

Master of Technology (IS)

Continuous Assessment 2

Project Report

Team Member:

CAO LIANG A0012884E

TAN CHIN GEE A0195296M

GENG LIANGYU A0195278M

Contents

[Executive Summary 3](#_heading=h.gjdgxs)

[Business Problem Background 3](#_heading=h.30j0zll)

[Project Objectives & Success Measurement 3](#_heading=h.1fob9te)

[Project Solution Design 3](#_heading=h.3znysh7)

[Project Implementation 4](#_heading=h.2et92p0)

[Project Performance & Validation 14](#_heading=h.tyjcwt)

[Project Conclusions: Findings & Recommendation 15](#_heading=h.3dy6vkm)

# Executive Summary

This project constructs deep-learning model to determine the public domain images classification on cat, bird and dog.

# Business Problem Background

The project requires the students to not use the pre-built image dataset from the public sources such as Kaggle, and the images should be downloaded from public domain without royalty and copyright. Also the number of images for each class should be at least 1000.

The project also requires the students to use only deep-learning model to resolve the classification problem.

# Project Objectives & Success Measurement

We decided to select 3 classification classes: cat, bird, and dog for the project. We proceeded to download the required public domain images and kept at least 1000 images for each classification class, and build the deep-learning model to train and validate the image classification model.

The measurement for the project will depend on the prediction accuracy on the test dataset, and it should show at least 10% or more improvement on classification accuracy over the initial unoptimized model to be successful, or it should reach at least 75% classification accuracy after tuning the model.

# Project Solution Design

The project should prepare the required image dataset, then filters out the outlier images, then construct the training dataset and test dataset, and then design and implement deep-learning model and perform hyper-parameter tuning, and finally perform the comparison and select the final classification model with the highest accuracy.

The project follows the supervised learning process as below.



# Project Implementation

The project performs the steps below.

1. Image Data Collection

We have tried three different ways to download the required images, and have settled on using the third method after the first two methods proved unsuccessful.

* 1. <https://www.pexels.com/api/>

We first try to download from pexels.com by using an existing python PyPexels package to download images. But there are some major drawbacks in this approach. Firstly, there is a limit on the number of images that can be downloaded every day (<= 200). Secondly, there was not enough images of our selected classes i.e. cat, bird and dog in this website. Third and most seriously, the image quality for our selected classes was not good. Especially for bird and cat, many (>50%) of the images were irrelevant images (e.g., bird-eye view images of scenery).

We have developed the “pyprexels.py” application for testing and downloading images from the website, but we gave up the solution based on the reasons above.

* 1. <https://snappygoat.com>

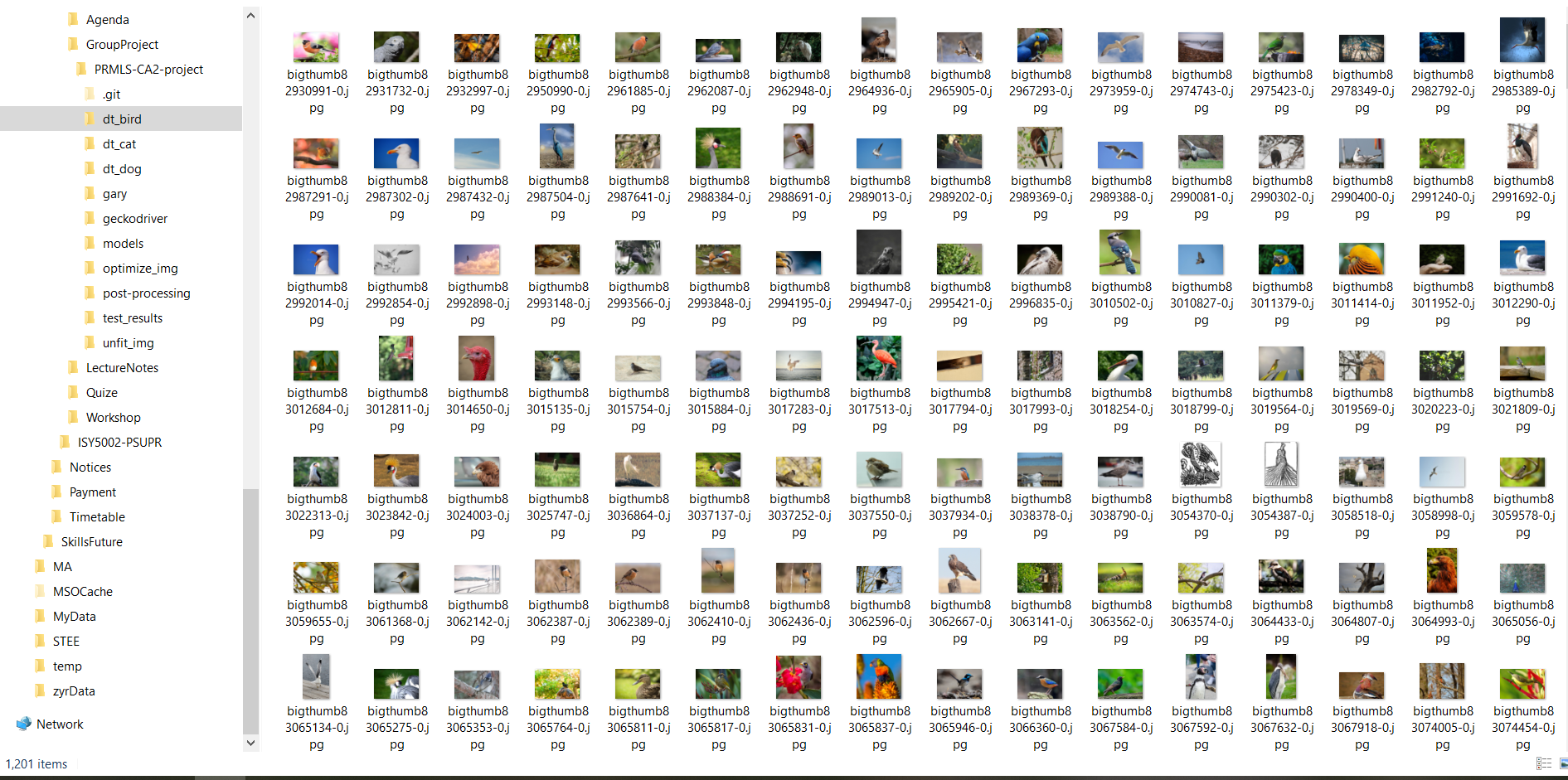
Next, we chose to try out the “snappygoat.com” web site and developed the “searchimg.py” application to download the images. We still encountered some issues in using this method. Firstly, the images for each class were not enough (<1000). Secondly, some of the images were irrelevant because people had given the wrong tag to their uploaded pictures and resulting in the wrong images being downloaded. Thirdly and most seriously, we have to pay the fee in order to use these images for our project. So we also give up this solution.

