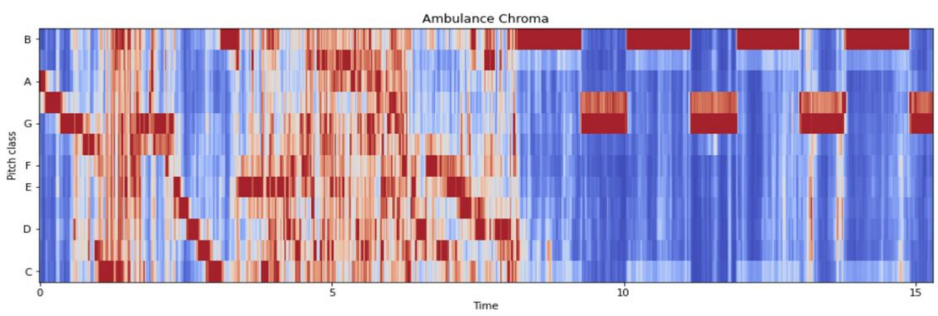
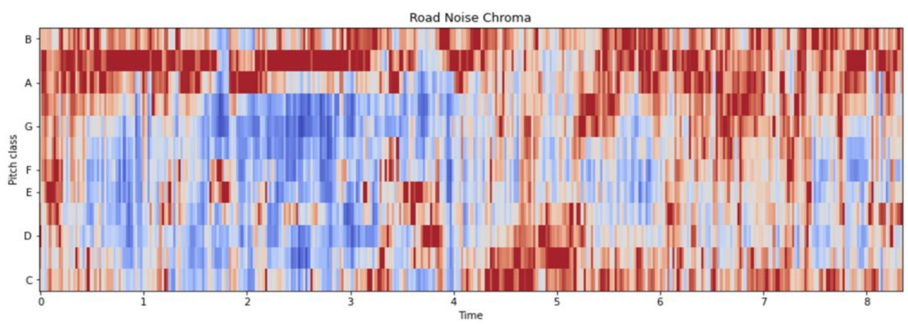
Sounds of Urgency: Investigating an Audio Dataset for Emergency Vehicle Sirens and Road Noises

Final Presentation

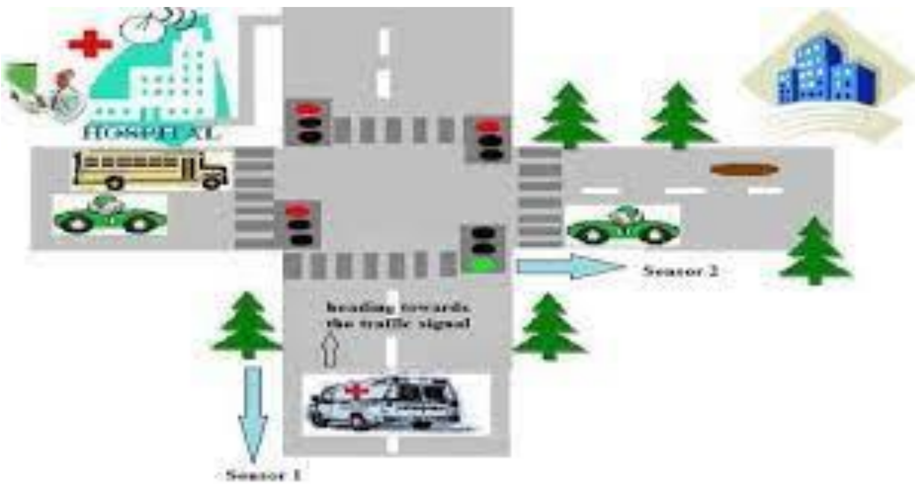




Authors: Frances Cue, Jerry Gonzalez, Nick Johnson, and Chi So

Berkeley
UNIVERSITY OF CALIFORNIA

What are the potential applications and benefits of analyzing the audio dataset?



1. Traffic congestion, accidents and pollution are a reality in today's society [1].
2. How do we solve the problem?

    a) Improve current infrastructure (Costly and not always practical) [2]

    b) Use the latest technology such as AI to make better use of the current infrastructure. [2]

3. Using AI techniques to distinguish emergency vehicles from traffic and road noise can improve traffic flow and reduce congestion [3].
4. Additionally, emergency response times can be improved for fire and health events.

