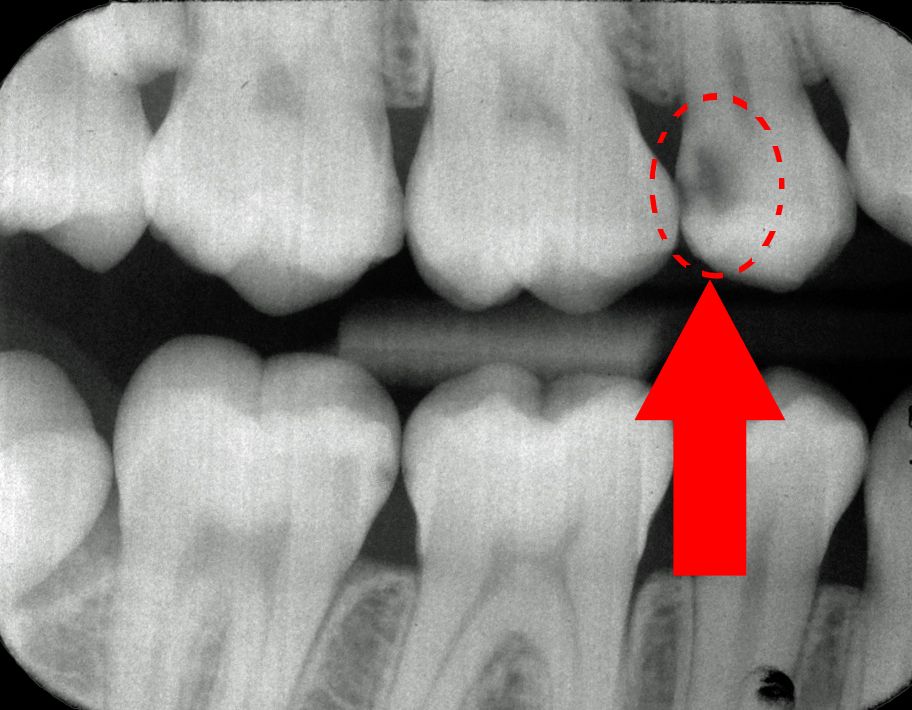
# Identifying Dental Cavities on Bitewing Images

Dental cavities are one of the most common diseases that people experience. Dental cavities can be diagnosed on x-rays and intraorally depending on the location of the cavities. For cavities developed in between adjacent teeth, they may not be clinically visible due to the location. They commonly appear as radiolucencies or dark shadows on certain areas of dental bitewing x-rays. The goal of this project is to train models to identify dental cavities on x-rays using bitewing images collected from Google image search. The model could help dentists save time by pre-screening the x-rays, and it could help patients better understand their own health.

1. Data collection

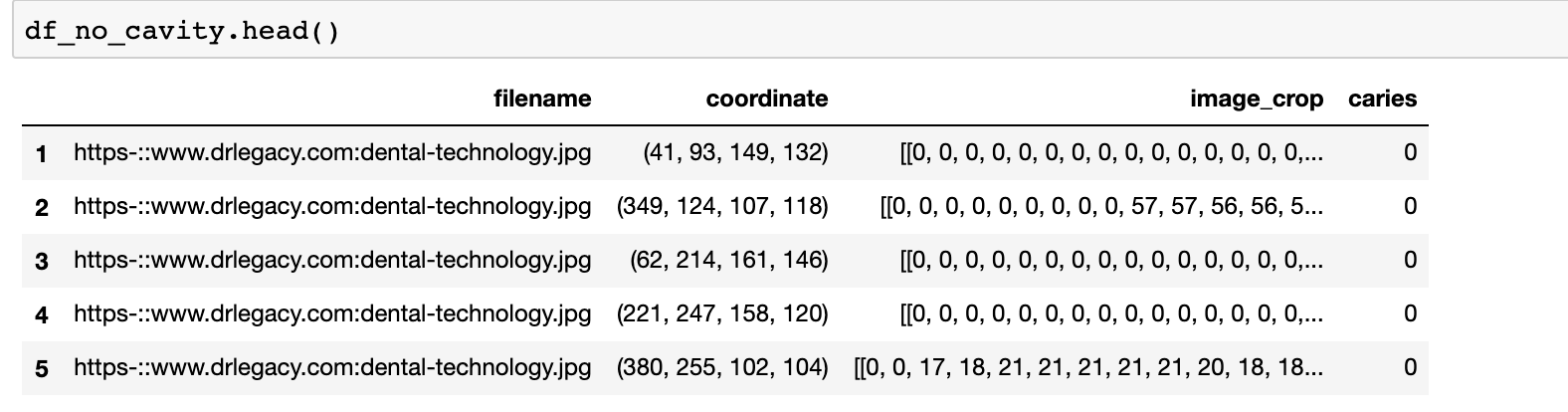
Bitewing images are collected from Google image search with keyword 'dental bitewing x-ray'. Unique bitewing images are saved and labeled with the name of the original website. The images are randomly selected into training set and test set. Training images that contain teeth with one or more cavities are saved in the folder named 'cavity'.



