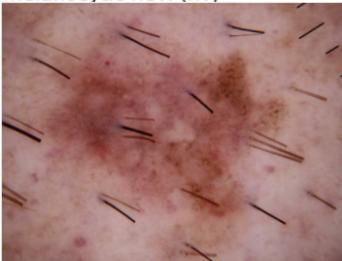


# IMAGE ANALYSIS

melanoma (mel)



melanocytic nevi (nv)



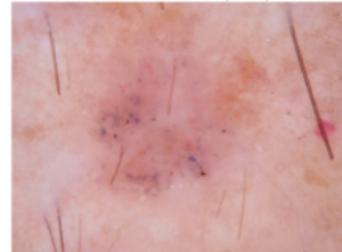
vascular lesions (vasc)



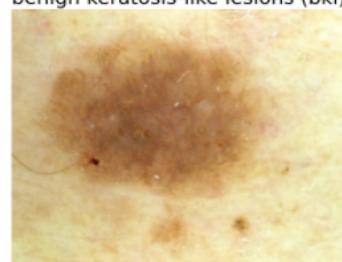
actinic keratoses (akiec)



basal cell carcinoma (bcc)



benign keratosis-like lesions (bkl)

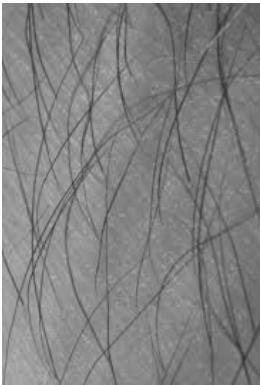


# IMAGE PREPROCESSING

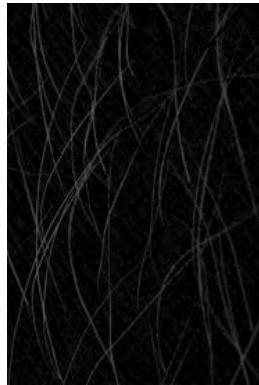
# HAIR REMOVAL



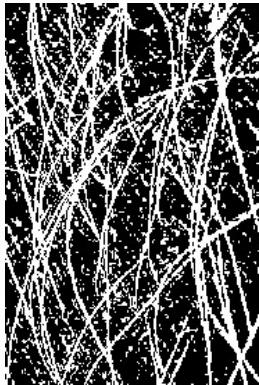
ORIGINAL IMAGE



GREYSCALED IMAGE



BLACK TOP-HAT  
TRANSFORMATION



INCREASING THRESHOLD



INPAINTING  
(image interpolation)

# BLACK TOP-HAT TRANSFORMATION

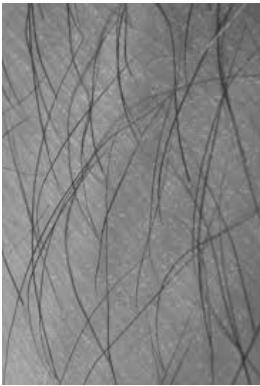
- Enhance dark objects of interest in a bright background



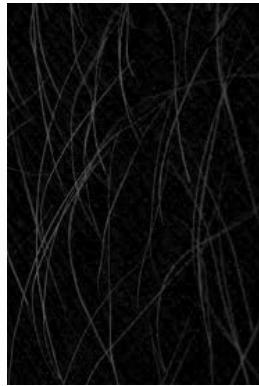
# HAIR REMOVAL



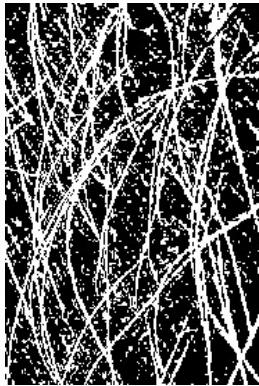
ORIGINAL IMAGE



GREYSCALED IMAGE



BLACK TOP-HAT  
TRANSFORMATION  
(morphological filtering)



INCREASING THRESHOLD



INPAINTING  
(image interpolation)

# INPAINTING

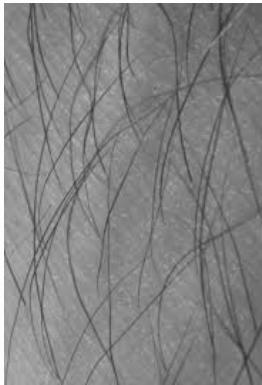
- Form of image interpolation
- Reconstruct missing parts of an image
- Replacing areas with pixels similar to the neighbouring ones



