DEEPCRAFT™ Ready Model for Cough Detection

Introduction

In this document, we describe the DEEPCRAFT™ Ready Model for Cough Detection, an audio AI-based model developed by Imagimob, an Infineon Technologies company. We provide details about the technical specifications of this machine learning model, its performance in common scenarios, and various test results for the model including the real-time testing on an Infineon PSOC™ 6 board.

Contents

Model Specification

Model Overview

The DEEPCRAFT™ Ready Model for Cough Detection is designed to detect coughs from adults with the purpose to identify users' health degradation. Such a model is developed with the intent to run on a wearable device like a smart watch, a bracelet, an armband, a necklace or even a smartphone or any other non-wearable device located in the vicinity of the person. There are no limitations concerning the environment, the user can be in any location so the model has to be robust to all typical sounds found at home, in restaurant, or work environment as well as in a noisy city setting given that the environmental sound is low enough. The model will not distinguish between cough types and it will be the input of a health app that checks the user's respiratory condition.

Expected Performance

The aim of this model is to detect the user's number of coughs per hour. The model focuses on serious coughs which are defined as those with prolonged acoustic components longer than 200 ms (see sample 1 in appendix). Multiple subsequent coughs can be detected as 1 cough event.

The model is expected to detect 84% of the cough events at distances between 10 cm and 2 metres. A cough, compared to its background, needs to have a Signal To Noise Ratio (SNR) which is at least 10 decibels (dB) in order to be detected. In addition, the model is expected to misclassify other sounds/noise with a rate of 6 per hour.

Besides different distances and angles, the model is expected to be robust against different gender, ethnicities, age etc. as well as different negative or background sounds like those found in homes, workplace, city, etc., as well as sounds from people, etc.

Operations

The DEEPCRAFT™ Ready Model for Cough Detection is designed to give the user an indication of their health condition, meaning that besides detecting coughs, sneezes and throat clearing sounds may be detected as coughing and contribute to the user's health status.

Since the model is developed to work from 10 cm up to 2 meters, it may show performance degradation if a person coughs less than 10 cm and more than 2 meters from the device and/or if there is an obstacle between the person and device (such as a wall).

This model detects any cough event in its vicinity without distinguishing the person who coughed, which means that the model may detect the cough of someone else in the vicinity.

In conditions where the SNR of a cough event is too low compared to its background, the cough model may not be able to detect it. For instance, it can happen that the noise of the device when covered by clothes can be louder than a cough event making it difficult to be detected. Take, for example, the case of a wrist-worn wearable; the user may keep their hands in a pocket and/or wear thick clothes over the wearable.

It is recommended that this model should be used to look at coughing statistics and count over a longer period, such as 1 hour, to get a good understand of a person's health rather than looking at if a single cough is detected or not.

Model Tech Specs and Deployment

The DEEPCRAFT™ Ready Model for Cough Detection is able to detect a cough from sound data with the following characteristics:

- Sample rate: 16000 Hz
- Channels: 1 (Mono)
- Bit Depth: 16bit

The C version of the model has the following memory footprint:

- Flash: 81.6 kB
- RAM: 26.6 kB

The inference time is about 150 ms when running on a PSOC™ 6 (model CY8CKIT-062S2-43012) mounting a Sense shield with a microphone (model CY8CKIT-028-SENSE). The model outputs a prediction every 172 ms, namely 5.8 times per second.

Data Properties

This Ready Model was built using various positive and negative sounds. The positive sounds are coughs from different individuals occurring in different environments and at different distances. The negative data represents different kinds of sounds that can occur both indoors and outdoors. Both kinds of sounds are listed in the following sections.

Positive Data

The positive data used to build the model consists of sound recordings with 1 or more cough events per file. Both dry and wet coughs have been used. The length of such files ranges from one second to about three minutes but most of them have a length below two minutes.

In the positive data, we can distinguish two different types of coughs:

- Short, single cough: its length is about 200-300 ms and it is separated from the previous cough and next one by at least 300 ms of non-cough data
- Long, multiple cough: it consists of multiple coughs, 2 or more, that are separated from the previous cough and next one by less than 200-300 ms of non-cough data

In the Appendix II, links to YouTube examples are provided for these two types of coughs.

