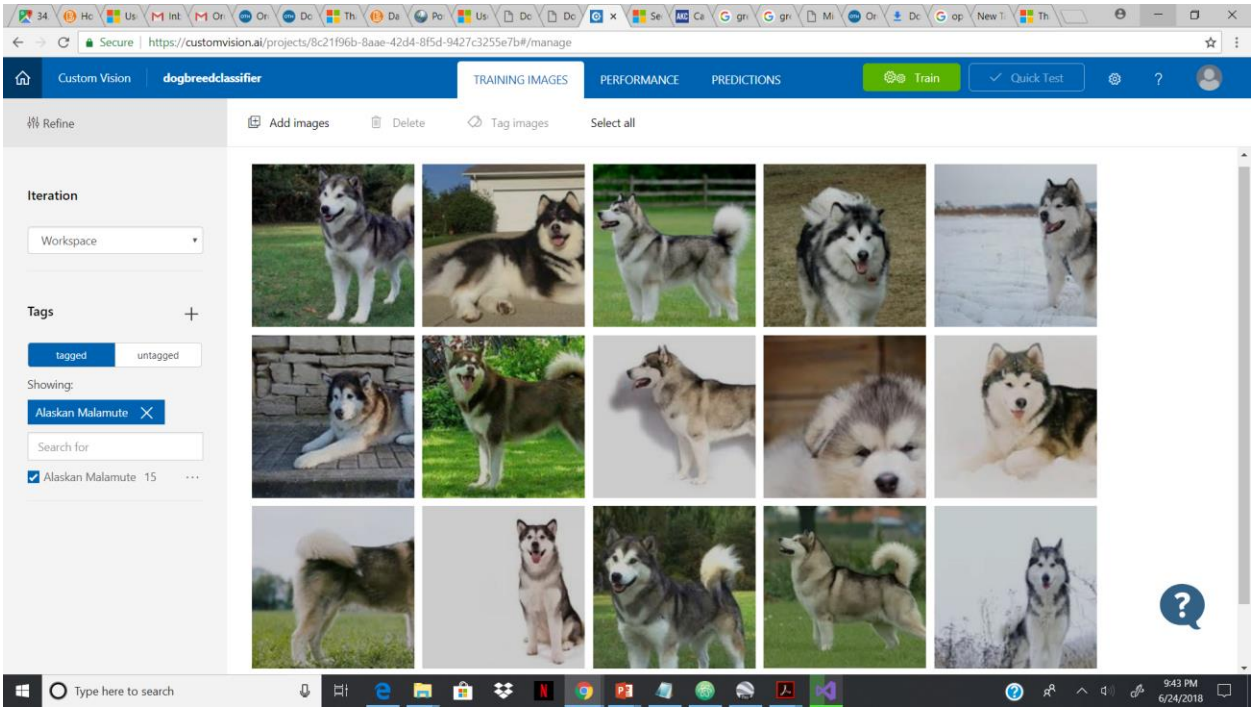
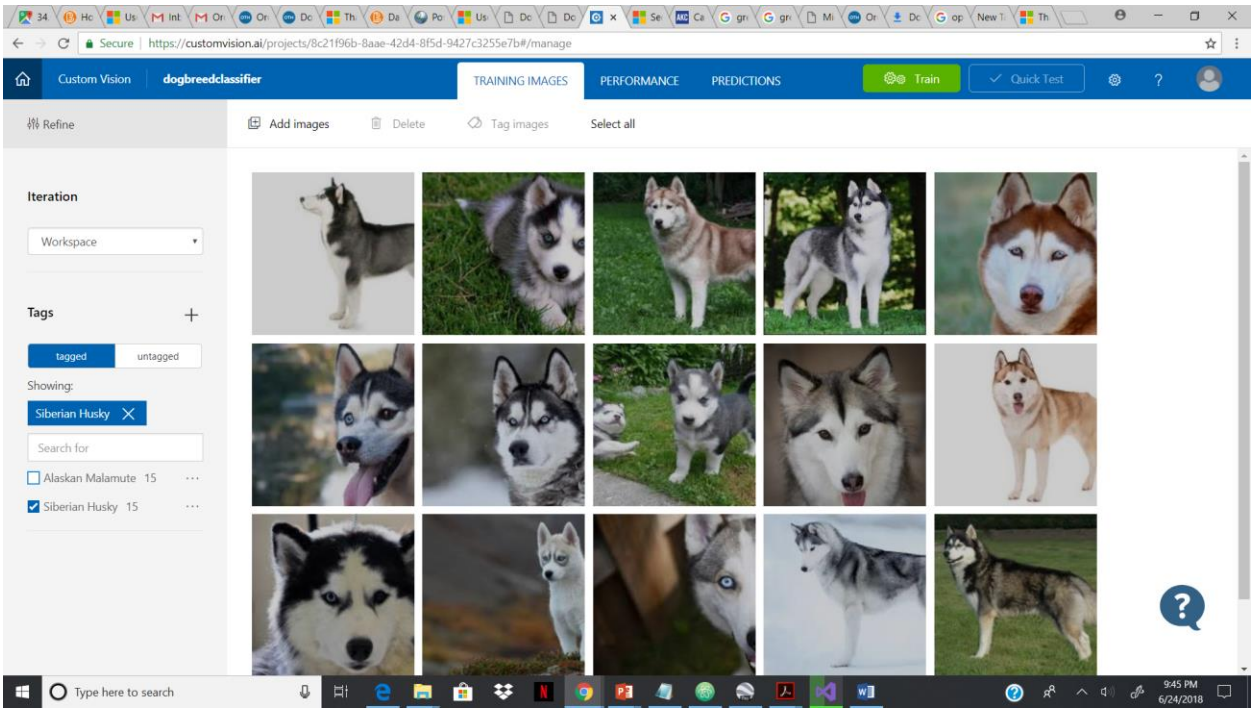
CSCI 585- DATABASE SYSTEMS HOMEWORK 4

