Audio & Speech
Week4: Baby Sounds Challenge

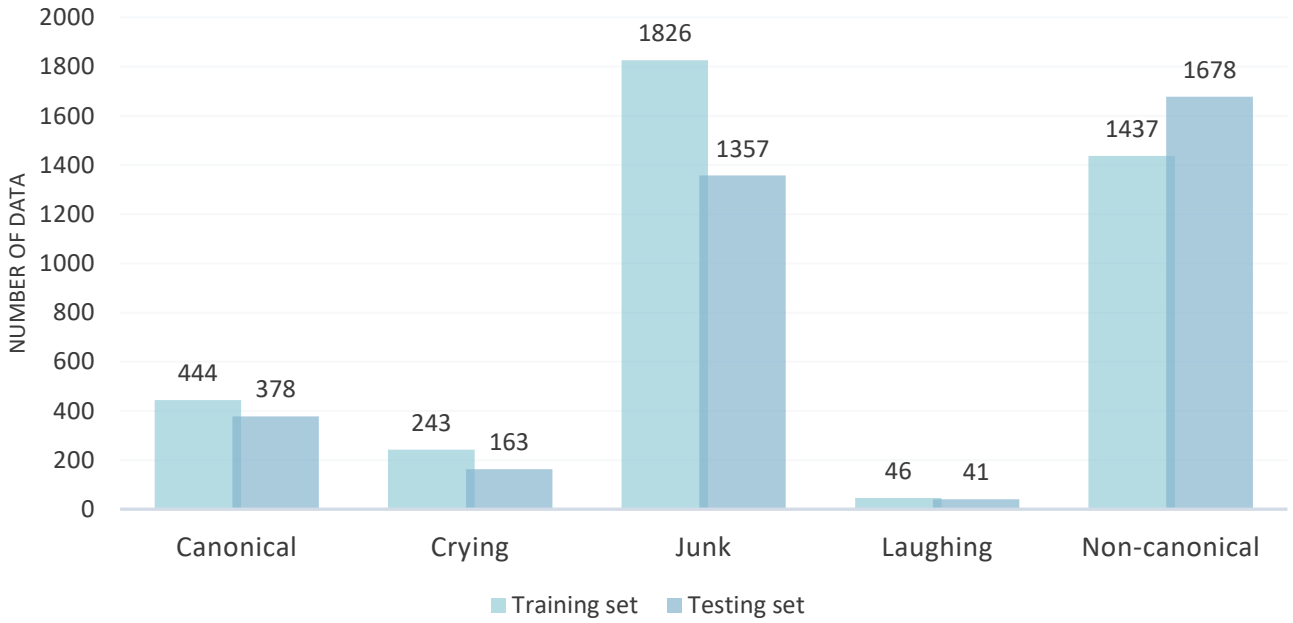
Prof. Chi-Chun Lee, Yi-Wen Liu
TAs: 楊晶宇、林蔭澤、鄭語芳、張爾云