Berkeley
UNIVERSITY OF CALIFORNIA

Summary of results

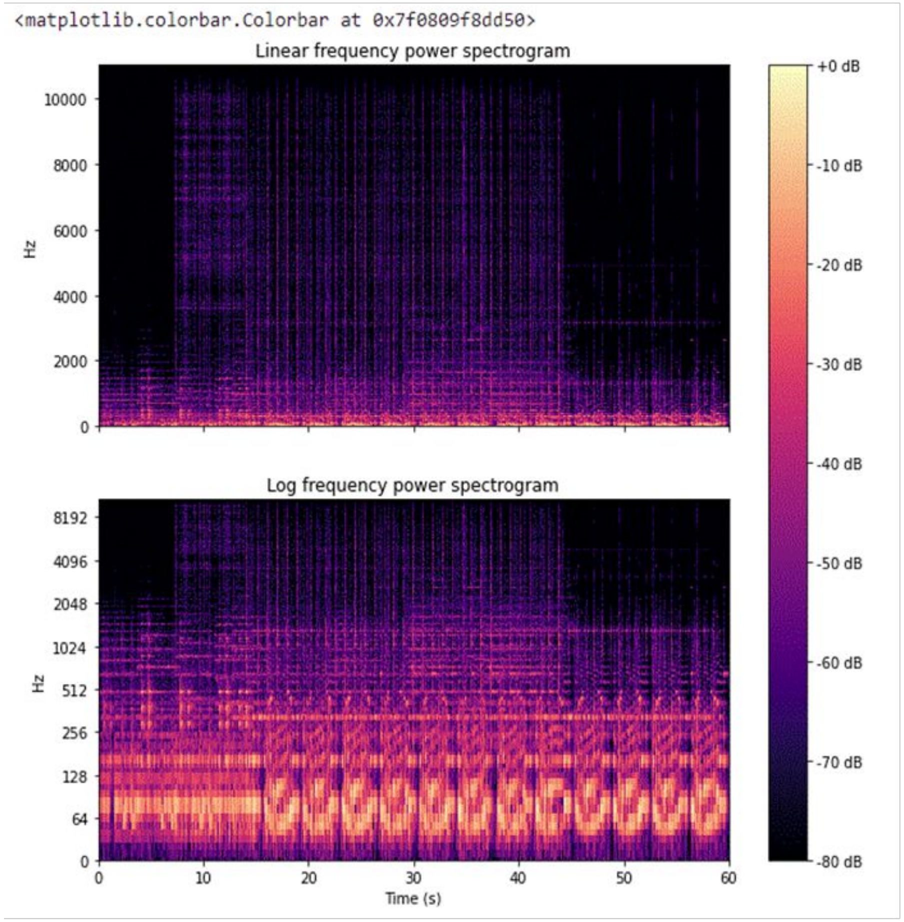
Model	Precision	Accuracy	F1 Score
Baseline (Logistic regression)	96.5%	95.3%	95.4%
KNN	86.2%	86.0%	86.2%
CNN	97.3%	98.2%	98.2%
LDA	95.3%	95.3%	95.3%
SVM	97.6%	97.5%	97.5%
FFNN	92.2%	92.0%	92.0%
RNN	93.5%	93.5%	93.5%
K-means + RNN	96.0%	96.0%	95.6%

Berkeley
UNIVERSITY OF CALIFORNIA

Data

Feature	Definition	High	Low
Chroma STFT	Represents the pitch content of a sound	Violin	Bass guitar
RMSE	Root Mean Square(Energy) , average energy of audio signal over certain timeframe.	Rock Concert	Whisper
Spectral centroid	Represents the center of mass of the frequency spectrum of a sound. “Brightness” of a sound	Trumpet	Tuba
Spectral bandwidth	Represents the width of the frequency spectrum of a sound. It is the variance from the spectral centroid.	Violin	Bass Drum
Rolloff	Measure of how quickly the energy of the sound decreases as it moves into different frequencies.	Guitar strum	One hit on a drum
Zero Crossing Rate	When a signal changes from positive to negative and vice versa.	Human speech	White noise
Mel Frequency Cepstral Coefficient 1- 20	MFCC is a compact representation of the spectrum of an audio signal. Each of the 20 coefficients extracted contains information about the rate change in each spectrum band.	Female voice	Male voice