ITERATION 1:

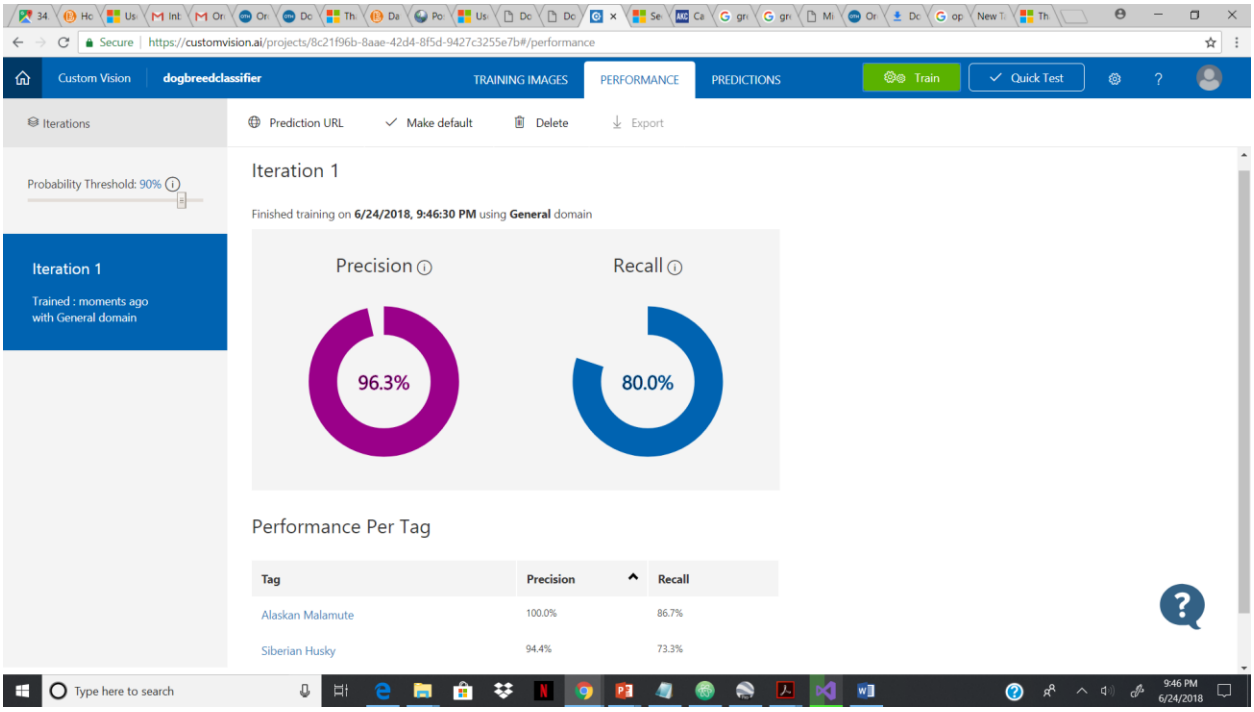
15 IMAGES OF ALASKAN MALAMUTE



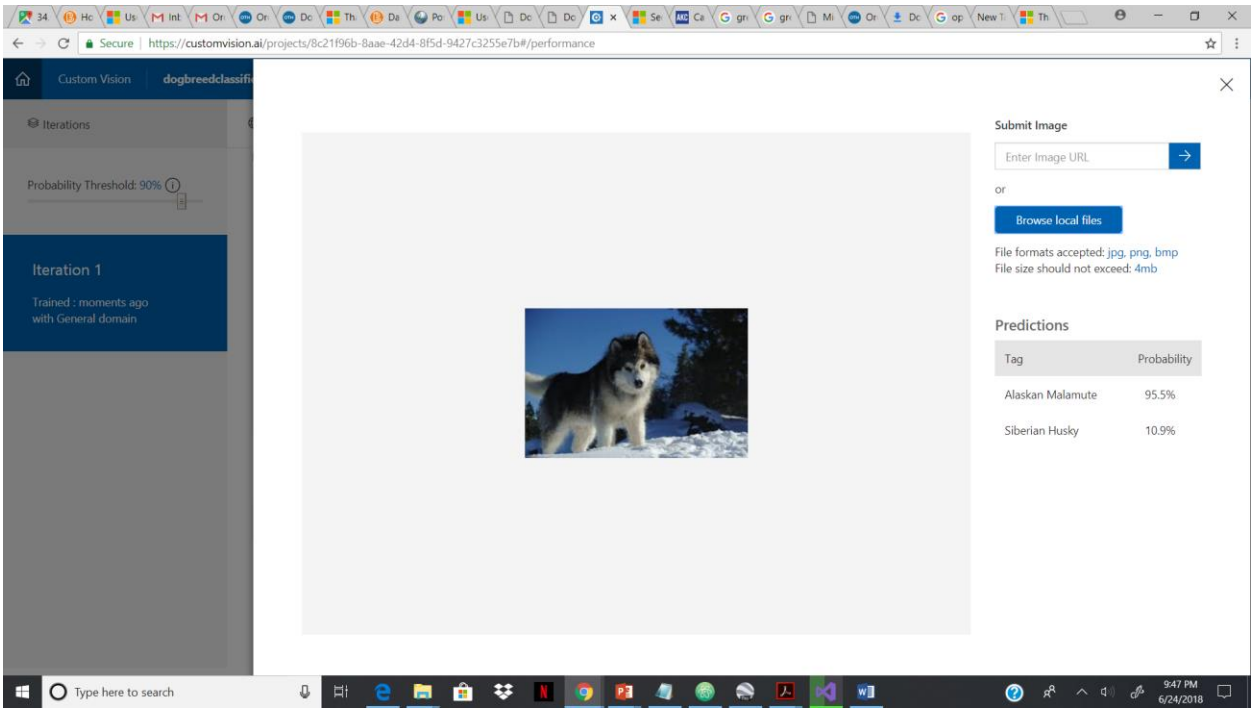
