Audio & Speech: Environment Sound Classification

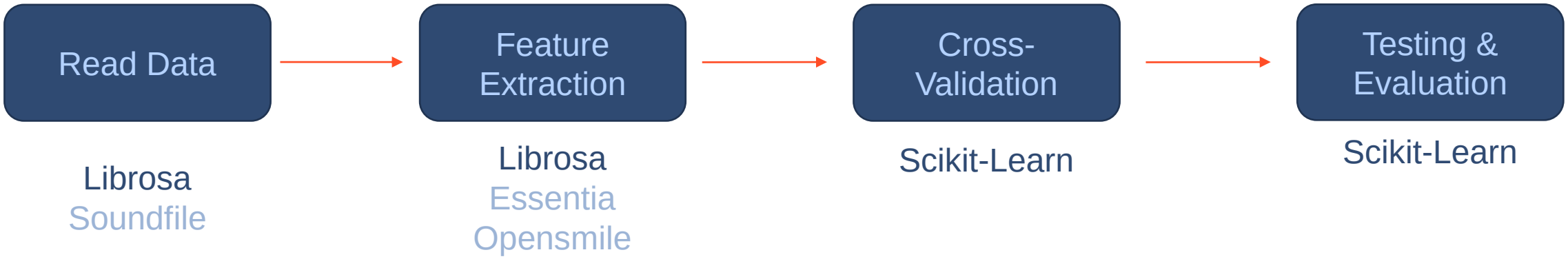
EE366200: Digital Signal Processing Lab

#Lab3 – Oct. 3, 2022

Prof. Chi-Chun Lee, Yi-Wen Liu

TAs: 林蔭澤、楊晶宇、鄭語芳、張藾云

Sound Classification Flow Chart



Data

- The ESC dataset is a collection of short environmental recordings available in a unified format (5-second-long clips, 44.1 kHz, single channel)
- Total: 400 Files, 40 in each class

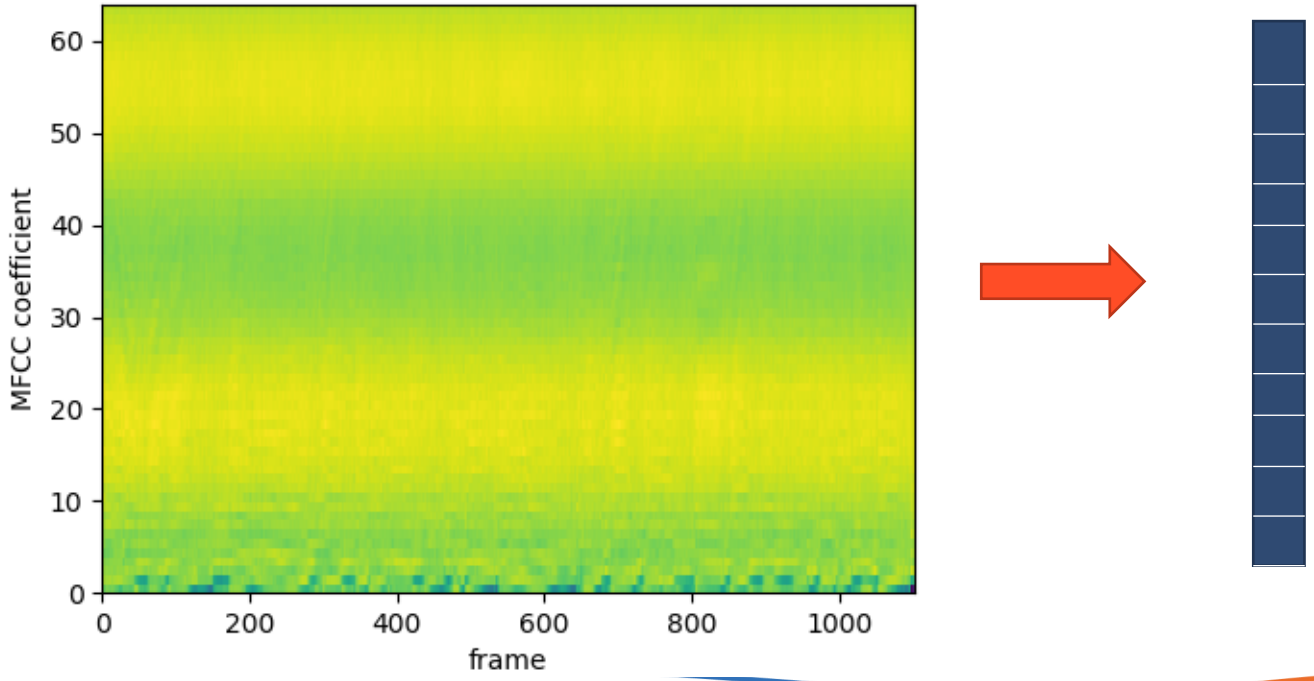
Karol J. Piczak, 2015, "ESC: Dataset for Environmental Sound Classification"