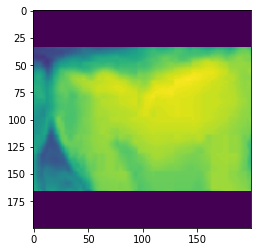
Sample dental bitewing x-ray with circled one area of cavity.

Open CV was used to label cavity and non-cavity on x-rays. For the 'cavity' folder, part of each tooth containing cavity is selected, and saved into a dictionary. If there is other cavity on the same image, any key other than the 'return' key is pressed to allow selection of a different part of the image. If there is no other cavity on this image, 'return' key is pressed and the next image will appear. After the last image is selected, the selections are saved into a dictionary. The cropped images are then resized and bordered to become square shape.

The dictionary is then saved into a pickle file. The dictionary contained 'r' - coordinate or location of the cropped image, 'filename'- original image name, 'image\_crop' - cropped image, 'caries' - 1 indicating cavity and 0 indicating non-cavity.

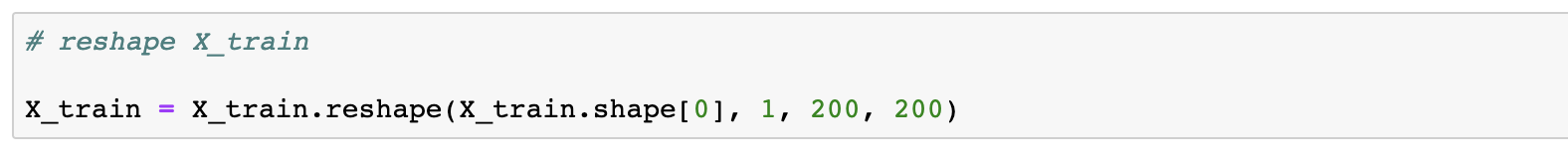


A total of 100 cropped cavities and 100 non-cavities were collected in the training set.

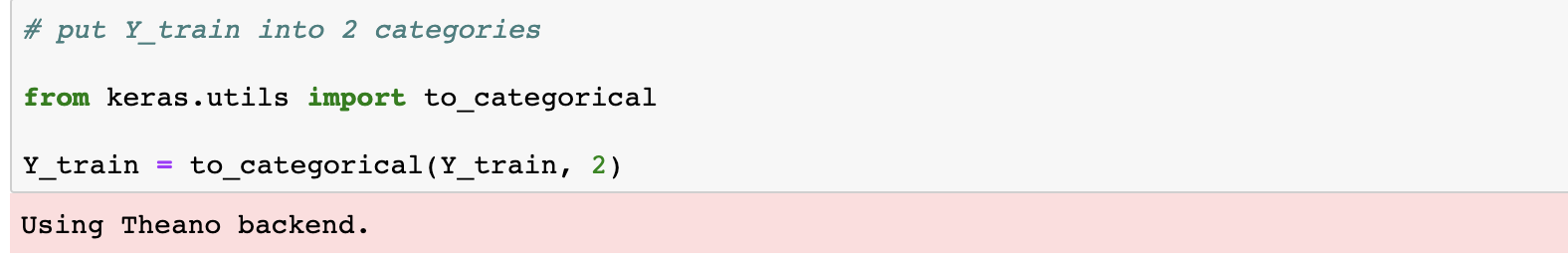
Cropped image containing cavity in between the teeth on the left side

1. Data processing

X\_train contains the cropped image with and without cavity. The dataset is reshaped.



