Robust Cough Detection With Out-of-Distribution Detection

Yuhan Chen, *Graduate Student Member, IEEE*, Pankaj Attri, Jeffrey Barahona, Michelle L. Hernandez, Delesha Carpenter, Alper Bozkurt, *Senior Member, IEEE*, and Edgar Lobaton, *Senior Member, IEEE*

***Abstract*—Cough is an important defense mechanism of the respiratory system and is also a symptom of lung diseases, such as asthma. Acoustic cough detection collected by portable recording devices is a convenient way totrack potential condition worsening for patients who have asthma. However, the data used in building current cough detection models are often clean, containing a limited set of sound categories, and thus perform poorly when they are exposed to a variety of real-world sounds which could be picked up by portable recording devices. The sounds that are not learned by the model are referred to as Out-of-Distribution (OOD) data. In this work, we propose two robust cough detection methods combined with an OOD detection module, that removes OOD data without sacrificing the cough detection performance of the original system. These methods include adding a learning confidence parameter and maximizing entropy loss. Our experiments show that 1) the OOD system can produce dependable In-Distribution (ID) and OOD results at a sampling rate above 750 Hz; 2) the OOD sample detection tends to perform better for larger audio window sizes; 3) the model's overall accuracy and precision get better as the proportion of OOD samples increase in the acoustic signals; 4) a higher percentage of OOD data is needed to realize performance gains at lower sampling rates. The incorporation of OOD detection techniques improves cough detection performance by a significant margin and provides a valuable solution to real-world acoustic cough detection problems.**

***Index Terms*—Cough detection, audio classification, out-of-distribution, bio-signal processing, machine learning.**

I. Introduction

DISEASES affecting the lung are among the most common medical conditions and are associated with a high mortality rate around the world [1]. Cough is a natural reflex of the human body to clear airways, but chronic coughing is often an indicator of lung diseases, such as asthma and bronchitis. Diagnosis of these diseases is often incorrect, and only 25–50% of these patients are known to their doctors [2]. Long-term monitoring of the frequency and type of cough can help patients track their chronic respiratory conditions and aid in the correct diagnosis. However, the devices used for clinical diagnosis are expensive. To reduce the financial burden on healthcare systems, in-home wearable devices embedded with machine learning models are designed to record and analyze biosignals including cough sounds [3], [4]. Machine learning models are trained to capture salient information from biosignals to use for medical purposes. However, the reliability of these systems is highly dependent on data quality which is not stable for data collected by wearable systems and needs to be taken into consideration in the system design.

One challenge is the lower performance expected for a machine learning model if it is applied to data that does not follow the distribution of the target classes used for training. This challenge happens when switching from a control setting for data acquisition to less structured settings in real-world deployments. Samples that do not follow the training distribution are often not learned appropriately by the models, and they are referred to as "Out-of-Distribution" (OOD) data. Data with the same distribution as the training set are designated "In-Distribution" (ID) data. Deep neural network classifiers can give high-confidence predictions to OOD inputs and lead to suboptimal results [5], [6], [7].

Most of the research related to cough detection focuses on specific sound classification problems and assumes the data are clean. However, low data quality caused by ambient environmental noise can complicate the ability of patients to monitor coughing at home using sensors, limiting the usefulness of home-based cough detection for patient management. Considering this limitation caused by sound diversity in the environment, we introduce an OOD detection method to tackle this issue. With the development of artificial intelligence, many

Manuscript received 20 December 2022; revised 21 February 2023; accepted 24 March 2023. Date of publication 5 April 2023; date of current version 3 July 2023. This work was supported by the National Science Foundation (NSF) under Awards IIS-1915599, IIS-1915169, IIS-2037328, and EEC-1160483 (ERC for ASSIST). *(Corresponding author: Edgar Lobaton.)*

Yuhan Chen, Jeffrey Barahona, Alper Bozkurt, and Edgar Lobaton are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27695 USA (e-mail: ychen239@ncsu.edu; jabaraho@ncsu.edu; aybozkur@ncsu.edu; edgar.lobaton@ncsu.edu).

Pankaj Attri is with the Department of Computer Science, North Carolina State University, Raleigh, NC 27695 USA (e-mail: pattri@ncsu.edu).

Michelle L. Hernandez is with the Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27559 USA, and also with the Center for Environmental Medicine, Asthma and Lung Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27559 USA (e-mail: michelle_hernandez@med.unc.edu).

Delesha Carpenter is with the Division of Pharmaceutical Outcomes and Policy, Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, NC 27559 USA (e-mail: dmcarpenter@unc.edu).

Digital Object Identifier 10.1109/JBHI.2023.3264783

2168-2194 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

cough detection algorithms are designed based on deep learning algorithms. However, these algorithms tend to under-perform when the distributions of training and testing data do not agree. For cough detection, OOD inputs can include body sounds (more prominent on wearable monitors) or environmental sounds that are not included in the training process such as breathing, heartbeats, dog barking, door opening, etc. False predictions in these classifiers can cause issues, such as overestimating how often a person coughs. For example, a model may falsely classify some sounds (e.g., the sound of a lawn mower) as a cough instead of flagging them as unknown sounds. Inaccurate data classification may lead to inaccurate designation of disease severity and the resulting treatment plan. Most existing neural network-based models address the OOD problem for computer vision tasks, with limited studies implementing OOD detection methods for audio classification tasks [8]. To the best of our knowledge, this study is the first of its kind to focus on using OOD detection techniques to boost cough detection performance.

In this study, we introduced OOD technique to reduce the effect caused by the environmental noise collected by wearable devices. Experiments are designed to analyze how much OOD can help in different settings. The main contributions of this paper include:

- Using publicly available datasets including the Musan [9], Coughvid [10], and FluSense [11] datasets to dynamically evaluate OOD detection performance at various frequencies of interest while balancing ID and OOD samples.
- Integrating OOD methodologies into the cough detection task to solve low data quality issues. We treated cough and speech as ID data because they compose the majority of the data in available sources, and the rest of the sound categories as OOD data. We implemented two OOD detection algorithms in this work; both give promising results in the cough detection problem.
- Evaluating the dependency of the OOD cough detection model to several parameters such as sampling rates and window size for feature extraction. We showed that the cough detection with OOD detection can produce reliable results at above 750 Hz sampling rate at 1.5–10 seconds window sizes. The lowest sampling rate to produce promising results for ID classification and OOD detection is 750 Hz. Lower sampling rate are important because they can save computational resources and protect user privacy.
- Demonstrating that models with OOD sample detection techniques improve overall cough classification accuracy as the percent of OOD samples increase. The gains become more prominent at higher sample rates and for higher proportions of OOD samples in the input audio signals.

II. Related Work

The ability to detect cough accurately has been well studied by researchers for a long time using statistical cough detection methods [12], [13], [14] but, in recent years, researchers have started to implement machine learning and deep learning methods to analyze the most suitable signal features used in these methods [15], [16]. These works investigated features including Short-time Fourier transform (STFT), mel-frequency cepstral coefficients (MFCC) and mel-scale filter banks (MFB) and classifiers including logistic regression, feed-forward artificial neural network, support vector machine, and random forest. Besides feature extractor and classifier, Monge-Alvarez et al. [17] proposed a system enhancing the performance of cough detection by adding high-level data representation steps, and Lee et al. [18] improved a cough detection system by adding data augmentation process. There are also some other works dealing with hardware device issues, the cross-device discrepancy, by using ensemble classifiers [19], [20]. In our prior work, Mahmoud et al. [21] demonstrated the effectiveness of a logistic regression classifier with mel-spectrogram input applied to cough and speech audio using various sampling frequencies and window sizes. Our work builds on that approach, wherein we use mel-spectrograms of audio signals as inputs to machine learning models based on Resnet-50 architecture with custom OOD detection modules to detect OOD samples in addition to cough classification.

Several contributions focus on detecting OOD data in the field of computer vision. [22], [23], [24], [25], [26], [27], [28], [29], [30], [31]. Some methods do not change the underlying pre-trained model architecture and use the maximum value of the softmax function to separate OOD inputs [22], [23], while others add an additional output indicating the confidence of the results to identify OOD inputs [24], [25], [26]. Generative models, like Variational Autoencoder (VAE), can also be used to do OOD detection by analyzing in latent space [27]. We implemented two approaches in our study. In the first approach we add a learning confidence output [24] and in the second we replace the SoftMax loss with IsoMaxPlus loss [32] to adhere with the maximum entropy principle. Based on IsoMaxPlus loss, each class has a prototype and the minimum distance of a sample to the prototypes is used to detect OOD inputs.