EE3662: Digital Signal Processing Lab

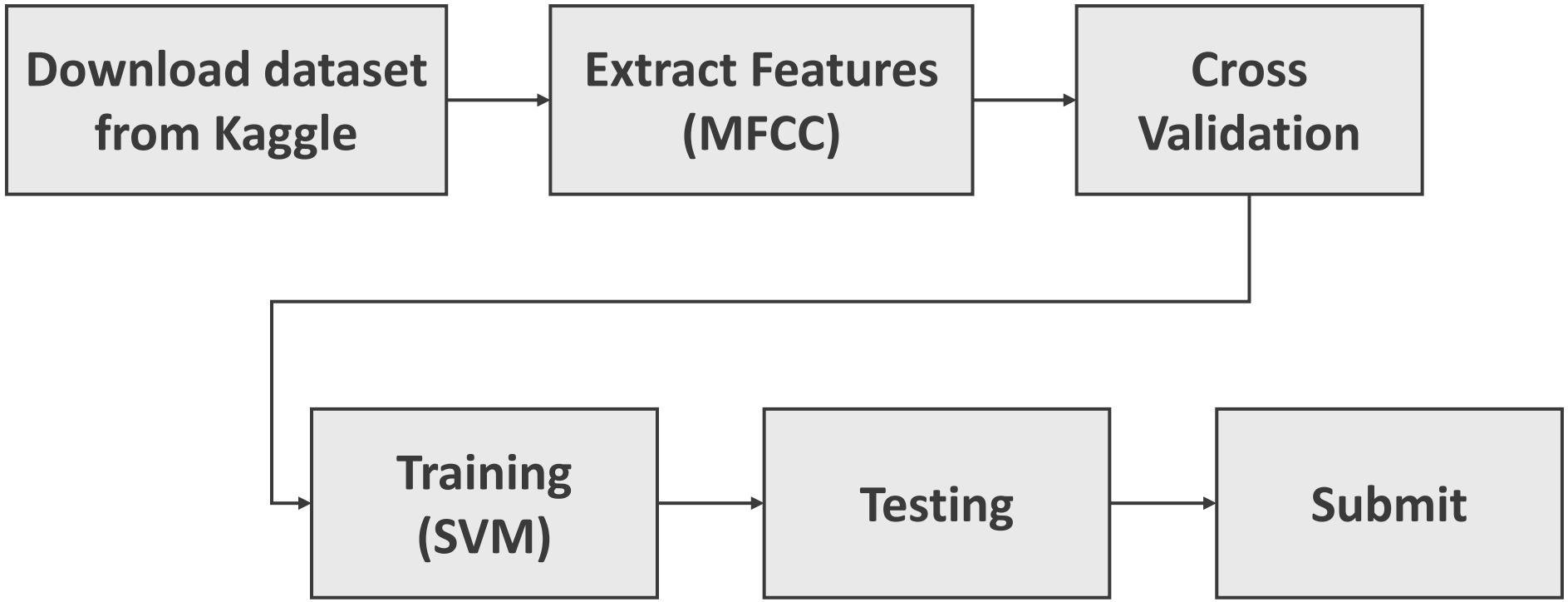
#Lab4– Oct. 17, 2022

Dataset Description

- **Baby Sounds Challenge**
  - The INTERSPEECH 2019 Computational Paralinguistics Challenge
- **Datasets**
  - Training set: 3996 files
  - Testing set: 3617 files
- **Download Datasets**
  - Kaggle → Data