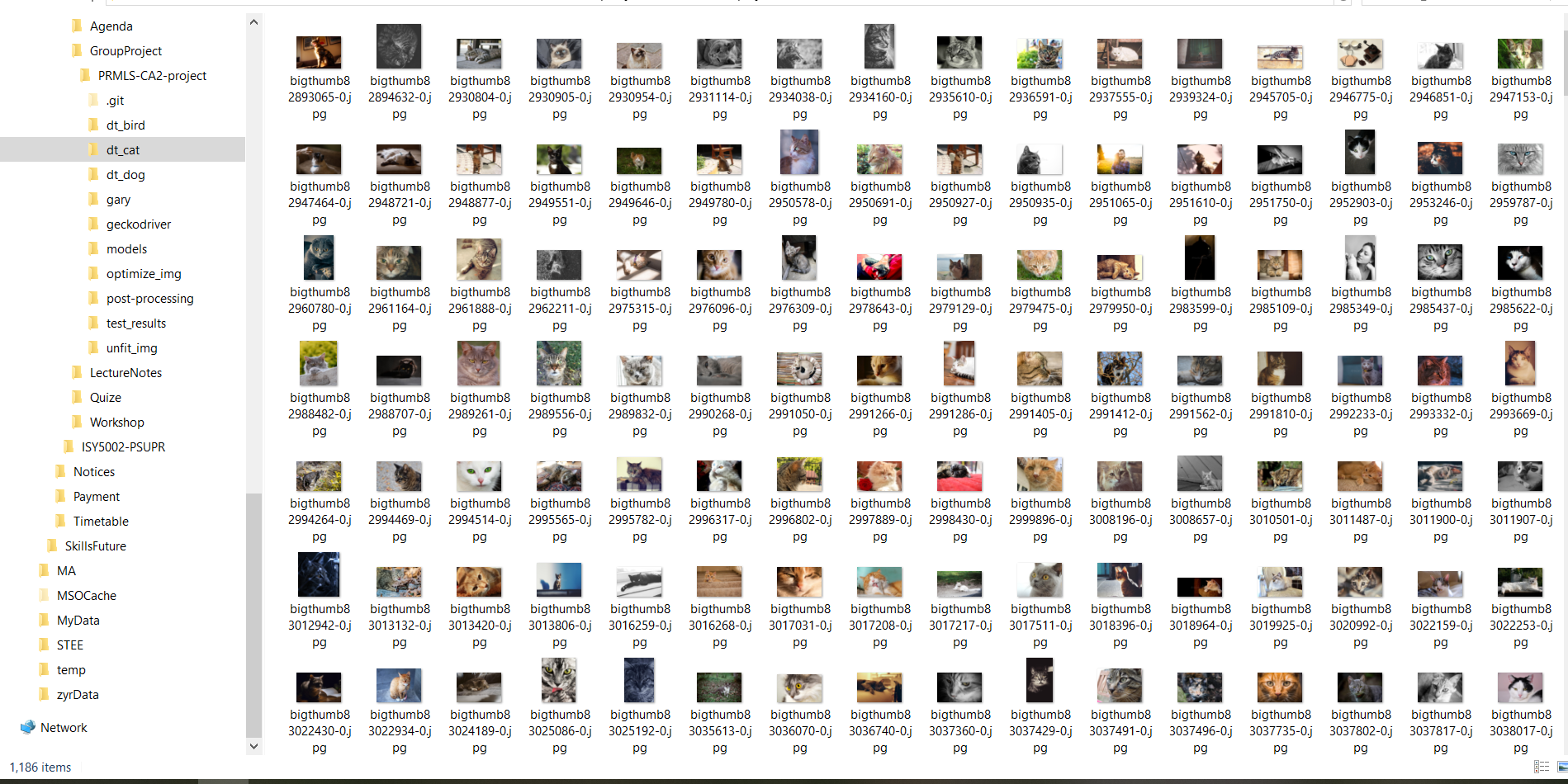
* 1. <https://www.dreamstime.com>

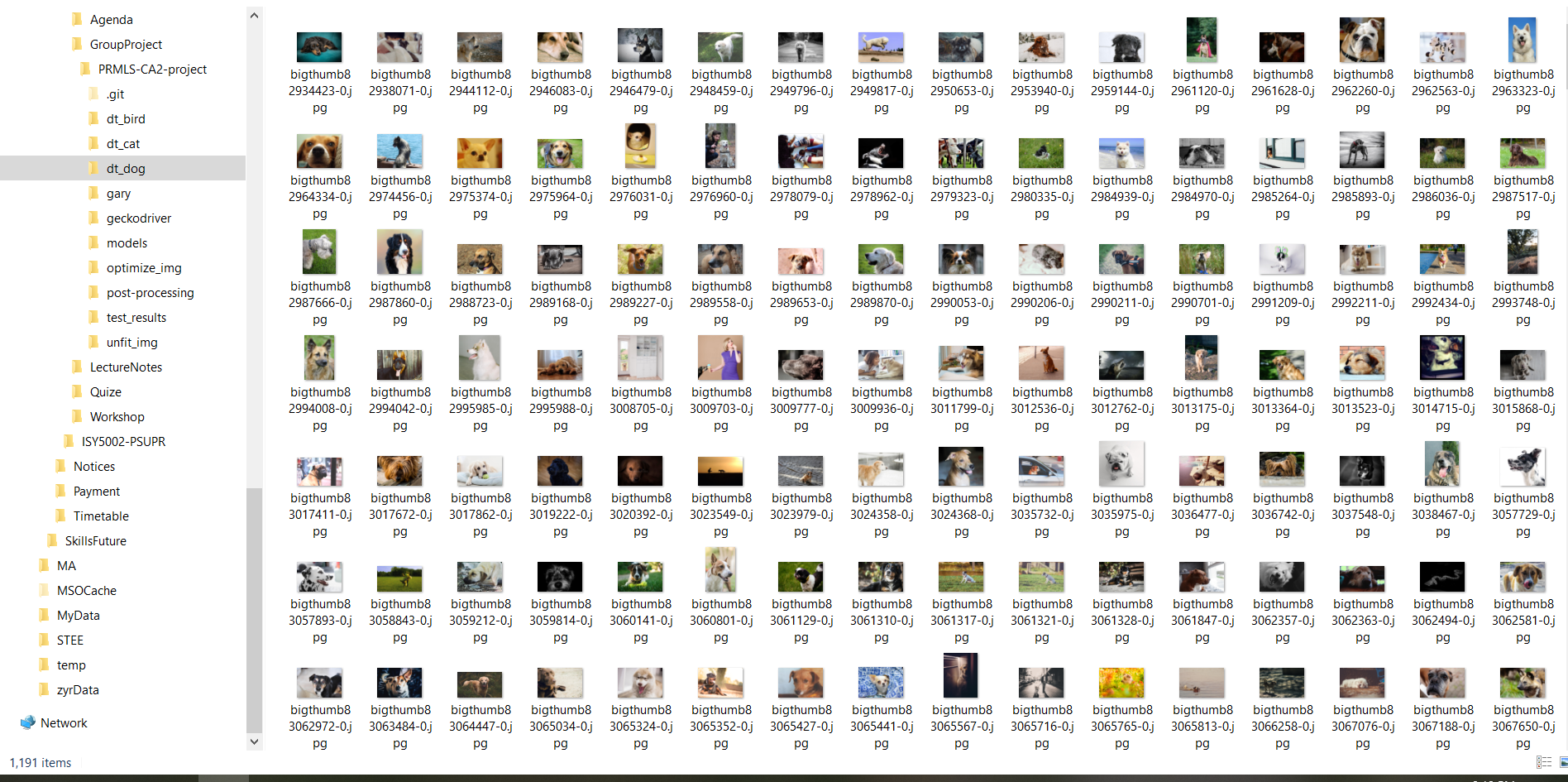
This is the solution we are using for the project. The site “dreamstime.com” not only provides enough images but also most of the images were good enough to be the candidate image data for the project. And we also learn that the images from “dreamstime.com” are truly royalty free for the project without any charge.

We developed the “getpubimg.py” application to download all the images for the project, and each class has 1200 images. We adopted selenium, a program to implement the solution and used firefox browser to simulate human actions. Since the web site html content is embedded in the script and cannot be parsed by html parser such as