III. Proposed Methods

Based on the system in the prior work [21], we propose a new pipeline that can address both cough detection and out-of-distribution detection problems. The work flow of the proposed pipeline is shown in Fig. 1. In our proposed pipeline, we use Mel-spectrogram [33] as audio inputs to a Convolutional Neural Network (CNN) based model to extract features for cough and Out-Of-Distribution (OOD) detection tasks. We experimented with different CNNs models as backbones along with two classifiers that have OOD detection capabilities. For CNNs, we test the Frequency Extraction Network (FENet) in [34], Residual Network 18 (ResNet18) [35], and VGG16 [36]. We test our pipeline on two datasets. We used a dataset from our earlier pipeline [21] to find the best CNN model and used newly generated datasets to find out the best model (for cough classification and OOD detection). We also investigated how different sampling rates and window sizes of the audio signals affect the results. Resources and code for our research can be found in the GitHub repository: https://github.com/ARoS-NCSU/OOD-CoughDet.

A. Preliminary Work

In the prior work, [21], Mahmoud et al. developed a framework for detecting cough among cough and speech audios. In

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

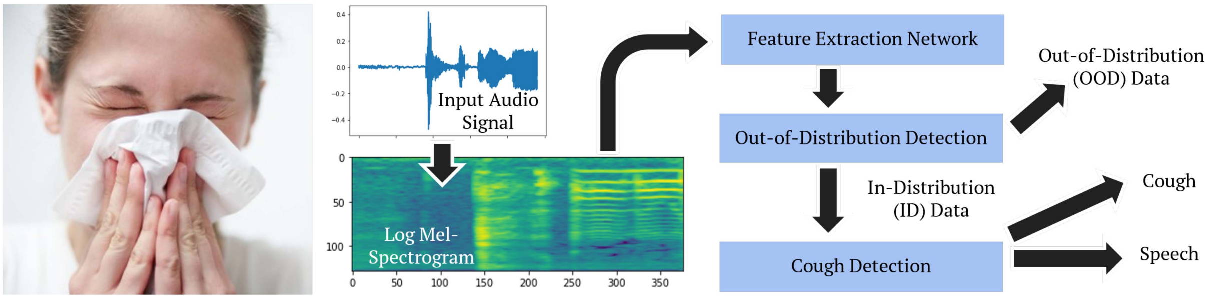
0.4
0.2
0.0
−0.2
−0.4
Input Audio Signal
0
50
100
Log Mel-Spectrogram
0
50
100
150
200
250
300
350
Feature Extraction Network
Out-of-Distribution Detection
Out-of-Distribution (OOD) Data
In-Distribution (ID) Data
Cough Detection
Cough
Speech

Fig. 1. Work flow of the robust cough detection with out-of-distribution detection. We first converted audio signals into log mel-spectrograms followed by a feature extraction network. Then, we used OOD detection to recognize ID and OOD data. The ID data is classified into cough and speech. Note that we used the same network to do OOD detection and cough detection, so no additional computation is involved. This pipeline produces robust cough detection by recognizing and removing unknown classes.

TABLE I
DATASET SUMMARY (THE NUMBER OF SEGMENTS OR THE RECORDING TIME ARE SHOWN IN BRACKETS)

		Dataset A	Dataset B
ID	Cough	ESC-50 [37](40) FSDKaggle2018 [38](273)	Coughvid [10]($\sim$37,000) FluSense [11](2486)
	Speech	LibriSpeech [39]($\sim$1 hr)	Musan (Speech) [9]($\sim$60 hrs) LibriSpeech [39]($\sim$1 hr)
OOD			Musan (Non-Speech) [9]($\sim$49 hrs)

the study, we found that 1.5 seconds window size gives the best performance on cough detection tasks and the model gives 92.5% accuracy even at very low sampling rate, 750 Hz. The work, however, was limited to detection in cough mixed with speech audio. In the real world, there will be other types of sounds, i.e., OOD data, in the input audio. Our work address the issue of identifying OOD data in input audio without affecting cough classification abilities of the model.

B. Datasets

Two datasets generated in different ways are used in this study. Dataset A was used in [21] and Dataset B is our new dataset. In Table I, the source data used for generating these two datasets are included. More details can be found in the following subsections.

1) Dataset a: Dataset A was collected from the ESC-50 [37], FSDKaggle2018 [38] and LibriSpeech [39]. All audio files are converted and saved in WAV format. The length of audio files varies from 5 to 30 seconds. We further extracted meaningful 1.5 seconds windows from these files, because 1.5 seconds is the majority of the cough lengths, containing the main part of each audio [21]. These 1.5-second clips were manually annotated for cough signals by student assistants. Dataset A has 8,046 data points labeled as "cough" augmented from both the ESC-50 and FSDKaggle2018 datasets, and 11,372 data points labeled as "speech" augmented from the LibriSpeech dataset [21]. The data was generated based on the manual annotations of the start time and the end time of each cough in ID inputs in Table I.We used a similar approach as [21] to extract data samples from these intervals. The dataset was split into training (80%), validation(10%), and testing (10%) subsets. Due to the manually annotated high-quality data samples, this dataset was used to select the best feature extraction algorithms.

2) Dataset b: Dataset B was generated from the Coughvid dataset [10], the FluSense dataset [11], the Musan dataset [9], and the LibriSpeech dataset [39]. We generated over 30,000 samples of 5 seconds cough segments from Coughvid and FluSense, and speech data from the Musan dataset and LibreSpeech as in-distribution dataset. Speech audios in the Musan dataset are used as in-distribution data, while music and noise are used to generate out-of-distribution data. Multiple data points can be generated automatically from each file using sliding windows. The training, validation, and testing sets are generated as follows. Training and validation data are generated from ID inputs datasets in Table I by using a sliding window with a specific overlap size to control the number of samples. The training data and the validation data have a proportion of approximately 4:1 containing only ID inputs. This simulates the real-world situation where only ID data is accessible for building a model. The testing data contains both ID and OOD data generated from ID inputs and OOD input in Table.I using the same way but with different overlap window sizes. The proportion of ID and OOD test data is approximately 1:2 when we tested the performance for different sampling rates. The final cough detection performance could be affected by this ID and OOD test data proportion which is tested in this work. We used dataset B to test how different sampling rates and window sizes affect the model's final performance.

C. Feature Extraction Models

FENet [34], ResNet [35], and VGG [36] are three outstanding neural network frames used to extract features from images and signals. To design a better system, we used the original system embedded with three models separately and compared the performance. FENet is a CNN-based model, which can extract different frequency features and is often used in audio data. ResNet is also a CNN-based model adding residual blocks to improve the training efficiency and avoid the gradient vanishing problem. This architecture is commonly used in image models with high performance. VGG Net also contains subsequent convolutional layers with a pyramidal shape which achieves promising results on computer vision tasks.

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

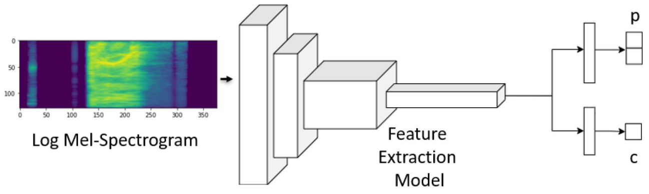

Fig. 2. Adding Learning Confidence [24] into cough detection algorithm. p represents the prediction probabilities of cough and speech, and c represents a confidence estimate.

D. Out-of-Distribution Detection

Two out-of-distribution architectures were added to our cough detection pipeline and we investigate how each performs on data with different window sizes and different sampling rates. The first system was adapted from the learning confidence out-of-distribution detection model [24] which estimates learning confidence for neural networks and produces intuitively interpretable outputs. The other approach incorporates an entropic out-of-distribution detection model [40] which replaces the SoftMax loss with a novel loss function dealing with the weaknesses of SoftMax loss anisotropy and tends to produce low entropy probability distributions which break the principle of maximum entropy. In OOD detection problems, only ID data is taken as training and validation dataset and combined ID and OOD data is taken as testing dataset.