Data Preprocessing: Librosa Library

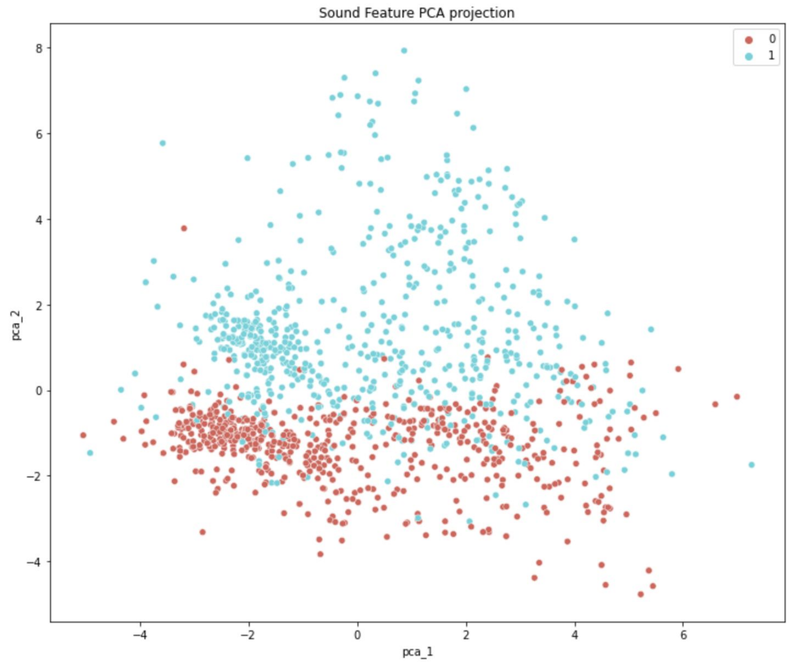


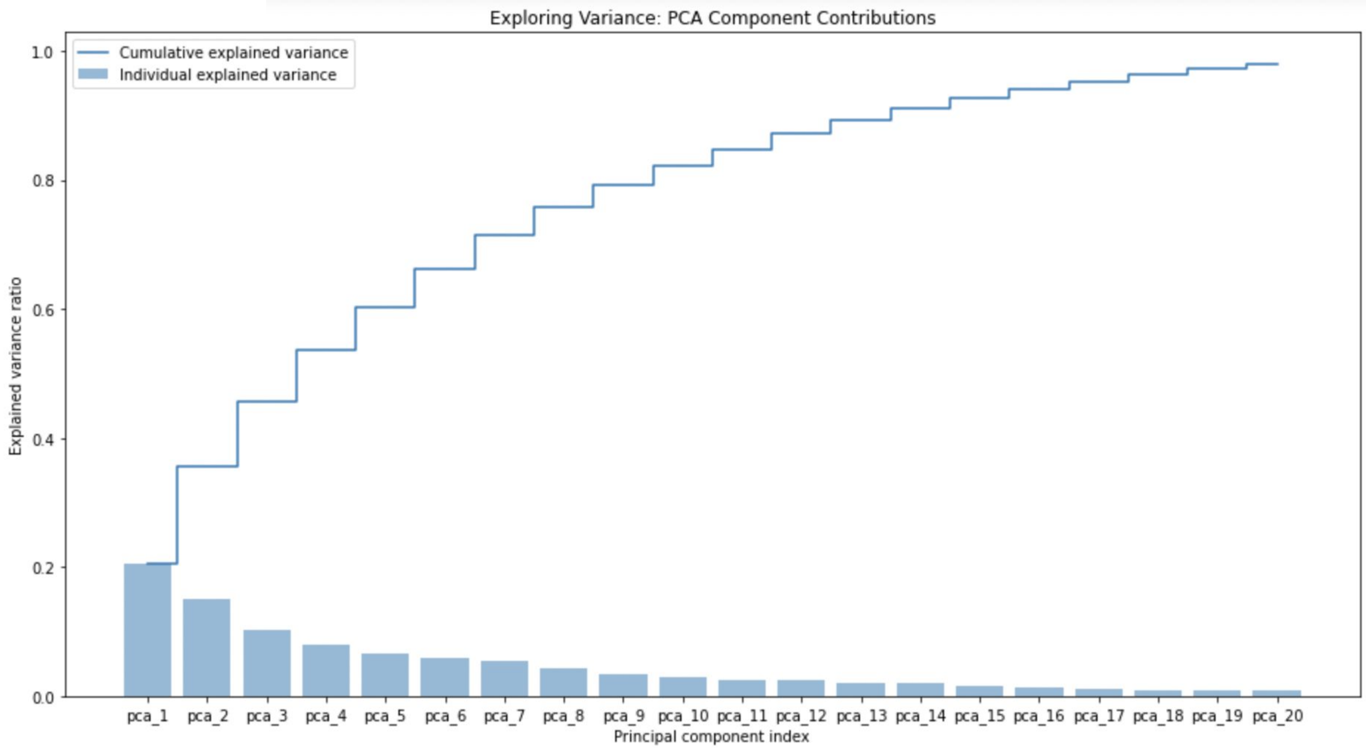
- Original dataset had an error in the creation of csv files. So we regenerated the emergency vehicle and noise csv files from the audio files.
- Used the Librosa library to extract the features used in our models.
- What is librosa?
- Features extracted

