ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746334

COUGHTRIGGER: EARBUDS IMU BASED COUGH DETECTION ACTIVATOR USING AN ENERGY-EFFICIENT SENSITIVITY-PRIORITIZED TIME SERIES CLASSIFIER

Shibo Zhang[⋆], *Ebrahim Nemati*[†], *Minh Dinh*[‡], *Nathan Folkman*[‡], *Tousif Ahmed*[†], *Mahbubur Rahman*[†],*Jilong Kuang*[†], *Nabil Alshurafa*[⋆], *Alex Gao*[†]

[⋆] Northwestern University, Chicago, IL, USA
[†] Samsung Research America, Mountain View, CA, USA
[‡] Samsung Design Innovation Center, San Francisco, CA, USA

ABSTRACT

Persistent coughs are a major symptom of respiratory-related diseases. Increasing research attention has been paid to detecting coughs using wearables, especially during the COVID-19 pandemic. Microphone is most widely used sensor to detect coughs. However, the intense power consumption needed to process audio hinders continuous audio-based cough detection on battery-limited commercial wearables, such as earbuds. We present *CoughTrigger*, which utilizes a lower-power sensor, inertial measurement unit (IMU), in earbuds as a cough detection activator to trigger a higher-power sensor for audio processing and classification. It runs all-the-time as a standby service with minimal battery consumption and triggers the audio-based cough detection when a candidate cough is detected from IMU. Besides, the use of IMU brings the benefit of improved specificity of cough detection. Experiments are conducted on 45 subjects and *CoughTrigger* achieved 0.77 AUC score. We also validated its effectiveness on free-living data and through on-device implementation.

Index Terms— Cough Detection Activation, Sensitivity-prioritized Classification, Multi-Center Classifier, Template Matching, Earbuds

1. INTRODUCTION

Persistent coughs can be a sign of serious lung diseases, such as Chronic Obstructive Pulmonary Disease (COPD), asthma, lung cancer, and COVID-19. Reliable automated detection of coughs using everyday wearable devices is especially desirable. In recent years, wearable devices such as smartphones and earbuds are becoming prevalent in our daily life. A body of work has emerged, with wearable sensors showing promise in detecting coughs and classifying different types of coughs [1–9], including earbuds-based devices [7–9]. Audio-based sensing has shown promise in detecting coughs on device, but requires higher battery consumption and introduces privacy concerns [1–3, 5, 6, 8]. Battery consumption is a major concern among commercial wearables, such as earbuds, where power is a limited resource. For example, Samsung Galaxy

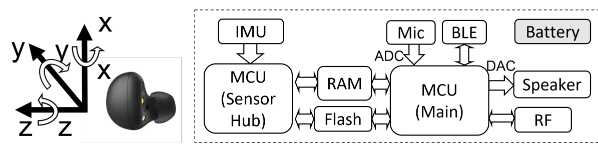

Fig. 1: Samsung Galaxy Buds2 hardware architecture

Buds2, as shown in Fig. 1, has a 61 mAh battery supporting up to 7.5 hours of play time per charge [10], and adding extra features will cause a drop in battery life, and ultimately utility and user interests. Some explored the use of an inertial measurement unit (IMU), due to its low battery usage and computational load, to detect coughs [7, 9]. It was soon realized that using a traditional classifier (XGBoost) to detect coughs could only yield a 47% sensitivity and 54% F1-Score [7]. The large number of confounding head movements made it challenging to distinguish between a cough and a non-cough.

Audio and IMU sensing have complementary characteristics: audio can better distinguish between coughs and non-coughs, while IMU enables battery-efficiency. Given that the majority of the time is often spent by people not coughing, the low-power IMU can be used to trigger the audio sensing pipeline when a candidate cough is detected. To do this, the IMU must yield high sensitivity, to ensure all coughs are ultimately passed onto the audio sensing. Since traditional classification methods yield low sensitivity, we propose a novel multi-center template matching algorithm to achieve high sensitivity in the IMU data. This algorithm is then used in a two-stage pipeline, where an always-on IMU triggers audio processing, only when needed, to reliably confirm the detection of a cough. It can further alleviate the privacy issue because it does not require constant collection of audio.

We summarize the contributions of this work: **(1)** A battery-efficient dual IMU-audio cough detection framework using earbuds; we define the requirements for the first IMU stage of the pipeline and formalize it as a sensitivity-prioritized classification problem. **(2)** We propose *CoughTrigger*, an IMU-based cough detection activator based on a novel

© IEEE 2022. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on November 19,2025 at 16:18:43 UTC from IEEE Xplore. Restrictions apply.