1) Confidence-Based Approach: This model provides a confidence value for each output predicted which specifies if this data is in-distribution. The confidence value ranges from 0 to 1. If the confidence value is close to 1, the data is more likely to be an in-distribution sample, and vice versa. Learning confidence is estimated by adding a confidence estimation branch along with the original class prediction branch after the second to last layer of the original network. In our case, we keep the same cough detection algorithm architecture but replace the last layer with two separated layers controlling the prediction task and confidence estimation task respectively. Fig. 2 shows the pipeline applied to our cough detection task. The log mel-spectrogram of each audio clip is the input and the feature extraction model can be any of the neural network architectures previously described. After feature extraction, the output features are passed through the prediction branch and confidence branch. For prediction logits, we apply a linear layer to map the output to a two-length vector, and then utilize a SoftMax function prediction to get cough and speech prediction probabilities p. Similar to the prediction branch, a linear layer, and a Sigmoid function is utilized in the confidence branch with one learning confidence output c. This process can be represented as:

$$p, c = f_\theta(x), \quad p_i, c \in [0, 1], \sum_{i=1}^{M=2} p_i = 1. \tag{1}$$

where x is the log-mel spectrogram input and θ represents the parameters for the neural network $f_\theta(\cdot)$.

In this model, the target probability distribution y helps to adjust the predict probability distribution from p to p' under the confidence value c:

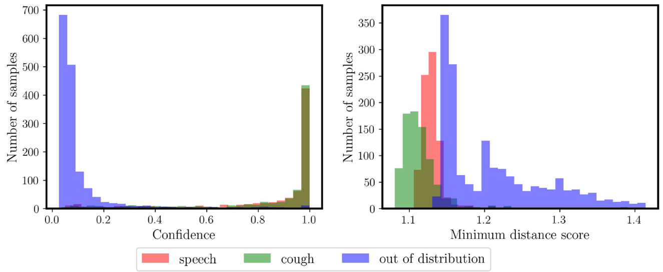

Fig. 3. Density Plots for Confidence-based OOD Detection and Entropy-based OOD Detection generated from the best performance settings with sampling frequency $f = 4$ kHz and $f = 16$ kHz respectively for each model and window size $\tau = 5$ s (Sections. IV-C and IV-D). The left plot is the confidence density and the right is the minimum distance score density.

$$p' = c \cdot p + (1 - c)y. \tag{2}$$

If c is close to 1, it means that the adjusted prediction is closer to the prediction output. This indicates that the original prediction results of this data are more convincing and the data is more likely to be an in-distribution data point. When c is close to 0, then the model tends to output the ground truth distribution which means that the original prediction of this data is suspicious, in other words, data could be out-of-distribution data.

Since the prediction probability is modified, the task loss is calculated using the adjusted distribution p' now. We use the negative log-likelihood as the classification loss and the modified task loss can be represented as:

$$\mathcal{L}_t = -\sum_{i=1}^{M=2} \log\left(p'_i\right) \cdot y_i. \tag{3}$$

If we were to minimize only this loss, the model will tend to set c to be zero. To solve this issue, we add a confidence loss, which is a log penalty:

$$\mathcal{L}_c = -\log(c). \tag{4}$$

Therefore, the total loss is the sum of the task loss and the confidence loss with a hyper-parameter λ to balance two losses:

$$\mathcal{L} = \mathcal{L}_t + \lambda \mathcal{L}_c. \tag{5}$$

The hyper-parameter λ is adjusted by a budget parameter β which is set to 0.3 according to [24]. During the training process, λ is dynamically adjusted following the rule: if $L_c > \beta$, then increase λ to $\lambda/0.99$, and if $L_c < \beta$ then decrease λ to $\lambda/1.01$.

In order to recognize out-of-distribution data, we still need to find a threshold value, which can be defined by finding the best threshold producing the minimum detection error (the ratio of misclassified samples) in the holdout/validation set based on the conclusion in [24]. The left plot in Fig. 3 exhibits the density of confidence in our data. Most of the ID sample confidences are close to 1 and on the contrary, most of OOD sample confidences are close to 0.

2) Entropy-Based Loss: The Entropic out-of-distribution detection proposed by David et al. in [40] solves the SoftMax

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

loss drawbacks by replacing it with the Isotropy Maximization (IsoMax) loss. The proposed IsoMax loss is isotropic and follows the maximum entropy principle.

The IsoMax loss is designed to be a drop-in replacement of the SoftMax loss. Therefore, the IsoMax loss $\mathcal{L}_I$ has the same structure of normalization, and cross-entropy (negative logarithm) as the SoftMax loss:

$$\mathcal{L}_I = -\sum_i \log\left(\frac{\exp\left(-d\left(\boldsymbol{f}_\theta(\boldsymbol{x}), \boldsymbol{z}_\phi^i\right)\right)}{\sum_j \exp\left(-d\left(\boldsymbol{f}_\theta(\boldsymbol{x}), \boldsymbol{z}_\phi^j\right)\right)}\right) \cdot y_i, \tag{6}$$

where $\boldsymbol{f}_\theta(\boldsymbol{x})$ denotes the embedded high level features and $\boldsymbol{z}_\phi^j$ denotes a learnable prototype of class j. These prototypes are learned during training process by minimizing loss function. The function $d(.,.)$ equals the non-squared Euclidean distance between sample features and class prototypes, and y stands for the correct class label. In the SoftMax loss, the part of $-d(\boldsymbol{f}_\theta(\boldsymbol{x}), \boldsymbol{z}_\phi^i)$ in (6) is replaced by the logits related to the i-th class. Therefore, IsoMax can be implemented easily as a replacement for SoftMax, and the inference probabilities can be written as

$$p_i^I(x) = \frac{\exp\left(-D_i(x)\right)}{\sum_j \exp\left(-D_j(x)\right)}, \tag{7}$$

where $D_i(x) := d(f_\theta(x), z_\phi^i)$.

In the training process, an entropic scale E_s is added to calibrate the IsoMax loss. The extended equation is:

$$\begin{aligned} d\left(\boldsymbol{f}_\theta(\boldsymbol{x}), \boldsymbol{z}_\phi^i\right) &= E_s \cdot D_i & (8)\\ &= E_s \left\|\boldsymbol{f}_\theta(\boldsymbol{x}) - \boldsymbol{z}_\phi^i\right\| & (9)\\ &= E_s \sqrt{\left(f_\theta(x) - z_\phi^i\right) \cdot \left(f_\theta(x) - z_\phi^i\right)}. & (10) \end{aligned}$$

Then the loss in training process can be represented as

$$\mathcal{L}_{\text{IM}} = -\sum_i \log\left(\frac{\exp\left(-E_s \cdot D_i\right)}{\sum_j \exp\left(-E_s \cdot D_j\right)}\right) \cdot y_i. \tag{11}$$

In [32], David Macêdo et. al. proposed an enhanced isotropy maximization loss (IsoMax+) OOD detection method considering normalized z_ϕ^j. Thus, they replaced $f_\theta(x)$ with its normalized version given by $\widehat{f}_\theta(x) = f_\theta(x)/\|f_\theta(x)\|$ and replaced z_θ^j with its normalized version given by $\widehat{\boldsymbol{z}}_\phi^j = \boldsymbol{z}_\phi^j/\|\boldsymbol{z}_\phi^j\|$. Therefore, the loss function and probabilities equation can be rewritten as

$$\mathcal{L}_{\text{IM+}} = -\sum_i \log\left(\frac{\exp\left(-E_s\beta\left\|\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \widehat{\boldsymbol{z}}_\phi^k\right\|\right)}{\sum_j \exp\left(-E_s\beta\left\|\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \widehat{\boldsymbol{z}}_\phi^j\right\|\right)}\right) \cdot y_i \tag{12}$$

with inference probability

$$p_i^{IM+}(x) = \frac{\exp\left(-\beta\left\|\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \widehat{\boldsymbol{z}}_{\boldsymbol{\phi}}^i\right\|\right)}{\sum_j \exp\left(-\beta\left\|\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \widehat{\boldsymbol{z}}_{\boldsymbol{\phi}}^j\right\|\right)}, \tag{13}$$

where β is the distance scale, which is a scalar learnable parameter which is optimized in the training process. The distance scale is used to avoid the unreasonable restriction introduced by the normalized version. All prototypes are initialized using a normal distribution with a mean of zero and a standard deviation of one and the distance scale is initialized to one.

