HAM10000: Prediction of common pigmented skin lesions

Pratyay Bandyopadhyay
Sandesh Kalambe
Satya Saurabh Mishra
Mitrup Kabi

Project Guide:
Dr. Amit Vishwakarma

Introduction

Dataset

The HAM10000 dataset consists of 10015 dermatoscopic images which can serve as a training set for academic machine learning purposes.

This dataset consists of 7 classes of different types of skin lesions namely ⟶

Pigmented Benign keratosis, Melanocytic nevi, Dermatofibroma, Melanoma, Vascular lesions, Basal cell carcinoma, Actinic keratoses.

Key Insights

- Small number of Missing Values for Age feature column where replacement with most frequent value(mode) was used.
- Similar distribution between Males and Females.
- Melanocytic nevi is the dominant class in the dataset (67%). It could result in a bias towards this type of skin lesion
- Most samples are from patients within 35 - 60 yrs old
- Melanoma, malignant skin lesion, seems to be more common in the ages of 45 to 70. Males represent 62% of the incidence of this type of lesion

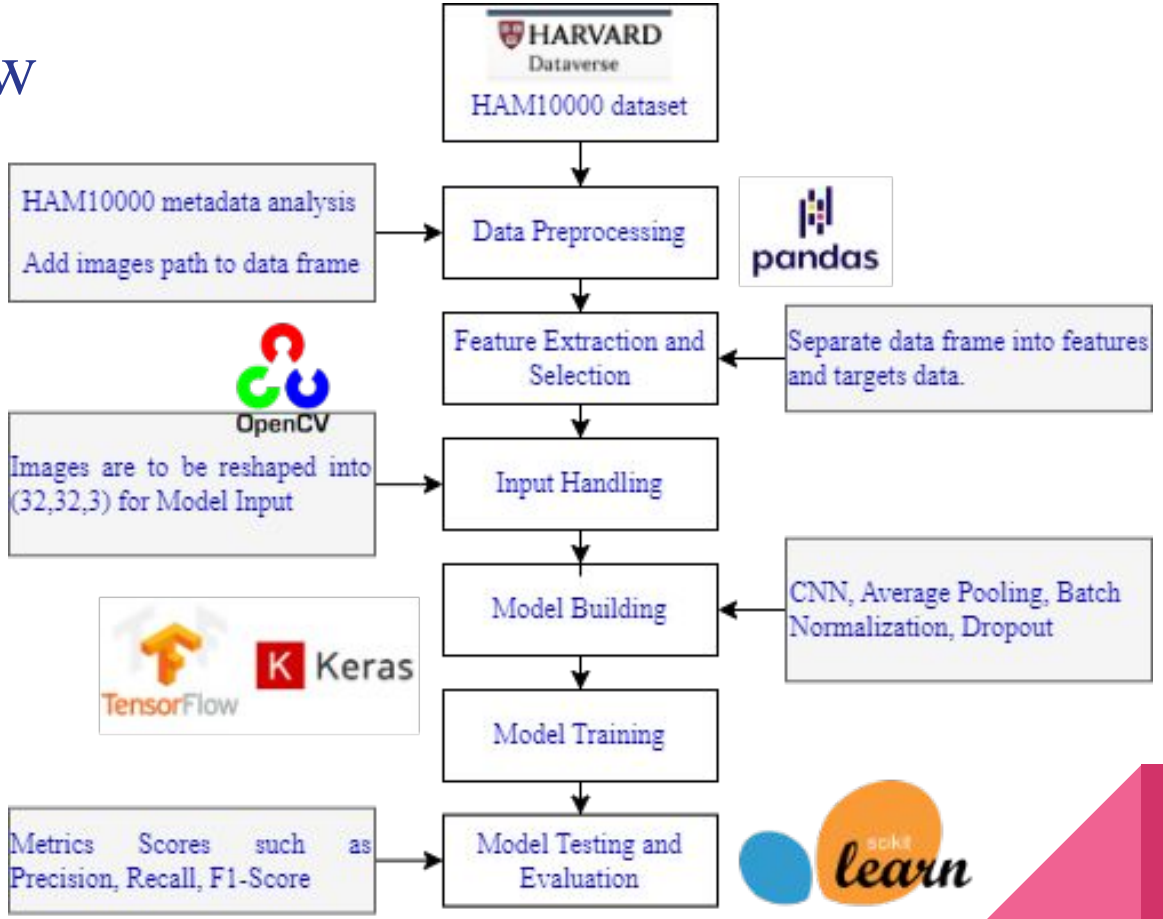
HAM10000 Metadata Analysis

The summary helps to understand the type of metadata collected.