1. Chroma stft or chromagram from wafeform
2. Rmse: root-mean-square energy
3. Spectral centroid
4. Spectral bandwidth
5. Roll off
6. Zero crossing rate
7. Mel-frequency cepstral coefficients (MFCCS)

Berkeley
UNIVERSITY OF CALIFORNIA

PCA


Sound Feature PCA projection
0
1
pca_2
pca_1


Exploring Variance: PCA Component Contributions
Cumulative explained variance
Individual explained variance
Explained variance ratio
Principal component index
pca_1 pca_2 pca_3 pca_4 pca_5 pca_6 pca_7 pca_8 pca_9 pca_10 pca_11 pca_12 pca_13 pca_14 pca_15 pca_16 pca_17 pca_18 pca_19 pca_20

Approach

Tabular Data

Logistic Regression (Baseline)

K-Nearest Neighbors (KNN)

Linear Discriminant Analysis (LDA)

Support Vector Machine (SVM)

Feed Forward Neural Network (FFNN)

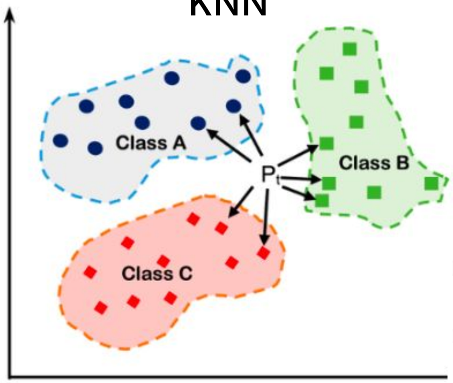
Spectrogram Data

Convolution Neural Network (CNN)

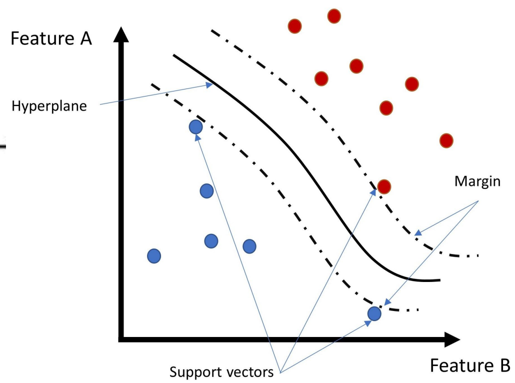
Recurrent Neural Network (RNN/LSTM)

