Assignment #2 - Audio Sensing
ECE 382V Human Signals: Sensing and Analytics - Spring 2026

*Due **February 19th 2026***

Q1. You have just been hired at Audio Intelligence Co. as a machine learning researcher. Your first job is to design an auditory scene analysis system that leverages ambient audio to identify whether a person is in one of four possible environments: *at a coffee shop*, *in the kitchen*, *at a soccer game, or at a party*. You are asked to build this system using Python, scikit-learn, and librosa (https://librosa.github.io).

To complete this task, four files have been made available to you: *coffee.wav*, *kitchen.wav, soccer.wav, and party.mp3.* These files are located inside the *'Q1 data' folder.* Evaluate the classifier using the split train/test metric (80% train, 20% test) with 3 learning algorithms of your choice. You are free to pre-process the files in anyway you believe would be most appropriate. Also, audio extraction and classification parameters such as frame size, whether you perform feature selection or not, are up to you.

Please submit a document (pdf format) explaining the rationale for the design choices you make along with your code. As detailed in the "Submitting the Assignment" section below, you may also turn in your code and comments as a jupyter notebook.

Q2. In this two-part problem, you will build a Cat Vocalization Recognizer (CVR) that consists of two components, a cat sound classifier and a cat vocalization classifier.

a) Cat Sound Classifier: YAMNet is a pre-trained deep neural network that can predict audio events from 521 classes. Using YAMNet, build and evaluate a binary classifier that can distinguish cat sounds vs. non-cat sounds.

b) Cat Vocalization Classifier: If the sound has been determined to be from a cat (based on output from part 'a'), you will run the sound through a 3-class vocalization classifier. This classifier must be built using features extracted through librosa, similarly to what you did in Q1.

The datasets for this problem are located inside 'Q2 data' folder. For part a), we are including a folder called 'animal-sounds-dataset' that includes sounds of cats and other animals. For part b), we are including a folder called 'cat-vocalization-dataset' that includes 440 cat vocalizations. These vocalizations fall in 3 categories: *brushing, isolation, and waiting*. More details about this dataset can be found in:

https://zenodo.org/record/4008297#.Y_QX4y-B1pQ

Like in Q1, please submit a document (pdf format) explaining the rationale for the design choices you make along with your code. As detailed in the "Submitting the Assignment" section below, you may also turn in your code and comments as a jupyter notebook.

Submitting the Assignment

Turn-in your assignment as a zip file including your code, data, and pdf files if applicable. You may also choose to turn in a jupyter notebook containing both your code and comments explaining your design choices. Report the performance of the classifiers in terms of accuracy, precision and recall metrics. Label which files correspond to which answers very carefully. Make it intelligible. If this is not clear, you will not receive credit for your solution.

Grading

For both Q1 and Q2, your grade will be based not only on whether your results are reasonable but also on the design choices that you make regarding features, pre-processing, frame extraction, classifier, etc.