We noticed that for some lesions there must be more than one image as the lesion and image ID do not match. The Uniques columns also indicate the number of classes (dx = 7), and how the age, sex and localization features were organized.

All the features are pretty much self-explanatory. To clarify, the **dx_type** column is the technique used to identify the type of skin lesion.

WorkFlow


HARVARD Dataverse
HAM10000 dataset
HAM10000 metadata analysis
Add images path to data frame
Data Preprocessing
pandas
Feature Extraction and Selection
Separate data frame into features and targets data.
OpenCV
Images are to be reshaped into (32,32,3) for Model Input
Input Handling
Model Building
CNN, Average Pooling, Batch Normalization, Dropout
TensorFlow
Keras
Model Training
Metrics Scores such as Precision, Recall, F1-Score
Model Testing and Evaluation
scikit learn

Data Preprocessing and Augmentation

- Add the images to the Dataframe
- Separate the data frame into Features and Targets data
- Create Training and Test sets (80 - 20 ratio)
- Normalise the training set data input.
- The test data should not be normalised, as it should remain unknown
- One Hot Encoding to transform the Target labels
- Separate the training set into Training and Validation sets (80 - 20 ratio)
- The CNN requires the images to be reshaped into 3 dimensions (height = 32px, width = 32px , channel = 3)

Model Building

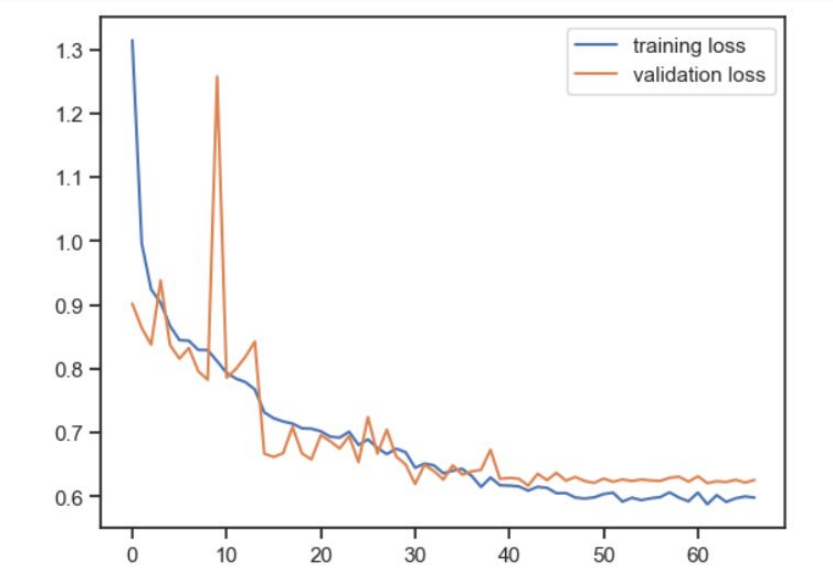
The Convolutional Neural Network is a specialised type of neural network, ideal for data that can be represented as a grid. CNN is most commonly used for image recognition tasks since this input can be perceived as a 2D grid of pixels. As described by Goodfellow et al., (2016), CNN are neural networks that use at least one of their layers the convolution operation.

For image classification tasks, Convolutional Neural Networks (CNN) are used most often.

