ML PROJECT

Arham Ali, Abhyudit Badhul, Vikhyat Yadav

Problem Statement

Melanoma is a deadly type of skin cancer whose primary cause is Ultraviolet light (UV) exposure and it spreads quickly. As a result, it is the worst skin cancer and the leading cause of mortality. When a patient is diagnosed, the classification of cancer stages is a time-consuming and tedious task. Cancer diagnosis at the time of surgical therapy is primarily determined by the cancer's stage or tumour thickness.We broadly classify the melanoma into two categories namely benign melanoma and malignant melanoma. We aim to propose a model which can successfully classify the cancer stage backed by strong accuracy.

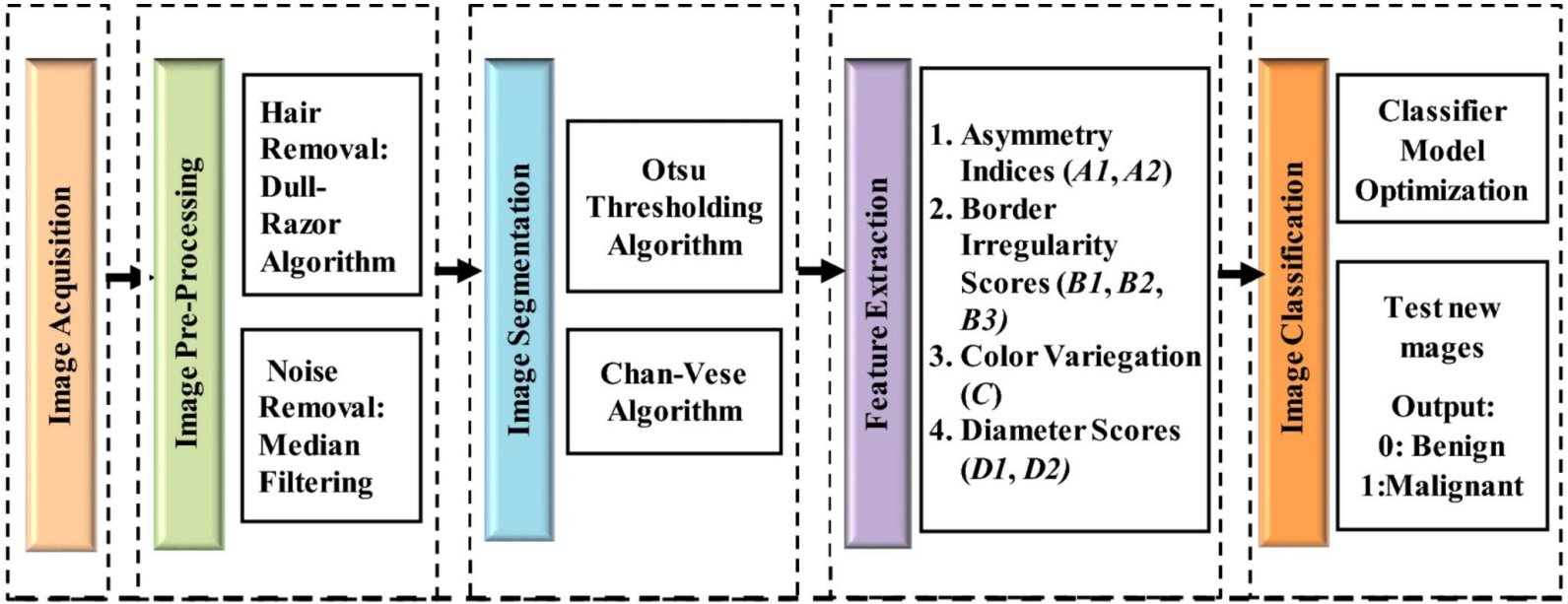
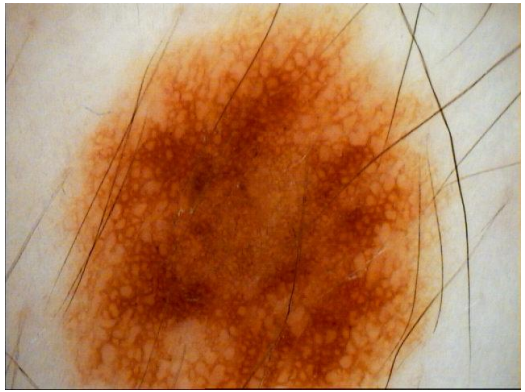


