Deliverable 3 (Gargi Singh and Jehan Dastoor)

Final Training Results (Dataset: https://challenge2018.isic-archive.com/task3/)

From Deliverable 2, a few notable changes have been made to the model.

1. An attempt was made to use a pre-trained model, however this seemed to always perform worse likely due to how specialised this classification task is. Even if it did perform better sometimes, the difference was that of 0.67 to 0.68, and so it was decided to use the original model and just try to tune that better.
2. The image sizes used were changed from 64x64px to other larger sizes. However, these seemed to perform just as well and only took longer to train. As such, the decision was made to stick to 64x64px.
3. Image augmentation was used to augment the original images slightly to give the model more robustness, i.e. including rotations, width and height shifts etc. However, originally this was done incorrectly as the modifications were also applied to the validation set, which is not what we want. This was changed so that augmentations only apply to the training set and not the other way.
4. The training was changed to include more epochs (50 with early stopping around 25 normally occurring) and also reduced the learning rate if the learning plateaued by a factor of 5.
5. The model was changed to have 3 layers with Conv2D (64, 128, 128 3x3 filters for each layer), Activation (relu), Max pooling (2x2). Then a flattening layer. I added another Dense Layer in between with 256 nodes. So a total of 3 dense layers, 256 (relu) then 64 (relu) then 7 (with softmax activation). Base learning rate was reduced to 0.0001.

One thing to note is that although there are 7 different classes for the type of lesion. Majority of the data falls under the label “NV”. As such, the model seemed to almost only predict “NV”, so changes were made. As such, the model might seem slightly worse performing on the data, however this is due to the mistake that was mentioned in 3. and the fact that the data being inputted does not accurately represent the other groups.

In order to provide a point of comparison, the Figure below shows the preliminary results given in the last deliverable.

