Feng Zongyao, 1st Year Ph. D Student
Human Interface Lab, The University of Aizu

Poster Session at Graduate School Information Fair

Cough Detection and Monitoring using DNN

Introduction

Cough is a vital respiratory defense mechanism, producing a characteristic sound, often associated with various respiratory illnesses. Traditional diagnosis relies on subjective patient reports like the Leicester Cough Questionnaire. Monitoring cough frequency aids in diagnosis, treatment, and respiratory research. Hence, automatic and precise cough detection has gained increasing research interest.

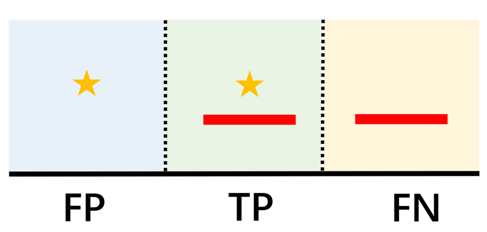
Cough detection is a subset of sound event detection, which identifies the sound source class and its temporal location within audio. Most studies focus on segment-based detection, which may not capture single event boundary. To comprehensively monitor coughs, precise single cough event identification is crucial, especially for continuous cough scenarios.

Cough detection also inherits sound event detection's challenges. Sound's additive nature could form unlimited combination of sound sources being active at the same time. Furthermore, additional challenges are posed by factors like the distance from sound source to recording device, device-related bias, and varies cough characteristics, such as differences in airways or lungs. To overcome these challenges and accomplish the task, the proposed system will leverage the power of DNNs to learn intricate cough patterns and representations from the audio data, enabling accurate cough detection and monitoring.

Data

All data were collected from 11 patients of various respiratory diseases by Fukushima Medical University. Each patient is asked to wear a portable audio recording device for 24, including their various daytime activities and sleeps. We have in total 6493 labeled coughs, each label contains the starting and ending time of a single cough event, the time is accurate to 10 milliseconds.

Evaluation



To mitigate the problem from unbalanced small amount of test data. We applied a three-fold leave-one-patient-out cross validation evaluation. When evaluating the performance of cough detection, we defined our own confusion matrix with regard to the predicted cough center.

Results

Cough detection result between HuBERT-RC and a XGBoost[2] based cough detection system.

	F1-Score (%)	Precision (%)	Recall (%)
XGBoost [2]	51.6	68.1	41.5
HuBERT-RC	**86.4**	**84.6**	**89.7**

Hourly Symmetric Mean Absolute Percentage Error ($sMAPE^{100}$)

Patient	Cough Counts	Hours	**HuBERT-RC**	LCM [1]	XGBoost [2]	Universal [3]
1	772	24	**1.21**	27.94	43.23	23.99
3	1016	24	**14.92**	19.62	44.10	22.99
4	475	6	**5.29**	33.62	57.63	30.51
5	448	24	**11.72**	16.99	43.61	43.46
6	599	24	**4.93**	47.29	49.26	50.25
7	234	6	**6.36**	41.05	29.35	33.93
8	749	1	**10.07**	26.95	42.67	39.61
9	220	15	**10.62**	25.76	43.74	48.31
10	215	24	**12.71**	68.86	36.98	50.98
Average Hourly $sMAPE^{100}$			**8.65**	34.23	43.4	38.23

When it comes to cough monitoring, the task is to predict the hourly cough count of each patient. We compared our model with other works using sMAPE to evaluate the performance.

References:
[1] Birring, S. S., et al. "The Leicester Cough Monitor: preliminary validation of an automated cough detection system in chronic cough." European respiratory journal 31.5 (2008): 1013-1018.
[2] Orlandic, L., Teijeiro, T. & Atienza, D. The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms. Sci Data 8, 156 (2021). https://doi.org/10.1038/s41597-021-00937-4
[3] Simou, Nikonas, Nikolaos Stefanakis, and Panagiotis Zervas. "A universal system for cough detection in domestic acoustic environments." 2020 28th European signal processing conference (EUSIPCO). IEEE, 2021.