Included Negative Data

The model has been built using sound recordings belonging to the following non-cough or negative data categories:

Indoor sounds

- Dog sounds
- Cat sounds
- Door sounds
- Generic kitchen sounds
- Blender
- Dishes, cups, cutlery sounds
- Alarms
- Dish washing
- Dishwasher
- Hammering on different materials
- Electrical shaver
- Vacuum cleaner
- Frying
- Glass break
- Jingle sounds
- Ice cubes sounds in a glass
- Classic and electric guitar
- Running water
- Carillon
- Phone ringtones
- Airplane cabin
- Cello and Violin
- Flute and saxophone
- Piano
- Drilling
- Sawing
- Hairdryer
- Microwave
- Washing machine
- Water boiler
- Water running
- Elevator sounds
- TV and radio
- Music

Outdoor sounds

- Fireworks
- Airplane sounds
- Horse sounds
- Cow sounds
- Church bell
- Construction site sounds
- Engine sounds from various vehicles
- Inside various vehicles sounds
- Forest sounds
- Sheep and goat sounds
- Shop and supermarket sounds
- Sport events sounds
- Fountain sounds
- Rain
- Thunderstorm
- Wind
- Birds
- Car horn
- Train horn
- Ocean waves
- Wood chopping
- Sirens

People sounds

- People talking in different environments like bars, restaurants, street, shops
- Clapping
- Laughing
- Screaming
- Shouting
- Breathing
- Sighing
- Baby cry and laughing
- Child screaming
- Children playing
- Inhaling
- Exhaling
- Sneezing
- Clearing throat sounds
- Whistling
- Foot steps
- Dinner sounds
- Snoring

Other

- White noise
- Pink noise

Testing

Validation Set Results

The performance of the model on the Validation set indicates that a good balance between true positives and false positives is achieved when using a confidence threshold equal to 80%. The model's performance is summarised in the two tables below, the so-called confusion matrices. The meaning of the percentages/values reported in both of them is as follows:

- Top Left Value (True Negatives): actual negative/non-cough data predicted as negative/non-cough data

- Bottom Left Value (False Positives): actual negative/non-cough data predicted as positives/cough data
- Top Right Value (False Negatives): actual positive/cough data predicted as negative/non-cough data
- Bottom Right Value (True Positives): actual positive/cough data predicted as positive/cough data

The sample-based confusion matrix is obtained by splitting the data in short recordings of about 300 ms. Even if this is a more accurate representation of the model's performance since it looks at samples short in time, it can overestimate False Negatives. In particular, this means that the True Positives percentage, equal to 69%, in the table below sets roughly a lower limit of the real True Positives of this model.

Samples-Based Confusion Matrix	**actual non-cough**	**actual cough**
predicted non-cough	99.68 %	31.05 %
predicted cough	0.32 %	68.95 %

However, in this case, another way to interpret the high value for the False Negatives is that the model may miss single, short cough events which are typically about 200-300 ms but perform better on multiple coughs (check the Data Properties section and Appendix II for the definition of these two types of coughs). This is confirmed by a more accurate analysis of the results.

On the other hand, the following confusion matrix looks at the coughs for a longer period of time and looking at if any coughs are detected within this timespan. This shows if at least 1 cough event is detected in cough data ranging from 10 seconds to 3 minutes in length with most being below 2 minutes.