BeautifulSoup. Also the web site takes certain measures to prevent auto-download scrapper, so we have use the timer to wait for 15 seconds after simulate the event to click next page link to bypass the restriction, and also we create a separate thread to download the images based on the image source URLs.

The images downloaded are stored at different folders (“dt\_cat”, “dt\_bird”, “dt\_dog”) which name represent their class. The screen shots for “dt\_bird”, “dt\_cat”, “dt\_bird” folder files are shown below.







1. Pre-Processing

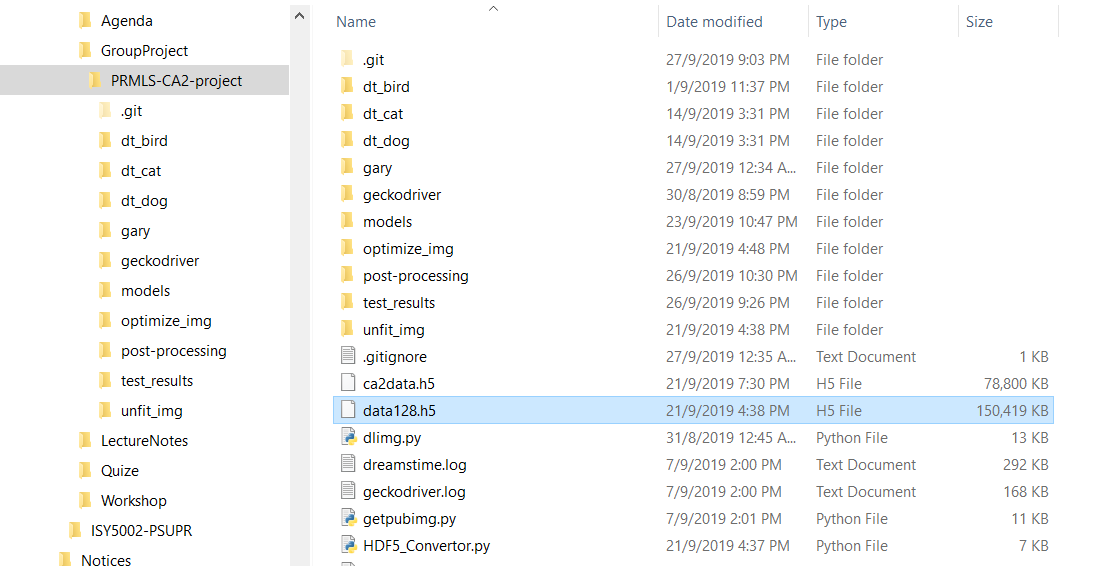
We have performed the following steps to filter irrelevant images.

2-1 Remove empty images

We noticed some of the image files contains only blank images, so all these images are removed.

