Comparison of the Performance of Statistical and Spectral Feature Based Models for Embedded Cough Detection Using Accelerometry Data

Maha S. Diab and Esther Rodriguez-Villegas
Wearable Technologies Lab, Department of Electrical and Electronic Engineering, Imperial College London
London, United Kingdom,
{m.diab21, e.rodriguez}@imperial.ac.uk

***Abstract*—The work presented in this paper compares the performance of different machine learning approaches, based on spectral and statistical features, in identifying coughs from accelerometry data sensed via a wearable attached to a shirt's collar. The extracted features are separately used to train and evaluate neural network models for cough detection - first as a multi-class problem, second as a binary problem. The models' performance was compared in terms of overall accuracy, cough sensitivity, specificity, and F1-score. It is concluded that the model using statistical features for a multi-class cough detection achieved the best cough sensitivity of 100% and F1-score of 0.95 with cough specificity of 97.3%. The four classification models were further evaluated for on-board performance as a TinyML cough system. They were all successfully deployed on Nordic Thing:53 using Edge impulse, and their memory requirements and estimated time per inference are reported. In terms of memory and time, the statistical- multi-class model was the smallest model occupying 13.7 KB of Flash memory and 1.1 KB of RAM; it was also the fastest model, requiring an estimated 1 ms per inference.**

Keywords—Accelerometer; cough detection; machine learning; TinyML; wearable device.

I. Introduction

Cough detection and monitoring has been one of the research areas with abundant work investigating improved detection and diagnosis algorithms [1]. This can be related to the fact that coughing is a common symptom of multiple respiratory diseases, and the detection and monitoring of its severity can provide information about diagnosis and a disease's progression [2]. Conventional methods into cough detection have mostly depended on audio recordings of cough sounds for training artificial intelligent (AI) algorithms, such as support vector machine (SVM), artificial neural networks (ANNs), and convolutional neural networks (CNNs) [1]. Some of these algorithms were successful and achieved cough detection accuracy and sensitivity above 90%, leading to their implementation and use. However, a major drawback to their use is the need for always-on microphones to record audio which is later used for cough detection on cloud servers. Aside from the cough sounds, these recordings also capture daily conversation and personal information which invade the user's privacy.

The main purpose of this research is to investigate the possibility of an alternative cough detection method, such that user's privacy is preserved. This requires a solution which can replace both the need for audio recordings, and cloud-based processing. In efforts to overcome this research challenge, we proposed the use of motion-based detection as an alternative to the conventional audio-based cough detection method; building on the hypothesis that audible sounds produced during a cough are the result of a forced expulsive motor pattern [2]. The use of accelerometer sensors for cough detection has not been investigated enough, and little insight has been found in the literature. Moreover, the use of motion-based cough detection, to the best of author's knowledge, was mostly reported using contact accelerometer with another type of sensor. In [3], a tri-axial accelerometer positioned around the epigastric region was used with a strain sensor placed around the neck for cough monitoring. This setup achieved binary cough detection using a variational autoencoder with sensitivity of 92% for both sensors and 91% for accelerometer only. In another case, a tri-axial accelerometer was used with an ECG front-end positioned around the chest [4] for a multiclass (6 classes) cough detection using custom decision tree reporting an accuracy of 96.9%. The work presented in [5] explored on-device cough detection implementing their cough detection algorithm on a Samsung Galaxy Buds2. Their detection algorithm was based on a two-stage framework using an inertia measurement unit (IMU) with a microphone in an earbud. A cough detection sensitivity of 90% was reported against stationary non-cough data and 86% against non-stationary non-cough data. As for the use of only an accelerometer for cough detection, the researchers in [6] investigated cough detection based on the position of the accelerometer. They used a tri-axial contact accelerometer at the chest, ear, stomach, and arm, and a non-contact accelerometer of a phone placed freely in a shirt's pocket. The spectral summation of the accelerometer data was used to train a CNN model which reported cough

This work was supported in part by the European Research Council (ERC) for the NOSUDEP project grant no 724334 and in part by the Engineeringand Physical Sciences Research Council (EPSRC), UK / grant agreement no. EP/P009794/1.

detection accuracy of approximately 99% for ear, 98% for chest, 97% for stomach, 95.8% for pocket, and 95.2% for arm.