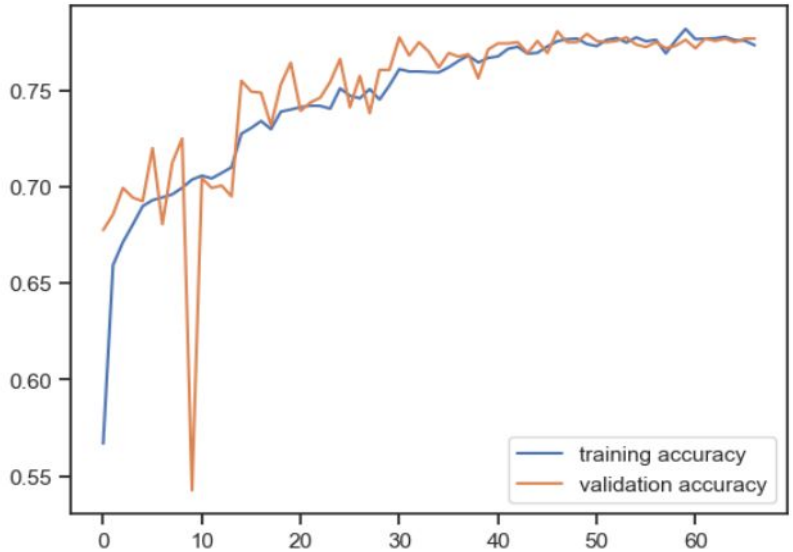
- We used **Average Pooling** in our model.
- A smaller pooling window (2,2) is used.
- We used **Batch Normalization** and **Dropout** to avoid overfitting.
- Convolutional Layer **>** Activation Function > Batch Normalization > Dropout was the best combination we found.