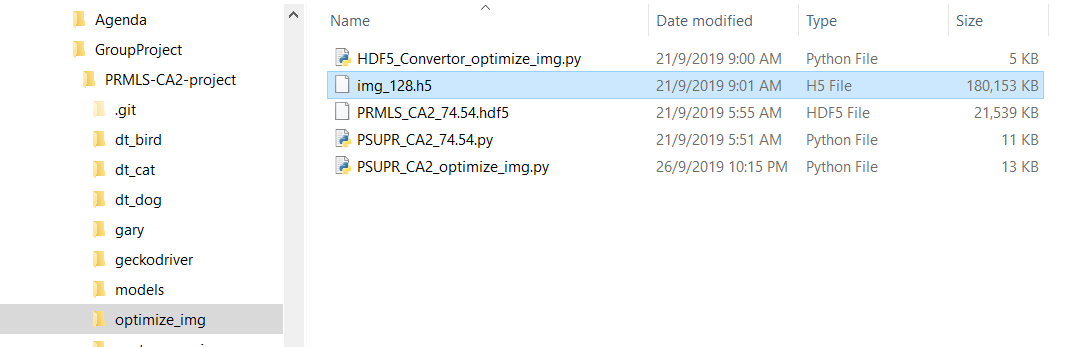
2-2 Save image data and class labels

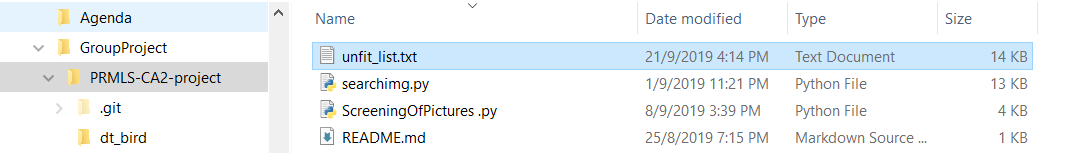
We develop “HDF5\_Convertor.py” to read images file, convert its resolution to 128x128, and save its content and class label to “data128.h5” hdf5 format file as shown below.



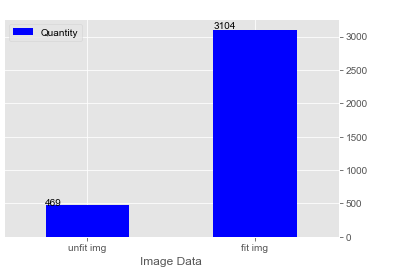
After several rounds of training and testing with the deep-learning base model, we notice the accuracy is not improved much and the model is unstable. So we decide to filter out the images which have low validation results.

We develop “HDF5\_Convertor\_optimize\_img.py” to read all available image files and store them into “img\_128.h5”, and “PSUPR\_CA2\_optimize\_img.py” to validate the image with pre-trained base model, and store the identification number list of all the unfitted images to “unfit\_list.txt”. The pre-trained model weight parameters are loaded from “PRMLS\_CA2\_74.54.hdf5” file, and all images (469) with below 0.1 prediction probability are recorded to “unfit\_list.txt”. We then modify the “HDF5\_Convertor.py” to read and filter out the files in “unfit\_list.txt”, and re-generate the “ data128.h5” file. The screen shots for these files are shown below.

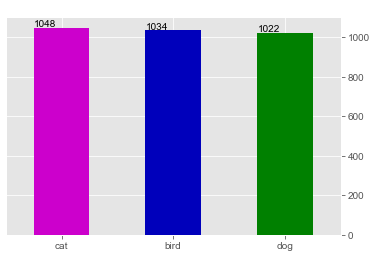




The figure below shows the number of unfitted images (469) and remaining images (3104) which are generated in “PSUPR\_CA2\_optimize\_img.py”.



The reminding images still have enough samples (>=1000) for each class to meet the assignment requirement, and can be validated with below figure which is also created by " PSUPR\_CA2\_optimize\_img.py”.



We also analyse the unfitted images and learn the reasons which make them to be hard to learn by the deep-learning model, and they are listed below.

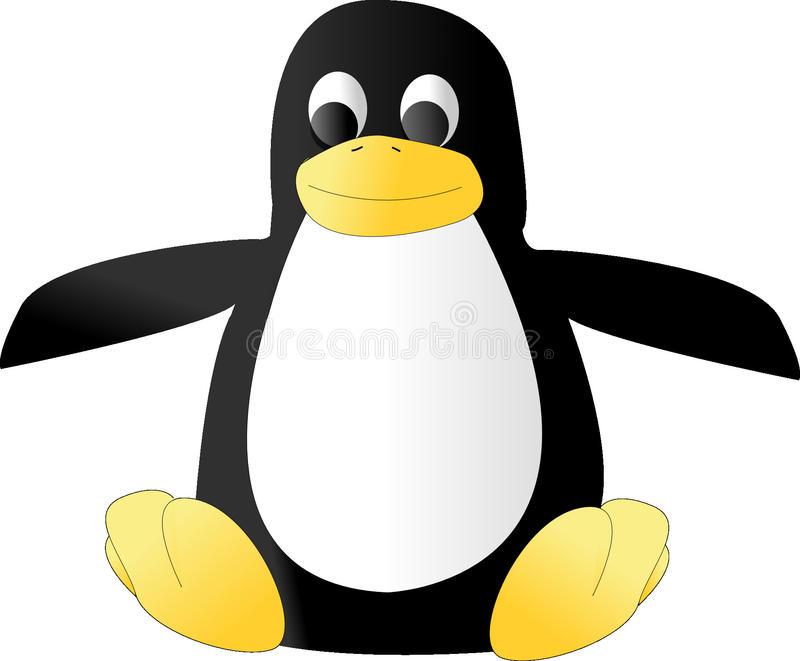
1. Targets blended into their environment because of their camouflaged coat colour