Though the reported results are acceptable, these works did not explore the final implementation and deployment of the cough detection systems, except for [5] which used an earbud. Moreover, considering user's acceptability towards the use of a wearable cough detection system, ease of use and comfort are two main deciding factors. The reported systems used contact accelerometers positioned at uncomfortable and impractical positions (chest, neck, stomach). Even the use of earbuds can still be irritating over long monitoring periods and hinder the main use of earbuds or hearing-aids. As for the only presented non-contact use of accelerometer in [6], placement of phone freely in a pocket is not enough of a reliable source and can be misleading based on the position of the phone freely place in the pocket.

The main contribution of this work is to overcome the limitations of these motion-based cough detection solutions through the investigation of a wearable non-contact motion-based cough detection system with on-board processing. In this work, a wearable non-contact accelerometer placed at shirt's collar was used to collect related data for the development of multiple cough detection algorithms for comparison. The use of non-contact accelerometer data for cough detection was explored using two different sets of extracted features – statistical and spectral, under two classification scenarios – multi-class and binary classification. The comparative study further compares the performance when the models are embedded on a TinyML platform - which allows embedded machine learning algorithm to run on a microcontroller for on-board inference [7].

The rest of the paper is organised as follows: The data collection procedure highlighting the materials and method the comparison builds on is described in Section II. This is followed by a description of the data processing and the feature extraction methods used, presented in Section III. The four different classification models trained for cough detection are then covered in Section IV. Finally, the performance of the models and their deployment for on-board inference are compared and discussed in Section V.

II. Data Collection Procedure

The data was collected using an IoT prototyping platform, the Nordic Thingy:53. The Nordic Thingy:53 is a prototyping board manufactured by Nordic Semiconductor including multiple sensors, one of which is a low power tri-axial accelerometer, built around the nRF5340 SoC- a dual core Arm Cortex M33 processor. The Nordic Thingy:53 was used together with Edge Impulse, an edge AI platform, for data collection, processing, model training, and deployment for on-board inference. This combined use allows embedded machine learning (ML) algorithms to run on-device (TinyML) for edge inference and evaluation.

The Nordic Thingy:53 enclosure was clipped to a shirt's collar for data collection as shown in Fig. 1. The data used for the comparison of the models included body movements corresponding to the subject coughing, sneezing, laughing,

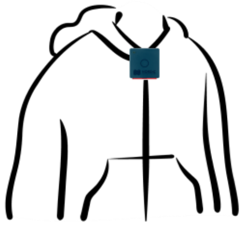

Fig. 1. A demonstrative figure showing the position of the Nordic Thingy:53 clipped to the collar during tri-axial accelerometer data collection.

talking, and resting while seated. Recordings for sneezing and laughing were relevant because of the close similarity in body movement to that of coughing, while talking and resting data were collected as the most common daily activities. The data contained ten recordings for each activity at a sampling frequency of 62.5 Hz for a duration of 10s per session. This was a total of 50 sessions (10 sessions/activity) to be used for multi-class cough detection.

As for the binary cough detection, the recordings belonging to the four classes (sneeze, laugh, talk, rest) were combined and re-labeled as 'non-cough'. This re-labeling resulted in an unbalanced dataset, which led to the collection of 30 additional cough recordings, making the total recordings used for the binary classification models reach 80 recordings. The data has been collected from a single subject but as this work's main objective was to compare models, examining the feasibility of the approaches, this was considered acceptable. Further to this work, additional data will be collected from a wider set of volunteers.

