

# About Dataset

This dataset contains 8732 labeled sound excerpts ( $\leq 4$ s) of urban sounds from 10 classes: `air_conditioner`, `car_horn`, `children_playing`, `dog_bark`, `drilling`, `engine_idling`, `gun_shot`, `jackhammer`, `siren`, and `street_music`. The classes are drawn from the urban sound taxonomy. For a detailed description of the dataset and how it was compiled please refer to our paper. All excerpts are taken from field recordings uploaded to [www.freesound.org](http://www.freesound.org). The files are pre-sorted into ten folds (folders named fold1-fold10) to help in the reproduction of and comparison with the automatic classification results reported in the article above. In addition to the sound excerpts, a CSV file containing metadata about each excerpt is also provided.

## AUDIO FILES INCLUDED

8732 audio files of urban sounds (see description above) in WAV format. The sampling rate, bit depth, and number of channels are the same as those of the original file uploaded to Freesound (and hence may vary from file to file).

## META-DATA FILES INCLUDED

UrbanSound8k.csv

This file contains meta-data information about every audio file in the dataset. This includes:

- `slice_file_name`:  
The name of the audio file. The name takes the following format: `[fs ID]-[class ID]-[occurrence ID]-[slice ID].wav`, where:  
`[fs ID]` = the Freesound ID of the recording from which this excerpt (slice) is taken  
`[class ID]` = a numeric identifier of the sound class (see description of class ID below for further details)  
`[occurrence ID]` = a numeric identifier to distinguish different occurrences of the sound within the original recording  
`[slice ID]` = a numeric identifier to distinguish different slices taken from the same occurrence
- `fs ID`:  
The Freesound ID of the recording from which this excerpt (slice) is taken.
- `start`  
The start time of the slice in the original Free sound recording.
- `end`:  
The end time of slice in the original Free sound recording

- salience:  
A (subjective) salience rating of the sound. 1 = foreground, 2 = background.
- fold:  
The fold number (1-10) to which this file has been allocated.
- Classic:  
A numeric identifier of the sound class:  
0 = air conditioner  
1 = car horn  
2 = children playing  
3 = dog bark  
4 = drilling  
5 = engine idling  
6 = gunshot  
7 = jackhammer  
8 = siren  
9 = street music
- class:  
The class name: air conditioner, car horn, children playing, dog bark, drilling, engine idling, gunshot, jackhammer, siren, street music.

## BEFORE YOU DOWNLOAD: AVOID COMMON PITFALLS!

Since releasing the dataset we have noticed a couple of common mistakes that could invalidate your results, potentially leading to manuscripts being rejected or the publication of incorrect results. To avoid this, please read the following carefully:

1. Don't reshuffle the data! Use the predefined 10 folds and perform 10-fold (not 5-fold) cross validation

The experiments conducted by vast majority of publications using UrbanSound8K (by ourselves and others) evaluate classification models via 10-fold cross validation using the predefined splits\*. We strongly recommend following this procedure.

- 2.

Why?

If you reshuffle the data (e.g. combine the data from all folds and generate a random train/test split) you will be incorrectly placing related samples in both the train and test sets, leading to inflated scores that don't represent your model's performance on unseen data. Put simply, your results will be wrong. Your results will NOT be comparable to previous results in the literature, meaning any claims to an improvement on previous research will be invalid. Even if you don't reshuffle the data, evaluating using different splits (e.g. 5-fold cross validation) will mean your results are not comparable to previous research.

1. Don't evaluate just on one split! Use 10-fold (not 5-fold) cross validation and average the scores

We have seen reports that only provide results for a single train/test split, e.g. train on folds 1-9, test on fold 10 and report a single accuracy score. We strongly advise against this. Instead, perform 10-fold cross validation using the provided folds and report the average score.

- 2.

Why?

Not all the splits are as "easy". That is, models tend to obtain much higher scores when trained on folds 1-9 and tested on fold 10, compared to (e.g.) training on folds 2-10 and testing on fold 1. For this reason, it is important to evaluate your model on each of the 10 splits and report the average accuracy. Again, your results will NOT be comparable to previous results in the literature.

## **Acknowledgements**

We kindly request that articles and other works in which this dataset is used cite the following paper:

J. Salamon, C. Jacoby and J. P. Bello, "A Dataset and Taxonomy for Urban Sound Research", 22nd ACM International Conference on Multimedia, Orlando USA, Nov. 2014.

More information at <https://urbansounddataset.weebly.com/urbansound8k.html>