1. Targets dressed in human attire



1. Target images are either drawings or cartoons



1. Images are not of the targets



1. Targets are obscured by objects (e.g., tree branches)



1. Target images contained other objects (e.g., human)



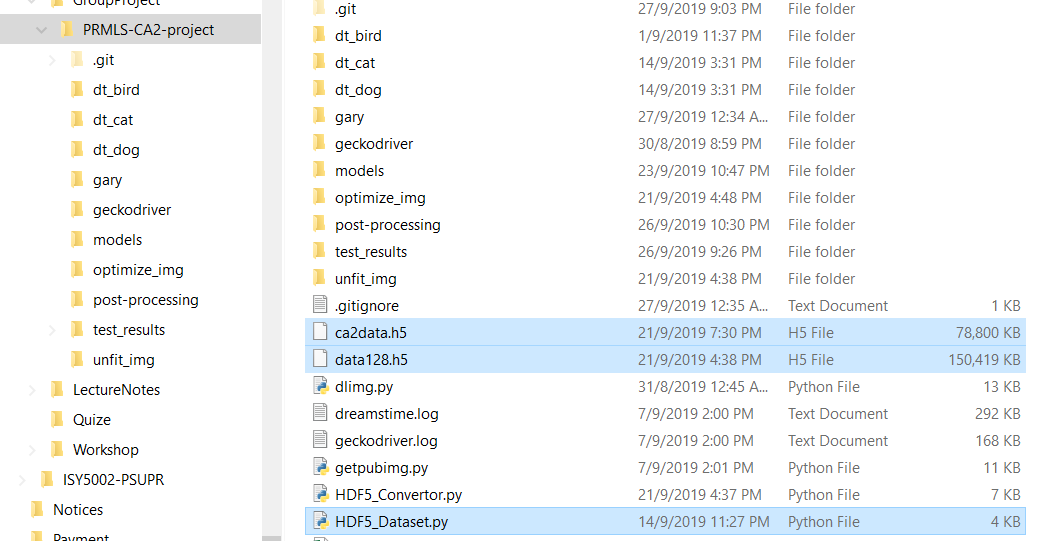
1. Targets are small inside their images



1. Sampling

We develop “HDF5\_Dataset.py” to read image data and class labels from “data123.h5” file, and then divide them into training dataset and testing dataset, next save to “ca2data.h5” hdf5 file.

The figure below shows these files.



1. Deep-Learning Model Training

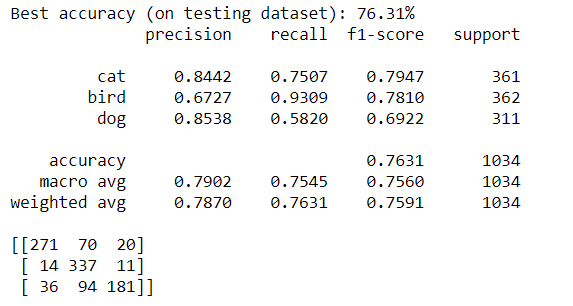
We develop “PSUPR\_CA2.py” and implement deep-learning ResNet model as the base model.

The model is shown at “PRMLS\_CA2\_model\_final.pdf” below, and the best training results are shown in “PRMLS\_CA2\_JT\_7631.pdf” which reaches 76.31% accuracy.

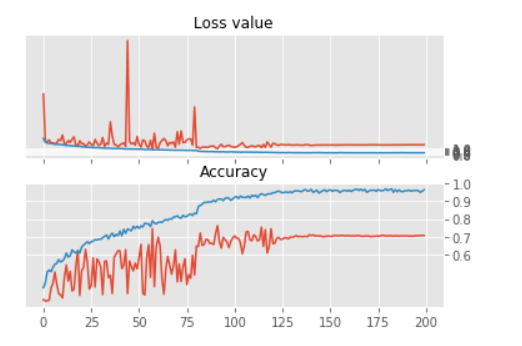


It contains total 2,674,819 parameters, and inside them there are 2,668,067 trainable parameters and 6,752 non-trainable parameters.

The figure below shows the accuracy on the testing data.



And the training loss value and accuracy history for the model are shown below.



1. Hyper-parameter Optimization

The very first time after we develop the base deep-learning model with all known best practice parameters and train the model, the accuracy on testing achieves 73.36%.

Then we adjust the parameters in the ways described in below sections, and keep to perform the adjustment to try to improve the model accuracy.

5-1 Select activation function

We choose Relu activation function so it will be better to update and prevent gradient vanish compared to other activation function such as SigMoid.

5-2 Set weights initialization value

We choose He initialization to set weight initialization values, so to avoid to assign value to be too small or too large.

5-3 Restrict weights

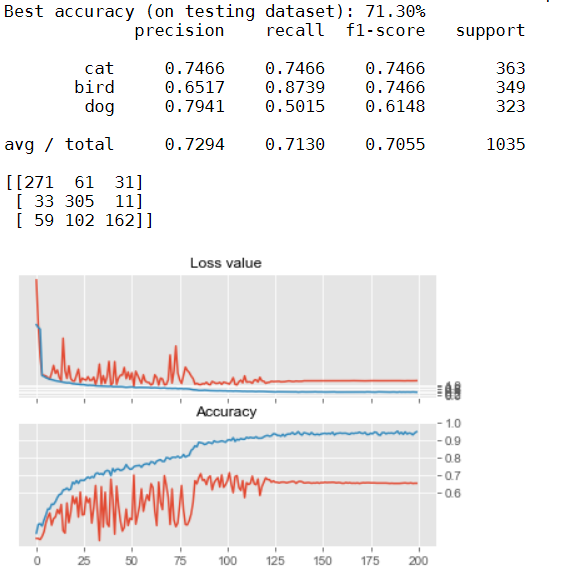
We select L2 regularizatio to control the magnitude of the weights so to keep the smaller weights value.

5-4 Add dropout

We have tested to add drop out between layers, and the accuracy results are improved. The following shows the performance before and after dropout layers are added.