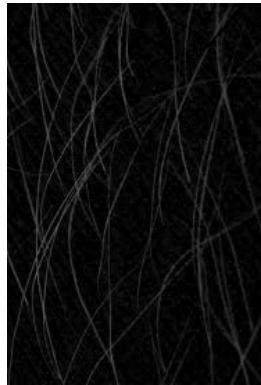
# HAIR REMOVAL



ORIGINAL IMAGE



GREYSCALED IMAGE



BLACK TOP-HAT  
TRANSFORMATION  
(morphological filtering)

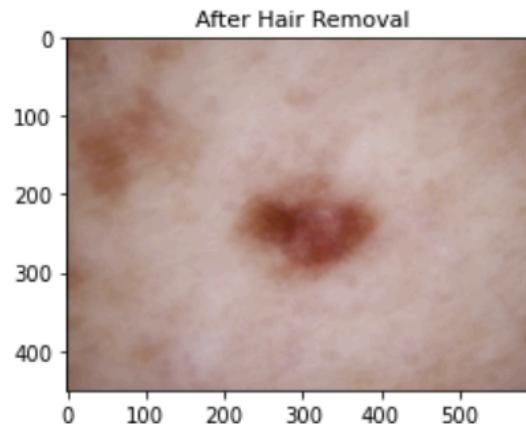
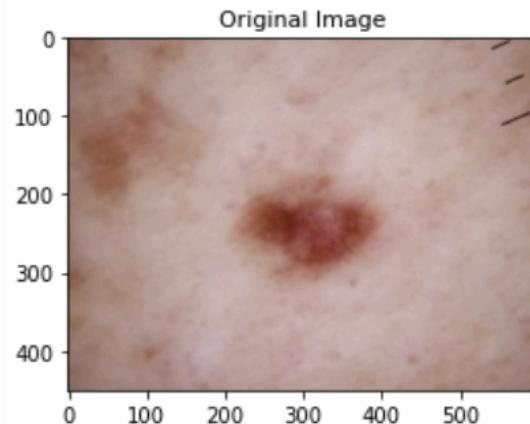
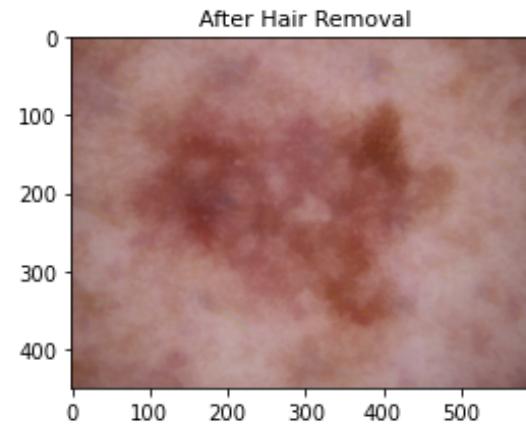
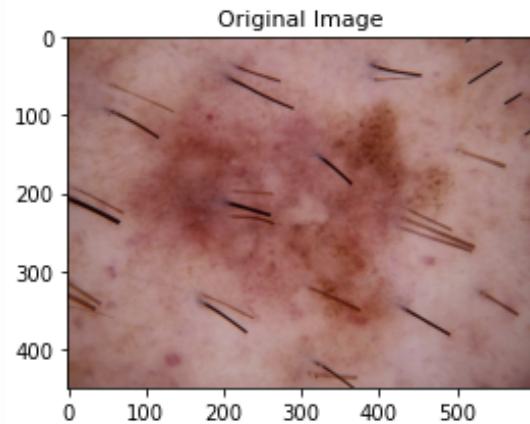


INCREASING THRESHOLD



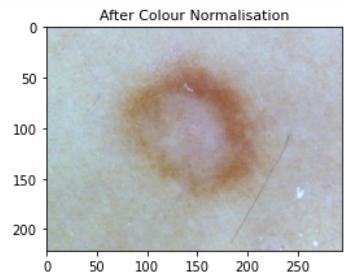
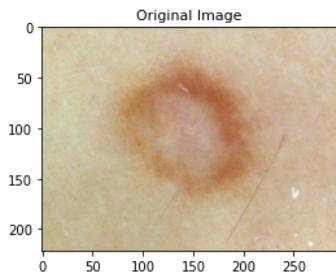
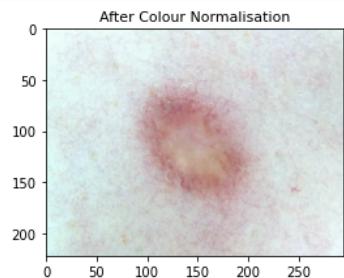
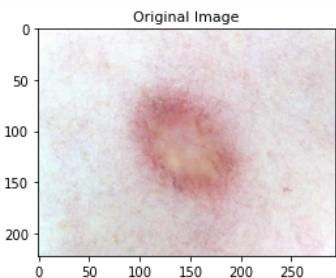
INPAINTING  
(image interpolation)