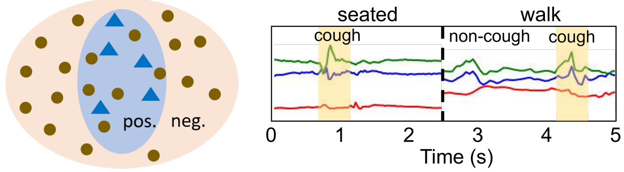

Fig. 2: Illustration of sensitivity-prioritized classification and 3-axis accelerometer cough data (x-axis, y-axis, z-axis).

template matching method – Multi-Nearest Center Classifier, to trigger audio processing and cough detection. We present superior experimental results compared with baselines. **(3)** We implement our method on a commercial device and prove the effectiveness of *CoughTrigger* with on-device result. In the future, towards all-day round continuous cough monitoring, an opportunistic sensing network composed of multiple devices is promising and feasible. Earbuds platform is a key component due to its consistent location and in-ear audio sensing enabling robust cough detection.

2. METHODOLOGY

2.1. Sensitivity-prioritized Classification

Unlike a traditional binary classification problem where the positive and negative samples are distinguishable either by human perception or machine learning techniques in a transformed feature space, the IMU cough data and non-cough ones are not fully separable due to an overlapping region in feature space, as illustrated in Figure 2. To guarantee the positive samples (coughs) can be detected to trigger audio-based cough detection in the next stage, we formalize the problem as a *sensitivity-prioritized* classification task, which means higher sensitivity is prioritized over specificity, and approach this problem with a novel template matching algorithm.

2.2. Multi-Nearest Center Classifier

Zhang *et al.* proposed a template matching algorithm [9], called Multi-Centroid Classifier, which aims at iteratively creating an increasing number of clusters, each of which has its own centroid and radius and all together cover all the positive samples while include as few negative samples as possible. When a satisfying accuracy is achieved, the derived centroids and radii will be used in the test set as templates and thresholds to classify positive samples from negative ones. The method has shown merits in accuracy, inference speed, and model size. In this work, we make substantial modifications and propose a new algorithm, Multi-Nearest Center (MNC) Classifier, on which we build *CoughTrigger*. The original algorithm is composed of three parts: discrepancy cost, discrepancy-based clustering, and cluster averaging, of which we have modified two parts significantly. Not only we change the manner of discrepancy-based clustering for increased robustness to regional density in feature space, but

Algorithm 1: Training a Multi-Nearest Center Classifier

input : Positive and negative training samples and stop criterion H
output: K templates C_k, $(k = 1, 2, \dots K)$ and K thresholds
Initialize number of clusters $K = 1$;
Assign all the positive samples to cluster R_1;
Randomly select seed center C_1 from positive samples of cluster R_1;
Do discrepancy-based clustering with C_1 to obtain total cost L, updated center C_1 and threshold;
while *total cost* $L >$ *stop criterion* H **do**
  Select R_t with the highest cost $t = \arg\max_i L_i$;
  From the positive samples of R_t, select the sample which brings the largest cost increase as the new seed centroid;
  Do discrepancy-based clustering using $K + 1$ centers $C_1, \dots, C_{K+1}$ to obtain total cost L, updated $K + 1$ centers, and $K + 1$ thresholds;
  Calculate cost L_k for each of the $K + 1$ clusters;
  $L = \sum_{k=1}^{K+1} L_k$;
  $K = K + 1$;
end

we totally replace the cluster averaging step with a center sample selection procedure for better convergence. Further, we show its superior classification capability in Section 4.1.

Training Phase We present the training procedure in Algorithm 1. It comprises three major parts: discrepancy cost, discrepancy-based clustering, and center selection. The discrepancy cost is used to measure and compare the purity of each cluster. Discrepancy-based clustering assigns positive samples into clusters. Center selection selects one center template and one threshold for each cluster. As the original method [9] is sensitive to regional density in the feature space, which prevents the derivation of a pseudo-optimal clustering result, we substantially modify the discrepancy-based clustering and cluster averaging. We also change the way of selecting new seeds for better convergence. Below are the three modifications introduced:

(1) We update the averaging step, making it easier to converge on noisy datasets. When deriving the centroid for each cluster, instead of averaging operation, we select as center the positive sample which has the minimum $Dist_{max}$ ($Dist_{max}$ is the max distance between the current sample and any other positive sample).

(2) When clustering, we require each sample to be assigned to only one cluster instead of allowing samples assigned to multiple clusters. By modifying the clustering step, we make the algorithm less sensitive to regional density in the feature space. We will showcase the effect in Section 4.1.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on November 19,2025 at 16:18:43 UTC from IEEE Xplore. Restrictions apply.

