# Dec 29, 2021

[Binh Duong Van](mailto:18520505@gm.uit.edu.vn)

* Kick-start the project
* Organize working folder: create document file
* Update full raw data
* Start to preprocess data

# Dec 30, 2021

[Binh Duong Van](mailto:18520505@gm.uit.edu.vn)

* Coswara raw data [done]

# Dec 31, 2021

[Binh Duong Van](mailto:18520505@gm.uit.edu.vn)

* COUGHVID raw data [done]
* Add country

# Jan 1, 2022

[Binh Duong Van](mailto:18520505@gm.uit.edu.vn)

* AICovidVN raw data [done]

# Jan 28, 2022

* Submitted pre-thesis document
* Define jobs will be done in the paper

# Feb 3, 2022

* Implement waveGAN model [WaveGAN: Learn to synthesize raw audio with generative adversarial networks](https://github.com/chrisdonahue/wavegan) : in process

# Feb 6, 2022

* Cough slicing completed: slice cough audio to separate coughs

# Feb 9, 2022

**Assign tasks:**

[An Le Tran Hoai](mailto:18520426@gm.uit.edu.vn)

* Label coswara chunk dataset
* Cough detection (ESC50):
  + Strategy to build models

[Binh Duong Van](mailto:18520505@gm.uit.edu.vn)

* Label Coughvid
* The baseline for covid recognition: Mobilenet for 3d data excluding metadata

# Feb 12, 2022

* Mel-spectrogram 2-d extraction for AICovidVN, Coswara, Coughvid

# Feb 13, 2022

* MobileNet architecture built successfully

# Feb 17, 2022

* Mel-spectrogram 1-d extraction for AICovidVN, Coswara, Coughvid
* MFCC 13, 26, 39 1-d extraction for AICovidVN, Coswara, Coughvid
* Build Light-GBM model with KFold
* Baseline cough detection:  
   MFCCs (n=39) - PCA (n =) - PLS (n=)  
   Mel 1d  
   SVM,RF,KNN: maximum score <85% accuracy

# Feb 22, 2022

* 2-d MFCC 13, 26,39 extraction for Coughvid
* Updated document file

# Feb 26, 2022

* 1d feature extraction for chunks completed

# Feb 27, 2022

* 2d feature extraction for full audio completed
* Baseline CNN model cough detection:

MFCC (n=39)  
 Mel 2d  
 Ensemble model (3 model)

Result: 20 epoch ~ 0.89% test auc

# Mar 1, 2022

* Building up a hybrid model

# Mar 7, 2022

* 1-d features extracted for full audio: AICovidVN, Coswara, Coughvid

# Mar 10, 2022

* Re-label and merge data for cough detection model (Coswara[:1000], Coughvid[:1000], ESC50).
* Cough detection datasets version 1:
  + Datasets version 1 have 3989 samples with:
    - Coughing: 2029 (from Coswara + Coughvid + ESC-50)
    - Non-Cough: 1960 (from ESC-50)
  + Result of CNN model with **3 indices shuffle, 20 epoch:**
  + **DATA[0]**
    - **Results: Loss = 0.3115939497947693 , AUC = 0.935832679271698**
    - **val: 0.9885155920139936**
    - **test 0.9790337173532782**
  + **DATA[1]**
    - **Results: Loss = 0.47223928570747375 , AUC = 0.9676210284233093**
    - **val: 0.9769978715170279**
    - **test 0.9761291975165584**
  + **DATA[2]**
    - **Results: Loss = 0.23795151710510254 , AUC = 0.9921346306800842**
    - **val: 0.9914255998452013**
    - **test 0.992182881959105**

# Apr 3, 2022

* Completed building mobilenet for coswara dataset.
  + Model performance in each training file, **Evaluation**

# Apr 20, 2022

* Fine tunning hyperparameter for Catboost, RF, LGBM, SVM