Training and Validation metrics during Model training

Loss Curve



Accuracy curve



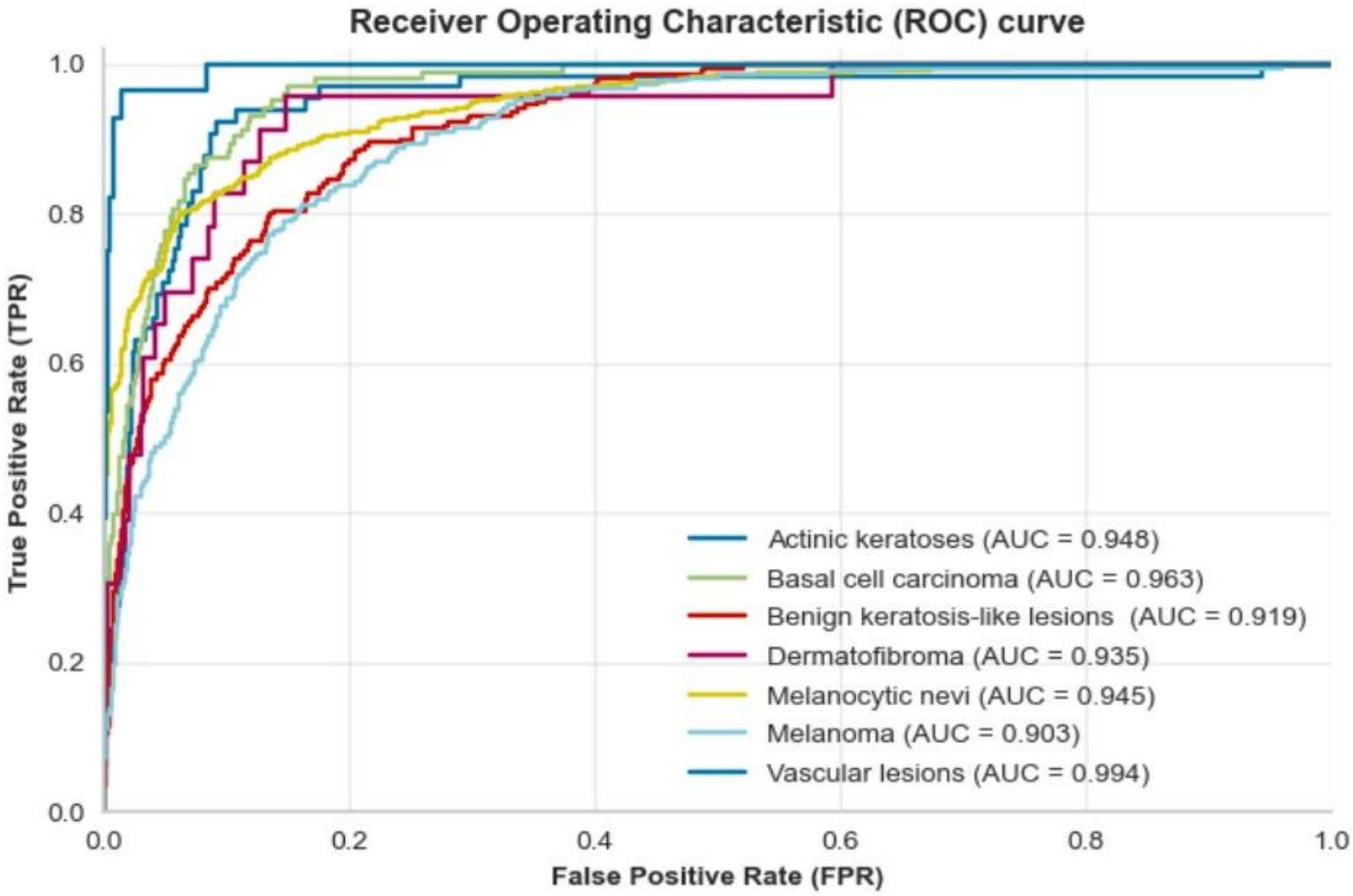
Model Testing

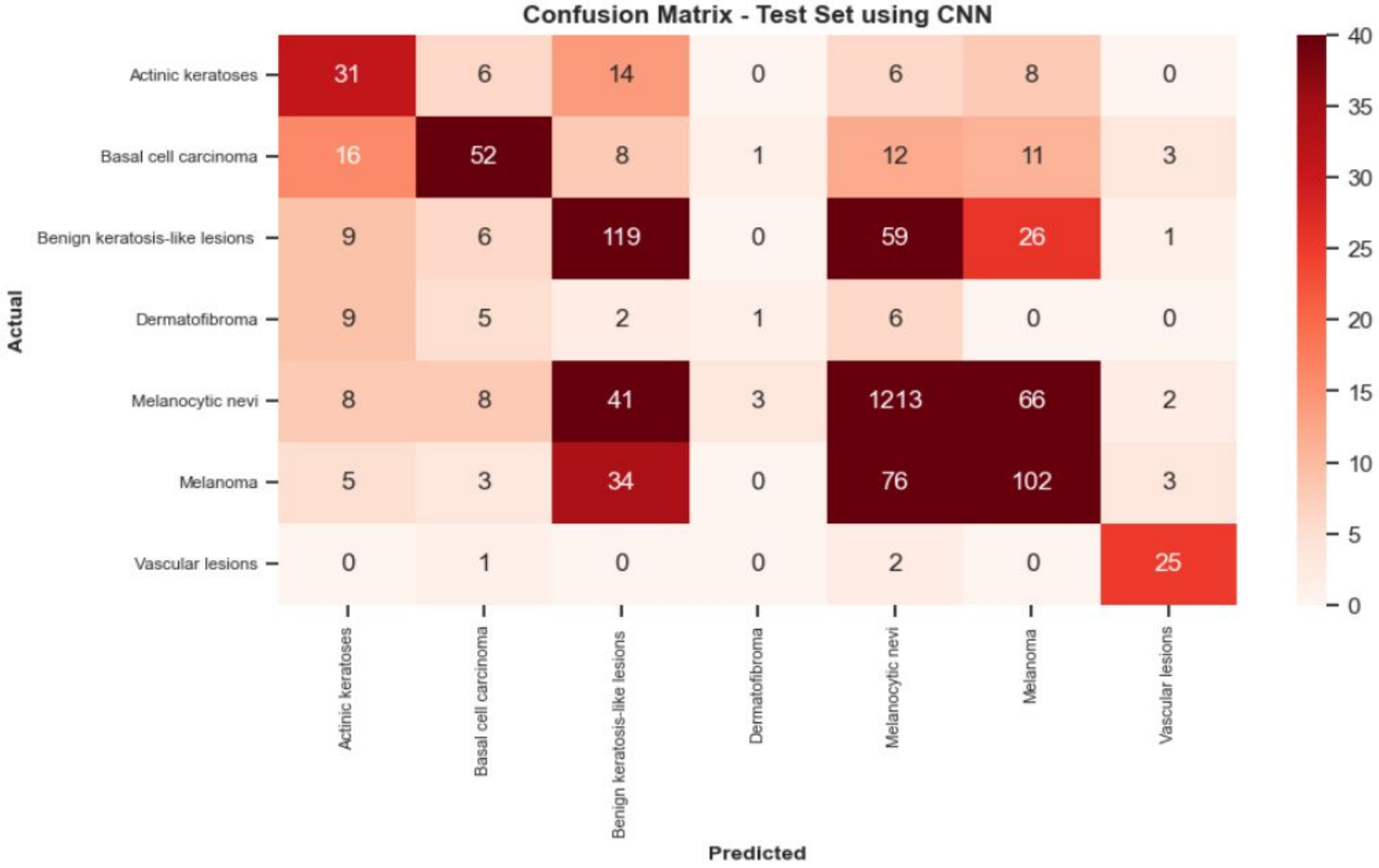
- The dataset was unbalanced, with the Melanocytic Nevi being the majority of the samples. For such cases, the accuracy metric can give us a false perception of the model reliability.
- Metric scores such as F1-Score, Precision, Recall can provide a better idea of our model behaviour.
- The Accuracy values are not as high as one wishes for cancer prediction. However, there is only a small difference between training and evaluation sets, which is a good indicator that the model is not overfitting