15 IMAGES OF SIBERIAN HUSKY



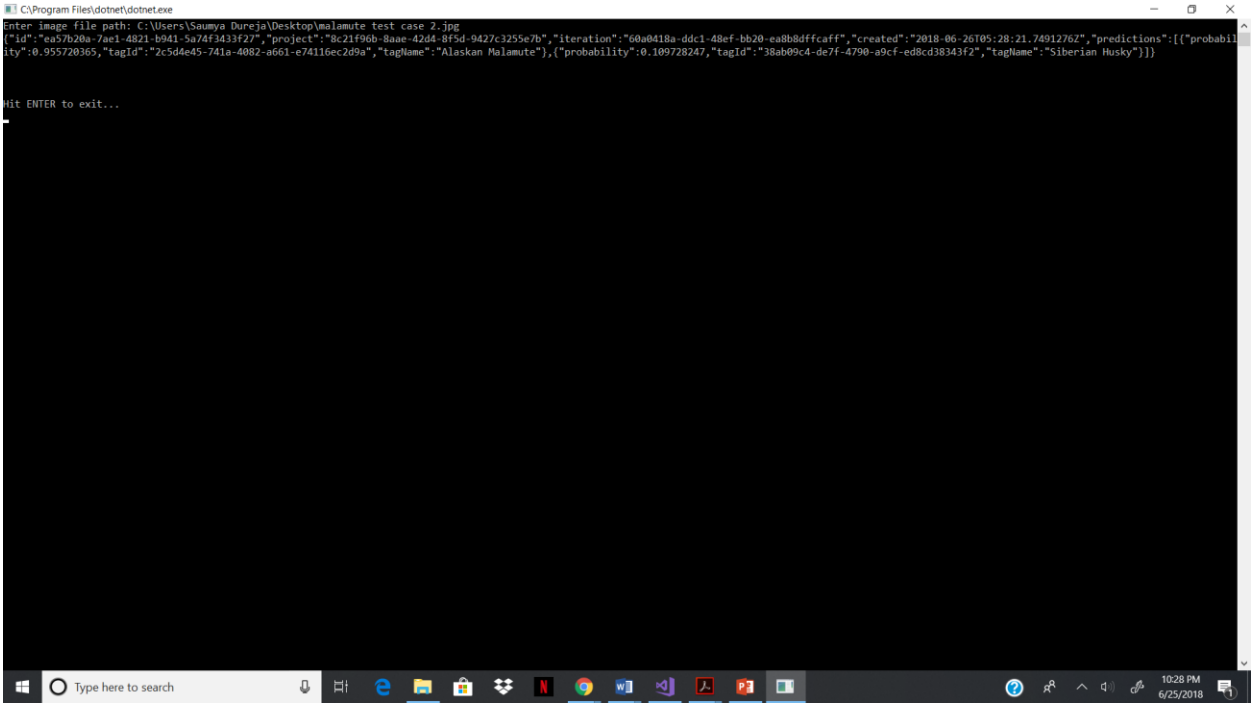
ITERATION 1 PERFORMANCE RESULTS



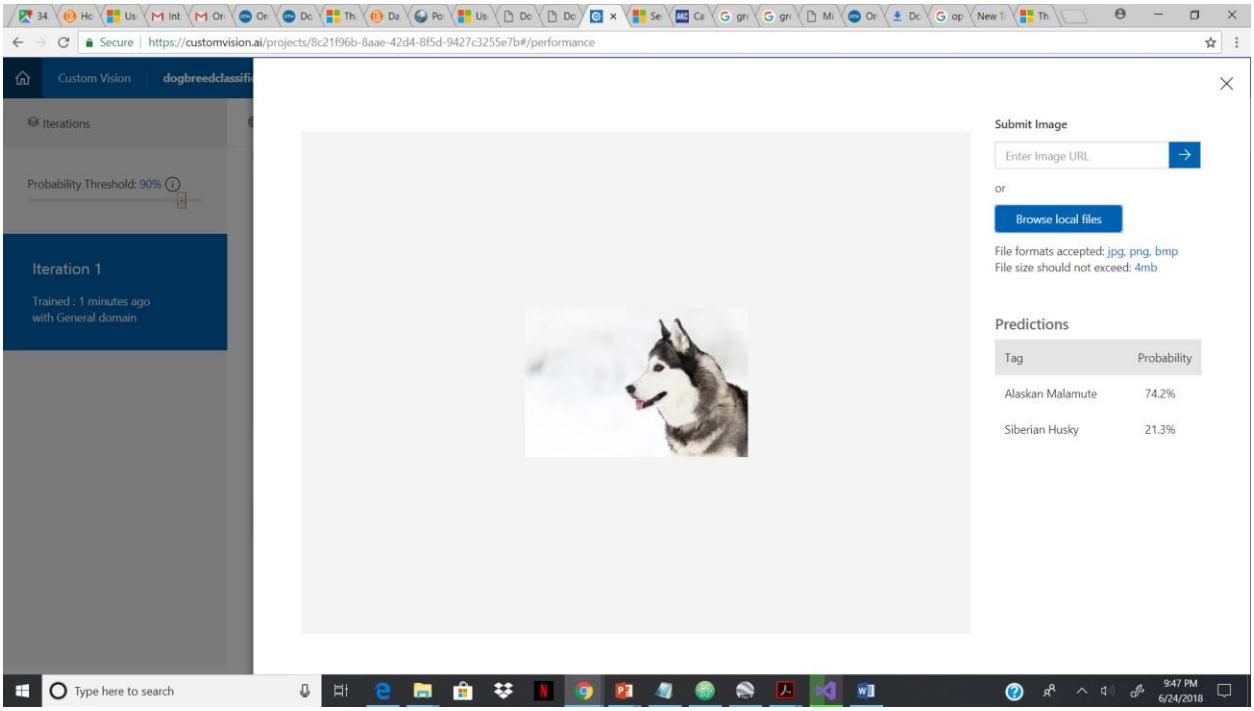