(3) When increasing the number of clusters, instead of adding one random centroid seed, we select the sample that brings the greatest increase to the cost function.

Inference Phase The inference steps are unchanged [9]. The distance between a test sample and each template is calculated and compared against the threshold. If the distance is smaller than the threshold, then the test sample is predicted as positive, otherwise it is negative.

3. EXPERIMENT

3.1. Data Collection

We generated an earbud-based cough dataset from 45 participants (15 with lung disease, 22 male, 41.4 $\pm$ 10.7 years old). In the experiments, one earbud was worn by each participant to collect IMU data at 50 Hz and audio at 16 KHz.

In-lab Experiment We collected eight cough sessions, including five stationary and three non-stationary periods. Each participant coughed continuously with a short pause between every two coughs. Stationary periods comprised: coughing while seated (30s), coughing while lying down (30s), coughing while listening to music from an earbud (30s), coughing with background fan noise (30s), and coughing with background music/TV noise (30s). Non-stationary periods comprised: coughing while performing yoga in quiet environment (45s), coughing while performing yoga in noisy environment (45s), and coughing while walking (1 min). To evaluate the specificity, non-cough activities that involve signals that could resemble cough motion were collected including: eating (30s), drinking (30s), laughing (30s), scripted speech (1 min), throat clearing (30s), free head motion while talking (30s), and one bystander coughing session (30s). On average, 10.5 coughs are captured in each cough session. We aim at detecting each single cough while preventing false alarms.

Free-living Experiment Enrolled participants also took part in a free-living experiment, where they were asked to cough naturally in the morning and in the afternoon for one week. At each time, data were collected for coughs while seated for 30s and while walking for 30s. Due to limitation of data logging app, there is an up-to-400ms random drift between audio and IMU data. We resolved the asynchronization by identifying peaks in accelerometer x-axis, which corresponds to the most probable motion of wearer while coughing. Using audio, we annotated every single cough from 45 in-lab participants and 15 free-living participants in lung disease cohort.

3.2. Model Development

Data Preprocessing The 3-axis accelerometer data are preprocessed using a moving average filter with a window size as 10 samples and a Butterworth high-pass filter ($\omega_c = 3\pi$). Positive samples are extracted from seated and lying down cough sessions using a 0.4s window centered around the annotated IMU cough.To increase the variety of cough data during training, we applied three time series augmentation methods, namely jittering, scaling, and magnitude warping [11] on cough data enlarging the positive sample size by its threefold. For negative samples, we segmented the non-cough session accelerometer data using a 0.4s sliding window with 0.1s stride. Then we randomly subsampled four times the positive samples from non-cough sessions as negative samples, in order to balance class ratio.

Model Training To expedite the training process, we trained one MNC model with stop precision criterion as 0.8 for each participant in the training set using multi-variate DTW distance measure, then aggregated the templates from each training participant and ran all the templates on the training participants. Afterwards, we used a greedy algorithm to rank the templates and select the top K templates based on template importance measured by how many new positive samples are hit by each template.

Model Testing We applied MNC with K templates on the accelerometer data of the test participant's cough and non-cough sessions, with 0.4s window size and 0.02s stride size. Then, we aggregated and merged all the predicted cough windows to determine the final cough event prediction. In our case, the capability of manually adjusting the trade-off between sensitivity and specificity is desirable. One benefit of MNC is that we have two ways to adjust it: to choose the number of top K templates used and to adjust the thresholds of templates. When more (less) templates are used, or when we increase (decrease) the thresholds of templates, it leans towards a higher sensitivity (specificity).

Evaluation Method In the in-lab experiment, we adopted Leave-One-Subject-Out Cross Validation (LOSOCV), a modified k-fold cross-validation method in human-centered studies [12, 13], where the number of folds k is equal to the number of participants, to evaluate the generalizability of a participant-independent model. During training, we excluded the in-lab non-stationary cough sessions due to a high volume of noise caused by a wide range of body movements. In the free-living experiment, we used the model trained on the in-lab stationary cough sessions and non-cough sessions and tested on the free-living dataset. In total, we utilized 10.25 hours of data. The summary of experiments is shown in Table 1. For cough sessions, we compared the predicted cough segments against the ground truth coughs. True positives are defined as predicted cough segments intersected by ground truth coughs. We calculated the sensitivity of cough sessions as a ratio of true positives to ground truth coughs. We used sample-level specificity to test how well it can specify non-cough events.

