**Research Project Report**

**Classifying the coat of cats using both categorical features and visual features**

Devon Hope – 101038344

Yichen Dou – 101003897

Prof. Alan Tsang

COMP 4107 B, Winter 2021

Apr 23, 2021

**1 Introduction**

**1.1 Research Question**

How neural networks help predict the coat of cats using both categorical features and visual features.

**1.2 Background**

Classification problems are a popular topic in areas of machine learning nowadays. The neural network is feed with several images and outputs a specific label or labels for classification. However, what if we have extra information about the labels? For example, if we want to predict the price of a house based on an image and numerical values such as the number of rooms, levels, and location, there is a problem that numerical, categorical, and image features are trained well on different neural network architectures respectively. In this research, we will make a neural network with two input layers, one is for categorical features, and another is for images when predicting the coat of cats.

**1.3 Datasets**

Cat Breeds Dataset from https://www.kaggle.com/ma7555/cat-breeds-dataset

It is a dataset which contains a csv data and a bunch of images. The csv data includes categorical features such as age, gender, breed, coat and size of a cat, where each instance belongs to a single cat with related images.

**1.4 Machine Learning Libraries Used In The Research**

Tensorflow: A well-known open-source library for machine learning.

Keras: An open-source library that provides extra features to tensorflow. It offers an interface user can access to make artificial neural networks with many toolkits.

Scikit-learn: A free machine learning library that offers many algorithms in data analysis.

**2 Methodology**

**2.1 Data Analysis**

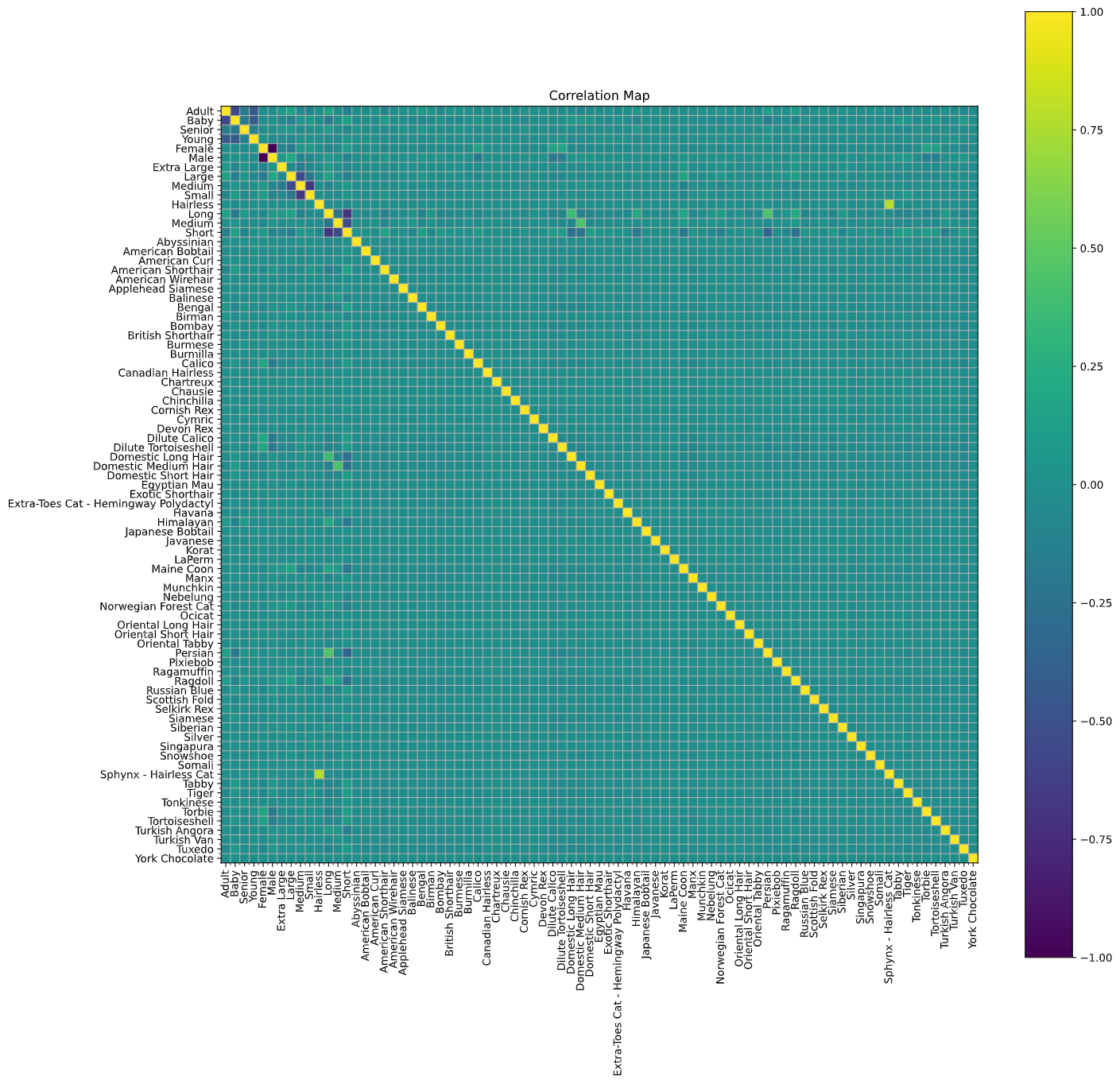
@TODO Find and describe problems in the datasets. Should we normalize data? Should we generate new attributes? How graphs look like? Analysis.ipynb

Missing values

Correlation map

A little number of images about a specific category (coat 4->3)

Etc.



**2.2 Data Preprocessing**

**2.2.1 Image Resizing**

When we load images from paths stored in the data file, we need to make sure all images should have the same dimension. Moreover, for a neural network model can successfully predict the coat of cats, the network should be present enough detail of images, such that patterns like hair edges can be detected. After many trials on a standard convolutional neural network, we found that resolutions above 128 pixels of width and 128 pixels of height give us an optimal result at an acceptable loss.

**2.2.2 Remove Instances That Contain Null Values**

Since the features we will train for the neural network are all categorical features, unlike numerical features, we may not fill empty spaces with mean values of a feature in this dataset. It is better for us to drop instances that contain null values in this case.

**2.2.3 Remove Labels That Contains Few Instances**

In our data analysis, we investigated several instances in each category of the coat. Then we determined that there are only 60 instances in hairless cats, comparing to more than thousands of instances in other coat categories. With a lack of data, the neural network may not learn its associated features well, so instances about hairless cats are removed.

**2.2.4 One-hot Encoding**

Our labels were represented as text, but the neural network needs them to be encoded before it can understand. Since our research problem is about making a single-label and multi-class classification network, we use the one-hot encoding as our technique to encode the text of coats.

**2.3 Neural Network Model**

@TODO NN architectures

**2.4 Training, Testing and Validation**

@TODO Plot some graphs, including testing and evaluation.

**2.4 Hyperparameters Tuning**

@TODO Testing different set-up for the net, including number of neurons, number of layers, learning rate, activation function etc.

**3 Results**

@TODO Figures

**4 Discussion**

@TODO Read from 3

**5 Conclusion**

@TODO Conclusion

**References**

APA Format