Dog bark	Rain	Sea waves	Baby cry	Clock tick
Person sneeze	Helicopter	Chainsaw	Rooster	Fire crackling

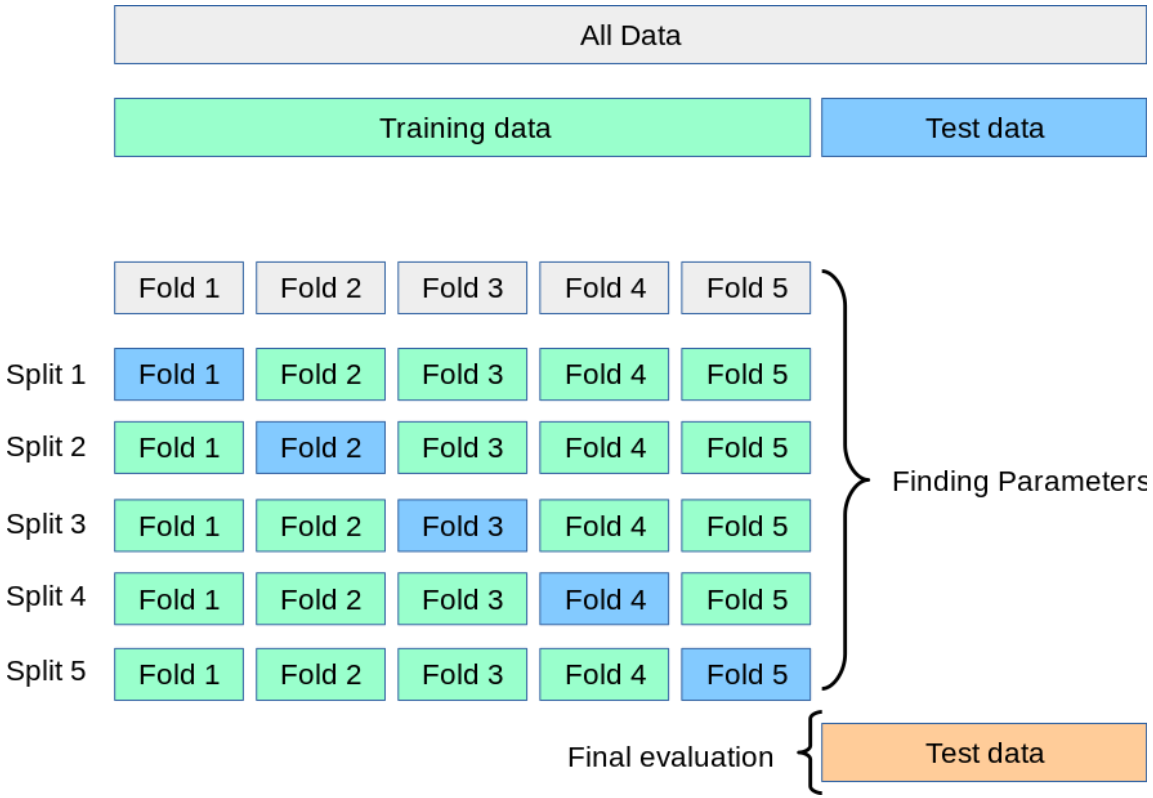
Component (1) – Acoustic Feature Extraction

