

Problem Solution Fit

Team id	PNT2022TMID42059
Name	D.BOOBALAN
Reg.no	623119104001

Analysis methodology

Our 2-phase analysis model for localization and classification is shown via the pseudocode in Algorithm. We decomposed the original image into its haemoglobin and melanin constituents using pre processing, to help our model extract valuable information from data that would have been otherwise unavailable. We provide these images as input to our segmentation model, the U-Net, which generated a segmented image. This segmented image was then analysed for clusters, which were subsequently cropped and input to our classification model, the Efficient Net, which then produced a classified label, thus completing our analysis model. The data for training and testing were obtained from Denet NZ, an archive of skin disease information launched and maintained by a group of dermatologists from New Zealand. The site provides open source images with labels. We selected 18 top-level categories each of which included enough data, besides including erythema as one of its common symptoms. Using a web crawler, we gathered a total of 15,851 images. Among the images obtained through Denet, the erythema of 100 images was masked by dermatologists, to be used as a ground truth. For segmentation, 60 images were used for training, and 40 images were used for testing. For classification, 13,473 images were used for training, and 2,378 images were used for testing. In addition, the test set for classification was split before segmentation cropping to prevent the subsections of one image from appearing in both the training and testing sets. Table 6 shows the distribution of data in greater detail. We chose the 100 images for segmentation in a balanced manner from each class, to minimize any bias that could occur during the classification phase. One of the significant merits of the Denet dataset is that it was created and is maintained by a diverse group of dermatologists. The images in each top-level category are independent as they are images of different patients at distinct locations taken with varying devices. This is evident in the diverse resolutions, lighting, and aspect ratios of the images. Regardless, it would be optimal to possess a similar dataset from an entirely separate association to truly validate the performance of our model.