**Baseline Implementation Documentation**

Baseline pipeline

The chosen dataset ([ISIC 2020 Challenge Dataset](https://challenge2020.isic-archive.com/)) has been provided by Kaggle, well established data science community, and was already half ready for the training, hence there has not been much what could have been done with data preparation. However, the team has decided to first resize the pictures and grey scale them.

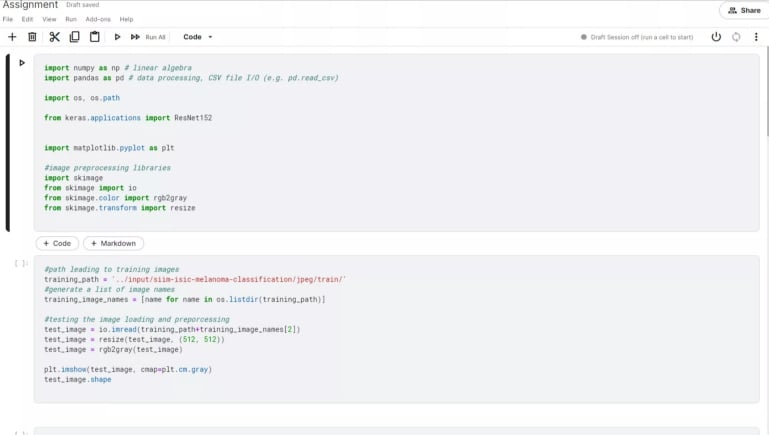
resize, greyscale, normalise values and reshape