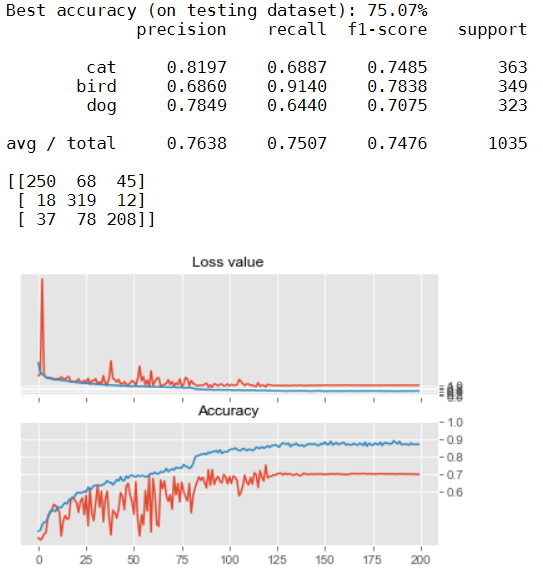
1. Without drop out layers

The accuracy without drop layers only reaches 71.30%.



1. Add drop out layers

The accuracy reaches 75.07% after adding drop out layers.



But the model with best accuracy results (76.31%) are achieved without dropout layers, so we utilize the model (76.31%) without dropout layers. From the above experiment, we learn the drop out layers improves the performance in general.

5-5 Perform batch normalization

We use the batch normalization to normalize the output for the convolutional layer.

5-6 Design residual network

We follow the ResNet design rules to design the residual network to improve the deep-learning network performance. We designed the residual layers by: (1) specifying the number of residual blocks from a range between 5 to 11, and (2) specifying the number of stages from between 3 to 4.

We kept the values of certain parameters constant in specifying the residual layers. Some of these parameters and their values are:

* Kernel size = 3
* Strides = 1
* Kernel initializer = he normal

The first stage of the residual layers does not contain a downsample block. But each of the subsequent stage comprises an initial downsample block with the reminding blocks being simple blocks.

We increased the number of filters from 16 for the initial stage of the residual layer to 32, 64, and 128 for each of the subsequent stages.

5-7 Set learning scheduler

We create and utilize the learning scheduler to vary the learning rate, so learning rate is reduced when the learning epochs increase.

5-8 Use image augmentation

We use the image augmentation to generate the randomly varied images in the beginning of each epoch, so to force the deep-learning model to learn the features that are relevant to the classification. The range of the image augmentation, e.g. the width\_shift\_range, height\_shift\_range and rotation\_range, have been adjusted to bigger value, but the accuracy is not observed to be improved. We also tested to change the range of brightness, but it lowers the test accuracy results so we do not include it in the final model.

5-9 Adjust batch size

We choose small batch size (32) to improve the accuracy, and also it is still allowed the laptop GPU used for training to work as normal.

5-11 Ensemble models

We ensemble 2 pre-trained models with best accuracy score using the integrated stacking method, which combines the 2 pre-trained models output and constructs common dense layer to finally output the classification results.

The final accuracy of the ensemble model achieves 79.61% and shows big improvement on the individual model 76.31% accuracy.

The ensemble model is described in more detail at section 7 “Final Deep-Learning Model”.

5-10 Problems encountered and challenge

1. Model Training Result Different

We notice that even for the same model design, the model accuracy after each training is very different, sometimes even reaches 4% to 5% accuracy difference. So we have to save the individual model weights in the separate file, then will be able to replicate the same accuracy with the model design, and avoid to lose the best model.

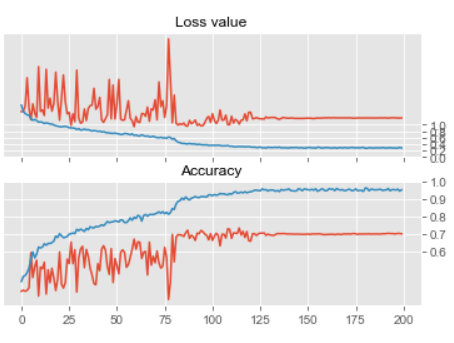
And we have encountered to lost some good models due to the weight files of these models have not been saved properly and thus cannot replicate the same accuracy result.

1. Determination on good parameters

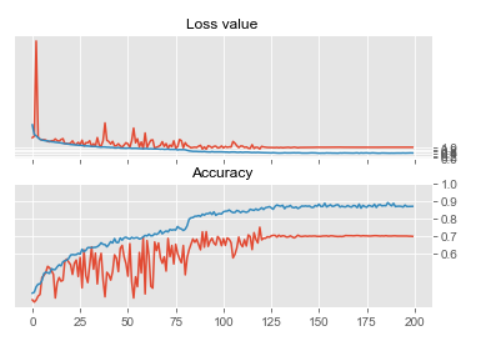
We observe that the model performance differs for each separate model training process. In order to determine the good parameters to be changed, we only choose single parameter to change at each time, for example, learning rate. Then perform several trainings. We will consider the model best accuracy across several trainings, and analyse the loss value and accuracy training diagram, then determine whether to keep the changes to the model. After that, we continue to pick the next available parameter and perform the same evaluation process again. But after several testing rounds, we find out that there are many factors to affect the good parameters, and so should consider the combination of the different parameters to make it work, therefore, we learn to understand the relation of parameters and adjust them together and adopt this as the better approach.

1. Unstable model

We notice that the base model is unstable from training loss value and accuracy chart below.



So in order to fix the issue, we have tried many different ways, e.g. change learning rate, add drop-out layers, add more layers, and find out that adding drop-out layers and adding one dense layer help to stabilize the model. The figure below shows the training result after adding drop-out layers and adding one dense layer.