To define the out-of-distribution data, the minimum distance score is used as OOD score. The minimum distance score is given by

$$\text{MDS} = \min_j \left(\left\|\widehat{\boldsymbol{f}}_{\boldsymbol{\theta}}(\boldsymbol{x}) - \widehat{\boldsymbol{z}}_\phi^j\right\|\right). \tag{14}$$

This represents the minimum distance between the prediction and one of the class prototypes which is also the predicted class. In Fig. 3, the plot on the right shows the density of minimum distance scores for the model trained on 16 kHz sampling rate with 5 seconds window sizes. From the plot, only the tail of ID sample minimum distance score overlaps with small part of the OOD sample minimum distance score distribution. Thus, a distance threshold could be used to distinguish OOD data.

E. Evaluation Metrics

To evaluate an OOD system, the results for not only OOD detection task but also ID cough classification task need to be measured. Typically speaking, a good OOD system can identify OOD data without dropping the ID detection performance in the meanwhile. Since OOD detection task can also be treated as a binary classification of ID class versus OOD class, classical classification metrics are both used in ID task and OOD task measurements. Other mainly used OOD metrics were also implemented to further illustrate the performance.

1) In-Distribution Metrics: Let TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives, respectively. True or false means the detection result and positive or negative means the class. For ID task, the positive class is cough and the negative class is speech. We take accuracy, precision, recall, F1-score, and AUROC as evaluation metrics.

2) Out-of-Distribution Metrics: OOD detection can be taken as a binary classification problem, therefore, all metrics for ID detection can be used in OOD detection by taking ID and OOD as two classes. In this case, we treated the ID class as the positive class and the OOD class as the negative class. Additionally, two more commonly used metrics proposed by Hendrycks et al. [22] were used in the evaluation.

FPR at 95% TPR: We denote FPR α as the value of the false positive rate at $\alpha\%$ true positive rate [30]. This metric is designed to test the detection performance at one strict threshold, which is different from the aim of AUROC. We can compare strong detectors clearly by evaluating performance at a certain strict threshold. FPR α represents the probability of predicting an ID example as an OOD example when the OOD examples detection achieves $\alpha\%$ positive detection. Note that the results may vary. It is possible that the rank of the detectors may be the same for all α but at different levels, or the rank can change. Therefore, the α in FPR α is problem-dependent and needs to be selected carefully. We use $\alpha = 95\%$.

Detection Error: Detection Error, just as its name implies, means the probability of error detection in an algorithm. DeVries

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

el al. [24] defined it as

$$\min_{\delta}\{0.5P_{in}(f(x) \leq \delta) + 0.5P_{out}(f(x) > \delta)\}, \tag{15}$$

where $\delta \in [0, 1]$ are all possible thresholds. $f(x)$ is the score assigned to the input sample x, which can be used to separate OOD and ID examples. P_{in} and P_{out} are the classification probability of ID and OOD examples respectively, and they are equally weighted as 0.5.

For the OOD detection task, we use the following rules to set any low confidence predictions as a negative detection (i.e., we prioritize OOD task AUROC value but other rules can be selected): (1) For the Entropy-based model, OOD samples are separated by a distance threshold which produces the highest OOD task AUROC on the test dataset [40]; (2) For the Confidence-based model, we selected a confidence threshold giving the highest OOD task AUROC on the test dataset. The results for detection error, F1, precision, and recall come from detection with these threshold selection rules.

3) Overall Performance: The overall performance evaluation in experiment 4 treats cough as positive class and all other instances including speech and OOD data as negative class. Based on this setting, we calculated accuracy, F1 score, precision, and recall values.

IV. Experiments

To figure out the best cough detection pipeline and investigate if OOD detection can help to improve the system, the experiments in the following sections are designed. The first experiment was designed to test different feature extraction models to find out the best features for the system. Other experiments were designed to investigate the effect of different window sizes and different sampling rates. To further prove the validity of the system, models with and without OOD detection were compared on different datasets consisting of different ratios of OOD data.

A. Feature Extractor Algorithm Comparison

To find out the most appropriate feature extractor structure, three backbone networks (FENet, ResNet18 and VGG16) are placed into the standard cough detection pipeline without OOD structure, and the ID task performances for each system are compared. We use data at 16 kHz sampling rate and 1.5 s window size with no overlap from Dataset A. The same optimization hyperparameters were used for a fair comparison, e.g., random seed and learning rate of 0.0001. Accuracy, Precision, Recall, and F1 score are taken as evaluation metrics for comparing three feature extraction models. We found that ResNet18 achieves the perfect performance (100% accuracy) implying that all clean samples selected manually are classified correctly meaning that ResNet has a strong ability to extract features from Mel-spectrogram images. We also notice that VGG16 gets a little higher performance with 97.8% accuracy than FENet with 96.6% accuracy meaning that image models are better than signal models when the input is signal Mel-spectrogram. Compared to VGG16, ResNet18 has a lighter structure size and better results. Based on these facts, ResNet is chosen to be a system feature extraction method in the following experiments. However, either one of these models would be an appropriate choice.

TABLE II
Number of Samples Generated While Varying the Window Size (τ)

	τ(s)	1.5	2	3	4	5	10
training data	cough	9522	8204	9918	9384	7074	6282
	speech	9522	8204	9918	9384	7076	6282
Validation data	cough	1602	1387	1669	1576	1124	1095
	speech	1602	1387	1669	1576	1125	1095
Test data	cough	1086	931	1057	961	802	748
	speech	1086	931	1057	961	802	748
	OOD	2172	1862	2114	1922	1604	1496

B. ID Detection Performance for Different Sampling Rates

In this experiment, ResNet50 was selected as backbone for both baseline model and OOD incorporated models since it had the best performance for feature extraction (as shown in the previous section). For all models, we set the learning rate to $1e^{-4}$. For the confidence-based model, the initial confidence parameter λ is 0.1 and a budget parameter is set to 0.3 [24]. For the entropy-based model, E_s is set to 10. To ensure the accuracy of our models, we trained each model with a batch size of 16 for a total of 5 epochs. The best model for each setting was selected based on its performance on the validation set. These settings are the same for the following experiments from Sections IV-C to IV-E.

We use Dataset B for this analysis. Cough data are generated using 5 s window sizes with 2.5 s overlap This setting is consistent with the best window size in Section IV-D. Speech data and OOD data are generated using 5 s window sizes without overlap to keep the number of cough data close to the number of speech data see Table II). In total there were 14150 training data samples, containing 7074 coughs and 7076 speech, 2449 validation data samples containing 1224 coughs and 1225 speech samples. In the test set, there were 1604 samples including 802 coughs, 802 speech, and 1604 OOD data samples.

By lowering the sampling rate, less information would need to be transmitted / processed, which would lower the power consumption in the case of a wearable device but it would make recognition more challenging. This is a trade-off. In our previous work [21], we showed that 750 Hz is the lowest sampling rate for which we observe a significant drop of Virtual Speech Quality Objective Listener (ViSQOL), which compares the original signal and target signal to provide a speech quality score using the Mean Opinion Score (MOS) [41], on the "test-clean" set in the LibriSpeech (with speech labels) [39]. Similarly, we further analyze signal quality on the same dataset by using World Error Rate (WER) [42], which is a common metric used in speech recognition representing the difference between two text sequences. WER can be represented as

$$WER = \frac{S + D + I}{N} = \frac{S + D + I}{S + D + C}, \tag{16}$$

where S represents the number of substitutions; D represents the number of deletions; I indicates the number of insertions; C stands for the number of correctly placed words; and N

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

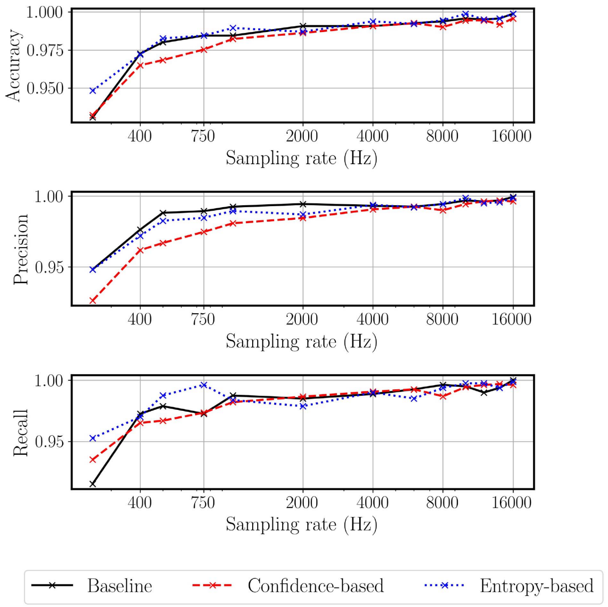

Fig. 4. Comparison of Baseline and OOD Detection on Classification Task for ID Samples using accuracy (top), precision (middle) and recall (bottom) metrics.

represents the total number of words in the reference, which is equal to the sum of substitutions, deletions, and correctly placed words (N=S+D+C).

