1) Domain Background

People are always in contact with audio data. The human brain is continuously processing and understanding this audio data, either consciously or subconsciously, giving us information about the environment around us. Environmental sound classification is a growing area of research with numerous real world applications. I would like to focus on solving automations problems while learning environmental sound classification. The machines running around us have a variety of sounds during their work. For example, when the press machine produces faulty parts during operation, the sound coming voice is different. Also, The sound of the strap from the hood while the car is running. In this project, I will investigate this problem. But I no avilable a enought data for my problem.

The goal of this capstone project, is to apply Deep Learning techniques to the classification of environmental sounds, specifically focusing on the identification of particular urban sounds.

There is a plethora of real world applications for this research, such as:

- Can be used for hearing impaired people.

- Smart home use

- Automotive where recognising sounds both inside and outside of the car can improve safety

- Industrial uses such as predictive maintenance

- Used for quality controls.

2) Problem Statement

The main objective of this project will be to use Deep Learning techniques to classify urban sounds.

When given an voice sample in a computer readable format (.wav file) of a few seconds duration, I want to be able to determine if it contains one of the target urban sounds with a corresponding likelihood score. Conversely, if don't detect of the target sounds, we will be presented with an unknown score.

3) Datasets and Inputs

For this project we will use a dataset called Urbansound8K [1]. The dataset contains 8732

sound excerpts(~4s) of urban sounds from 10 classes, which are:

- Air Conditioner

- Car Horn

- Children Playing

- Dog bark

- Drilling

- Engine Idling

- Gun Shot

- Jackhammer

- Siren

- Street Music

The accompanying metadata contains a primary key ID for each sound excerpt along with it's given class name. These sound excerpts are digital audio files in .wav format. Sound waves are digitised by sampling them at discrete intervals known as the sampling rate. The data we will be analysing for each sound excerpts is essentially a one dimensional array or vector of amplitude values. (dataset taxonomy)[3] is explain to dataset which With the exception of “children playing”and “gun shot”which were added for variety, all other classes were selecteddue to the high frequency in which they appear in urbannoise complaints

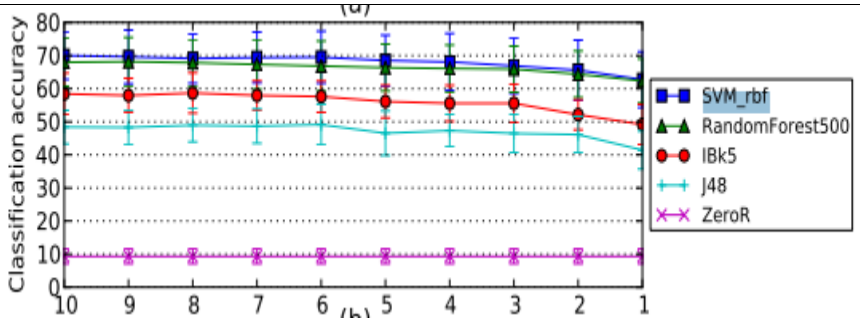
4 Solution Statement

The proposed solution to this problem is to apply Deep Learning techniques. (MFCC) [2] is the audio samples on a per-frame basis with a window size of a few milliseconds. The MFCC summarises the frequency distribution across the window size, so it is show to analyse both the frequency and time characteristics of the sound. These audio representations will allow us to identify features for classification. After that, we will be to train a Deep Neural Network with these data sets and make predictions.

I think this will be very effective at finding patterns within the MFCC's much like they are effective at finding patterns within images. I will use the evaluation metrics described in later sections to compare the performance of these solutions against the benchmark models in the next section.

5 Benchmark Model

For the benchmark model, I will use the algorithms outlined in the paper " A Dataset and Taxonomy for Urban Sound Research " (Salamon, 2014) [3]. The paper describes five different algorithms with the following accuracies for a audio slice maximum duration of 4 seconds.



Algorithm	Accuracy
SVM_rbf	68%
RandomForest500	66%
IBk5	55%
J48	48%
ZeroR	10%

6 Evaluation Metrics

Our dataset balanced is quite relatively. [3] (THE URBANSOUND DATASET) .

The evaluation metric for this problem is simply the Accuracy Score.

7 Project Design

Data Preprocessing

First identify the different data types in our dataset and what preprocessing needs to be done to make it uniform.

● resample so all audio had the same sample rate and bit depth

● make sure the sample duration is uniform

● Consider any data augmentations, such as adding background noise (though this maybe a nice to have)

Data Splitting

Split the data into a training set and validation set with an 80-20 split.

Model training and evaluation

I will start with the simple model architecture first before training and evaluating it. Then iterate this process trying different architectures and hyper-parameters to reach an accuracy score we are happy with.

8 References

[1] Justin Salamon, Christopher Jacoby and Juan Pablo Bello, “Urban Sound Datasets”, “UrbanSound8K” https://urbansounddataset.weebly.com/urbansound8k.html
[2] Mel-frequency cepstrum Wikipedia page
https://en.wikipedia.org/wiki/Mel-frequency_cepstrum
[3] J. Salamon, C. Jacoby, and J. P. Bello, “A dataset and taxonomy for urban sound research”
http://www.justinsalamon.com/uploads/4/3/9/4/4394963/salamon_urbansound_acmmm14.pdf