Y\_train is the corresponding label of the cropped image. Y\_train is put into categorical. Similar processing is done for test set.

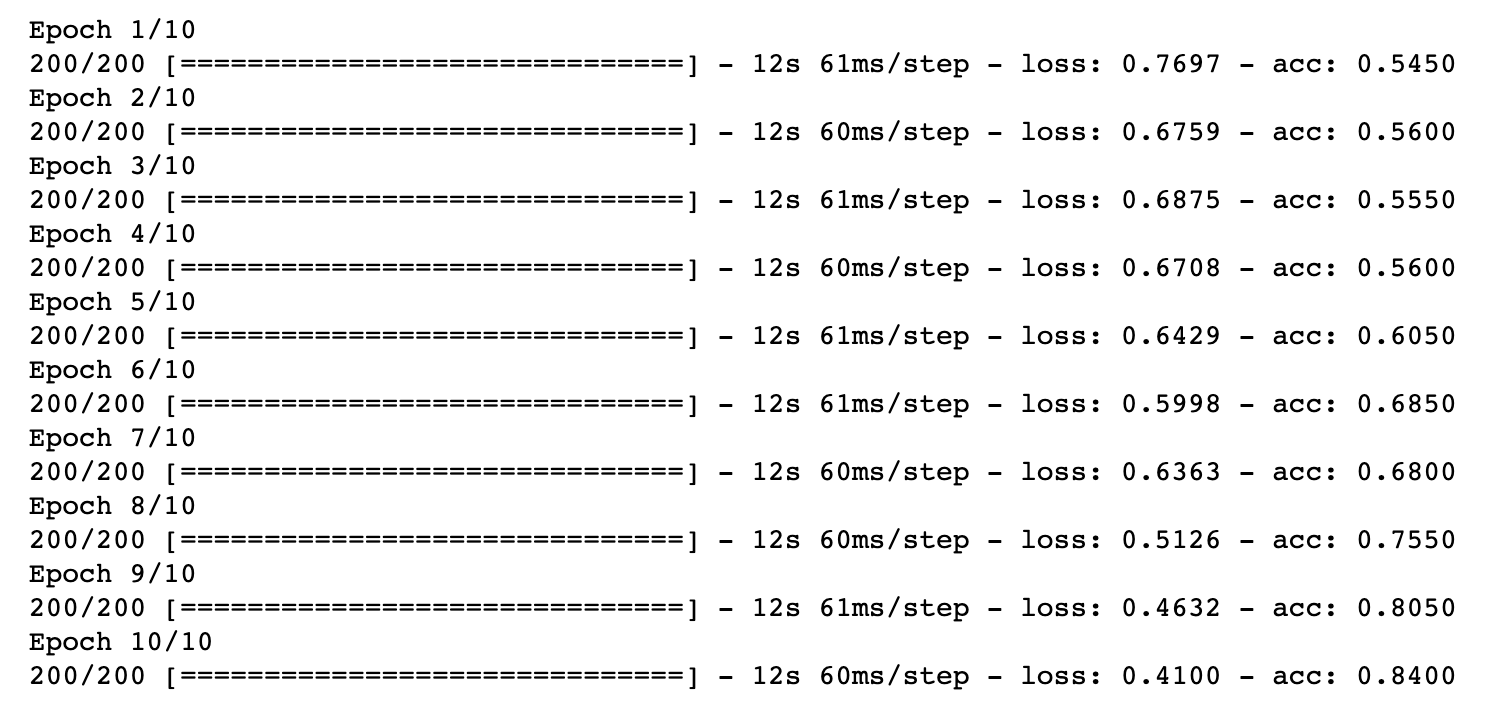


1. Modeling

Modeling directly on this small dataset was performed first using Sequential model.