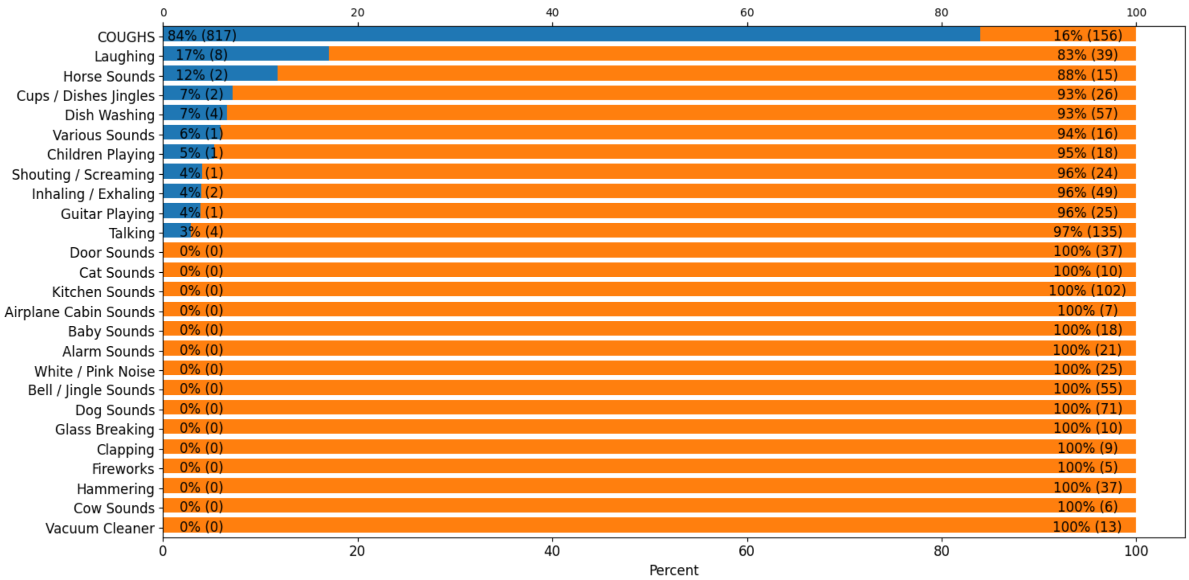
Files-Based Confusion Matrix	**actual non-cough**	**actual cough**
predicted non-cough	96.96 % (830 files)	16.03 % (156 files)
predicted cough	3.04 % (26 files)	83.97 % (817 files)

We see that the True Positives percentage is about 84%.

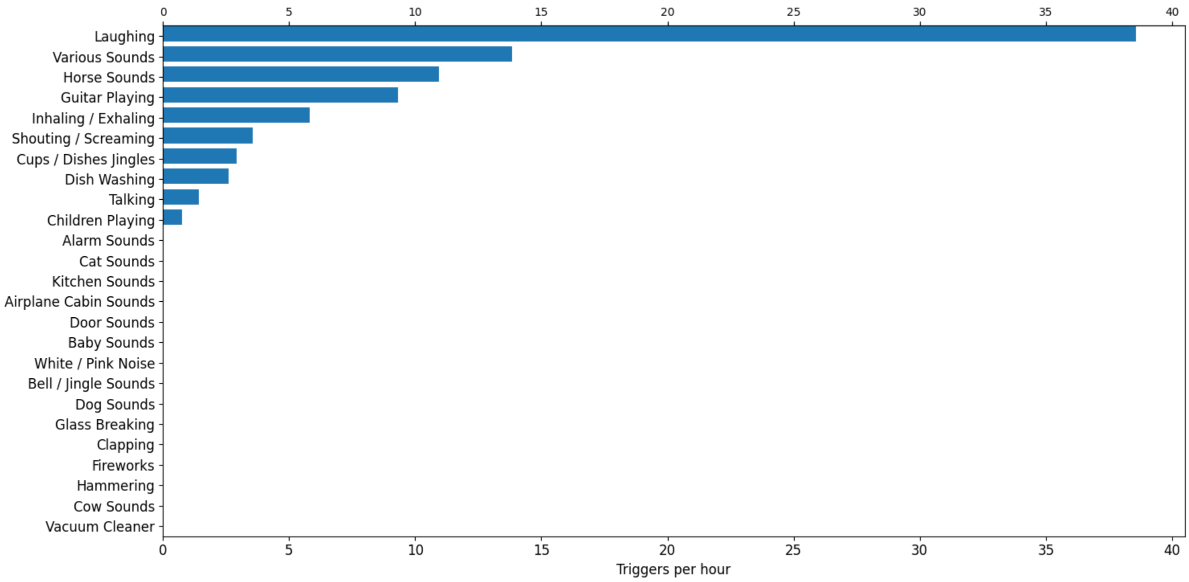
From the files-based confusion matrix, we see also that the False Positives occur in the 3% of the non-cough files and they are caused by the following sound categories:

- Laughing
- Screaming / shouting
- Inhaling / exhaling
- Talking
- Horse neighing
- Dish / cups / cutlery sounds
- Guitar playing
- Children playing

All other categories included in the training (see the list in the previous section) do not trigger any cough. The complete list of false positives per category is shown in the histogram plot below. There the blue bar and the values on the left side refer to the number of files in each category where at least one cough event is detected. The orange bar and the values on the right side indicates in how many files no coughs are detected.



From this histogram, it is clear that the Laughing category contributes the most. To understand better the impact of the False Positives on each hour when the model is running, we refer to the histogram plot below showing the False Positives per hour per category.