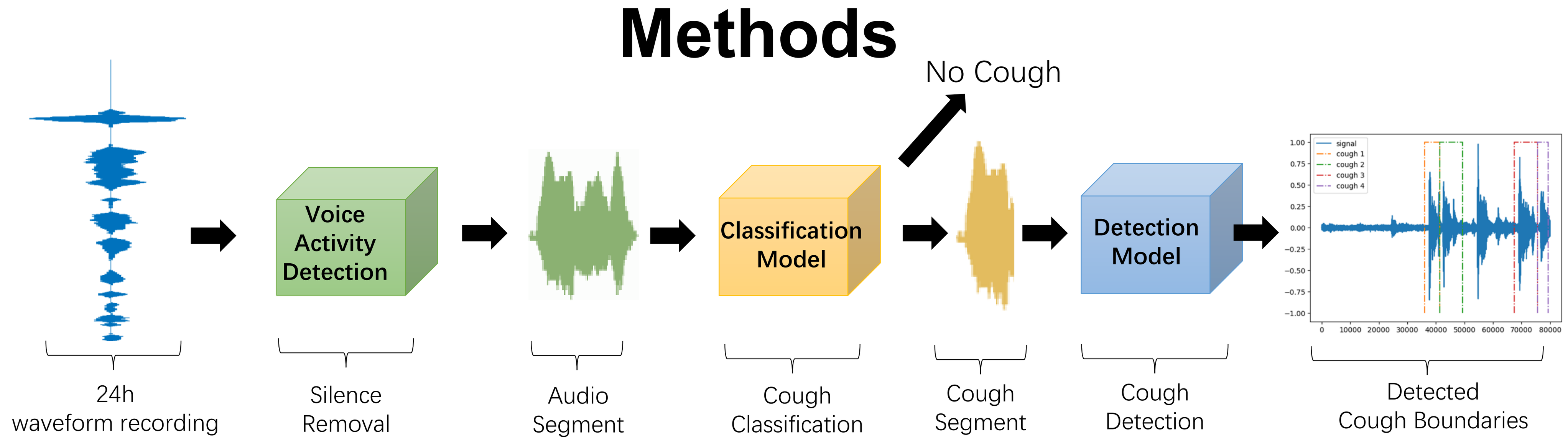
Conclusion

Overall, this research contributes to the field of machine learning and audio signal processing by presenting novel approaches for cough detection and monitoring.

The proposed models, HuBERT-C, HuBERT-F and HuBERT-RC, demonstrate improved performance and provide valuable insights for the development of more effective systems in healthcare, research, and clinical applications.

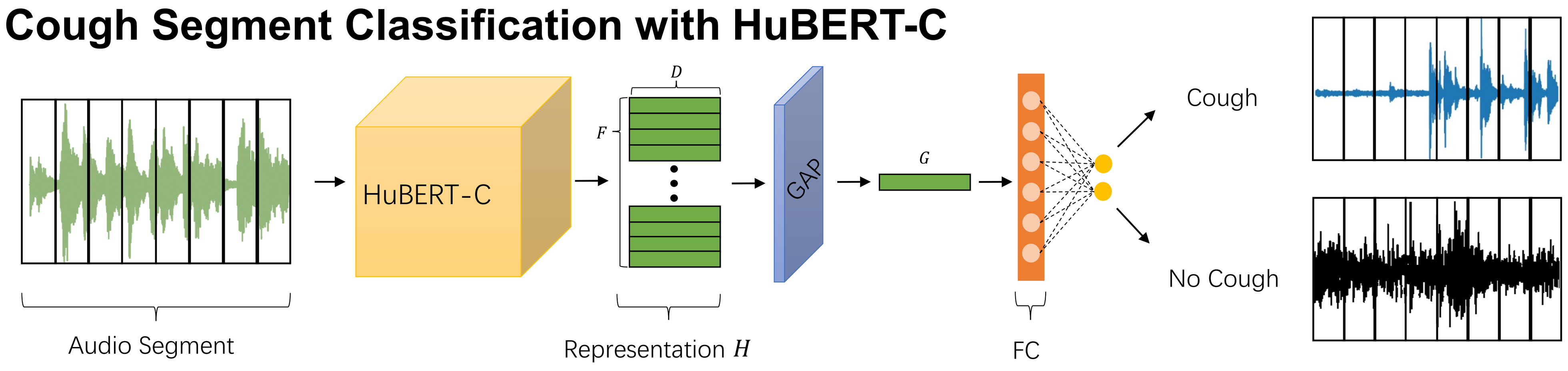
Future research can further explore the potential of deep learning techniques and extend the application of these models to broader sound event detection tasks.