To investigate the consistency of sampling rate between cough detection task and OOD task, we tested the system detection performances and speech recognition performance at the sampling rates $f \in \{250, 400, 500, 750, 1\text{k}, 2\text{k}, 4\text{k}, 6\text{k}, 8\text{k}, 10\text{k}, 12\text{k}, 14\text{k}, 16\text{k}\}$. The reason why we consider the speech recognition is that a lower sampling rate can also inhibit speech recognition, which can ensure that privacy is maintained for the users.

For our evaluation, we select the model with the best ID task accuracy on the validation set during training, i.e., the epoch that yielded the best accuracy. We do not include the OOD task as part of the selection criteria, because we assume that the OOD data is not available during training or validation. This is to ensure generalization and to avoid ending up with a model that is tailored to a specific OOD sample. The performances reported below correspond to averages of 4 runs of training with differential random seeds for each model in order to reduce model variability.

We compared ID task results between the baseline model and the two OOD detection approaches in Fig. 4. The curves for all three models almost overlap on the three metrics. The accuracy (top plot) of the baseline- and the entropy-based models are the closest over all sampling rates, while confidence-based model is a slightly lower at some sampling rates, e.g. 750 Hz, 8 kHz, 14 kHz, and 16 kHz. In the precision (middle) plot, the entropy-based model is slightly worse than the baseline model at recognizing coughs at sampling rates lower than 4 kHz, and the confidence-based model shows a more significant drop. In the recall (bottom) plot, the entropy-based model got the highest recall at a sampling rate of 750 Hz, and all models got similar recalls at higher frequencies. Overall, the OOD detection models did not show any significant drop in ID task performance. We observe a drop in performance from all approaches with frequencies below 500 or 750 Hz, which is consistent with our previous observations.

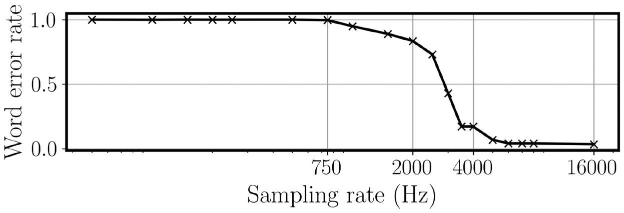

Fig. 5. Speech quality for different sampling rates.

When it comes to the speech recognition, as seen in Fig. 5, the WER is close to 100% at 750 Hz where the cough/speech classification performance remains relatively high. WER reaches 50% at around 3 kHz giving us a range of 750 Hz–2 kHz to protect user privacy while still correctly classifying cough and speech.

C. OOD Detection Performance for Different Sampling Rates

In this section, we use the same setting for the models and Dataset B for our analysis. Fig. 6 shows the results plots of mean and standard deviation of 4 confidence-based and entropy-based models separately, trained with different starting points.

From the plots in Fig. 6, we found that when the sampling rate is 4 kHz, the confidence-model is able to achieve the best performance and the model maintains relatively acceptable results when the sampling rate is 750 Hz or higher. We could observe that the model becomes less consistent at sampling rates higher than 6 kHz and gets an outlier at 12 kHz sampling rate with high standard deviation. This implies that the confidence model may not be good at dealing with high-frequency information.

The entropy-based model achieves the best performance at the highest sampling rate 16 kHz and yields less consistent results at rates higher than 4 kHz with larger standard deviation for AUROC, FPR95, and detection error plots. When looking more closely into F1, precision, and recall, we found that the model keeps high recall from 2 kHz to 8 kHz and the results are less consistent for precision which means that the models tend to focus more on retrieving ID data.

Overall, 750 Hz is the lowest sampling rate with acceptable results and 8 kHz is usually when results become less consistent for both models. Compared to confidence-based models, entropy-based models overall obtain better results at sampling rates higher than 750 Hz. From Fig. 6, the AUROC value from the entropy-based model at 16 kHz is slightly higher than the confidence-based model at 4 kHz with lower standard deviation. For FPR95, detection error and F1, the confidence-based model at 4 kHz obtains better values than entropy-based model at 16 kHz but with higher standard deviations. In conclusion, 1) OOD detection models give less consistent results at the high frequency range from 4 kHz to 16 kHz; 2) the confidence-based model at 4 kHz performs better than the entropy-based model

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

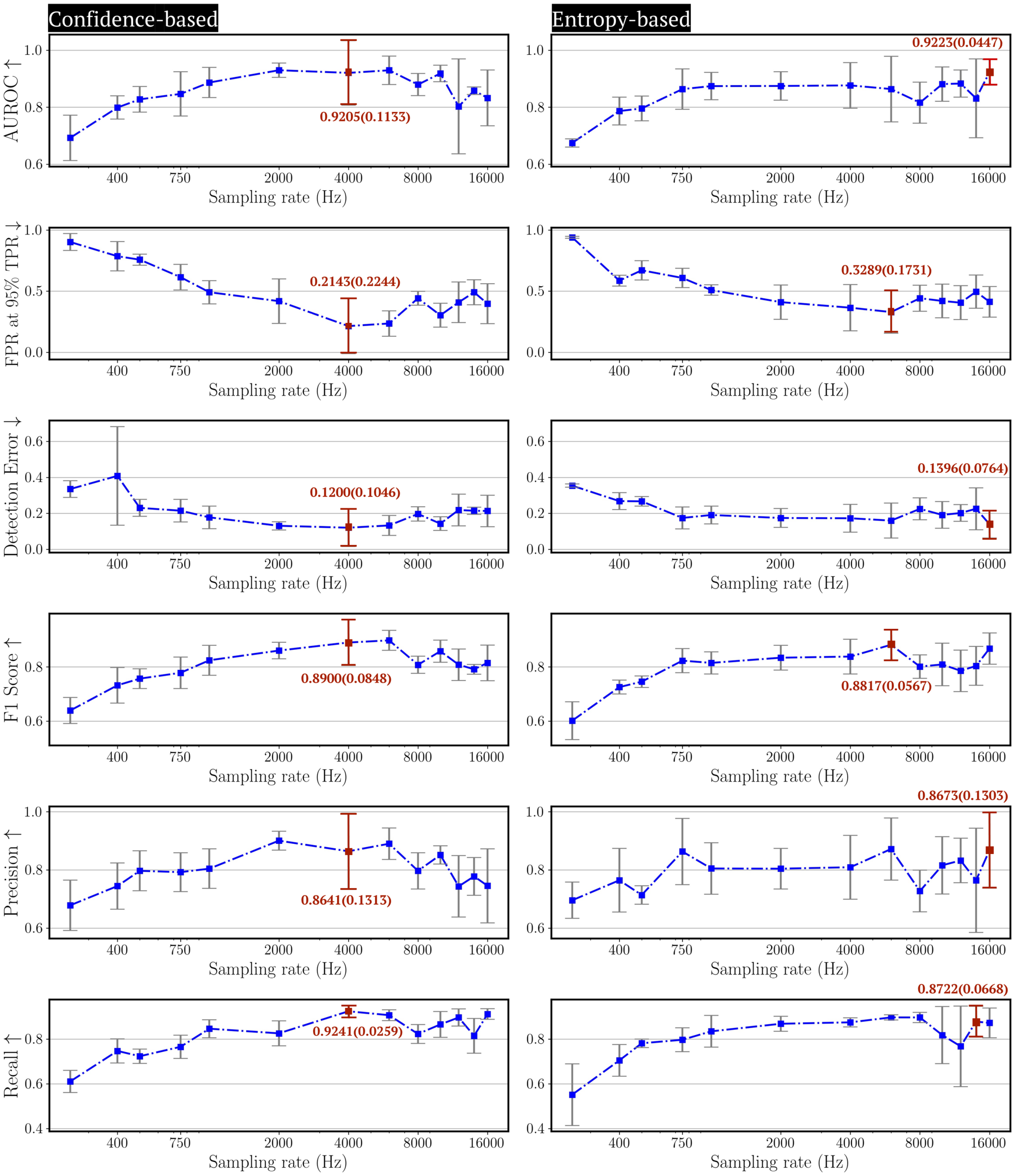

Fig. 6. The Comparison of Learning Confidence-based OOD and Entropy-based OOD Over Variant Sampling Rates. Blue lines are average results over 4 runs and gray lines are the corresponding standard deviation. The best performance over sampling rates is highlighted in red.

at 16 kHz; and 3) the entropy-based model is slightly better at dealing with higher frequency information than the confidence-based model.