# HAIR REMOVAL

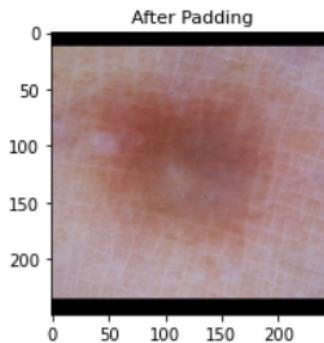
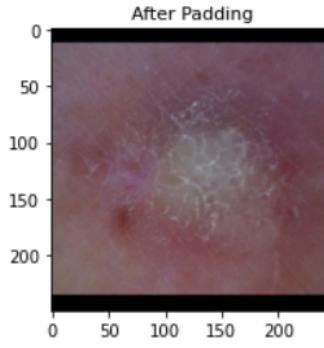
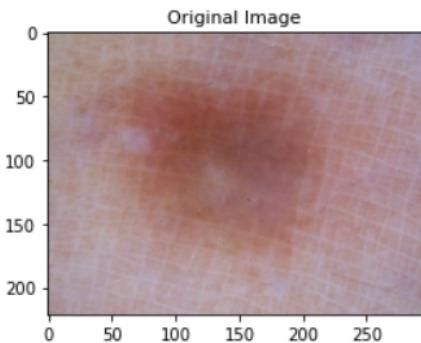
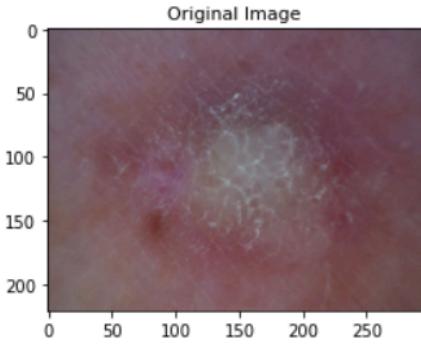


# COLOUR NORMALISATION

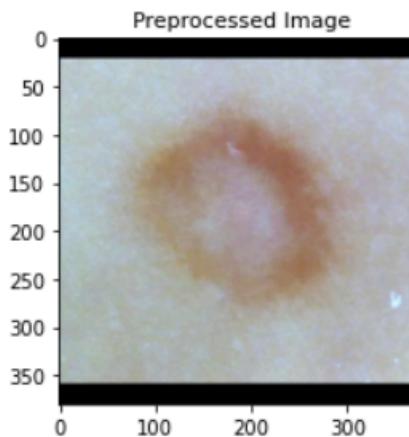
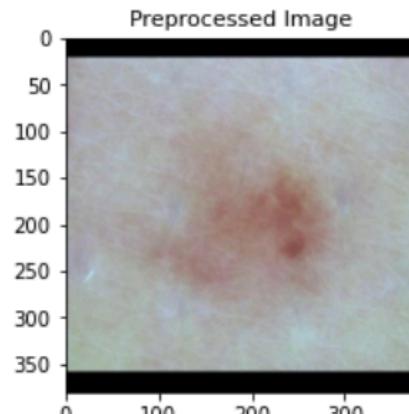
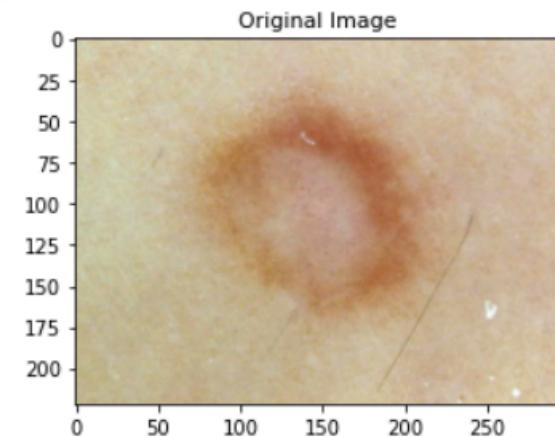
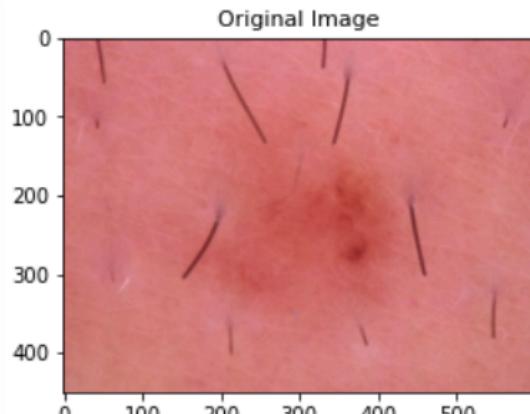
- **White-balancing Grey World Algorithm**
- Assumes equal average intensities of R, G, B channels
- R and B channels are multiplied by their respective gains
  - Gains = Average of G channel ÷ Average of respective channel