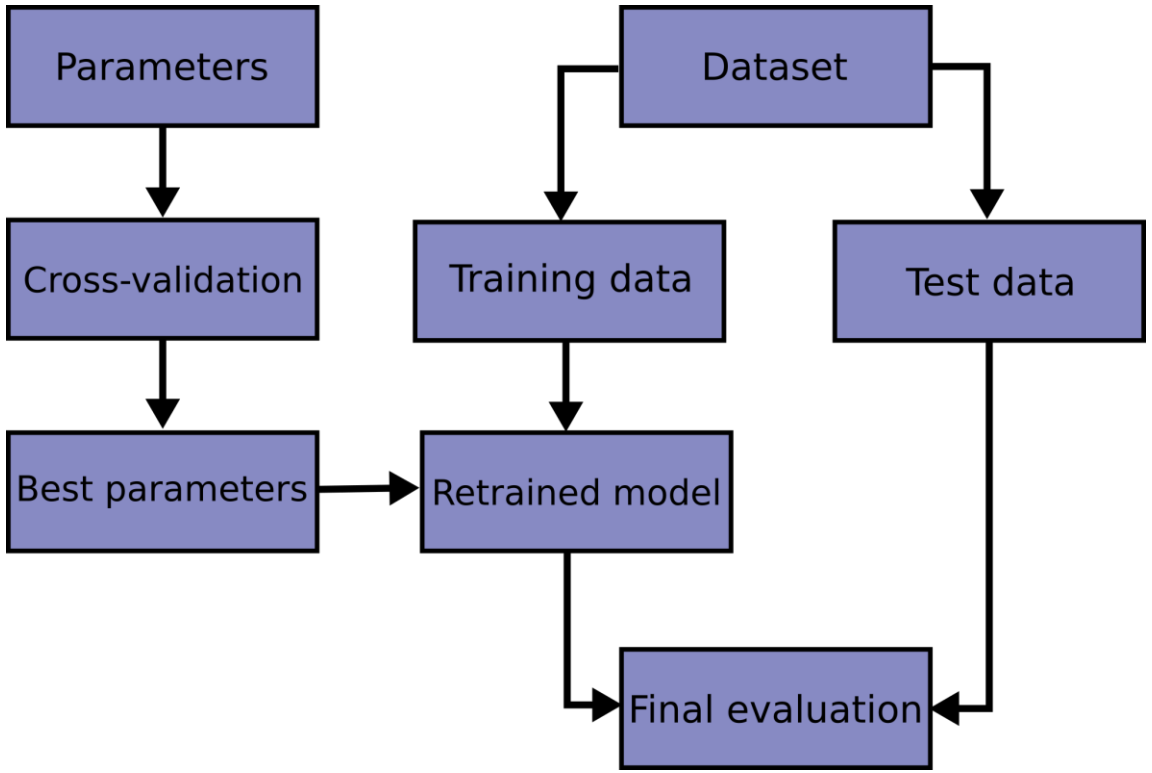
Use librosa.load to read sound files

Use librosa.feature.mfcc to calculate MFCC (returns 2D N*T array)
Get statistic (mean, std, …) along time axis to reduce it into 1D vector