D. Detection Performance for Different Window Sizes

The mean of a single cough instance is 1.5 s and larger window sizes usually contain more information producing a lower performance which was shown in our previous research [21]. To further investigate the influence of window sizes, we tested OOD models with variant window sizes $\tau = [1.5, 2, 3, 4, 5, 10]$ seconds on $f = [400, 750, 8\text{k}, 16\text{k}]$ Hz. The number of samples is listed in Table II. Different overlap sizes are used in different window size settings to keep the fraction of cough and speech as 1:1, and the fraction of ID data and OOD data constant.

For ID task performance, OOD detection models tend to get the best performance at 4 s or 5 s window size at all sampling frequencies and the higher sampling frequency produces better results as shown in Fig. 7. For the OOD task, Fig. 8 shows the

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

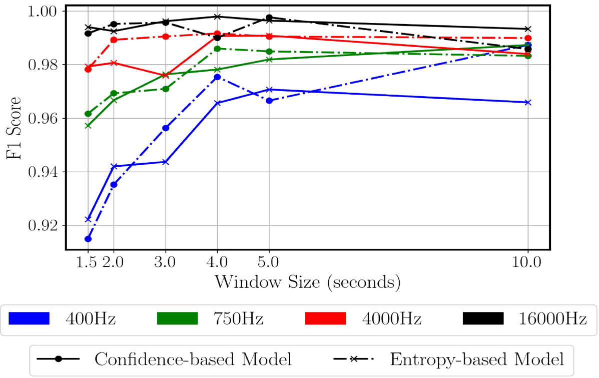

Fig. 7. ID Task Results while Varying the Window Size.

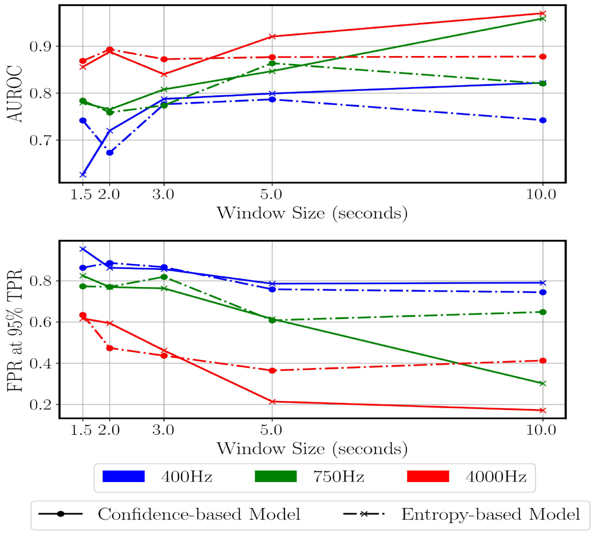

Fig. 8. OOD Task Results for Varying Window Sizes. The results from 16 kHz are removed for clear visualization due to its inconsistency.

results for the OOD models for two metrics as a function of window size that gave the most consistent results. That is, we only show the plots for which a trend was observed in performance as a function of window size. Overall, the models tend to get better performance for greater window sizes. In general, the models in Fig. 8 at 10 seconds window size achieve the highest AUROC and the lowest FPR95. Results using a 5-second window are comparable. When it comes to the comparison between the confidence-based and entropy-based models, there is no obvious trend that shows that one is superior.

E. Detection for Different OOD Proportions

Up to this point, we have discussed performance separately for ID and OOD tasks. However, in the real live setting, we would expect to see data that is a mixture of ID and OOD samples. Hence, we analyze the performance of all models as a function of the proportion of the OOD data introduced. From this study, we found that the baseline model performance is drastically reduced by OOD data and the OOD detection models surpass the baseline model once a high enough proportion of OOD data is present. The proportion at which this trade-off occurs varies depending on the sampling rate.

TABLE III
Recall Values (%) for Each Model

Sampling Rate (Hz)	Recall (%)		
	Baseline (without OOD)	Confidence-based	Entropy-based
400	97.26	78.18	77.18
750	97.26	73.07	83.92
16000	100	86.53	90.40

We use the best performing models for $f = [400, 750, 16\text{k}]$ Hz with $\tau = 5$ s for this study and changed the proportion of OOD data from 0% to 50% of the total test data set. We focus purely on the detection of cough instances as our positive samples and consider any other type of sample as negative. For OOD detection, the same strategies are used to select confidence and distance thresholds for confidence- and entropy-based models, that is, using the threshold which produces the best performance (the highest AUROC). Thus, thresholds are variant at different OOD data proportions.

Fig. 9 shows the comparison between baseline and OOD detection models. The x-axis represents OOD data proportion ranging from 0% to 50%, where $n\%$ indicates the number of OOD samples introduced in the test set. This quantity is given by

$$n\% = \frac{\text{The \# of OOD Samples}}{\text{The \# of Samples in the Test Data}} \cdot 100\%. \tag{17}$$

In the accuracy plot, although the initial accuracy of baseline model is higher than that of OOD detection models, the baseline model accuracy keeps going down while the OOD detection models increase slightly when adding more OOD data. At 16 kHz sampling rate (left), OOD models surpass the baseline model at an OOD proportion of around 10%, and the initial accuracy of the baseline drops from 99.89% to around 90%. At 750 Hz (middle), our models outperform the baseline at an OOD proportion of around 15%, and the accuracy of the baseline drops from 98.44% to around 81%. At 400 Hz (right), our models outperform the baseline at an OOD proportion of around 22%, and the accuracy of the baseline drops significantly to around 71%. On the other hand, we observe for all sampling rates that the OOD approaches maintain a consistent slightly increasing performance. Overall, we observe decreases in all model for the F1 score and precision as OOD data is introduced. However, the decay in performance for F1 score and precision for the OOD models is less drastic. Compared to confidence-based model, the entropy-based model has an overall lower drop in performance in all metrics (except for accuracy at 400 Hz). The entropy-based model seems to be equipped with a stronger mechanism to deal with OOD data.

Since we are using the same ID data in all these experiments, the recall does not change for any of these models. Table III summarizes these values. For example, at 16 kHz, the baseline model