K-means clustering + LSTM

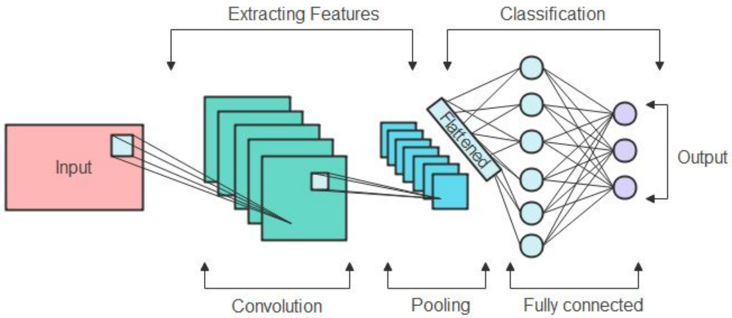
KNN



SVM



CNN



Berkeley
UNIVERSITY OF CALIFORNIA

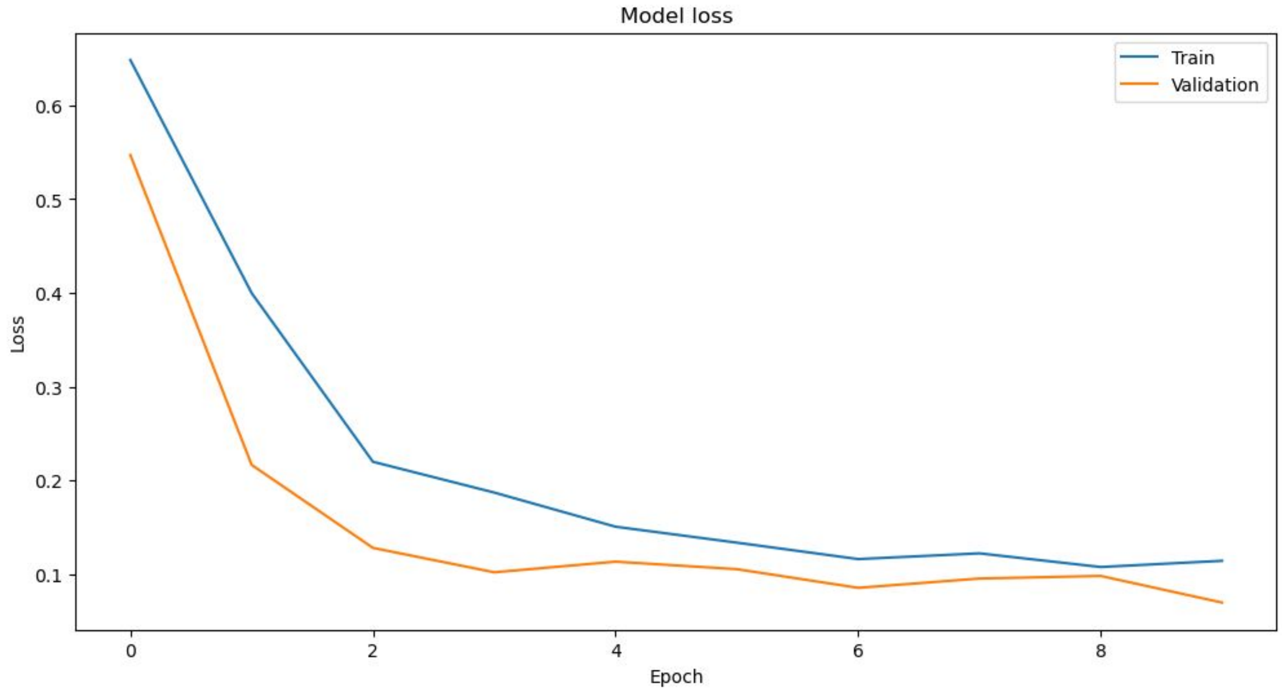
Experiments

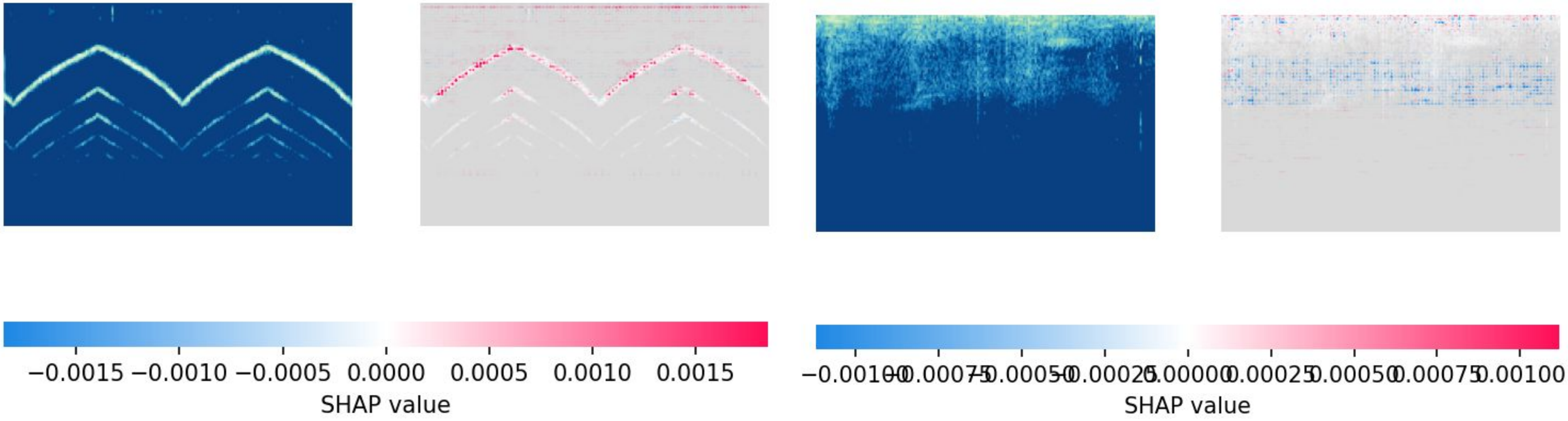
Implemented RandomizedSearchCV to fit 100 models with following options:

```
param_distributions = {
    'filters': [[64, 48], [32, 32]],
    'kernel_size': [(5, 5), (3, 3)],
    'strides': [(1, 1), (2, 2)],
    'pool_size': [(2, 2), (3, 3)],
    'activation': ['relu', 'tanh'],
    'dropout_rate': sp.stats.uniform(scale = 1),
    'hidden_layer_sizes': [[500], [1000], [500, 500]],
    'learning_rate': [0.001, 0.01]
}
```