PREDICTION RESULT: CORRECTLY PREDICTED ALASKAN MALAMUTE



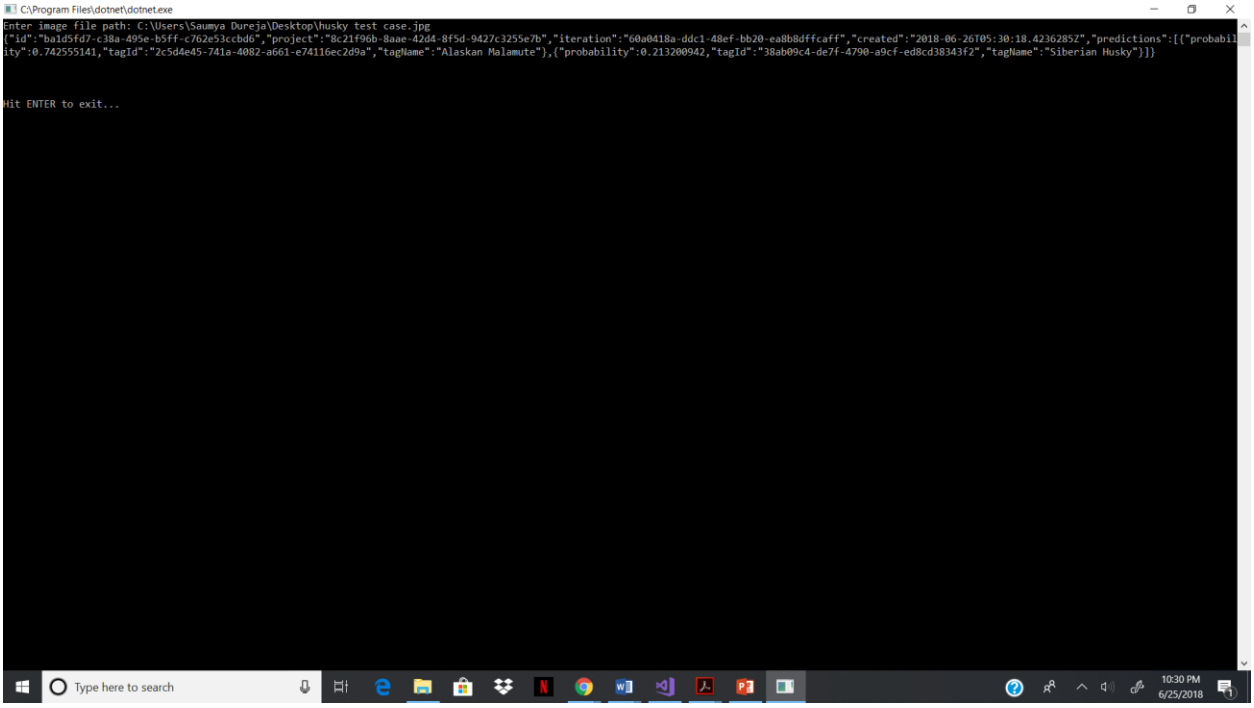
COGNITIVE API RESULT:



PREDICTION RESULT: INCORRECTLY PREDICTED SIBERIAN HUSKY

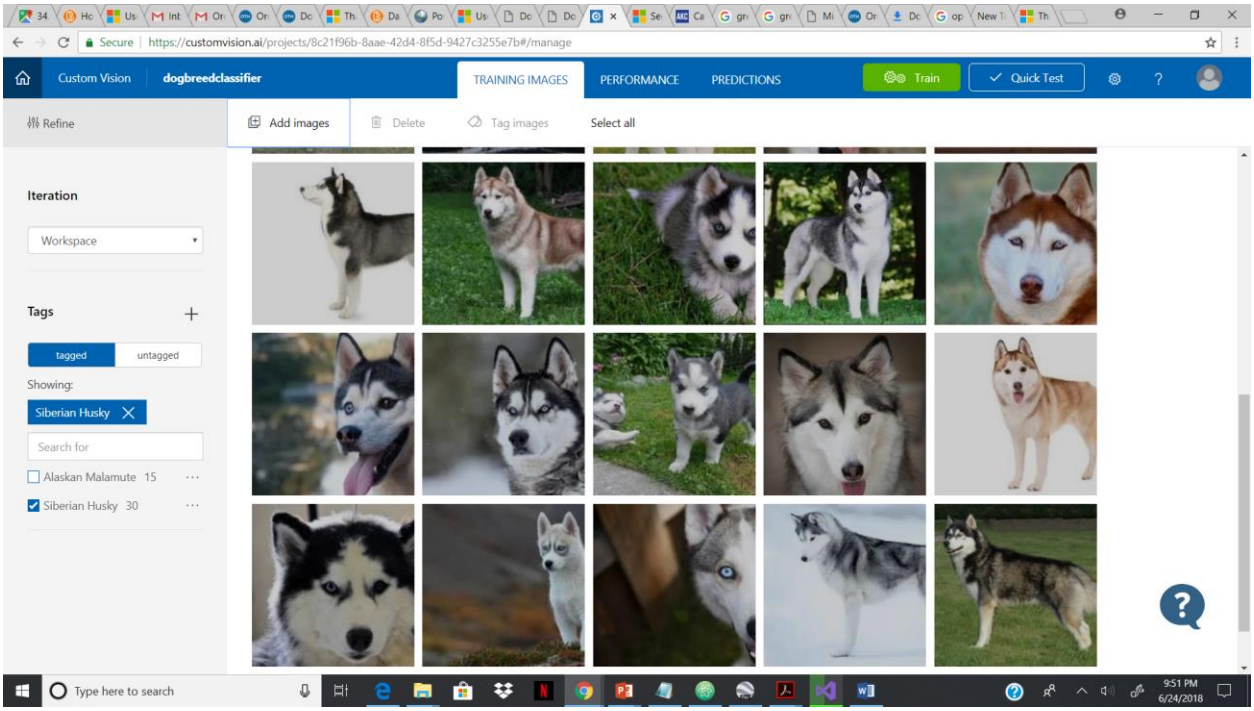


COGNITIVE API RESULT:

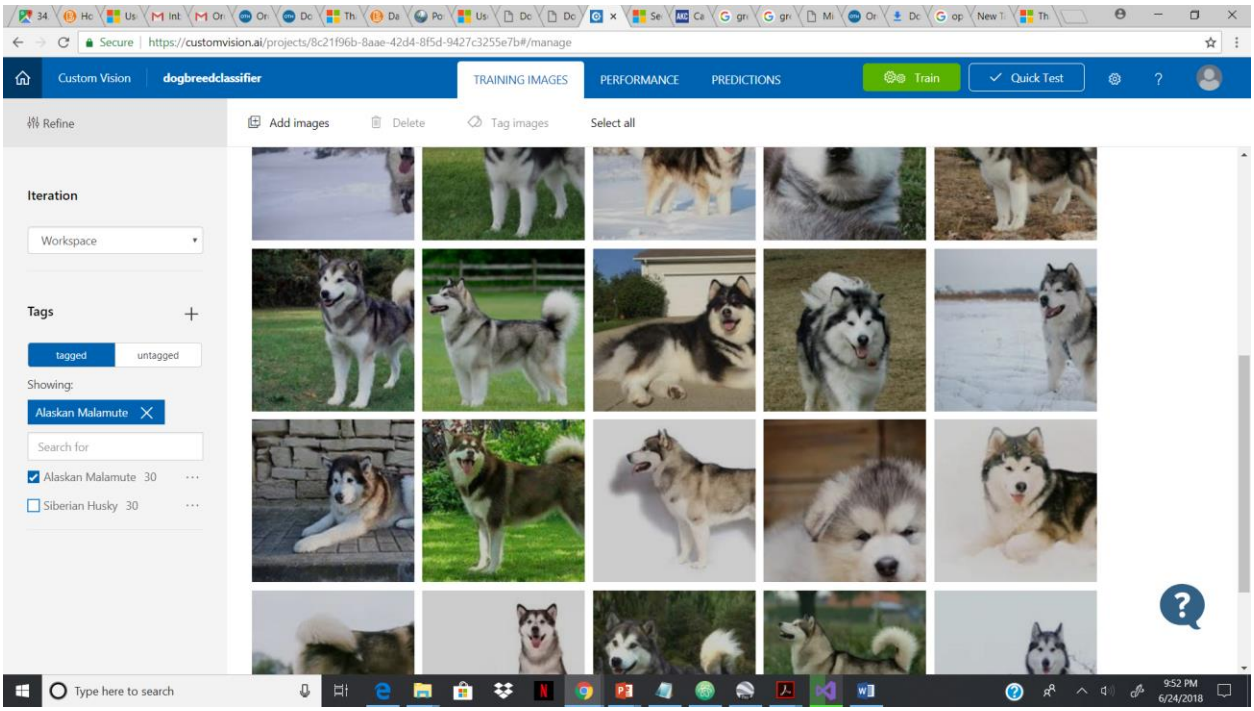


ITERATION 2:

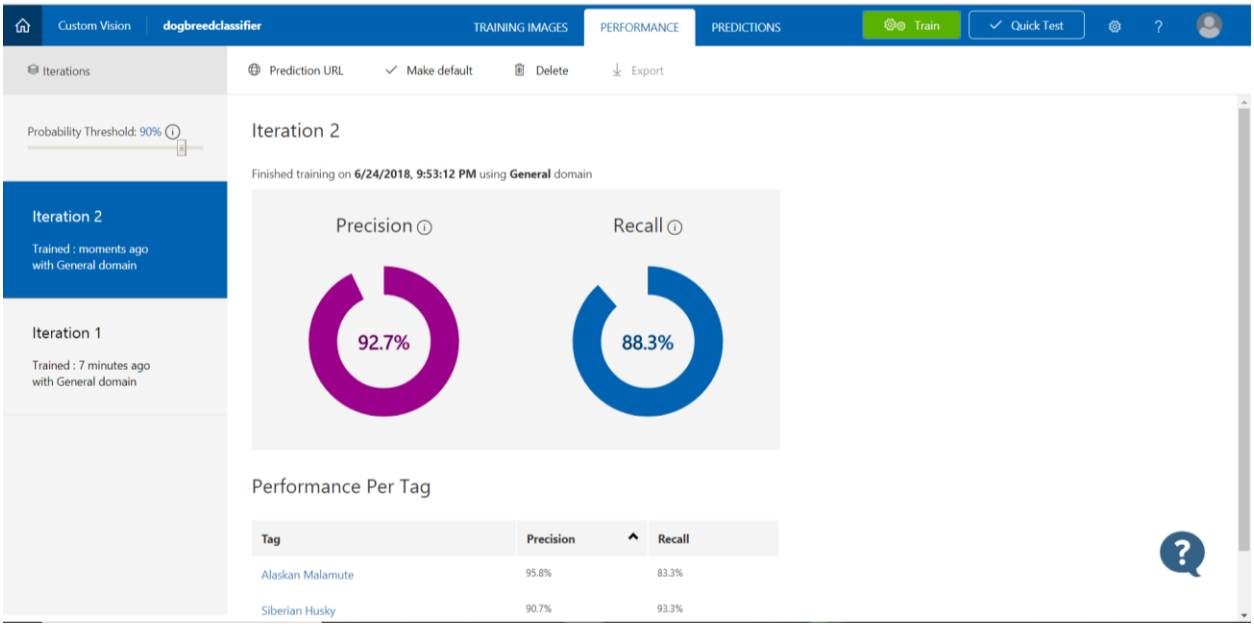
30 IMAGES OF SIBERIAN HUSKY



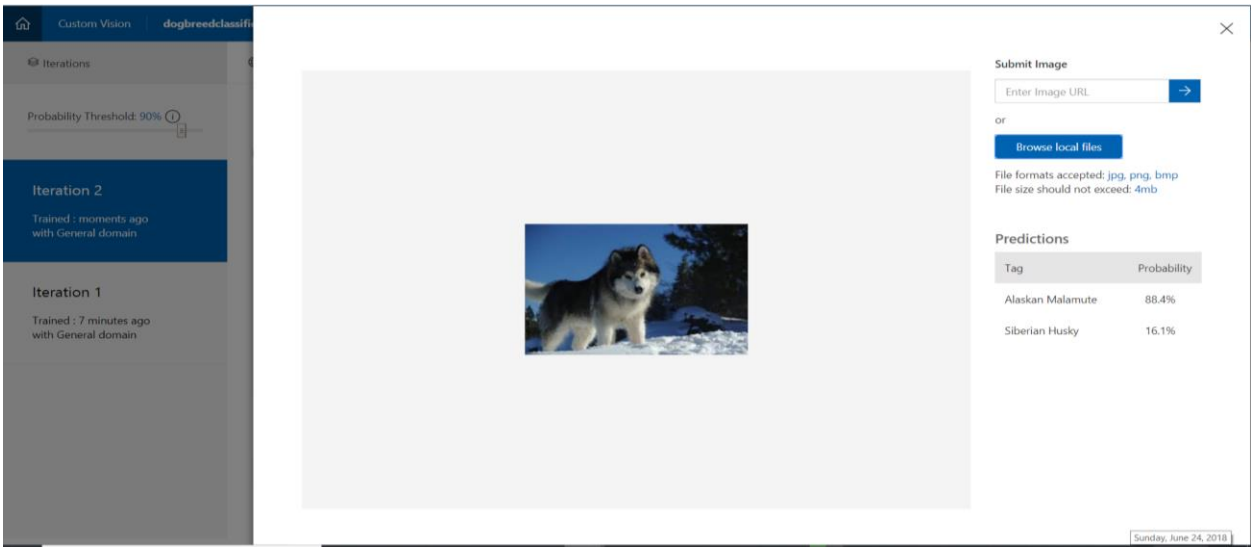
30 IMAGES OF ALASKAN MALAMUTE



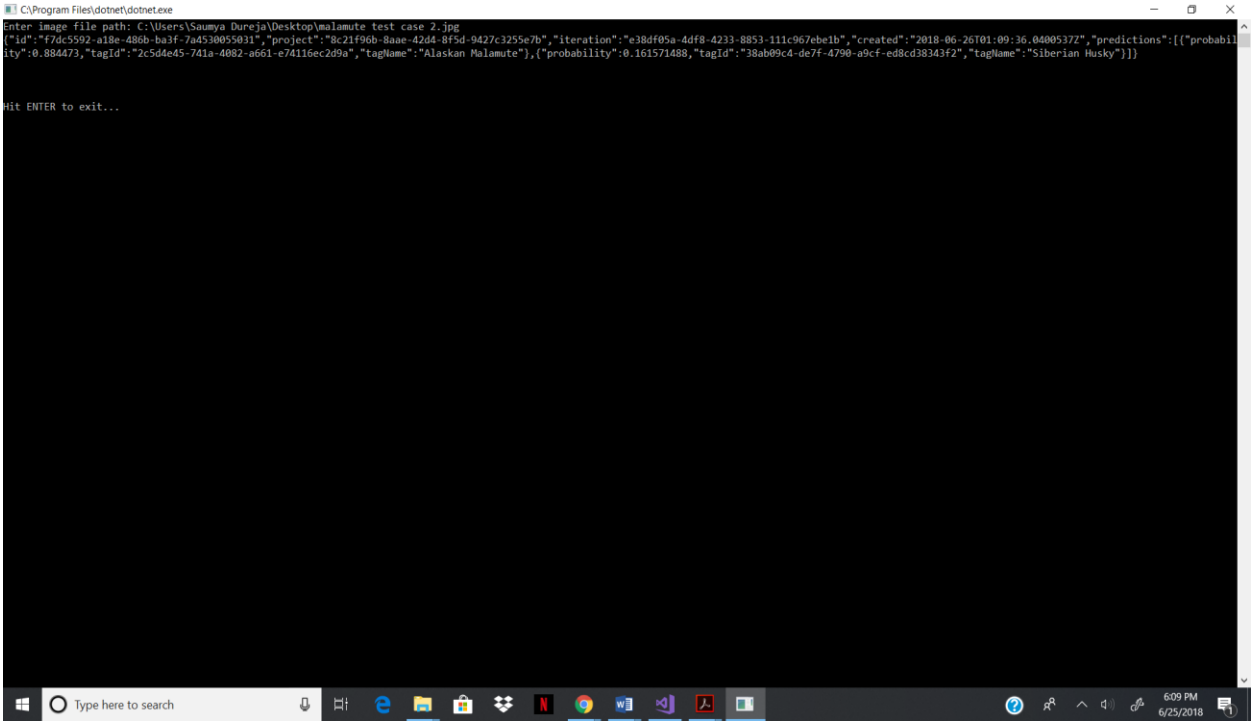