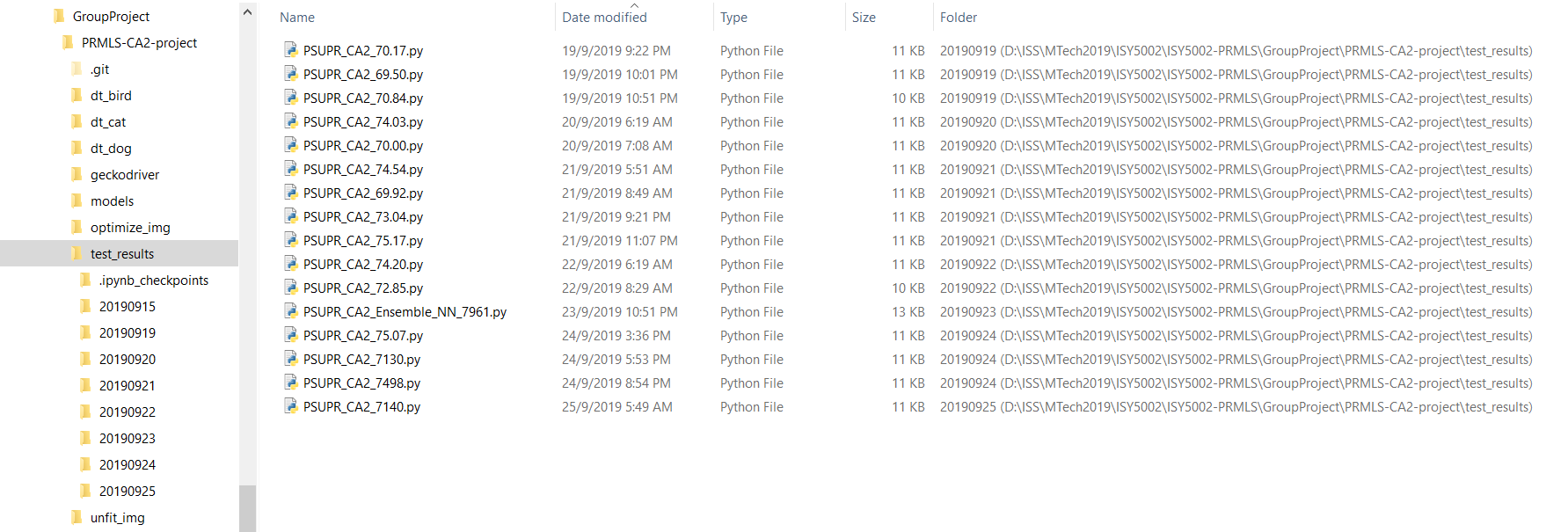


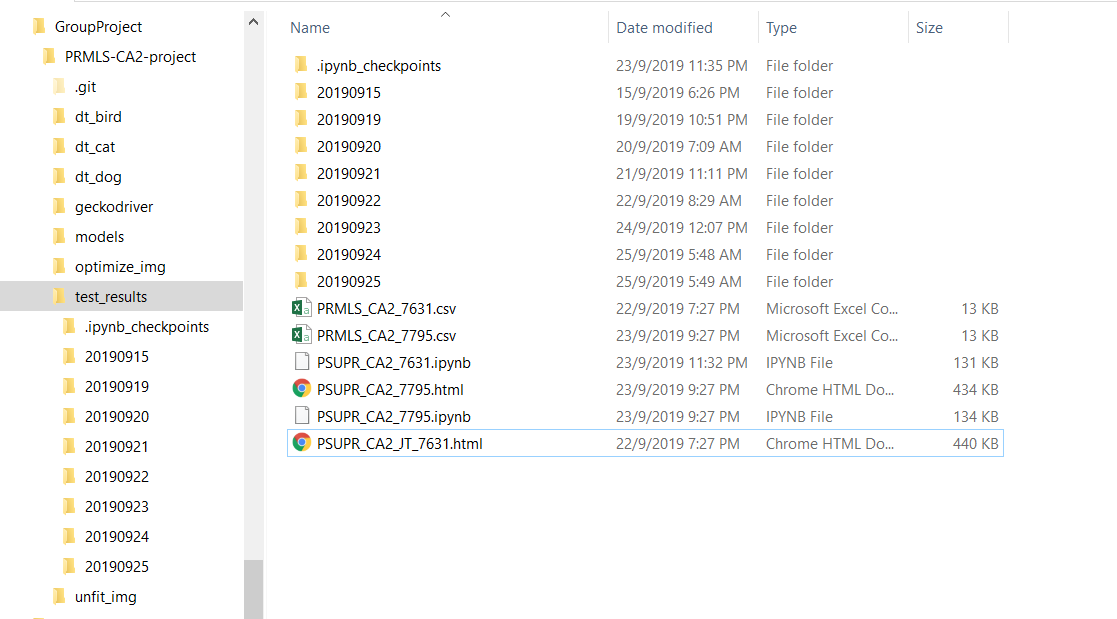
And we notice that the most significant improvement to stabilize the mode is to use the integrated stacking models. After integrating 2 same models, the training loss value and accuracy chart shows as below.



1. Challenge

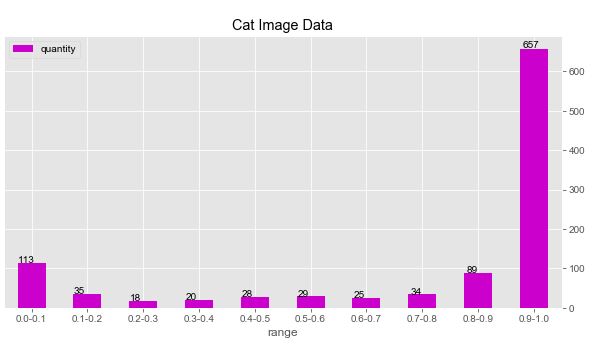
The biggest challenge we are facing is the computer resource constrains. The normal training time for our deep-learning model takes 2 hours to 4 hours. And we do not have enough time to improve the model. So the best model and best performance is the best one after we have performed many tests. The python application files with different models which have been tested are listed below to show the tests we have performed.