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

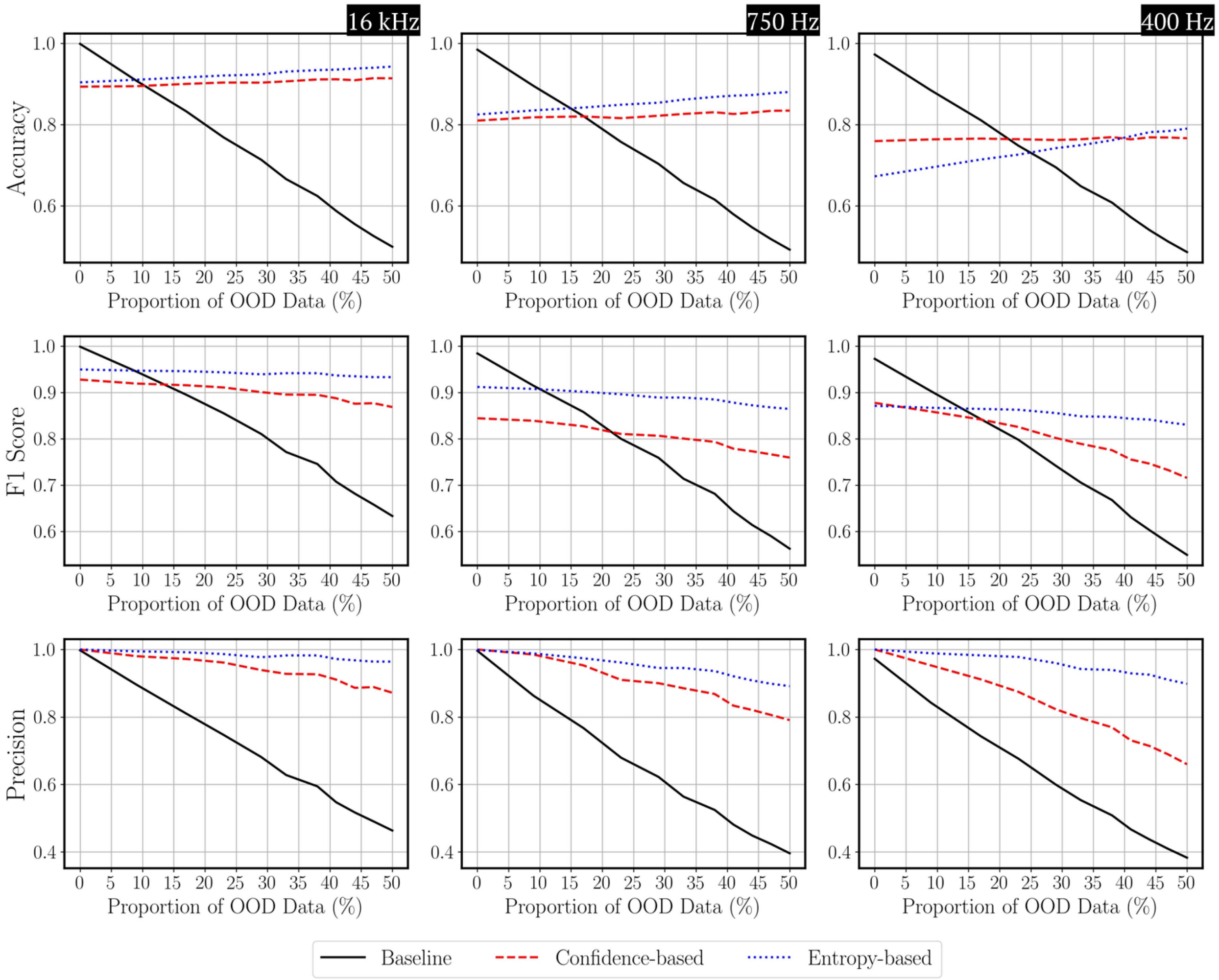

Fig. 9. Comparison for three models with variant OOD Data proportion at 16 kHz, 750 Hz, and 400 Hz sampling rates.

can retrieve all cough samples while OOD detection models retrieve 86.53% and 90.40% for confidence and entropy-based models respectively. Note that there is a more significant drop on the OOD models recall performance when going to lower sampling rates. The lower performance in OOD models is due to our choice on the rule for turning low confidence samples into OOD samples. This can be tuned depending on the use case to get higher recall with the trade-off of having a higher sensitivity to higher proportions of outliers.

F. Discussion

To get convincing results, we use the following strategy to ensure that the cough samples in the train, validation, and test sets come from different individuals. We use all segments from the same audio sample as either part of the train, validation, or test sets. That is, there is no overlap of audio samples between these sets. This a valid scheme for the Coughvid dataset [10] since the author mentioned a diverse range of participant demographics in the data collection process. Regarding the FluSense dataset [11], this dataset was collected "in four public waiting rooms within the university health service of the University of Massachusetts Amherst, a research and land-grant university with more than 30,000 students.". Even though it is hard to guarantee that each cough audio comes from unique individuals, given the collection location and the large student population, the probability of the same person's coughs appearing in multiple sets is very low.

For hyper-parameters selection, specified hyper-parameters for each model follow the work in which they were proposed [24], [40]. The authors of [24] mentioned that the budget hyper-parameter β will not affect results a lot, and the authors of [40] observed stable improvement (but not significant) in results when E_s is equal to or higher than 10.

In Section IV-A, image feature extraction models, ResNet and VGG, outperform frequency feature extraction model FENet when the input is signal Mel-spectrogram. This indicates that Mel-spectrogram can be a strong image feature representing signal information. In this case, models used or pretrained on image data have potential ability to deal with bio-signal processing problems by transferring input signal into Mel-spectrogram. The vision transformer is a popular topic in the computer vision area and many transformer models are proven to be powerful [43], [44], [45]. These models can embed input features into a latent space where the OOD detection can be processed in a more

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

interpretable way. Besides, the attention mechanism in the transformer could be another potential tool to detect OOD data.

In Sections IV-B through IV-D, the OOD models are tested on inputs with different sampling frequencies and audio window sizes. Fig. 4 in Section IV-B proves that the ID performances of OOD models are close to that of the baseline model. Even at 750 Hz where the user privacy is protected by low speech quality (Fig. 5), OOD detection models achieve convincing performance and the entropy-based model reaches almost the same performance as the model without OOD detection. Therefore, we can infer that the OOD models are more stable than the baseline system when the data is not ideal. Sections IV-C and IV-D (Fig. 8) show that the confidence-based OOD model produced the highest performance at 4 kHz and the entropy-based model produced some of the highest performance at 16 kHz; however, the results were less consistent (i.e., there was more variability) at higher sampling rates. Not like ID performance, the higher sampling, the better performance, OOD performance does not rely on sampling rate indicating that different OOD models have different sampling rates to produce the best results. Based on this fact, we can use different models while different sampling frequencies to improve the overall results. From Section IV-D, we learned that the OOD prediction quality was better with larger window sizes with little change observed between 5 and 10 seconds window sizes. The conjecture is that longer window sizes may contain more cough instances leading to a more recognizable spectrogram that is easier to separate from OOD samples.

In Section IV-E, we investigated the limitation of OOD detection models and Fig. 9 showed that the baseline model is very sensitive to OOD data and the performance drops with the increase of OOD data. We used the best window size setting (5 s) in this experiment to better show the improvement and the limitation of OOD - Cough Detection methods. The limitation of the OOD method is that when there is not enough OOD data, the detection accuracy is lower than the baseline detection accuracy. As the proportion of OOD samples increases, this gap is reduced and OOD models end up performing better. The points at which OOD models become better than the baseline is at OOD proportions of 10%, 15%, and 22% for 16 kHz, 750 Hz, and 400 Hz respectively. In our work, we only tested proportions ranging from 0% to 50% because 50% is enough to see the trends in improvement, and the cough detection model had dropped below 50% accuracy by then on the standard model.

In this paper, the strategy used to select the threshold for OOD detection makes use of the test set. We use the test set instead of the validation set because, for more general settings, we may not have access to a representative sample of OOD data for validation. For OOD problem, it is common to not use OOD data during the training and fine-tuning process, which is a simulation of real-world scenarios where only ID data is available. Therefore, an optimal threshold is important for OOD detection model to produce convincing results, however, in the real-world problem, we only have the access to ID, leading to threshold selection difficulty. To measure the overall performance, we also used the AUROC metric, which is not affected by the threshold selection. For the confidence-based model, Terrance DeVries [24] has proved that "misclassified in-distribution examples from a validation set can serve as a conservative proxy for out-of-distribution examples when calibrating the detection threshold." For the entropy-based method, the density plot for the minimum distance between samples and prototypes in Fig. 3 clearly shows that the tail of the ID distance density overlaps with part of OOD distance density. Based on this fact, a 95% single-sided confidence interval could be used as a threshold.

V. Conclusion and Future Work

We have presented a robust cough detection algorithm with out-of-distribution detection. An image feature extractor was embedded into the pipeline because it outperformed the frequency feature extraction neural network. We proved that the new algorithm is able to detect OOD samples without sacrificing ID task performance. A wide range of experiments was designed to analyze the performance of this new algorithm. We found that the new algorithm produces trustful results when the sampling rate is greater than 750 Hz and the window size is between 4–10 seconds. We also investigated the limitation of the OOD cough detection algorithm. If there is virtually no OOD data present in the test set, the baseline model performs slightly better than the OOD models. In the best-case scenario (16 kHz sampling rate), the OOD models perform better than the baseline once the test set consists of more than 10% of OOD data. Cough detection with OOD can be useful in the real world due to more noisy data and more acoustic classes.

One main challenge is threshold selection for OOD detection. In the future, we will investigate different methods to choose thresholds including 95% confidence interval, clustering, and the relation between ID task and OOD task. We also plan to investigate other architectures such as vision transformers to extract features that can capture attention while possibly providing more explainability.

References

[1] K. S. Alqudaihiet al., "Cough sound detection and diagnosis using artificial intelligence techniques: Challenges and opportunities," *IEEE Access*, vol. 9, pp. 102327–102344, 2021.

[2] C. Van Schayck and N. Chavannes, "Detection of asthma and chronic obstructive pulmonary disease in primary care," *Eur. Respir. J.*, vol. 21, no. 39 suppl, pp. 16s–22s, 2003.