III. Data Processing and Feature Extraction

The recorded accelerometry data captured the body movements as a tri-axial signal. Movements in the horizontal direction (left-right) were represented by the x-axis signal, movements in the vertical direction (up-down) by the y-axis, and outward movement (forward/backward) by the z-axis. Visual inspection of the raw signals showed more noticeable changes in the y- and z- axes than the x-axis during a cough, a sneeze, and a laugh. This is expected as the body tends to move forward during these activities, causing movements in the vertical and outward direction of the wearable with less likely sideways movements. Moreover, it was noticed that cough events within a recording lasted for an approximate of 2s each. Building on this observation, to correctly identify the presence of a cough, the raw signals were segmented into samples of window size 2s and using a sliding window of 200ms. This resulted in a total of 2100 samples from the 50 recordings for the multi-class cough detection, and a total of 3360 samples from the 80 recordings for the binary cough detection. In both classification cases, the samples were split into 80% training and 20% testing with a balanced distribution of samples per class.

Two different feature extraction methods were used and compared in this work. Using the Edge Impulse processing block, the first method extracted statistical features using the flatten block, while the second method extracted frequency and power characteristics using the spectral analysis block. The flatten block was used to compute the following 5 statistical features per axis per window: average value for the window; minimum value in the window; maximum value in the window; root-mean square (RMS); and standard deviation of the window; resulting in a total of 15 input statistical features. While the spectral analysis block was used to calculate the Fast Fourier Transform (FFT) of the signal, after normalisation by subtracting the mean, using an FFT length of 64. This resulted in a total of 111 features, 37 features per axis where 32 features are the log of the spectrum, while the remaining 5 features are the root-mean square (RMS), skewness, kurtosis, spectral skewness, and spectral kurtosis.

IV. Classification Models

Statistical and spectral features were extracted to train and evaluate multiple classification models, for the sake of comparing them. In this section, the best performing models using both feature extraction methods are reported under their respective subsections, under which each case further presents two models targeting cough detection as a multi-class problem in the first model, and as a binary problem in the second model.

A. Stastistical Features

Using the 15 statistical features extracted from the accelerometry data, two neural network models are presented here. The first model was trained to detect cough among a total of 5 classes (Model T1), while the second model was trained to differentiate between cough and non-cough samples (Model T2). The neural network for Model T1 was trained using a single hidden layer, having an architecture of 15 – 10 – 5. This model was trained for 100 epochs under a learning rate of 0.005 using 20% of training windows for validation. The model achieved an overall accuracy of 92.6% and a sensitivity of 89.3% for cough on the validation set. Evaluation on the test set resulted in an overall test accuracy of 90.24%. The resulting confusion matrix on the test set is given in Table I. The second model, Model T2, was trained using the same 15 statistical features to classify the input samples as either cough or non-cough. The neural network was trained for 100 epochs with a learning rate of 0.005 having an architecture of two hidden layers (15 – 10 – 5 – 2). The resulting validation accuracy was 93.5%, and the test accuracy was 86.6%. The confusion matrix on the test set is given in Table II.

B. Spectral Features

Similarly, to the classification models using the statistical features, two classification models are evaluated using the extracted spectral features. The first model for multi-class cough classification (Model F1), was trained using three hidden layers with an architecture of 111 – 60 – 40 – 20 – 5. The model was trained for 100 epochs under a learning rate of 0.01, achieving an accuracy of 96.4% and a cough sensitivity of 98.6% on the validation set. The evaluation of the model on the test set resulted in an overall accuracy of 90.71% with the resulting confusion matrix given in Table III.

The second model using the spectral features targeted cough classification as a binary problem, differentiating between input samples as cough or non-cough. The model was trained using two hidden layers with an architecture of 111 – 50 – 25 – 2, for 100 epochs using learning rate of 0.005. The resulting validation accuracy was 97.80% and the test accuracy was 92.41%. The confusion matrix evaluating the model's performance on the test set is given in Table IV.

TABLE I. Confusion Matrix For Model T1 Multi-Class Classification On The Test Set Using The Statistical Features.

A / P	Cough	Laugh	Rest	Sneeze	Talk	U
Cough	84	0	0	0	0	0
Laugh	0	81	0	0	0	3
Rest	2	0	78	0	4	0
Sneeze	6	15	0	61	0	2
Talk	1	0	0	0	75	8
F1 Score	0.95	0.90	0.96	0.84	0.92	

A: actual; P: predicted; U: uncertainty

TABLE II. Confusion Matrix For Model T2 Binary Classification On The Test Set Using The Statistical Features.