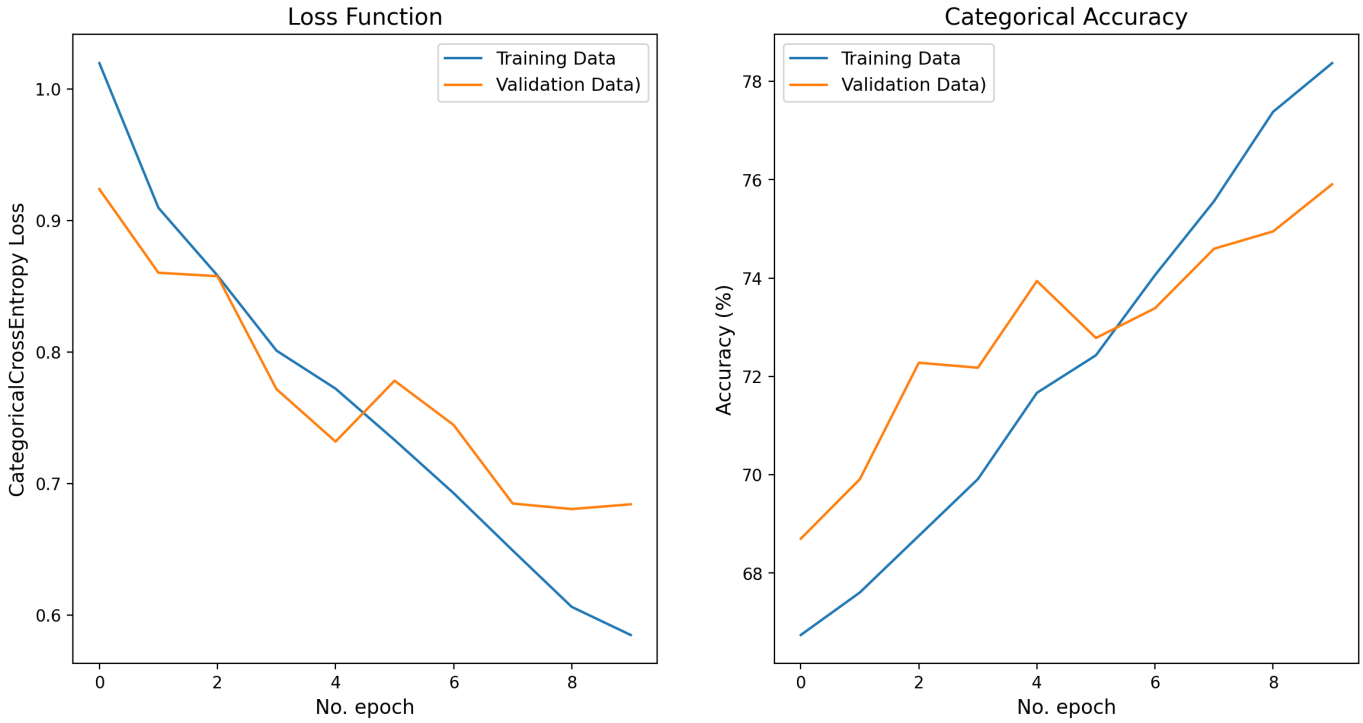


Figure 1: Preliminary plots for the accuracy of the model over the number of epochs trained for the validation and training dataset. Results have only been shown for the model that detects lesion type, not diagnosis type (even though the diagnosis accuracy is higher).

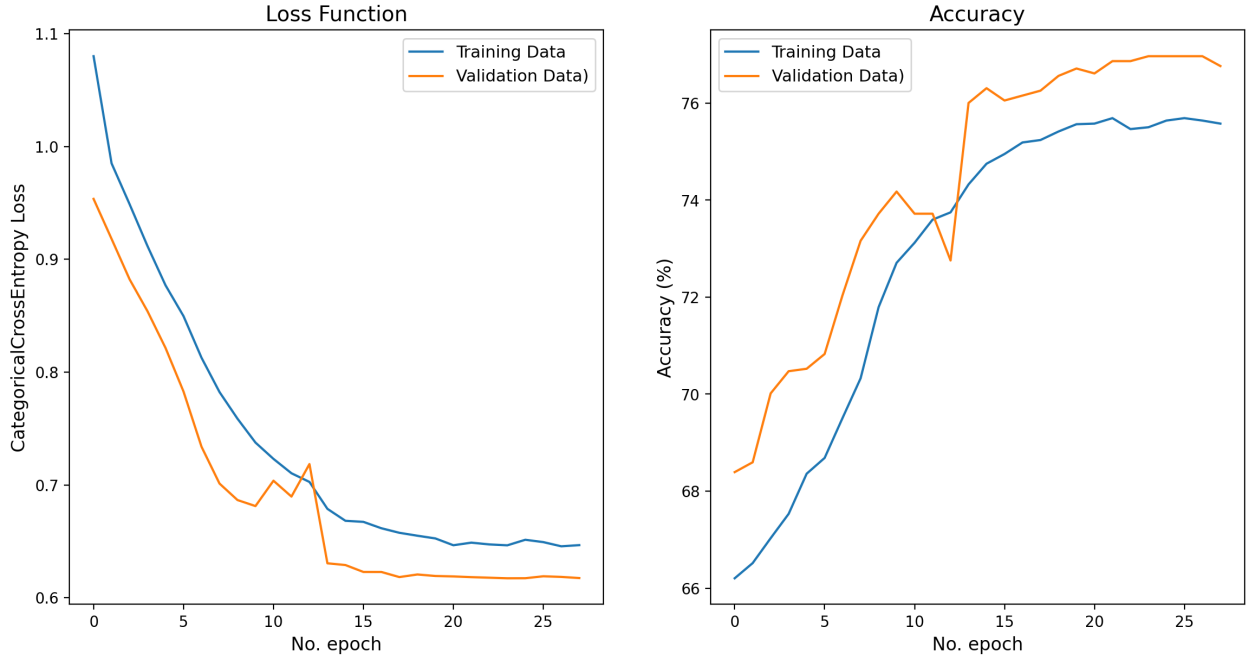
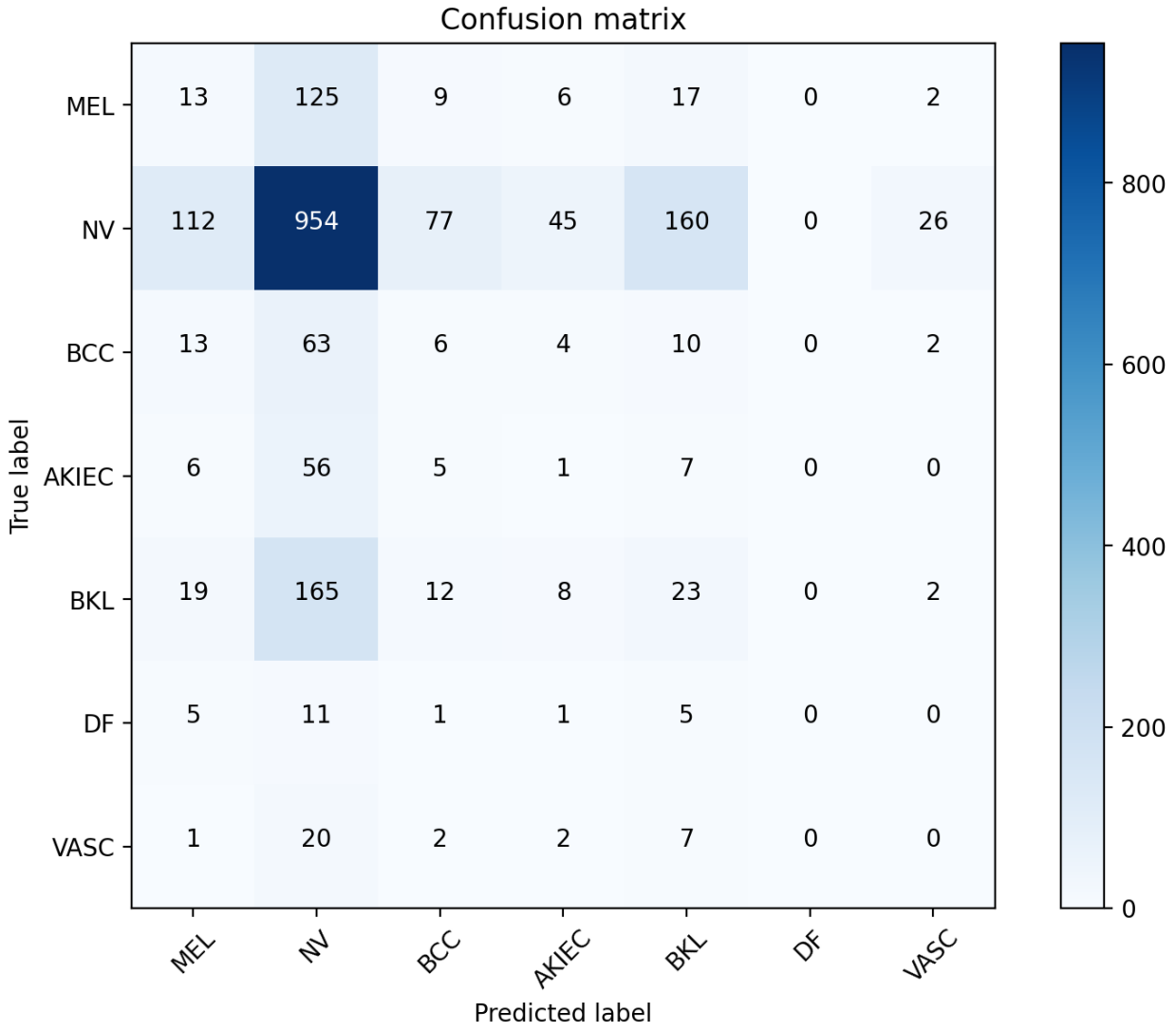