PERFORMANCE RESULTS:



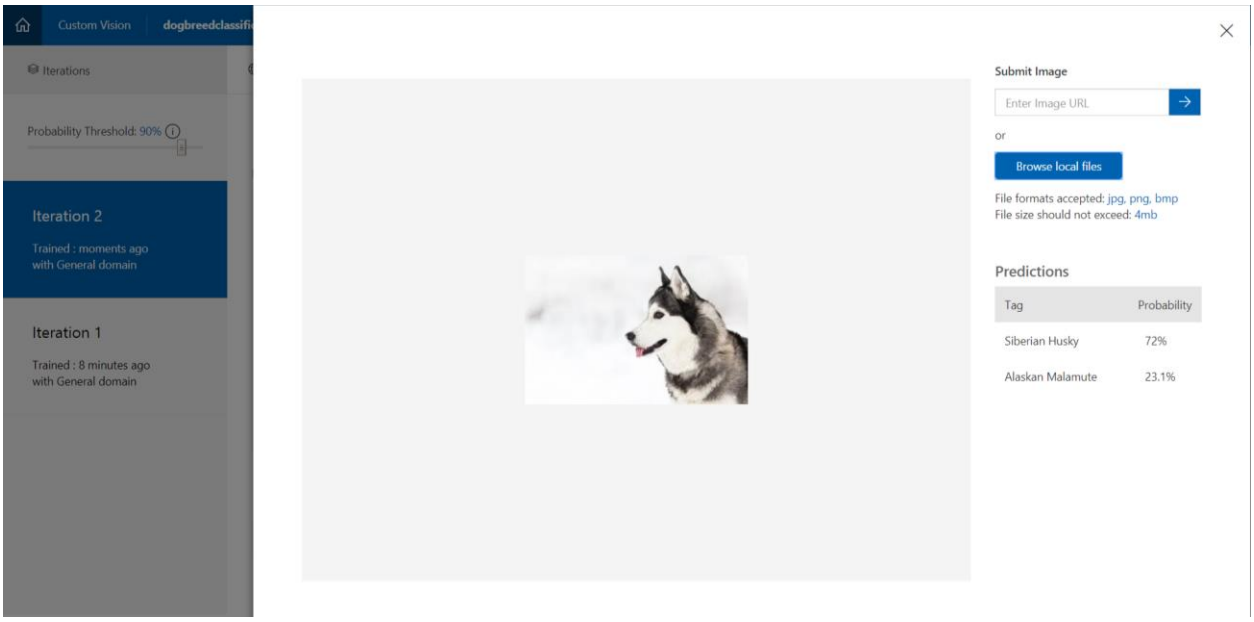
PREDICTION RESULT: CORRECTLY PREDICTED ALASKAN MALAMUTE BUT WITH LESSER PROBABALITY AS COMPARED TO ITERATION 1