[3] V. Misra et al., "Flexible technologies for self-powered wearable health and environmental sensing," *Proc. IEEE*, vol. 103, no. 4, pp. 665–681, Apr. 2015.

[4] Y. Chen, M. D. Wilkins, J. Barahona, A. J. Rosenbaum, M. Daniele, and E. Lobaton, "Toward automated analysis of fetal phonocardiograms: Comparing heartbeat detection from fetal doppler and digital stethoscope signals," in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 975–979.

[5] J. Ren et al., "Likelihood ratios for out-of-distribution detection," in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2019, pp. 14707–14718.

[6] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in *Proc. Int. Conf. Learn. Representations*, 2015, *arXiv:1412.6572*.

[7] A. Nguyen, J. Yosinski, and J. Clune, "Deep neural networks are easily fooled: High confidence predictions for unrecognizable images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 427–436.

[8] T. Iqbal, Y. Cao, Q. Kong, M. D. Plumbley, and W. Wang, "Learning with out-of-distribution data for audio classification," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 636–640.

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.

[9] D. Snyder, G. Chen, and D. Povey, “MUSAN: A music, speech, and noise corpus,” 2015, *arXiv:1510.08484*.

[10] L. Orlandic, T. Teijeiro, and D. Atienza, “The COUGHVID crowdsourcing dataset, a corpus for the study of large-scale cough analysis algorithms,” *Sci. Data*, vol. 8, no. 1, pp. 1–10, 2021.

[11] F. Al Hossain, A. A. Lover, G. A. Corey, N. G. Reich, and T. Rahman, “FluSense: A contactless syndromic surveillance platform for influenza-like illness in hospital waiting areas,” *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 1, pp. 1–28, 2020.

[12] M. Al-Khassaweneh and R. Bani Abdelrahman, “A signal processing approach for the diagnosis of asthma from cough sounds,” *J. Med. Eng. Technol.*, vol. 37, no. 3, pp. 165–171, 2013.

[13] J.-M. Liu, M. You, Z. Wang, G.-Z. Li, X. Xu, and Z. Qiu, “Cough detection using deep neural networks,” in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2014, pp. 560–563.

[14] S. Matos, S. S. Birring, I. D. Pavord, and H. Evans, “Detection of cough signals in continuous audio recordings using hidden Markov models,” *IEEE Trans. Biomed. Eng.*, vol. 53, no. 6, pp. 1078–1083, Jun. 2006.

[15] I. D. Miranda, A. H. Diacon, and T. R. Niesler, “A comparative study of features for acoustic cough detection using deep architectures,” in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 2601–2605.

[16] R. X. A. Pramono, S. A. Imtiaz, and E. Rodriguez-Villegas, “Automatic cough detection in acoustic signal using spectral features,” in *Proc. IEEE 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2019, pp. 7153–7156.

[17] J. Monge-Álvarez, C. Hoyos-Barceló, L. M. San-José-Revuelta, and P. Casaseca-de-la Higuera, “A machine hearing system for robust cough detection based on a high-level representation of band-specific audio features,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 8, pp. 2319–2330, Aug. 2019.

[18] G.-T. Lee, H. Nam, S.-H. Kim, S.-M. Choi, Y. Kim, and Y.-H. Park, “Deep learning based cough detection camera using enhanced features,” *Expert Syst. Appl.*, vol. 206, 2022, Art. no. 117811.

[19] S. Jokić et al., “Tripletcough: Cougher identification and verification from contact-free smartphone-based audio recordings using metric learning,” *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2746–2757, Jun. 2022.

[20] F. Barata, K. Kipfer, M. Weber, P. Tinschert, E. Fleisch, and T. Kowatsch, “Towards device-agnostic mobile cough detection with convolutional neural networks,” in *Proc. IEEE Int. Conf. Healthcare Inform.*, 2019, pp. 1–11.

[21] M. Abdelkhalek, J. Qiu, M. Hernandez, A. Bozkurt, and E. Lobaton, “Investigating the relationship between cough detection and sampling frequency for wearable devices,” in *Proc. IEEE 43rd Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2021, pp. 7103–7107.

[22] D. Hendrycks and K. Gimpel, “A baseline for detecting misclassified and out-of-distribution examples in neural networks,” in *Proc. Int. Conf. Learn. Representations*, 2017.

[23] S. Liang, Y. Li, and R. Srikant, “Enhancing the reliability of out-of-distribution image detection in neural networks,” in *Proc. 6th Int. Conf. Learn. Representations*, 2018.

[24] T. DeVries and G. W. Taylor, “Learning confidence for out-of-distribution detection in neural networks,” 2018, *arXiv:1802.04865*.

[25] G. Shalev, Y. Adi, and J. Keshet, “Out-of-distribution detection using multiple semantic label representations,” in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, 2018, pp. 7375–7385.

[26] A. Subramanya, S. Srinivas, and R. V. Babu, “Confidence estimation in deep neural networks via density modelling,” 2017, *arXiv:1707.07013*.

[27] T. Denouden, R. Salay, K. Czarnecki, V. Abdelzad, B. Phan, and S. Vernekar, “Improving reconstruction autoencoder out-of-distribution detection with mahalanobis distance,” 2018, *arXiv:1812.02765*.

[28] V. Abdelzad, K. Czarnecki, R. Salay, T. Denounden, S. Vernekar, and B. Phan, “Detecting out-of-distribution inputs in deep neural networks using an early-layer output,” 2019, *arXiv:1910.10307*.

[29] K. Lee, H. Lee, K. Lee, and J. Shin, “Training confidence-calibrated classifiers for detecting out-of-distribution samples,” in *Proc. 6th Int. Conf. Learn. Representations*, 2018.

[30] D. Hendrycks, M. Mazeika, and T. Dietterich, “Deep anomaly detection with outlier exposure,” in *Proc. Int. Conf. Learn. Representations*, 2019.

[31] S. Vernekar, A. Gaurav, V. Abdelzad, T. Denouden, R. Salay, and K. Czarnecki, “Out-of-distribution detection in classifiers via generation,” 2019, *arXiv:1910.04241*.

[32] D. Macêdo and T. Ludermir, “Enhanced isotropy maximization loss: Seamless and high-performance out-of-distribution detection simply replacing the softmax loss,” 2021, *arXiv:2105.14399*.

[33] M. Huzaifah, “Comparison of time-frequency representations for environmental sound classification using convolutional neural networks,” 2017, *arXiv:1706.07156*.

[34] A. Kumar, M. Khadkevich, and C. Fügen, “Knowledge transfer from weakly labeled audio using convolutional neural network for sound events and scenes,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 326–330.

[35] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[36] I. K. Niazi, N. Naseer, and H. Santosa, *Signal Processing Using Non-invasive Physiological Sensors*. Basel, Switzerland: Multidisciplinary Digit. Publishing Instit., 2022.

[37] K. J. Piczak, “ESC: Dataset for environmental sound classification,” in *Proc. 23rd ACM Int. Conf. Multimedia*, 2015, pp. 1015–1018.

[38] E. Fonseca et al., “General-purpose tagging of freesound audio with audioset labels: Task description, dataset, and baseline,” 2018, *arXiv:1807.09902*.

[39] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, “Librispeech: An ASR corpus based on public domain audio books,” in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2015, pp. 5206–5210.

[40] D. Macêdo, T. I. Ren, C. Zanchettin, A. L. Oliveira, and T. Ludermir, “Entropic out-of-distribution detection: Seamless detection of unknown examples,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2350–2364, Jun. 2022.

[41] R. C. Streijl, S. Winkler, and D. S. Hands, “Mean opinion score (MOS) revisited: Methods and applications, limitations and alternatives,” *Multimedia Syst.*, vol. 22, no. 2, pp. 213–227, 2016.

[42] A. C. Morris, V. Maier, and P. Green, “From WER and RIL to MER and WIL: Improved evaluation measures for connected speech recognition,” in *Proc. 8th Int. Conf. Spoken Lang. Process.*, 2004.

[43] N. Parmar et al., “Image transformer,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4055–4064.

[44] Z. Liu et al., “Swin transformer: Hierarchical vision transformer using shifted windows,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10012–10022.

[45] K. Han et al., “A survey on vision transformer,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 1, pp. 87–110, Jan. 2023.

Authorized licensed use limited to: N.C. State University Libraries - Acquisitions & Discovery S. Downloaded on August 22,2023 at 13:12:35 UTC from IEEE Xplore. Restrictions apply.