Actual / Predicted	Cough	Non-Cough	Uncertainty
Cough	316	9	11
Non-cough	58	266	12
F1 Score	0.89	0.87	

TABLE III. Confusion Matrix For Model F1 Mullti-Class Classification On The Test Set Using The Spectral Features.

A / P	Cough	Laugh	Rest	Sneeze	Talk	U
Cough	75	4	0	2	1	2
Laugh	1	78	0	2	0	3
Rest	0	0	76	0	7	1
Sneeze	1	4	0	76	1	2
Talk	3	0	3	0	76	2
F1 Score	0.91	0.92	0.93	0.93	0.90	

A: actual; P: predicted; U: uncertainty

TABLE IV. Confusion Matrix For Model F2 Binary Classification On The Test Set Using The Spectral Features.

Actual / Predicted	Cough	Non-Cough	Uncertainty
Cough	335	1	0
Non-cough	36	286	14
F1 Score	0.95	0.92	

V. Results and Discussion

The objective of this work was to compare the performance of statistical and spectral features-based models under two different classification scenarios - multi-class and binary - in the context of identifying coughs using accelerometry data, rather than the more conventionally used audio. In terms of model's overall accuracy, the use of spectral features outperformed the use of statistical features, where the maximum accuracy achieved was 92.4% by the spectral-binary - Model F2, and the worst accuracy achieved was (86.60%) by the statistical- binary - Model T2. The models' performance based on overall accuracy follow the trend F2 (92.4%) > F1 (90.71%) > T1 (90.24%) > T2 (86.60%).

The use of overall accuracy to compare the performance of models in cough detection is not a reliable measure for a fair comparison between multi-class models and binary models. Therefore, three other performance metrics were calculated to evaluate a model's performance in detecting cough event, our main objective. These metrics are sensitivity, specificity, and F1 score for the 'cough' class of the model. A summary of the calculated metrics for the four models is presented in Fig.2.

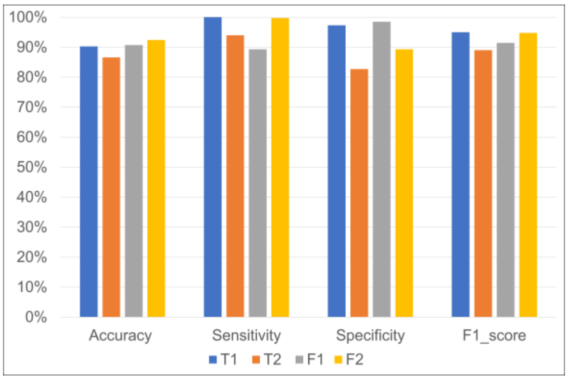

Fig. 2. Performance comparison of the four trained models on test set based on model's accuracy, cough sensitivity specificity and F1 score.

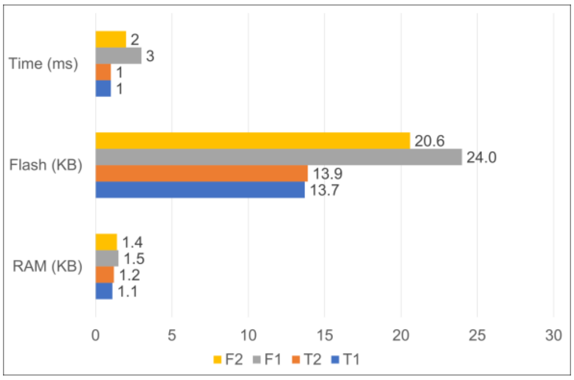

Fig. 3. Summary of edge performance for the four quantized models deployed on Nordic Thingy:53 comparing memory and time per inference.