Performance Metrics

Classification Report

	precision	recall	f1-score	support
Actinic keratoses	0.40	0.48	0.43	65
Basal cell carcinoma	0.64	0.50	0.57	103
Benign keratosis-like lesions	0.55	0.54	0.54	220
Dermatofibroma	0.20	0.04	0.07	23
Melanocytic nevi	0.88	0.90	0.89	1341
Melanoma	0.48	0.46	0.47	223
Vascular lesions	0.74	0.89	0.81	28
accuracy			0.77	2003
macro avg	0.55	0.55	0.54	2003
weighted avg	0.76	0.77	0.77	2003


Receiver Operating Characteristic (ROC) curve
True Positive Rate (TPR)
False Positive Rate (FPR)
0.0
0.2
0.4
0.6
0.8
1.0
Actinic keratoses (AUC = 0.948)
Basal cell carcinoma (AUC = 0.963)
Benign keratosis-like lesions (AUC = 0.919)
Dermatofibroma (AUC = 0.935)
Melanocytic nevi (AUC = 0.945)
Melanoma (AUC = 0.903)
Vascular lesions (AUC = 0.994)


Confusion Matrix - Test Set using CNN
Actinic keratoses
Basal cell carcinoma
Benign keratosis-like lesions
Dermatofibroma
Melanocytic nevi
Melanoma
Vascular lesions
Actual
Predicted
31 6 14 0 6 8 0
16 52 8 1 12 11 3
9 6 119 0 59 26 1
9 5 2 1 6 0 0
8 8 41 3 1213 66 2
5 3 34 0 76 102 3
0 1 0 0 2 0 25
40
35
30
25
20
15
10
5
0

Conclusion

- From the Confusion Matrix, the main interest was to evaluate how the different skin lesion samples were being classified.
- An expressive number of samples were mistakenly classified as Melanocytic Nevi, almost the same number as the ones that were properly classified.
- It is not good that so many malign cancer samples are being classified as benign. For this reason, Accuracy is a dangerous metric for this application.

Future Scope

- Optimization methods to find optimal Data Augmentation, number of Convolutional layers and other hyperparameters.
- We can use additional data. The ISIC website contains additional pictures that could be used to improve the detection of the least represented classes.
- We can use **Generative adversarial network (GAN)** to generate more samples and improve model generalisation.
- As a final note, **Ensemble methods** are a good way to improve model accuracy.