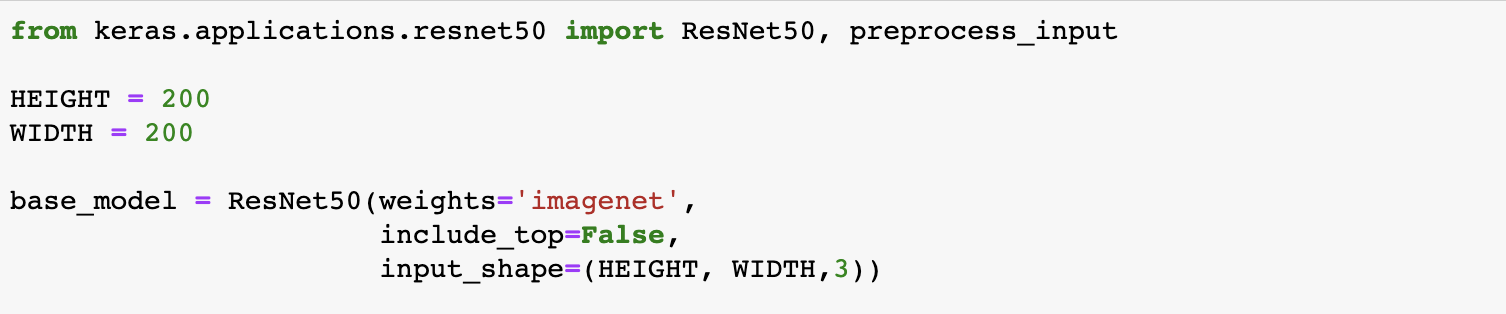
The following is the result:

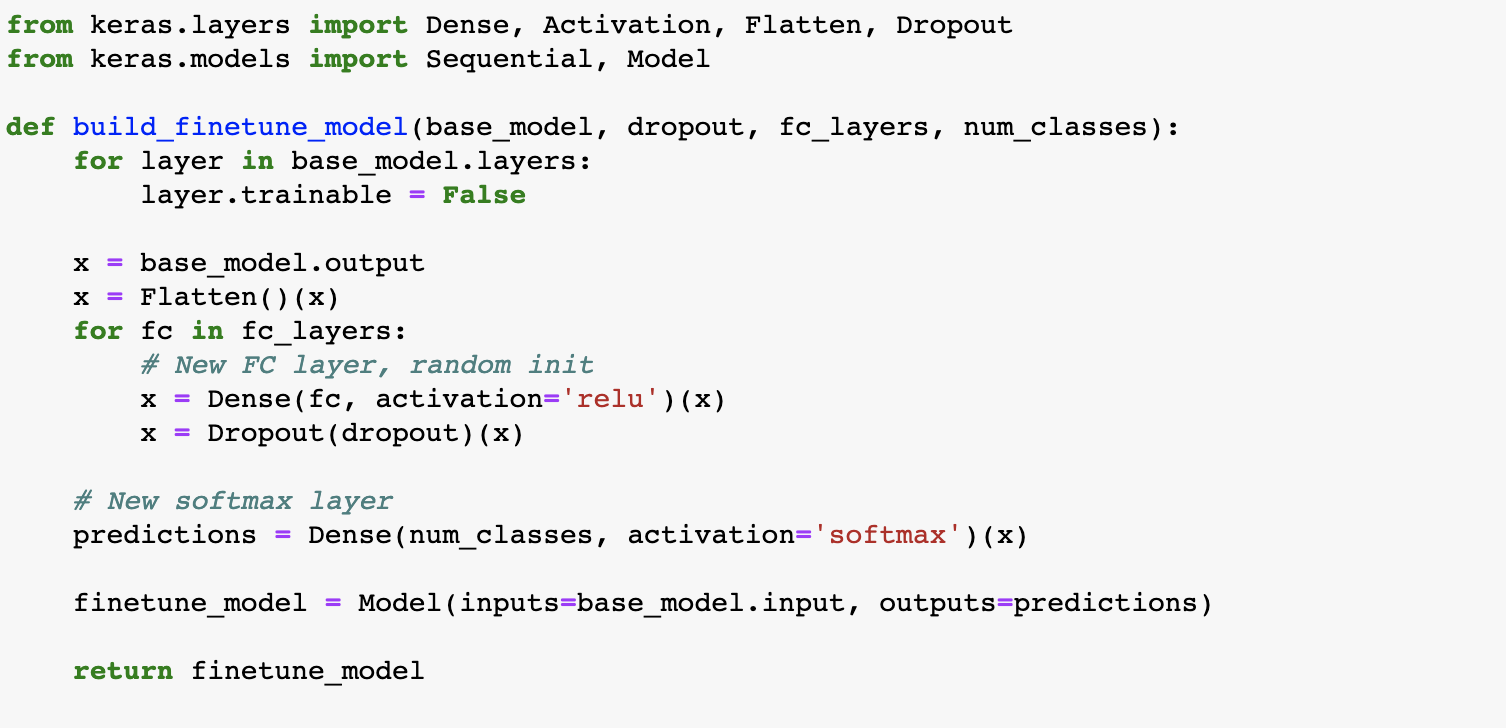


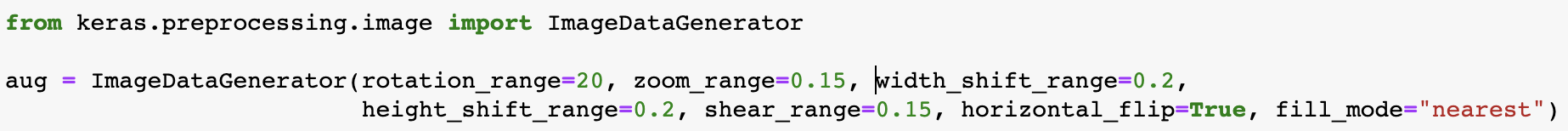
Towards the last Epoch, the accuracy has increased to 0.84, which is very accurate considering the small size of the data set. However, checking the test set gives 0.5 accuracy. In the test set, half of the images are cavity and the other half are non-cavities. The testing result suggests that the model is labeling all images as either cavity or non-cavity. The model is clearly overfitting for the training set.

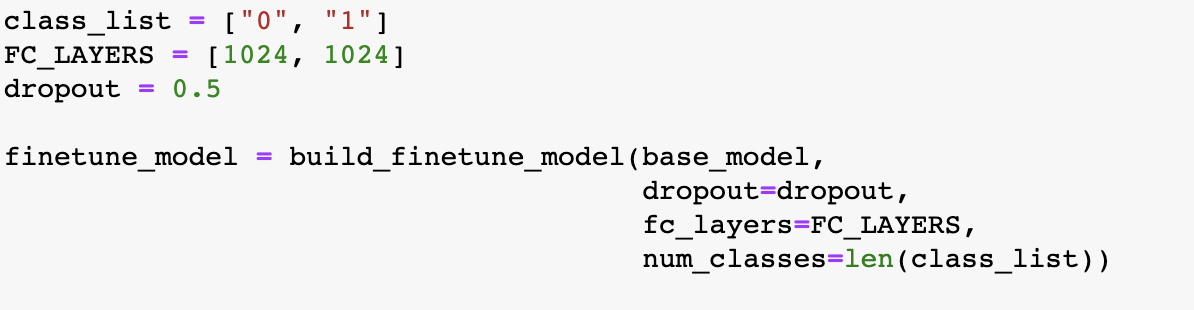
1. Transfer learning using ResNet50

Due to the small image size, we next tried transfer learning using a commonly used pre-trained model ResNet50.