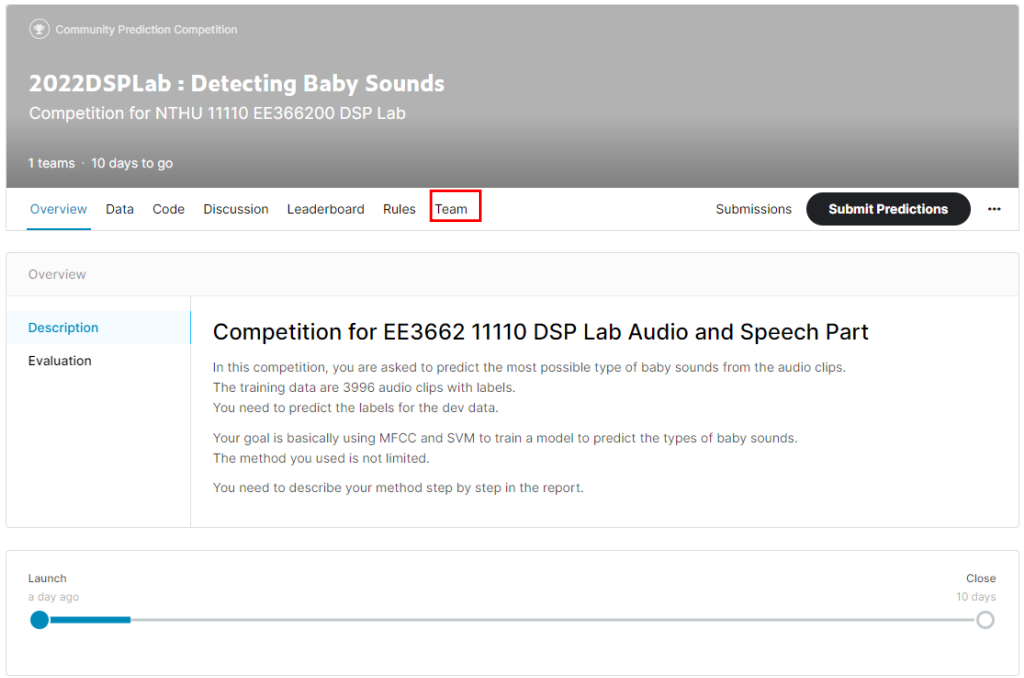


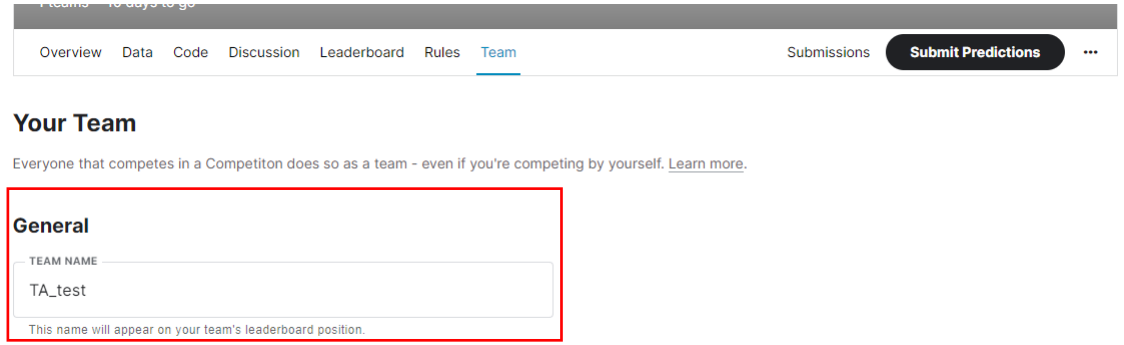
Flow Chart


Download dataset from Kaggle
Extract Features (MFCC)
Cross Validation
Training (SVM)
Testing
Submit

Kaggle

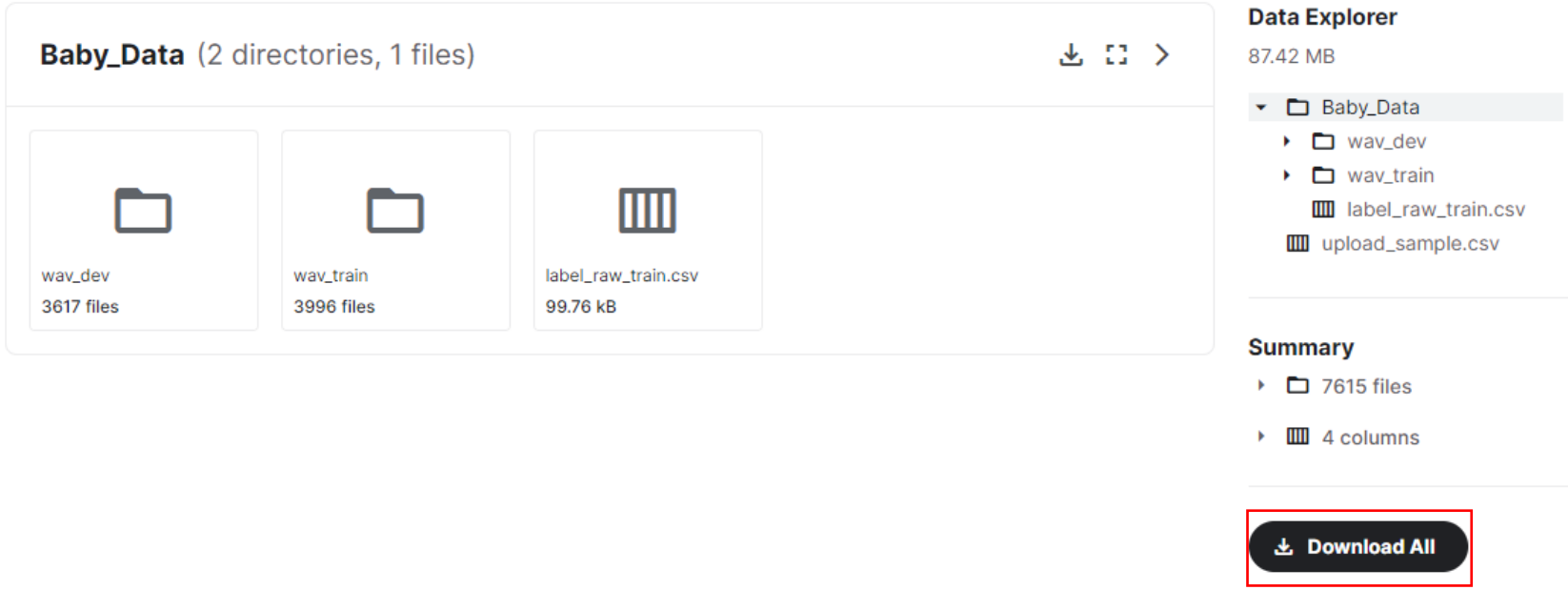
- Join this link and create your own account:
  - https://www.kaggle.com/t/5109c86420ca6360dfb1b2e891fc3bc7
  - Create your **Team** name: using your student ID





Kaggle (Download Datasets)