# PADDING, CROPPING AND RESIZING



# ORIGINAL VERSUS PREPROCESSED



# SYNTHETIC MINORITY OVERSAMPLING TECHNIQUE (SMOTE)

- Generates synthetic samples for the minority classes
- Obtain a synthetically class-balanced training set
- Split data into a 60-20-20 ratio and used SMOTE on the training set

TASK 1:  
SKIN LESION CLASSIFICATION

20,965

images

TASK 2:  
DIAGNOSIS CLASSIFICATION

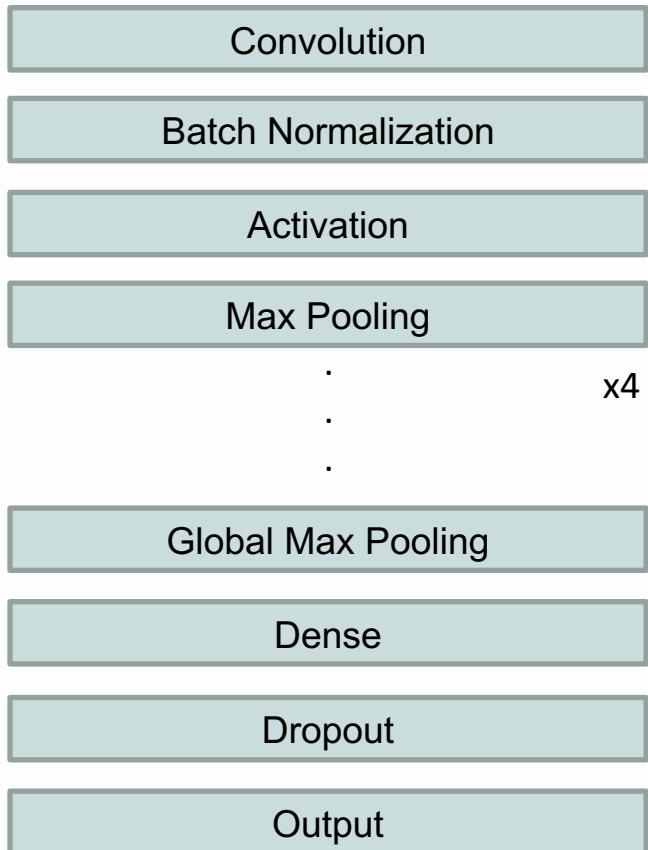
12,240

images

# MODELLING AND EVALUATION

# CONVOLUTIONAL NEURAL NETWORKS

- Similar model architecture for both tasks
- **Batch Normalization**
  - Standardize inputs of each layer
  - Faster training times
  - Better model performance
  - Better regularization effect (compared to Dropout)





