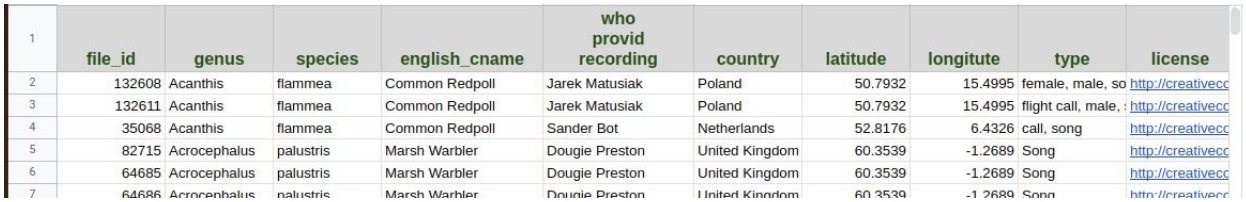
Introduction to Artificial Intelligence
Sima Faraji

What is the chosen dataset?
The chosen dataset contains 264 audio files (.flac) of the birdsongs of 88 unique species which usually are heard in the United Kingdom. This dataset also holds some of the species' specifications such as their English name, countries, genus, species, so forth; however, just some of which are used to solve the undermentioned problem.

What is the addressed problem?
The problem which is addressed by one of the Machine Learning models is to classify and identify these birds by processing the audio files. As can be seen in Figure1, there are ten columns in this dataset which most of them cannot help to solve this classification problem. In other words, the column of 'file_id' which is X, and 'english_name' that will be y are useful in this problem.



	file_id	genus	species	english_cname	who provid recording	country	latitude	longitute	type	license
2	132608	Acanthis	flammea	Common Redpoll	Jarek Matusiak	Poland	50.7932	15.4995	female, male, so	http://creativecc
3	132611	Acanthis	flammea	Common Redpoll	Jarek Matusiak	Poland	50.7932	15.4995	flight call, male,	http://creativecc
4	35068	Acanthis	flammea	Common Redpoll	Sander Bot	Netherlands	52.8176	6.4326	call, song	http://creativecc
5	82715	Acrocephalus	palustris	Marsh Warbler	Dougie Preston	United Kingdom	60.3539	-1.2689	Song	http://creativecc
6	64685	Acrocephalus	palustris	Marsh Warbler	Dougie Preston	United Kingdom	60.3539	-1.2689	Song	http://creativecc
7	64686	Acrocephalus	palustris	Marsh Warbler	Dougie Preston	United Kingdom	60.3539	-1.2689	Song	http://creativecc

Figure1

The data preprocessing and missing?
As can be seen in the image above, most of the columns such as genus, species, English name and country contain alphabetical values which cannot be suitable for feeding to any of classification models. Therefore, before feeding into the model, all of these columns' values must be converted to numerical unique values using OneHotEncoding method. For example, there are 18 unique values in the column of "country", and this means that an array with 18 elements should be created to encode this column, so that the first unique value, which is "Poland", must be something like the following array:
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1].
Likewise, the "United Kingdom" as the second unique value becomes:
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0].
And so on. All the alphabetical columns of the dataset must be converted to numerical through this encoding method by using OneHotEncoding class of Scikkit-Learn library. However, due to the fact that just the column of 'english_cname' will be used in this problem, the OneHotEncode method will be used only for this column. Moreover, as to be indicated in the Figure2, there are two columns which have three missing values in the whole dataset. To cope with this issue, the method called impute the missing value for continuous variable which replaces the missing values with the mean or median ones can be used because either the dataset is small and the columns with missing values are numerical (Kumar, 2020). Of course, this step also will not be required in this classification problem.

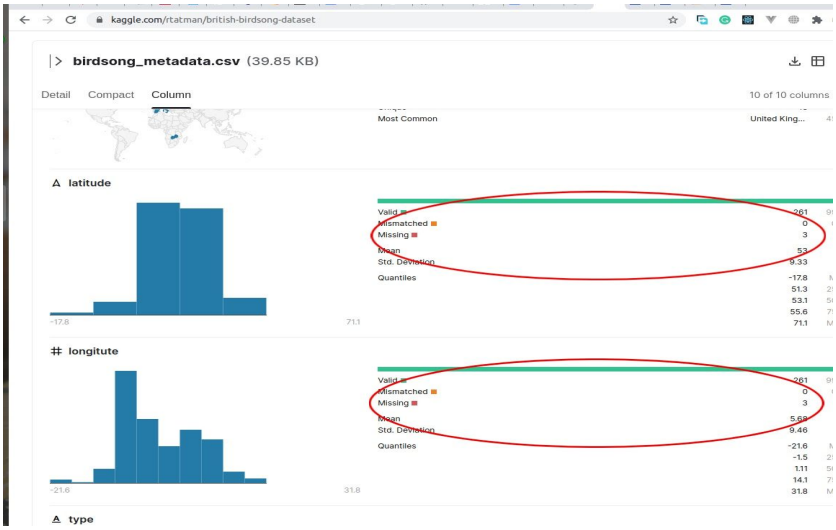


Figure2

Nevertheless, the final and most important step is to extract the corresponding Spectrograms and the features such as Spectral Centroid, Spectral Rolloff, Mel-Frequency Cepstral Coefficients, and so forth of the sound files which here are '.flac'. However, at first, let see what the spectrogram are. The spectrogram is an image in which the frequency - the number of a sound's repetition by itself at a given time (Natural Park Service, 2018) - is showed in a time-domain (The Experimental Writer, 2019) that its x-axis is time, the y-axis is the frequency of the given time, as shown in Figure3.