- Click ***Data***

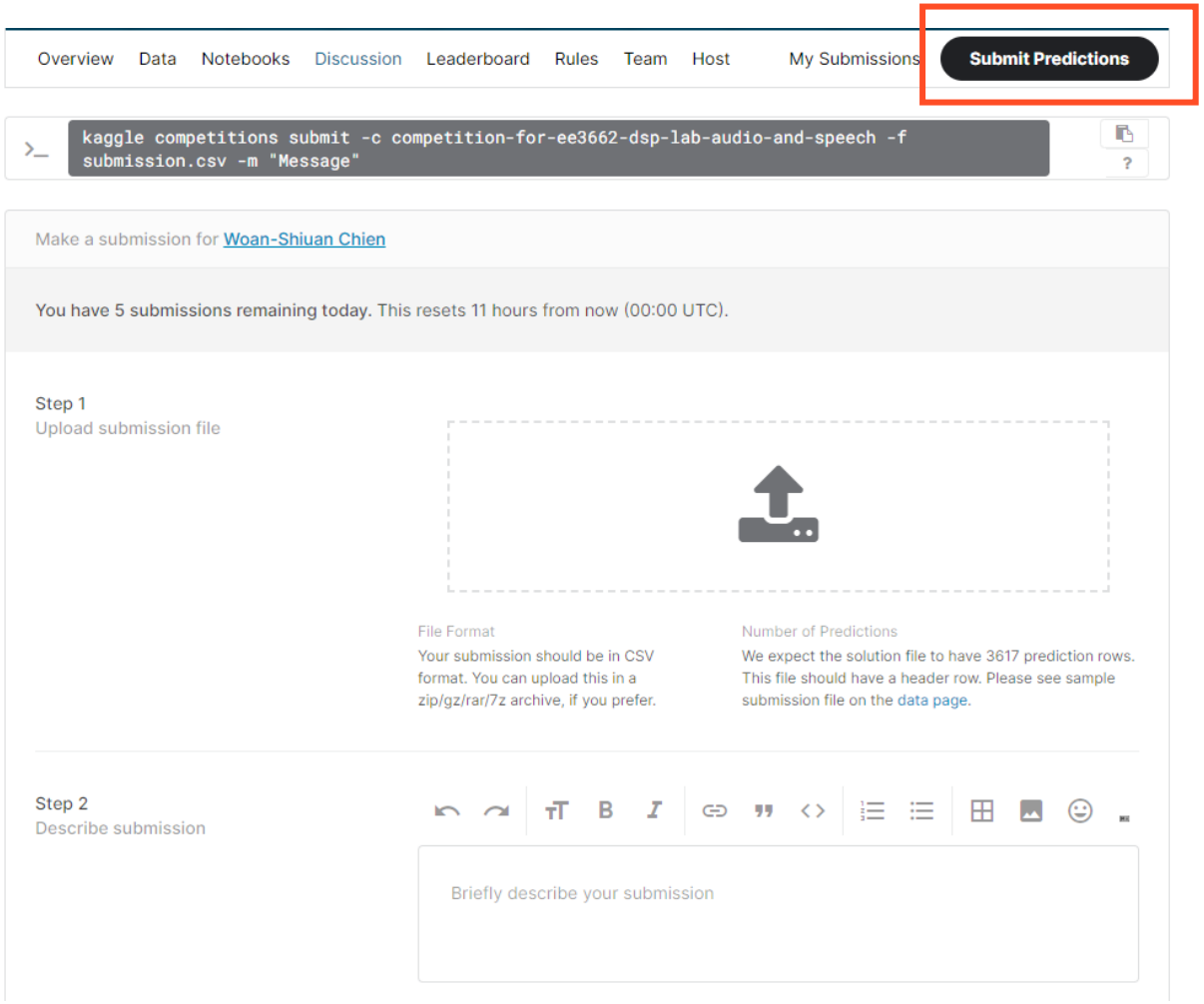


Kaggle (Upload results)

- Click ***submit predictions***
- Upload your results.csv
  - Results format:

	A	B
1	file_name	Prediction
2	devel_0001.wav	Junk
3	devel_0002.wav	Junk
4	devel_0003.wav	Junk
5	devel_0004.wav	Non-canonical
6	devel_0005.wav	Junk
7	devel_0006.wav	Non-canonical
8	devel_0007.wav	Junk
9	devel_0008.wav	Crying
10	devel_0009.wav	Junk

You can refer to upload_sample.csv



Lab12 Grading

Kaggle Challenge Results (40%) + Report (60%)

Kaggle Challenge Grading

- **Baseline:** Using MFCC and SVM → **63.23%**

#	Team	Members	Score	Entries	Last	Code	Join
	score_90.csv		0.69809				
1	**TA_test**		0.69809	2	1d		
	Your Best Entry! Your submission scored 0.61736, which is not an improvement of your previous score. Keep trying!						
	score_80.csv		0.64445				
	score_70.csv		0.61736				

Challenge Hints

- Kaggle has maximum daily submissions (5 times one day), so you can use SVM and Cross Validation in Lab3 to verify the training data
- Evaluation is based on Mean F1-Score
- Try different parameters (kernel, c, degree) to tune your model
- Try different features, statistics functions or classifiers to achieve higher performance

The evaluation metric for this competition is Mean F1-Score. The F1 score, commonly used in information retrieval, measures accuracy using the statistics precision p and recall r.