Image Acquisition
Image Pre-Processing
Hair Removal: Dull-Razor Algorithm
Noise Removal: Median Filtering
Image Segmentation
Otsu Thresholding Algorithm
Chan-Vese Algorithm
Feature Extraction
1. Asymmetry Indices (A1, A2)
2. Border Irregularity Scores (B1, B2, B3)
3. Color Variegation (C)
4. Diameter Scores (D1, D2)
Image Classification
Classifier Model Optimization
Test new mages
Output:
0: Benign
1:Malignant

DataSet https://www.isic-archive.com/#!/topWithHeader/onlyHeaderTop/gallery?filter=%5B%5D

First DataSet : 2000 Images of Benign and 2000 images of Malignant, manually extracting 7 features
Second DataSet : 3661 combined Images of Benign and Malignant having 80 features



Pre-Process

- Dull Razor -

  Convert the original image to grayscale, blackhat-> to enchance dark objects in the light background. (hair are dark, skin is light in color), intensify the hair countours in preparation for the inpainting algorithm.

- Median Blur -

  Adding median filter.

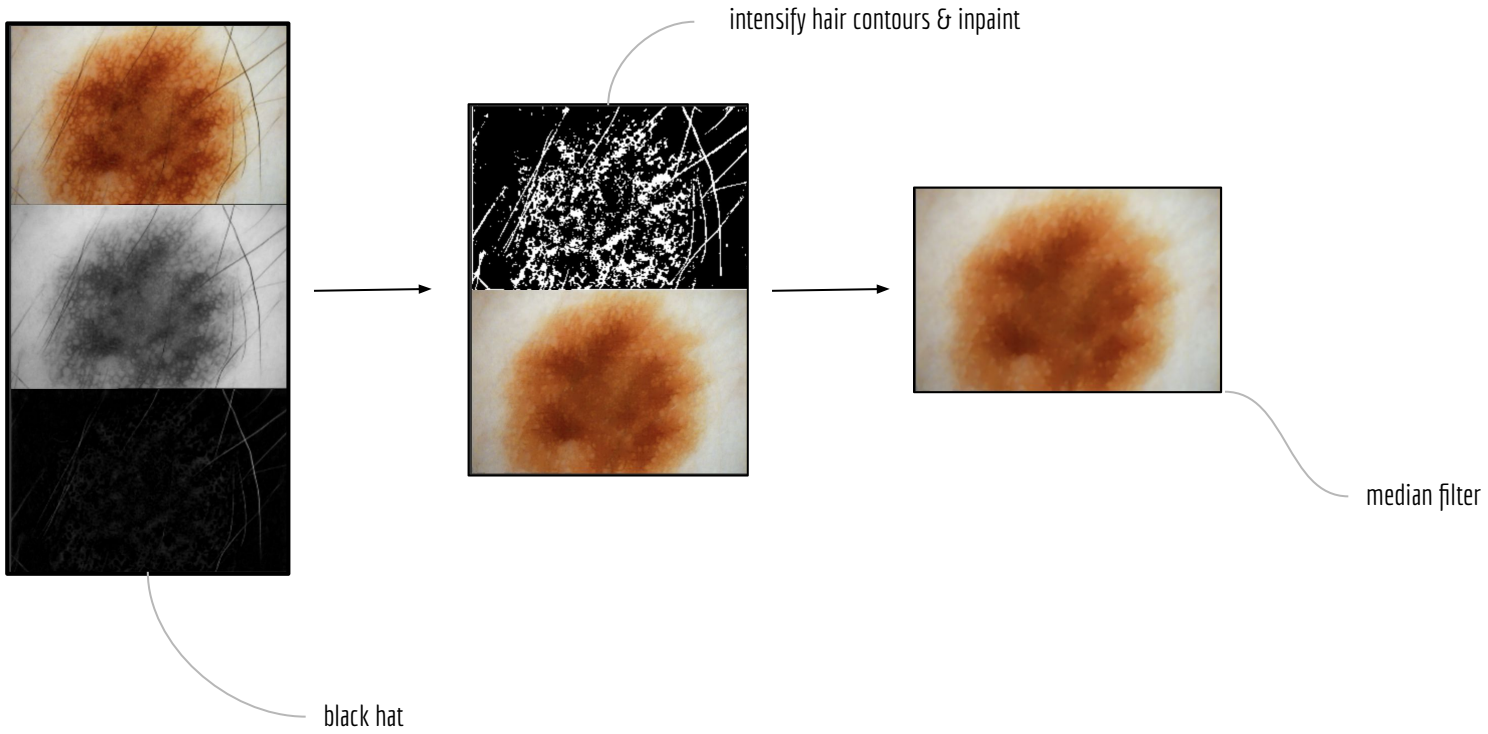
- Otsu Algorithm -

  used to perform automatic image thresholding.