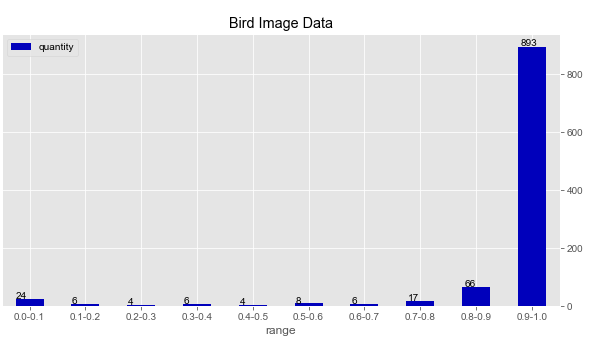


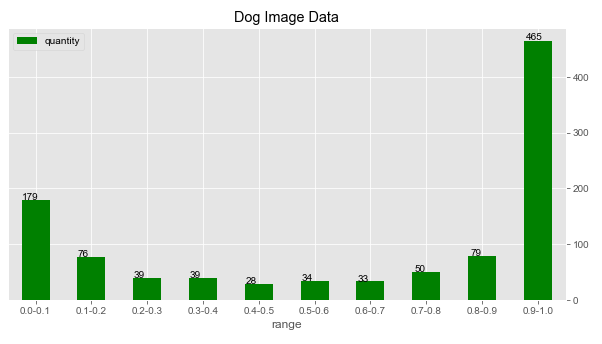


1. Post-Processing

After processing with best ensemble model, we develop “PSUPR\_CA2\_img\_analysis.py” to analyse the image classification results which are shown as below.







As the charts shown, there are still many low-quality images for cat (113) and dog (179) which prediction probabilities are below 0.1. Therefore, it is possible to improve the classification accuracy by removing some of them.

1. Final Deep-Learning Model

We choose the ensemble deep-learning ResNet model as the final dee-learning model, and develop “PSUPR\_CA2\_Ensemble\_NN.py” to implement the model.

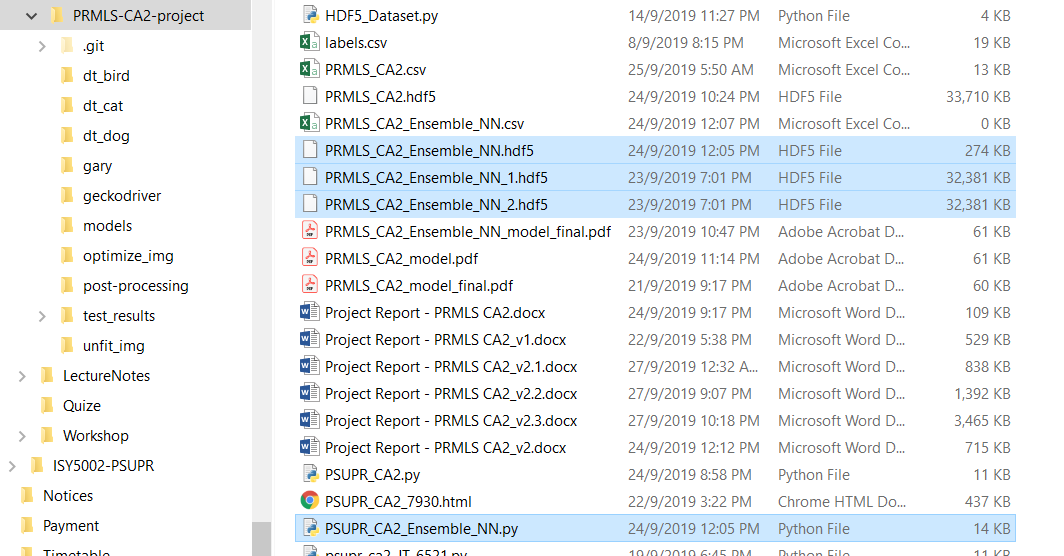
The 2 models inside the ensemble model are pre-trained deep-learning ResNet models with accuracy 76.31%.

The model diagram is shown at “PRMLS\_CA2\_Ensemble\_NN\_model\_final.pdf”, and the application to run is “PSUPR\_CA2\_Ensemble\_NN.py”.



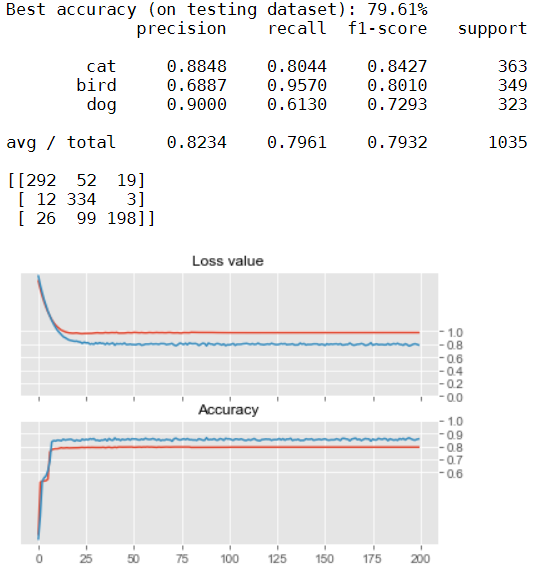
It contains total 5,349,701 parameters, and inside them there are 63 trainable parameters and 5,349,638 non-trainable parameters.

The figure below shows the relevant files for the ensemble model.



# Project Performance & Validation

The project final deep-learning ensemble model archives 79.16% accuracy on the test dataset, so it meets the project performance requirement. The validation result is shown below.



# Project Conclusions: Findings & Recommendation

The project shows that the deep-learning ResNet model is able to archive good image classification results, but adjusting hyper-parameter to improve the classification results are not easy. The ensemble model is one of the recommended ways which achieves the better accuracy on the individual model. Also another recommended way to improve the performance is to adjust the hyper-parameters together and not just take them one by one when performing hyper-parameter tuning.

We also notice that the original dataset has vast impact on the final accuracy results. Without picking up the adequate dataset, the deep-learning model becomes unstable and hard to improve. So the effort spent on the dataset preparation is justifiable, even if it seems like a waste of time. And we need to understand the dataset before performing the model training.