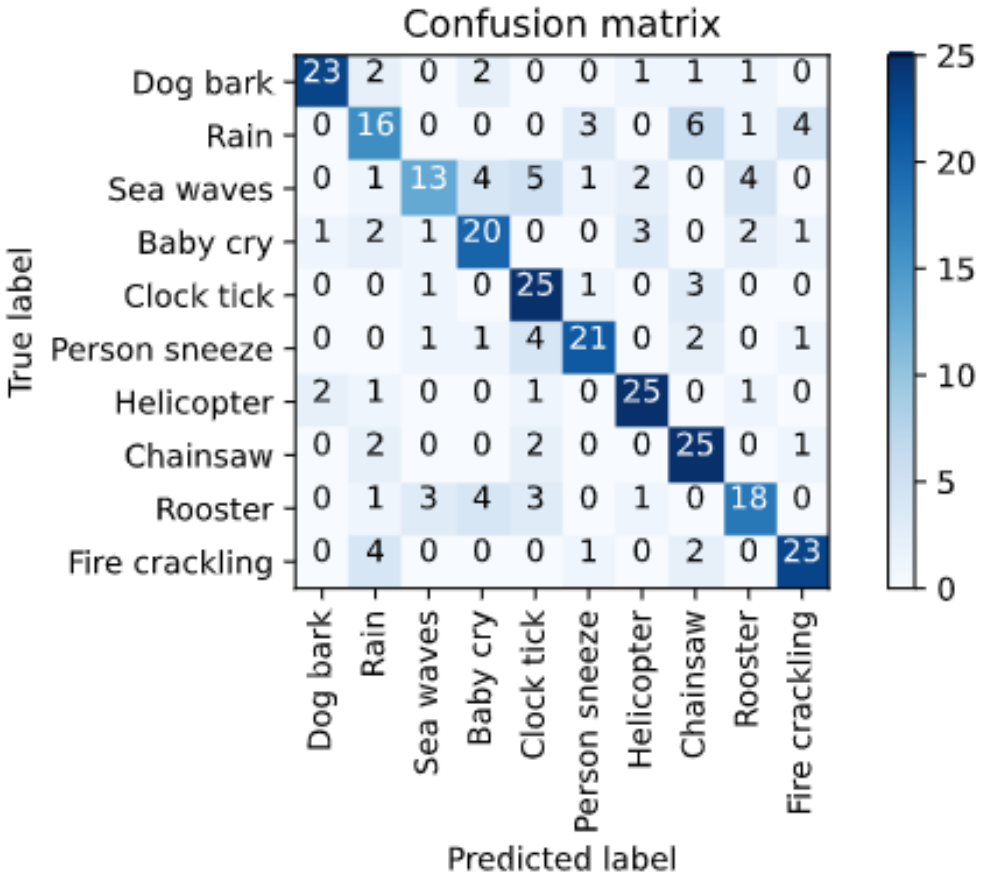


Component (2) – Cross-Validation



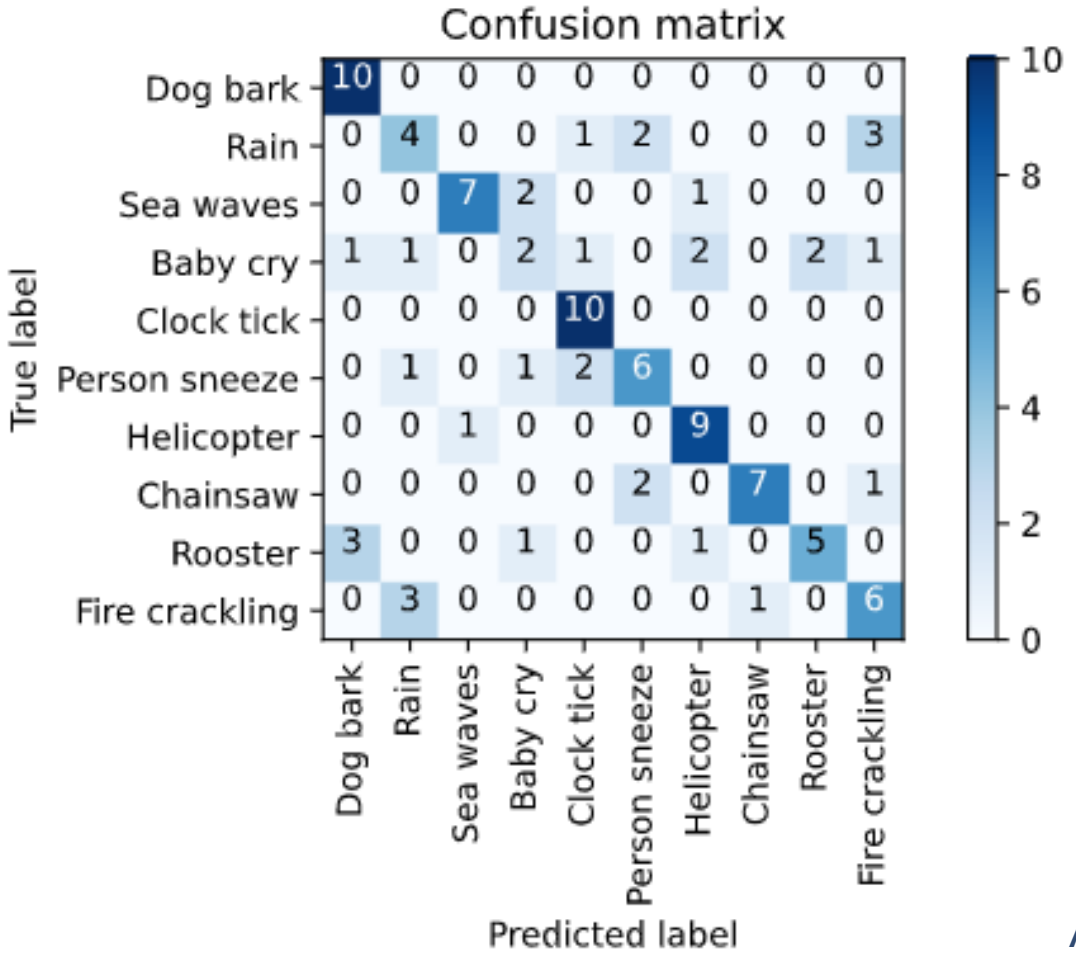


Demo (1) – Cross Validation Results



ACC = 0.697

Demo (2) – Testing Evaluation



ACC = 0.66

Report Questions

- (Bonus) Try different features, statistics functions, classifiers to achieve higher performance

- (Bonus) You get extra credit if you achieve 75+% accuracy on testing set. Briefly explain what and how you managed to make the prediction better.