Figure 2: Improved model plots for the accuracy of the model over the number of epochs trained for the validation and training dataset. Results have only been shown for the model that detects lesion type, not diagnosis type (even though the diagnosis accuracy is higher).

The confusion matrix is shown below,



The confusion matrix is not diagonal as we would want. It seems to preferentially predict "NV" as this is what most the images are, even when the true label is different. We see that the model never predicts "DF" when the true value is "DF" and tends to preferentially predict "NV" as we would expect given our dataset. Sadly, this seems to be the best that I could get after what seemed like an entire weekend of tuning parameters, trying padding, strides, more layers, pretrained models, different learning rates and so on. With a better data set this model might perform better, however it would also take much longer to train (the 2020 dataset I think has over 30k images compared to 8k now). Notably, the validation set is performing better than the original which is a good sign.

Final Demonstration Proposal

The goal is to make a React.js with Flask website that can run on my local machine. Jehan has experience using Figma to design websites and integrating them in React using the Anima extension from Figma. We do not have much experience in React, but I have example projects completed in past hackathons that I can look at, as well as online tutorials and documentation that I can learn from. Due to my lack in experience, the website will be fairly basic and will have brief intro text discussing what the website will do and how our model was trained. The user then uploads their photo (at least larger than 222x222px) and our website will predict what kind of skin lesion is shown in the picture and the best method for diagnosis of this type of skin lesion. As well as this, the website will also show the user the accuracy of the prediction so that they can judge best whether to believe the model or not. If time permits, the website can be made to look cleaner and include appropriate animations when the prediction is being made/the image is being uploaded.