- Chanvese Algorithm -

  used to segment objects without clearly defined boundaries.

Dull Razor Process:



Feature extraction

We have extracted 8 features :

- Asymmetry Index
- Eccentricity
- Border Irregularity
- Diameter
- Correlation
- Homogeneity
- Energy
- Contrast

Basic and Advanced Baseline Approaches

Model Used

- Logistic Regression
- Random Forest
- MLP Classification
- KNN Classification
- Stacking Classification
- CNN with Binary Cross Entropy Loss

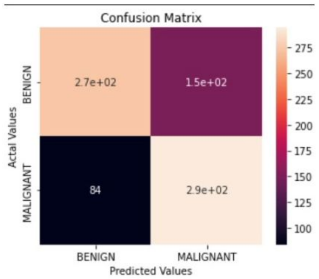


Figure 2. Confusion Matrix for logistic regression.

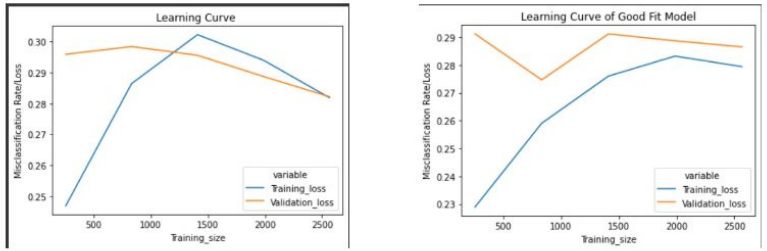


Figure 3. Loss vs Experience for logistic regression.

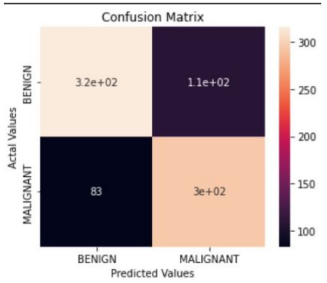


Figure 4. Confusion Matrix for Random Forest.

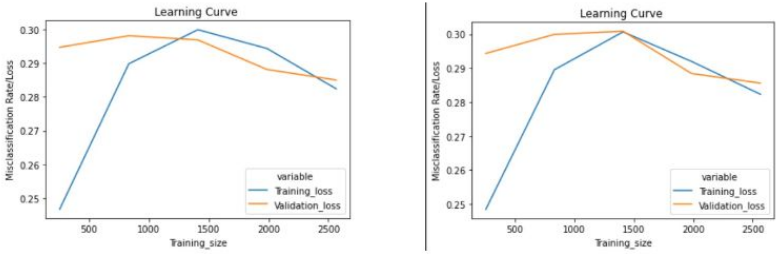


Figure 5. Loss vs Experience for Random Forest.

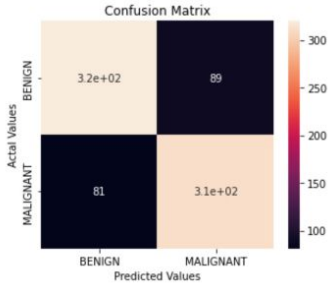


Figure 6. Confusion Matrix KNN

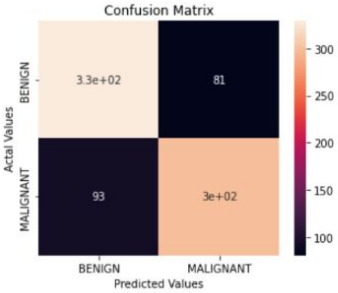


Figure 7. Confusion Matrix MLP

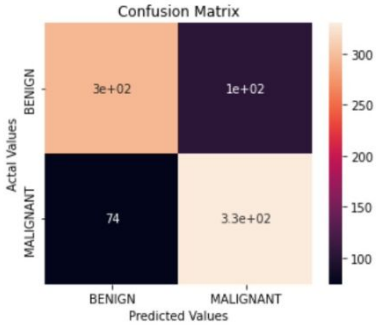
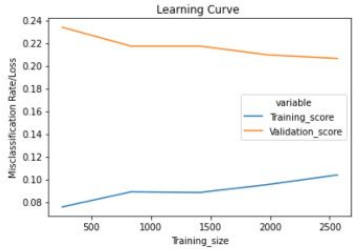