4. RESULT

4.1. Improvement of MNC classifier

We used a synthetic dataset generated from Gaussian distribution to validate the MNC classifier, as in Fig. 3. The original MCC achieved 0.57 test accuracy with 10 centroids. Us-

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on November 19,2025 at 16:18:43 UTC from IEEE Xplore. Restrictions apply.

Table 1: Summary of experiments. A: In-lab stationary cough sessions (112.5 min); B: In-lab non-stationary cough sessions (112.5 min); C: In-lab non-cough sessions (180 min); D: Free-living sessions (210 min).

Experiment	Evaluation	Training Pool	Test Pool
In-lab	LOSOCV	A + C	A + B + C
Free-living	Train&test set	A + C	D

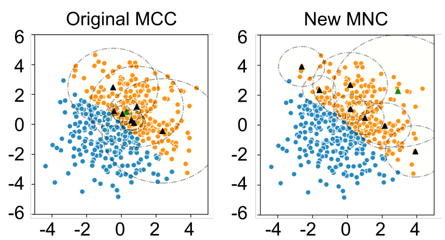

Fig. 3: Multi-Nearest Center Classifier improvement.

ing MNC, 0.99 test accuracy was obtained with 8 centers. The first reason is that the original method allows one sample to belong to multiple clusters, increasing the possibility of highly overlapped clusters. Instead, the new method can separate clusters towards different directions. Second, the new clustering averaging method makes MNC less sensitive to regional density.

4.2. *CoughTrigger* Results

Choice of Input Data We compared all the combinations of the 3-axis accelerometer data, including three using one axis, three using two axes, and one using three axes. The best result is achieved with both the x- and y-axis. That aligns with the IMU direction in Fig. 1 as z-axis only contributes to sensing motion in the left/right direction, which is not as useful in detecting coughs. We tried with only 3-axis gyroscope and combining both accelerometer and gyroscope data but observed lower accuracy.

In-lab Experiment Result Under LOSOCV, when training MNC, on average 14.2 templates were generated for each participant in each fold. When using top five templates, we achieved 90% average sensitivity for stationary cough sessions, and an average sensitivity of 86% on all cough sessions. The average specificity is 53% across all non-cough sessions. When the number of adopted templates ranges from 1 to 30, the Receiver Operating Characteristic (ROC) curve is shown in Fig. 4 Left. We achieved 0.77 Area Under the ROC Curve (ROC AUC) with all the 15 sessions. We achieved 0.59 Precision-Recall AUC in the in-lab experiment, which is as expected since we would receive a higher precision after the next audio-based detection pipeline. Our specificity evaluation was designed for the worst case scenario with a variety of activities. In real life, as for most of the time the wearer is stationary, the specificity is expected to be higher.

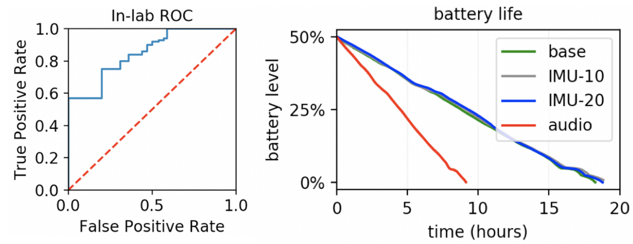

Fig. 4: Left: In-lab experiment ROC curve; Right: Battery life comparison.

Free-living Experiment Result When we applied top 10 templates, we received an average sensitivity of 87% during the stationary periods, and an average sensitivity of 82% on both seated and walking cough periods. Although we only have cough sessions in the free-living setting, we found a 55% average specificity across both sessions for all the participants. When we adjusted the thresholds of each templates, we received 0.80 ROC AUC for stationary coughs and 0.73 ROC AUC for both stationary and walking sessions.

Baseline Methods Since there is no out-of-box sensitivity-prioritized classification method, we investigated ways that may apply, including one-class classifier and adjusting the decision boundary of a traditional classifier. As one-class classifier only models the distribution of positive samples, the overlapping of two classes should not interfere with the modeling of positive class. We tested OC-SVM which identifies the smallest hypersphere consisting of all positive samples [14, 15]. We used the same preprocessing and concatenated x- and y-axis into one vector as input. After adjusting the hyperparameters in a large range with different kernels (linear, RBF, and Sigmoid), we received 0.51 AUC, which is no better than random guess. For a traditional classifier, we implemented a 3-layer NN with 20, 10, and 5 neurons in each layer and observed the same result as OC-SVM.