Precision is the ratio of true positives tp to all predicted positives tp + fp. Recall is the ratio of true positives tp to all actual positives tp + fn. The F1 score is given by:

$$[\mathrm{F1} = 2\frac{\mathrm{p}\cdot\mathrm{r}}{\mathrm{p}+\mathrm{r}} \quad \text{where} \quad \mathrm{p} = \frac{\mathrm{tp}}{\mathrm{tp}+\mathrm{fp}}, \quad \mathrm{r} = \frac{\mathrm{tp}}{\mathrm{tp}+\mathrm{fn}}]$$

The F1 metric weights recall and precision equally, and a good retrieval algorithm will maximize both precision and recall simultaneously. Thus, moderately good performance on both will be favored over extremely good performance on one and poor performance on the other.

p:陽性, n:陰性
t:真　, f:假

		真實狀況	
		事實為真	事實為假
預測狀況	預測為 陽性	tp	fp (Type I error)
	預測為 陰性	fn (Type II error)	tn

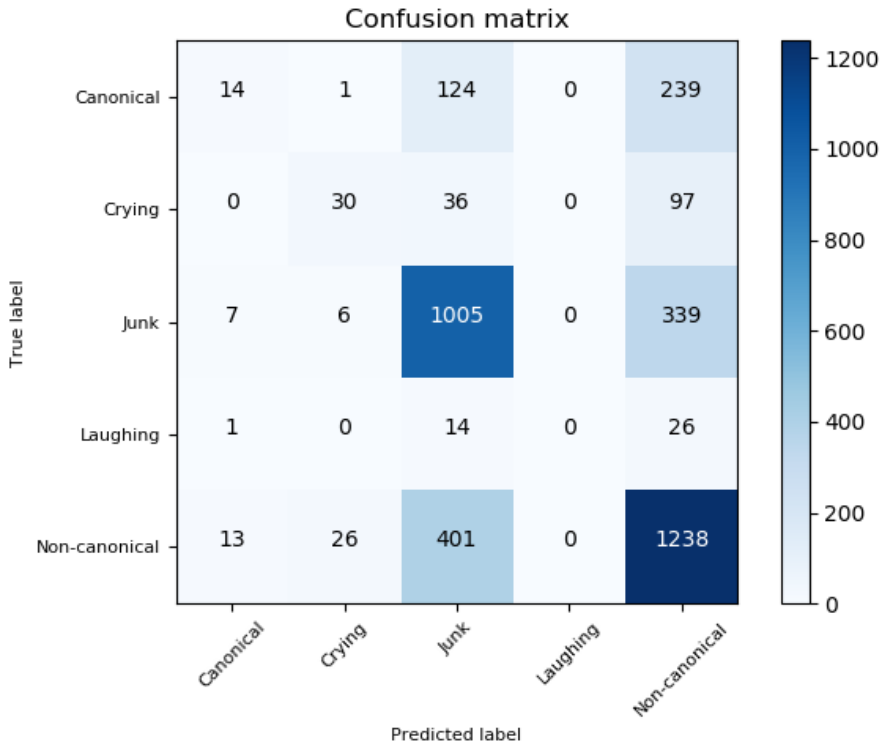
Challenge Rules

- Don’t create another account to get more submission chances!
- Don’t submit results from your classmates. We will run your code and compare the result with your score on leaderboard.

Report

- Try to explain your accuracies from your results table
- Using your best results to plot the confusion matrix (lab11)

Baby Sounds (%)						
	Canonical	Crying	Junk	Laughing	Non-canonical	UAR
SVM (rbf)	3.7	18.4	74.06	0	73.78	63.23
SVM (linear)						
SVM (poly)						
…						
…						



Report

- (Bonus) Analyses
  1. Try to explain why some categories perform worse? The data distribution seems quite imbalanced, try to solve the problems and compare the experimental results.
  2. What other problems did you meet? What is your solution? Any interesting findings?