Figure 9. Confusion Matrix Stacking



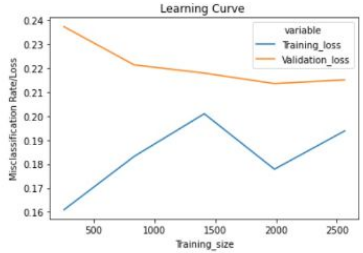


Figure 8. Loss vs epoch (i)KNN on left (ii)MLP on right

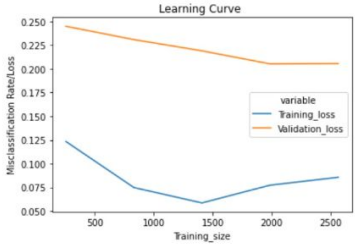


Figure 10. Loss vs epoch Stacking

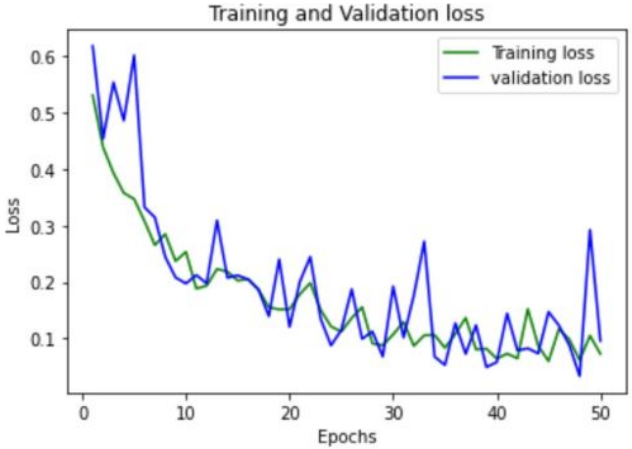


Figure 12. Loss vs epoch curve

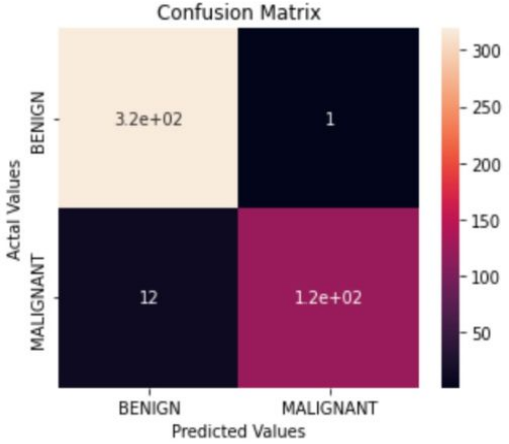


Figure 11. Confusion Matrix CNN

Model	Accuracy	Precision	Recall	F1
Logistic Regression	0.7319	0.7341	0.7319	0.7321
Random Forest	0.7618	0.7629	0.7618	0.7616
MLP Classifier	0.7830	0.7836	0.7830	0.7831
KNN Classifier	0.7643	0.7643	0.7643	0.7643
Stacking Classifiers	0.7793	0.7820	0.7793	0.7796
CNN 7 features	0.7755	0.7804	0.7755	0.7760
CNN 80 features	0.9713	0.9734	0.9713	0.9717

Contribution

Delivered
Arham and Vikhyat Collected Data and pre-processed it using multiple techniques to remove hair and other noise from the image. Vikhyat and Abhyudit visualised the data and Feature Extraction selection. Abhyudit and Arham per- formed EDA Training model and different Ensemble techniques for better accuracy.

Not Delivered
CNN with SMTP Loss function is not delivered, instead we did it using binary cross entropy

Individual Contribution
Arham : Data collection , Pre-processing & Applying Models likes Logistic Regression, Random Forest, KNN.
Abhyudit : Pre-processing, Feature Extraction, Stacking Classifier , MLP Classifier.
Vikhyat : Pre-processing, CNN with binary cross Entropy, K-fold cross validation.

Thank You