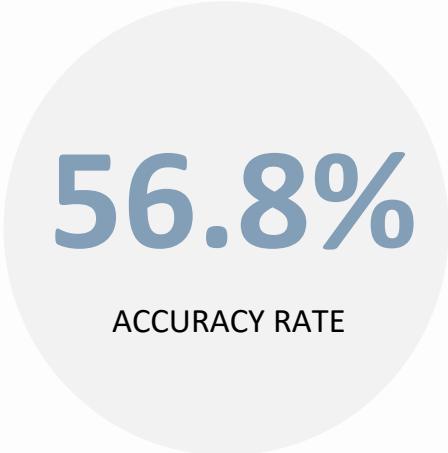
# TASK 01

## SKIN LESION CLASSIFICATION

7-way Classification

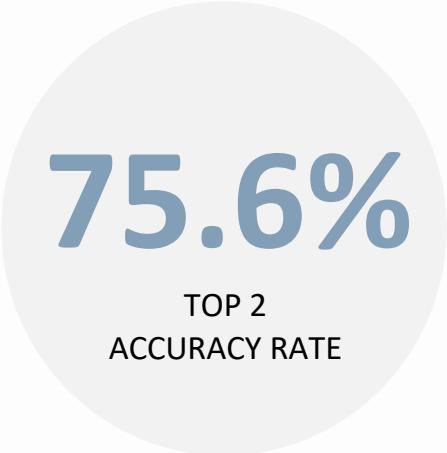
## BASELINE MODEL:

### Artificial Neural Networks



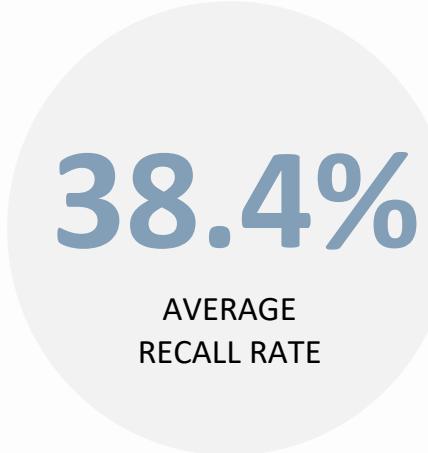
**56.8%**

ACCURACY RATE



**75.6%**

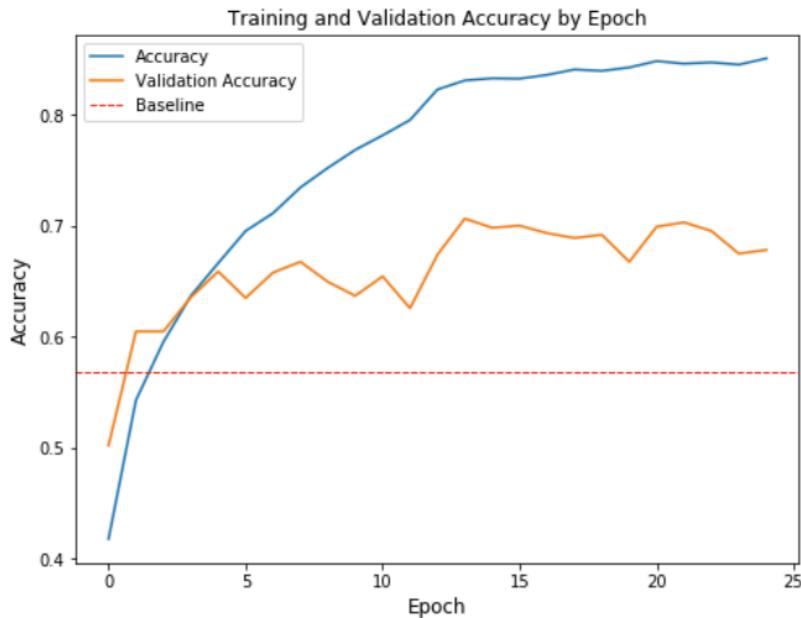
TOP 2  
ACCURACY RATE



**38.4%**

AVERAGE  
RECALL RATE

# CONVOLUTIONAL NEURAL NETWORKS



**70.7%** > **56.8%**

ACCURACY RATE

BASELINE ACCURACY

**86.2%** > **75.6%**

TOP 2  
ACCURACY RATE

BASELINE TOP 2  
ACCURACY

**57.7%** > **38.4%**

AVERAGE  
RECALL RATE

BASELINE RECALL

# MODEL EVALUATION

**70.5%**

TEST  
ACCURACY RATE

**86.2%**

TEST TOP 2  
ACCURACY RATE

**60.7%**

TEST AVERAGE  
RECALL RATE

**0.1%**

DIFFERENCE

**0%**

DIFFERENCE

**3%**

DIFFERENCE

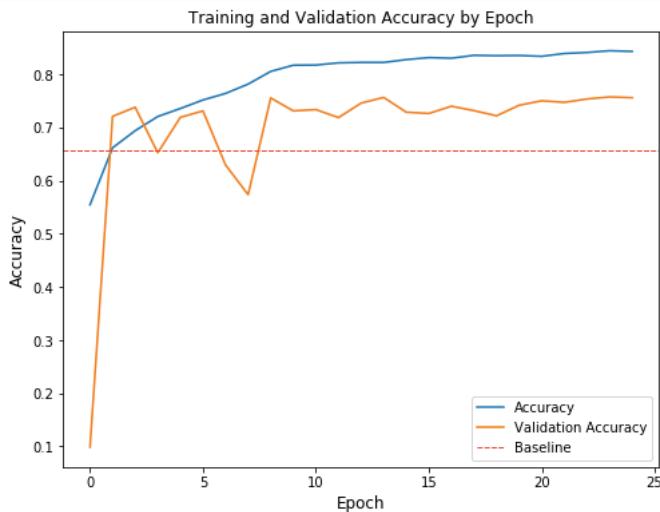


## TASK 02

### DIAGNOSIS CLASSIFICATION

3-way Classification: Benign, Malignant or Precancerous

# CONVOLUTIONAL NEURAL NETWORKS



75.8% >  
ACCURACY RATE

65.8%  
DERMATOLOGISTS'  
ACCURACY RATE

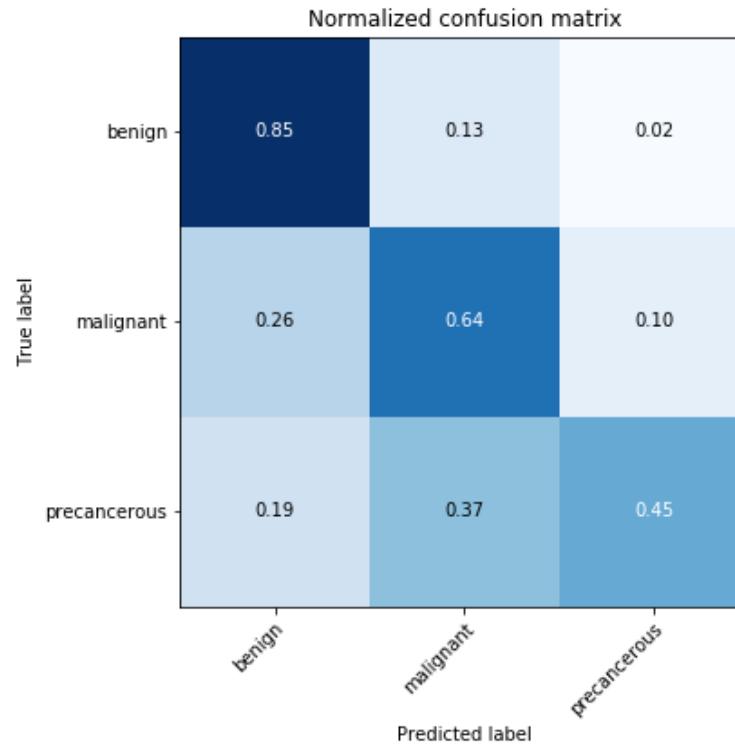
63.6% >  
AVERAGE  
RECALL RATE

49.2%  
DERMATOLOGISTS'  
AVERAGE  
RECALL RATE

# MODEL EVALUATION

	Accuracy	Average Recall Rate
Validation	75.8%	63.6%
Test	76.4%	64.4%
<b>Difference</b>	<b>0.6%</b>	<b>0.8%</b>

# MODEL EVALUATION



# MODEL EVALUATION

**76.4%**

ACCURACY RATE

**64.4%**

AVERAGE  
RECALL RATE

**63.5%**

AVERAGE  
PRECISION RATE

**64.0%**

AVERAGE  
F1 SCORE

# CONCLUSION

# CONCLUSION

- Moderately high accuracy at 70.5% for the skin lesion classification task, and 76.4% for the diagnosis classification task
- Recall rates surpassed baseline scores
- Performed well for most benign skin lesions (especially for diagnosis classification)
  - o Serve as great models specifically for benign skin lesions in the future

# RECOMMENDATIONS

1. Collect more data for minority classes
2. Build a web application for secure in-house usage at hospitals and skin cancer clinics
3. Add additional clinical data (demographic data of patients, and characteristics of lesions)
  - Feed into fully connected network
  - Merging outputs
4. Expand classification task to include other type of lesions
  - Non-neoplastic lesions (diagnosis classification)
  - Cystic lesions (skin lesion classification)

# THANK YOU