	mean_test_score	std_test_score	params	filters	kernel_size	strides	pool_size	activation	dropout_rate	hidden_layer_sizes	learning_rate
86	0.964156	0.014429	{'activation': 'relu', 'dropout_rate': 0.47746...	[64, 48]	(3, 3)	(2, 2)	(2, 2)	relu	0.477461	[500]	0.001
44	0.962591	0.003268	{'activation': 'relu', 'dropout_rate': 0.28445...	[32, 32]	(5, 5)	(2, 2)	(3, 3)	relu	0.284453	[500, 500]	0.001
12	0.961034	0.012252	{'activation': 'tanh', 'dropout_rate': 0.13780...	[32, 32]	(5, 5)	(2, 2)	(3, 3)	tanh	0.137800	[1000]	0.001
69	0.960258	0.008285	{'activation': 'relu', 'dropout_rate': 0.52536...	[32, 32]	(3, 3)	(2, 2)	(2, 2)	relu	0.525369	[500]	0.001
79	0.959478	0.008574	{'activation': 'relu', 'dropout_rate': 0.59552...	[32, 32]	(5, 5)	(1, 1)	(3, 3)	relu	0.595525	[1000]	0.001

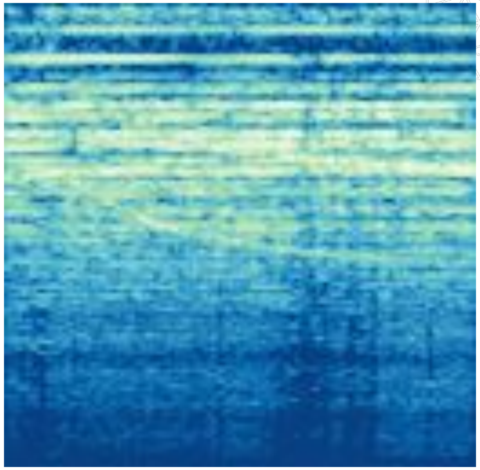
Experiments



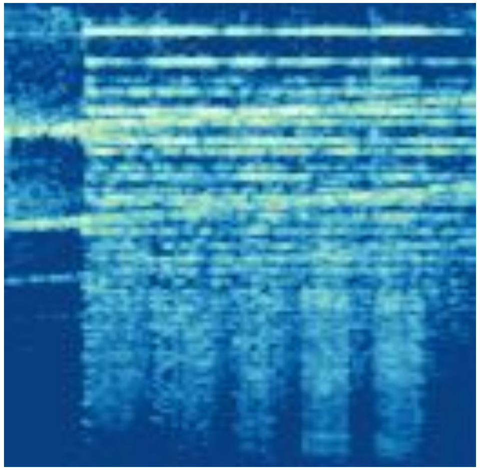


Fairness

ambulance438.wav



road536.wav

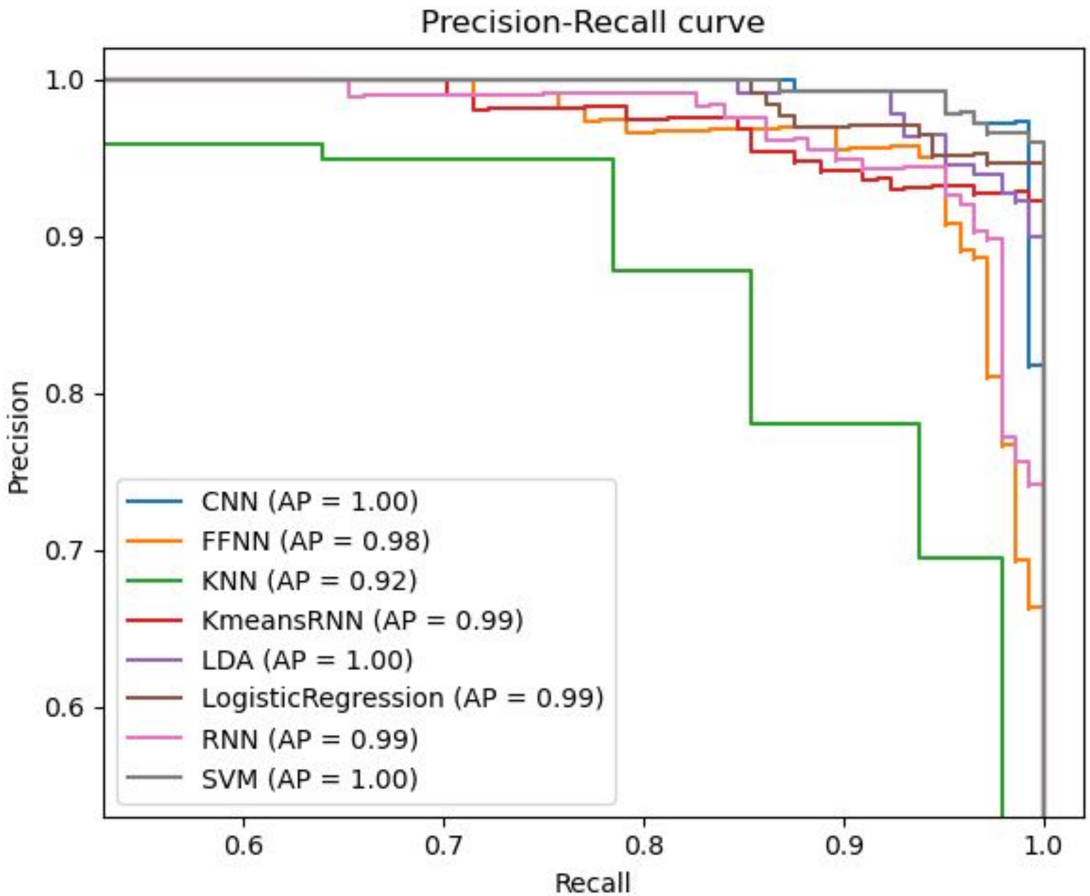


Evaluation

- Balanced
  - Logistic Regression
  - FFNN
  - KNN

- Precision favored
  - CNN
  - LDA
  - SVM
  - RNN
  - Kmeans_RNN