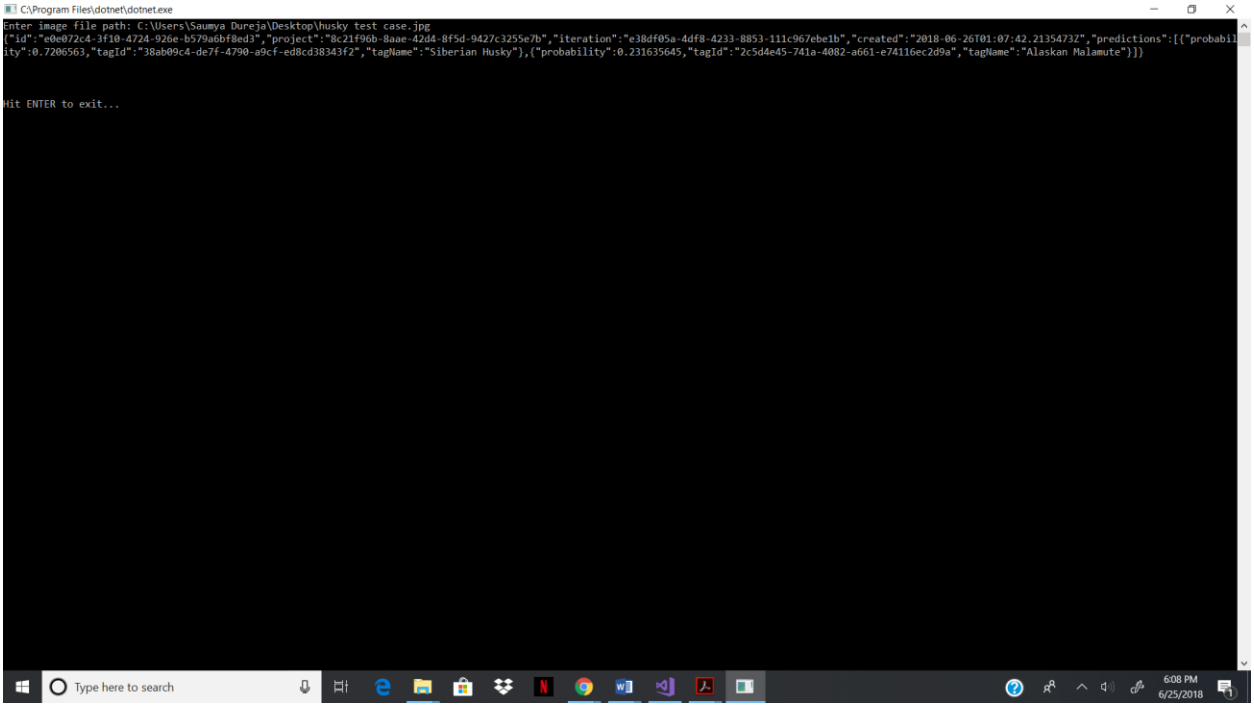
COGNITIVE API RESULT:



PREDICTION RESULT: CORRECTLY PREDICTED SIBERIAN HUSKY



COGNITIVE API RESULT:



Try to discuss your findings and the possible reasons, such as why your iteration 1&2 are different (better or worse), and the difficulty you met in this assignment.

In iteration 1, the classifier correctly predicts the Alaskan Malamute with 95.5 percent probability but incorrectly predicts the Siberian husky to be Alaskan Malamute. This might be due to **too few features** of Siberian Husky being identified in the dataset containing just 15 images of each. The size of the dataset was too small for it clearly distinguish between the 2 breeds.

In iteration 2, the classifier correctly predicts the Siberian Husky with 72 percent probability. The performance, in this case, has improved as **more features/relevant features** of Siberian Husky were identified now that the dataset size was increased to 30 images which have helped the classifier in distinguishing between the 2 breeds and correctly predicting the Siberian husky.

However, in iteration 2, the classifier predicted the Alaskan Malamute correctly but this time the probability reduced from 95.5 percent to 88.4 percent. Therefore, the performance, in this case, became worse. This might be due to the fact that in iteration 2, the size of the dataset was increased from 15 to 30 which must have resulted in **identification of more relevant features** of the two breeds, and since each of the breeds have quite **similar appearances** (wolfish appearances), this must have resulted in decrease in the probability when predicting the Alaskan Malamute. The model is not able to differentiate between the two breeds and requires more training (**Insufficient training data**).

Difficulty:

- Problem while installing Visual Studio: Didn't know .Net core development module was to be chosen for c#. Had to uninstall Visual Studio and then reinstall it.

- While executing Visual Studio code, got "System.AggregateException". The mistake made was not entering the correct image path. After correcting the path, the problem was resolved.