Methods

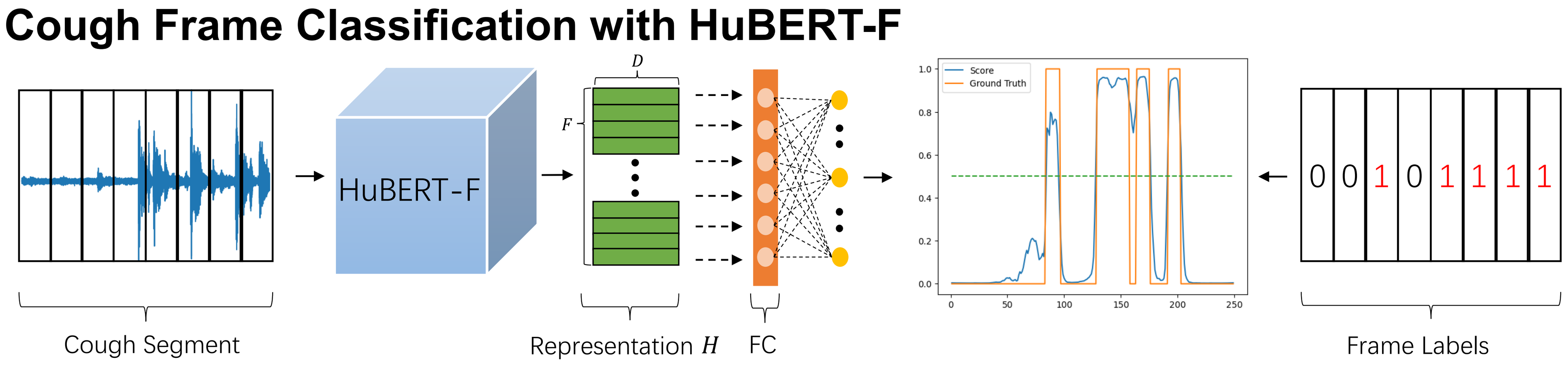


Our cough detection pipeline consists of three key steps. Firstly, we perform voice activity detection to eliminate audio silences and focus on relevant segments. Subsequently, we employ cough classification to identify the presence of cough within the remaining audio segments. Finally, we implement the event level cough detection, by dividing the cough segments into smaller frames and performing classification and regression tasks on these frames. The aggregation of the frame-level results then form the final detection outcome.

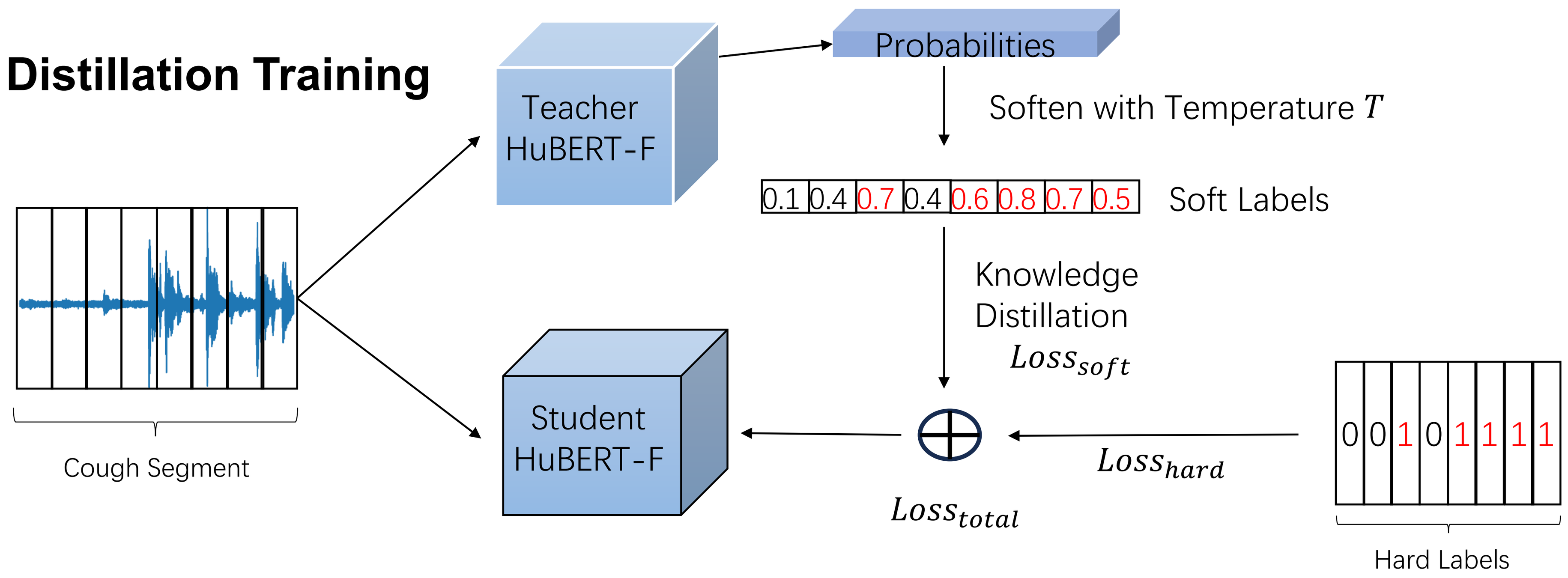
Cough Segment Classification with HuBERT-C