- Our recommendation
  - CNN
  - SVM



References:

1. Litman T (2013) Congestion Costing Critique: Critical Evaluation of the "Urban Mobility Report"

2. Lakshmi Shankar Iyer, AI enabled applications towards intelligent transportation, Transportation Engineering, Volume 5, 2021

3. Shamsi, M., Rasouli Kenari, A. & Aghamohammadi, R. Reinforcement learning for traffic light control with emphasis on emergency vehicles. *J Supercomput* **78**, 4911–4937 (2022)

Daly, Heather, "The influences of musical training and spectral centroid on perceptual interactions of pitch and timbre" (2017). Masters Theses. 516. https://commons.lib.jmu.edu/master201019/516

https://devopedia.org/audio-feature-extraction#:~:text=The%20spectral%20bandwidth%20or%20spectral,correlation%20with%20the%20perceived%20timbre.

Berkeley
UNIVERSITY OF CALIFORNIA

Contributions

Areas	Nick	Frances	Jerry	Chi
Data Set Discovery	x	x	x	x
Preprocessing		x	x	x
Meeting Logistics	x	x	x	x
PCA	x	x		x
EDA		x		x
Logistic Regression, CNN	x			
KNN			x	
LDA, SVM, RNN, Kmeans RNN				x
Experimentation/Hyper Param Tuning	x	x	x	x
Slide Creation	x	x	x	x