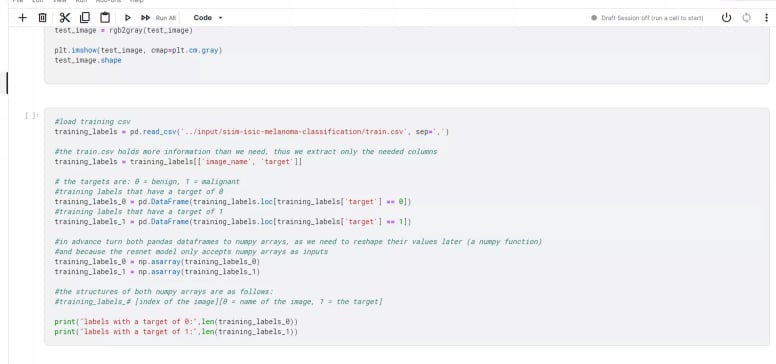
load images

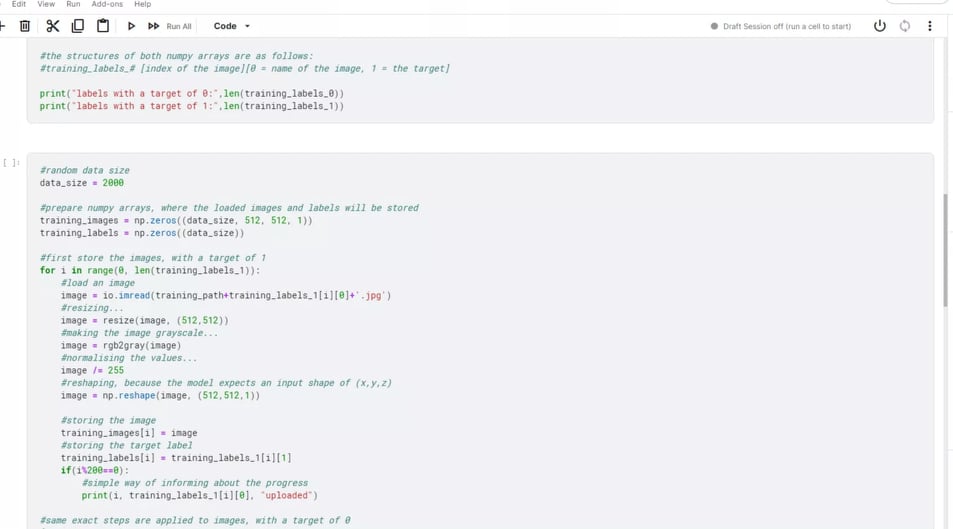
predict

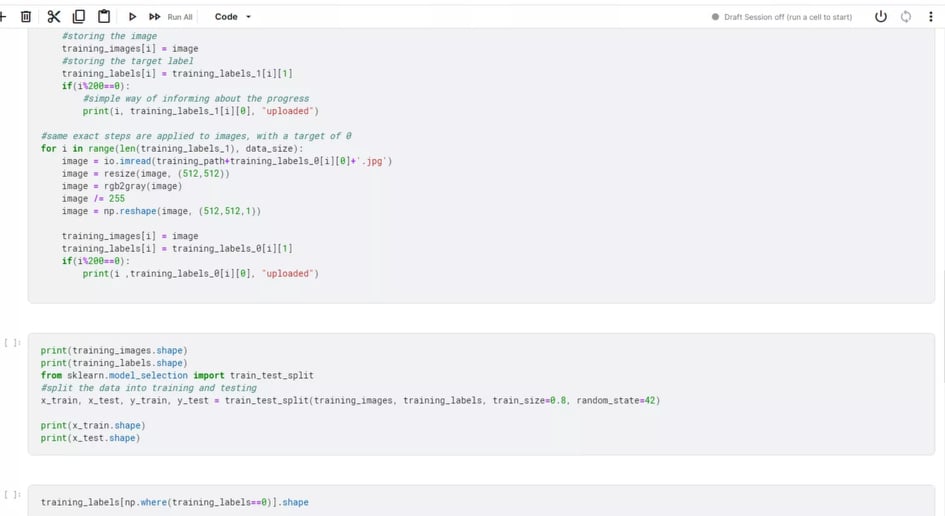
train

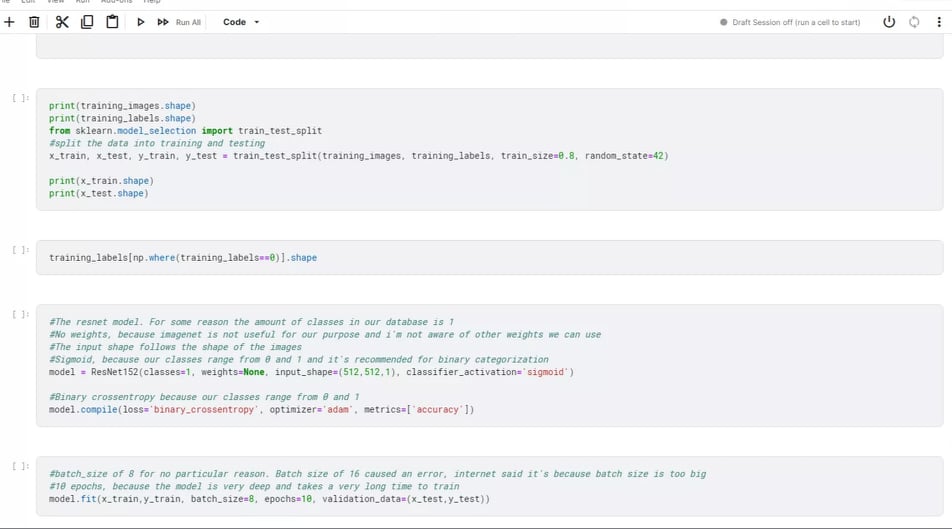
Snapshots of baseline solution (the code)

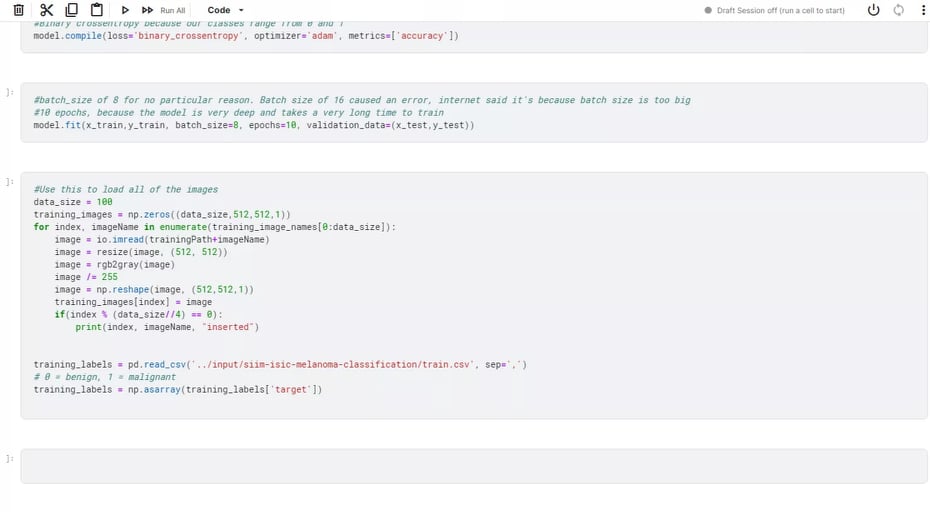












The code was written on Kaggle. The dataset used is ISIC 2020 Challenge Dataset which is provided by Kaggle. For easier and faster access, the code has also been uploaded on our group’s GitHub repo.

Baseline Evaluation

The goal of this project is to make an application that would help in diagnosing skin cancer. The idea was that it would scan the pictures of the skin and would detect if the skin shows symptoms of skin cancer. The baseline solution written tried to do exactly that.

In the code written, it is seen that both pandas data frames are turned to numpy arrays in advance because there was a need to reshape their values later and also because the ResNet model only accepts arrays as inputs.

The pictures in the dataset were already pre-processed. Then for the basic processing, the pictures had to be resized, decision was made to resize them to 512x512, this was necessary because the images in the dataset varied in size. Then the images were turned into greyscale pictures. Following that, the pictures’ pixel values were normalised. This is all good practice because by doing this processing, the model’s accuracy should increase. The model used was Keras ResNet152 for the sake of simplicity.

In the code written, it is seen that both pandas data frames are turned to numpy arrays in advance because there was a need to reshape their values later and also because the ResNet model only accepts arrays as inputs.

The value accuracy did not get above 0.6875. The team decided that it was not too bad for a start as the code would be altered in the future for better performance. The team decided to not settle with these results and while using the baseline solution, to vary data size and other parameters to reach better results. The group has also decided that maybe the chosen ResNet was not needed for the amount of data the group had and decided to try out other ResNets, such as ResNet50 or ResNet100.

The baseline solution, while it was good enough to start to code the project, it also brought a lot of doubts and confusion on why some things are working while others are not working as well.