Cough Frame Classification with HuBERT-F

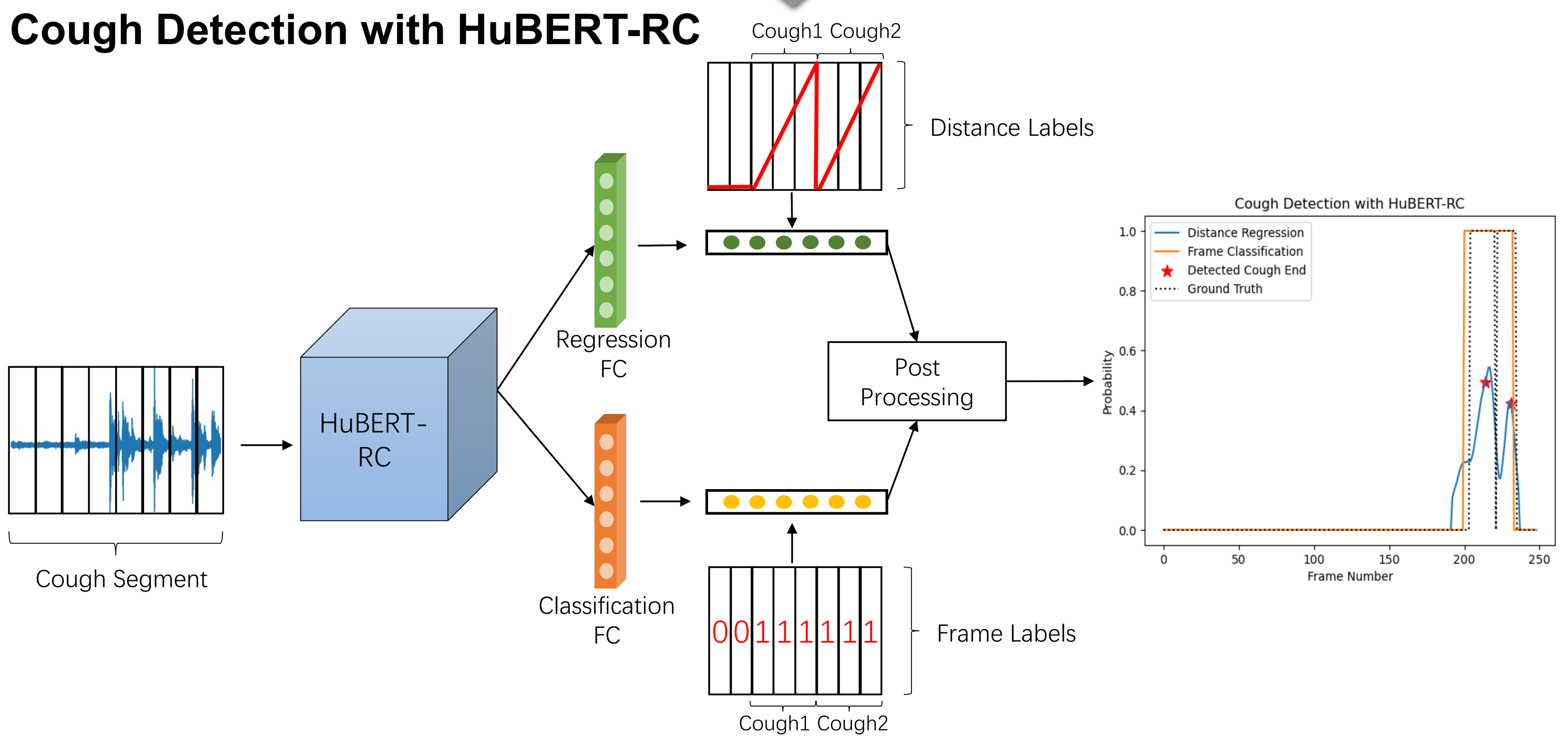


Distillation Training



The performance of HuBERT-F increases, if the learning process is guided by a probability-distribution-like labels i.e., the soft labels, instead from hard coded zeros and ones i.e., the hard labels. Therefore, the knowledge distillation training is applied here aiming to generate the soft labels.

Cough Detection with HuBERT-RC



The HuBERT-RC fuses both regression and classification branches, where the regression branch predicts the relative distance from current frame to its corresponding cough start frame. And the classification branch performs the cough frame classification. By combining the results from two branches, we will be able to locate every single cough event.