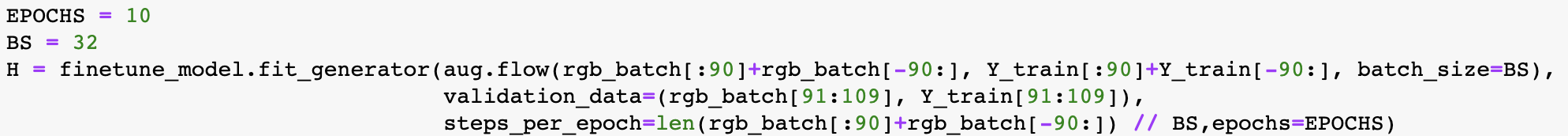




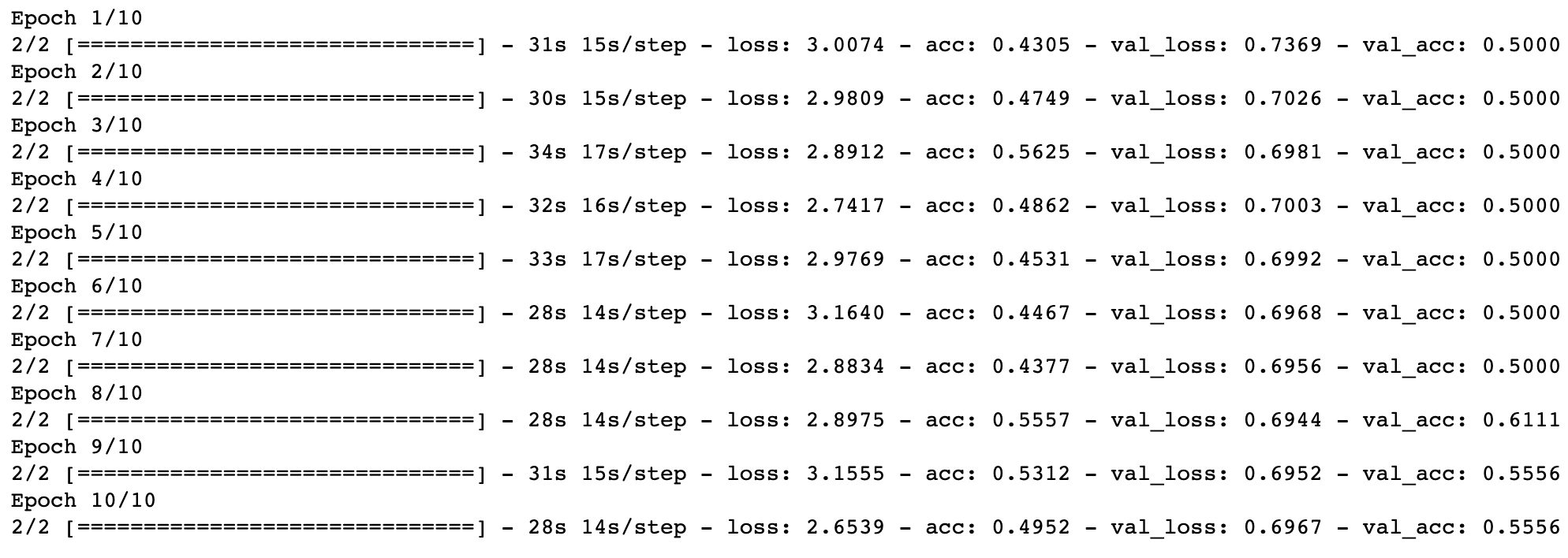








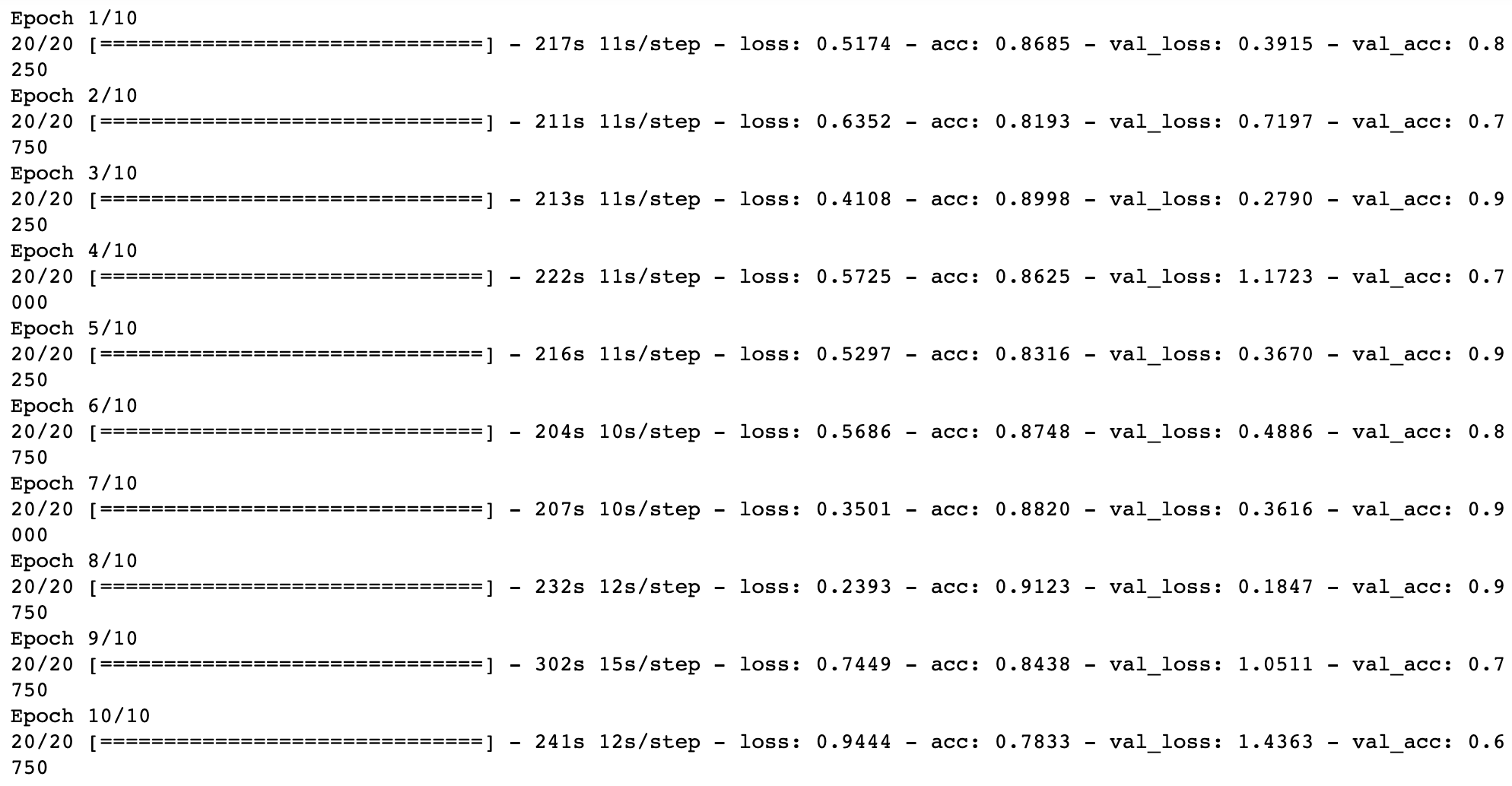
The following is the result:



The result also suggests overfitting of the training data and labeling all testing data into one category. This is likely due to the dataset was too small, the cavities were not very apparent for the models, or the pre-trained model doesn’t work well with x-ray images.

1. Cats vs dogs on ResNet50

In order to check that the model doesn’t work well with x-ray data but could work with other types of images, we use the same model on cats vs dogs dataset of the same amount. From Kaggle we obtained 200 images of cats or dogs. Here we used the same ResNet50 pre-trained model to identify cats vs dogs.



The results show improvement compared to the x-ray data. It suggests that the cavity data could be too difficult to identify, or ResNet50 is not working so well with x-ray data.

1. Future direction

In the future I will try transfer learning using a model that has been pre-trained on x-ray data. The models will learn features of x-ray data that could be helpful for the cavity dataset.