In terms of cough sensitivity, the two top performing models achieved a very close measure of sensitivity of 100% and 99.7% for Model T1 and Model F2 respectively. This is followed by a sensitivity of 94% achieved by Model T2 and 89.29% by Model F1. As for cough specificity, it is reflected that the spectral models outperform statistical models under either classification scenarios - multi or binary - and that the multi-class models perform better than the binary models. This results in the specificity trend of F1 (98.51%) > T1 (97.30%) > F2 (89.29%) > T2 (82.70%). In terms of F1-score, both Model T1 and Model F2 take the lead, similar to the sensitivity, achieving an equal score of 0.95, followed by Model F1 at 0.91, the Model T2 at 0.89. In general, it is noticed that under the case of multi-class cough detection, the use of statistical features outperforms the use of spectral features in terms of sensitivity and F1-score. While under the case of binary cough detection, the use of spectral features performs better. Therefore, for the use of statistical features in cough detection it is best to use the multi-class model (Model T1), while for the use of spectral features, it is best to use the binary classification model (Model F2).

As all four models presented acceptable results, all four models were further evaluated based on their performance when running on the microcontroller for comparison. The models were deployed separately on Nordic Thingy:53 using Edge Impulse for on-board estimation of model performance and hardware requirement. The Nordic Thingy:53 has 1 MB of available Flash memory, and 512 KB of RAM for the application core Cortex-M33 processor running at 128 MHz. For each of the deployed models, the occupied RAM and Flash memory as well as the time estimated per inference summarised in Fig.3. Overall, it is clearly noticed that in terms of hardware requirement the statistical-based models used less resources and were able compute an inference faster. This is expected as the number of input features and model size of the statistical models are smaller in comparison to that of the spectral models. In terms of RAM, all models required around the same amount of memory ranging between 1.1 KB (Model T1) and 1.5 KB (Model F1). As for the Flash memory, the smallest model was Model T1 (13.7KB) followed by Model T2 with a difference of just 0.2 KB. However, the spectral models required almost double the amount of memory. Where multi-class model (Model F1) occupied 24.0 KB of Flash (1.75 more than the smallest model); and Model F2 occupied 20.6 KB of Flash. In terms of inference time, in comparison to that of the statistical models, Model F1 required three times the amount of time, and Model F2 double the time for an inference.

As mentioned earlier, in terms of cough detection performance, the two best models were Model T1 and Model F2. This is also reflected in the deployment of model for edge inference, where Model T1 required less memory than Model T2 for the use of statistical features; and Model F2 occupied less memory and was faster than Model F1 for the use of spectral features. In terms of model's performance after quantization both Model T1 and Model F2 maintained their sensitivity (100% and 99.7%) and F1-score (0.95). Therefore, it can be concluded that either models can be deployed for the cough TinyML system with bias towards the use of Model T1 which is two times faster and occupies 6.9 KB less Flash and 0.3 KB less RAM.

References

[1] A. Ijaz, *et al.*, "Towards using cough for respiratory disease diagnosis by leveraging artificial intelligence: A survey," Informatics in Medicine Unlocked, vol. 29, p. 100832, 2022.

[2] K. K. Lee, *et al.*, "Global physiology and pathophysiology of cough: part 1: cough phenomenology–chest guideline and expert panel report," Chest, vol. 159, no. 1, pp. 282–293, 2021.

[3] T. Otoshi, *et al.*, "A novel automatic cough frequency monitoring system combining a triaxial accelerometer and a stretchable strain sensor," Scientific Reports, vol. 11, no. 1, p. 9973, 2021.

[4] J. Liu, *et al.*, "Activity recogni-tion in wearable ECG monitoring aided by accelerometer data," in 2018 IEEE International Symposium on Circuits and Systems (ISCAS), 2018,pp. 1–4.

[5] S. Zhang, *et al.*, "Coughtrigger: Earbuds imu based cough detection activator using an energy-efficient sensitivity-prioritized time series classifier," in ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 1–5.

[6] K. Doddabasappla and R. Vyas, "Spectral summation with machine learning analysis of tri-axial acceleration from multiple wearable points on human body for better cough detection," IEEE Sensors Letters, vol. 5, no. 9, pp. 1–4, 2021.

[7] M. S. Diab and E. Rodriguez-Villegas, "Embedded machine learning using microcontrollers in wearable and ambulatory systems for health and care applications: a review," IEEE Access, vol. 10, pp.98450 - 98474, 2022.