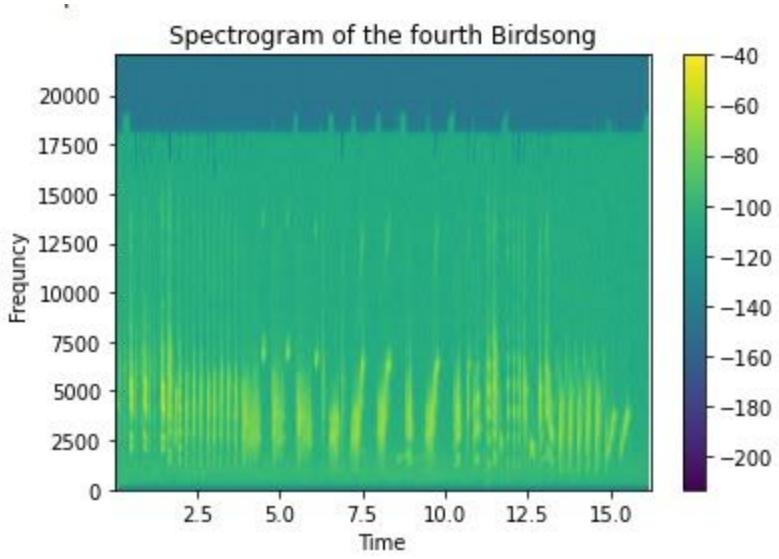


Figure3

After extracting the corresponding spectrograms for each one of the '.flac' files, they will be categorized in the folders with the English common name of the species, as indicated in Figure4.

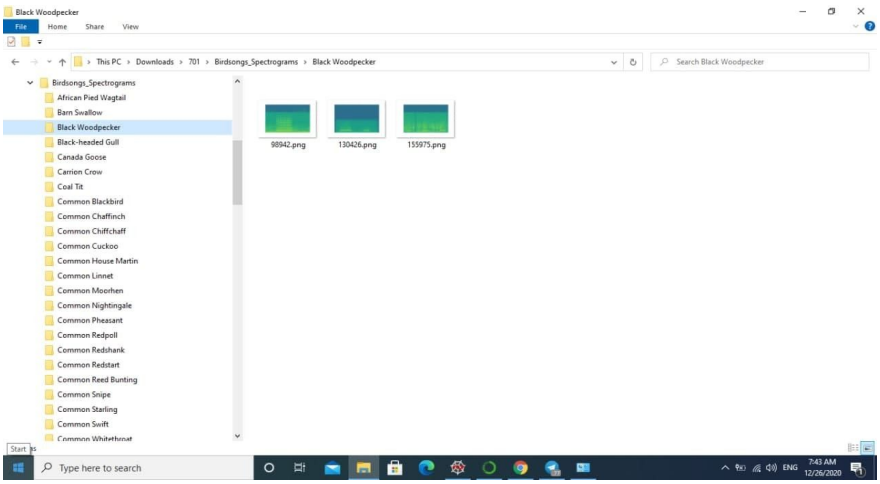


Figure4

The approaches for this Audio classification problem

There are two general approaches for classification of audio files:

1. Artificial Neural Network approach:
   In this method, the extraction of audio features will be carried out by the manual method using the feature function of the 'librosa' library, and the extracted features will be saved in a new csv file, categorised by the birds' English name. Then, the new csv file will be loaded and fed into the ANN model after performing the required preprocessing steps which here are the LabelEncoding, Normalization, and splitting the training set and test sets. In fact, the advantage of this approach is working with a lower amount of data and higher accuracy in this specific problem. Nevertheless, its disadvantage is the complexity of its algorithm and the requirements of more lines of code.
2. Convolutional Neural Network approach:
   The extraction of the audio features is an automatic task and performed by CNN in this method. In other words, CNN behave

References:

1. The Experimental Writer. (2019, March 22). *Audio Classification Using CNN — An Experiment*. Medium.com. https://medium.com/x8-the-ai-community/audio-classification-using-cnn-coding-example-f9cbd272269e
2. Natural Park Service. (2018, July 03). *Understanding Sound*. Natural Park Service. https://www.nps.gov/subjects/sound/understandingsound.htm
3. Kumar, S. (2020, July 24). *7 Ways to Handle Missing Values in Machine Learning*. TowardDataScience. https://towardsdatascience.com/7-ways-to-handle-missing-values-in-machine-learning-1a6326adf79e
4. Nagesh Singh Chauhan. (n.d.). *Audio Data Analysis Using Deep Learning with Python (Part 1)*. KDnuggets.com.

https://www.kdnuggets.com/2020/02/audio-data-analysis-deep-learning-python-part-1.html

5. Nagesh Singh Chauhan. (n.d.). *Audio Data Analysis Using Deep Learning with Python (Part2)*.KDnuggets.com. https://www.kdnuggets.com/2020/02/audio-data-analysis-deep-learning-python-part-2.html
6.