On-device Implementation We implemented *CoughTrigger* on Samsung Galaxy Buds2. Fig. 4 Right shows the battery life of base firmware without cough detection, *CoughTrigger* using 10 and 20 templates integrated in the base firmware, and an integrated audio-based cough detection method [8]. The base firmware without cough detection has around 18 hours battery life, and integrating *CoughTrigger* makes no significant change, which shows the feasibility of leveraging *CoughTrigger* to reduce battery consumption.

5. CONCLUSION

We introduce a battery-efficient earbuds-based two-stage IMU-audio cough detection framework and formalize the first stage as a sensitivity-prioritized classification problem. We propose using a novel multi-nearest center classifier as a first-stage cough detection activator and demonstrate its effectiveness via in-lab, free-living, and on-device experiments.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on November 19,2025 at 16:18:43 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] Eric C. Larson, TienJui Lee, Sean Liu, Margaret Rosenfeld, and Shwetak N. Patel, "Accurate and privacy preserving cough sensing using a low-cost microphone," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, New York, NY, USA, 2011, UbiComp '11, p. 375–384, Association for Computing Machinery.

[2] Xiao Sun, Zongqing Lu, Wenjie Hu, and Guohong Cao, "Symdetector: Detecting sound-related respiratory symptoms using smartphones," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, New York, NY, USA, 2015, UbiComp '15, p. 97–108, Association for Computing Machinery.

[3] Xuhai Xu, Ebrahim Nemati, Korosh Vatanparvar, Viswam Nathan, Tousif Ahmed, Md Mahbubur Rahman, Daniel McCaffrey, Jilong Kuang, and Jun Alex Gao, "Listen2cough: Leveraging end-to-end deep learning cough detection model to enhance lung health assessment using passively sensed audio," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 5, no. 1, Mar. 2021.

[4] Yanzhi Ren, Chen Wang, Jie Yang, and Yingying Chen, "Fine-grained sleep monitoring: Hearing your breathing with smartphones," in *2015 IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 2015, pp. 1194–1202.

[5] Jesús Monge-Álvarez, Carlos Hoyos-Barceló, Paul Lesso, and Pablo Casaseca-de-la Higuera, "Robust detection of audio-cough events using local hu moments," *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 1, pp. 184–196, 2019.

[6] Daniyal Liaqat, Salaar Liaqat, Jun Lin Chen, Tina Sedaghat, Moshe Gabel, Frank Rudzicz, and Eyal de Lara, "Coughwatch: Real-world cough detection using smartwatches," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 8333–8337.

[7] Tobias Röddiger, Michael Beigl, Michael Hefenbrock, Daniel Wolffram, and Erik Pescara, *Detecting Episodes of Increased Cough Using Kinetic Earables*, p. 111–115, Association for Computing Machinery, New York, NY, USA, 2021.

[8] Ebrahim Nemati, Shibo Zhang, Tousif Ahmed, Md Mahbubur Rahman, Jilong Kuang, and Alex Gao, "Coughbuddy: Multi-modal cough event detection using earbuds platform," in *2021 IEEE 17th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*. IEEE, 2021, pp. 1–4.

[9] Shibo Zhang, Ebrahim Nemati, Tousif Ahmed, Md Mahbubur Rahman, Jilong Kuang, and Alex Gao, "A novel multi-centroid template matching algorithm and its application to cough detection," *arXiv preprint arXiv:2109.00630*, 2021.

[10] Samsung, "All about your samsung earbuds' battery life," [Online]. Available: `https://www.samsung.com/us/support/answer/ANS00078081/`, Accessed: 2021-09-23.

[11] Terry T. Um, Franz M. J. Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, New York, NY, USA, 2017, ICMI 2017, pp. 216–220, ACM.

[12] Atesh Koul, Cristina Becchio, and Andrea Cavallo, "Cross-validation approaches for replicability in psychology," *Frontiers in Psychology*, vol. 9, 2018.

[13] Davoud Gholamiangonabadi, Nikita Kiselov, and Katarina Grolinger, "Deep neural networks for human activity recognition with wearable sensors: Leave-one-subject-out cross-validation for model selection," *IEEE Access*, vol. 8, pp. 133982–133994, 2020.

[14] Yunqiang Chen, Xiang Sean Zhou, and T.S. Huang, "One-class svm for learning in image retrieval," in *Proceedings 2001 International Conference on Image Processing (Cat. No.01CH37205)*, 2001, vol. 1, pp. 34–37 vol.1.

[15] Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

Authorized licensed use limited to: UNIVERSITAETSBIBL STUTTGART. Downloaded on November 19,2025 at 16:18:43 UTC from IEEE Xplore. Restrictions apply.