We see that the Laughing category is expected to have the biggest rate per hour, while the impact of the other categories is much less. Notice that the numbers of False Positives per hour reported in the histogram above are an overestimate of the actual values since they do not represent a real case scenario.

Test Results on 25 People’s Recordings

We tested the model on data from 25 people that were not part of the training. We recorded cough events at distances ranging from 10 cm to 2 m and using both a phone and the PSOC™ 6. We report the results of this testing in the table in the Appendix III.

The recall value for this test turned out to be 77.5% using a 80% confidence threshold as the only post processing. Not all participants of this testing were sick/ill, only 8 people actually had a cough. The rest of the participants “faked” the cough as close as the way they usually do.

A part of the recorded data were non-cough sounds belonging to the following categories:

- Laughing
- Talking
- Playing piano
- Typing on laptop
- Kitchen sounds / washing dishes
- Tv playing
- Music playing
- Dryer running
- Whistling
- Sighing
- Dishwasher

Only one false positive triggered by people talking on a total of 442 seconds of non-cough data which means eight false positives per hour.

On-device testing

To test the model, proceed in the following steps:
Loading a model using the hex file for the [PSOC™ 62S2 Wi-Fi BT Pioneer Kit]:

1. Flash the board with the hex file
2. Open a serial terminal to observe the prints. Terminal settings:
    a. COM port is dependent on the computer being used, check device manager to find the port number
    b. Speed: 115200
    c. Data: 8 bit
    d. Parity: none
    e. Stop bits: 1
    f. Flow control: none

Loading model using the static library:

1. Load your target firmware whether this is a code example or target firmware. In the case of code example, we recommend a basic PDM/PCM Audio example
2. Add the provided static gcc library to your project
3. Use the API and code examples provided in the header file
4. Trigger a UI event based on the flag raised by the library. I.e. printf statement that the event occurred

Testing the model:

1. Place the device 0.5 meters away from a person
2. Produce a long cough as described in the Positive Data section
3. Prediction is generated:
4. If using custom firmware: check whatever output you have created yourself
5. If using hex file: check the terminal for prints

Test Results

We performed on-device testing with 6 of the 25 people involved in the testing discussed in the previous section. For this test, we deployed the model on a PSOC™ 6 (model CY8CKIT-062S2-43012) mounting a Sense shield with a microphone (model CY8CKIT-028-SENSE) and set the confidence threshold to 80%. We summarise the results in the table below. We provide the definition of multiple and single coughs in the Data Properties section. In Appendix II, we list examples from YouTube.

People	True Positives	False Negatives	False positives
Person 1	Loud coughs	Weak and short coughs and when music is playing in background	0 on 1 hour of music playing, kitchen sounds, walking
Person 2	Loud / multiple coughs	Short / single coughs and coughs further than 2 m	0 on 2.5 hours of talking on phone, minimal talking, phone message tones, music from phone, TV, sneeze 4 on 1 hour of tea party with 12 people talking and laughing, plates and cups sounds, dog barking 0 on 15 min of baby fussing and people talking
Person 3	1 short / single cough detected from less than 1 m	6 short / single coughs missed from less than 1 m	0 on 25 min of talking at phone with speaker ON and kids talking and singing 0 on 1 hour of family dinner with 2 adults and 2 kids with various kitchen sounds and people talking
Person 4	Loud and/or multiple coughs, especially at less than 1 m	Weak and short coughs and/or close to and further than 2 m and/or certain types of cough	0 on 1 hour in office environment 0 on some minutes of clapping, laughing, music playing from phone at 20 cm 0 on 30 min of home sounds
Person 5	Loud and/or multiple coughs with no pause in between, especially	Quick single coughs with pause in between and/or close to and further	0 on 10 minutes of clapping, laughing, growling, talking and circular sawing,

	at less than 1 m	than 2 m	drilling and leaf blowing coming from a speaker
Person 6	Loud coughs and/or in a small room	Weak and/or muffled coughs and/or in big room, even with multiple coughs	Negative testing was not performed

This test with the cough model running on-device shows that it is more likely to detect multiple coughs than short, single coughs. The detection increases if the cough is louder and/or closer to the device. This means that this model is more suitable for the detection of serious coughing conditions.

We notice also that the model is robust against False Positives in different real case scenarios. During this test, the model triggered only 4 in about 7 hours of non-cough sounds, namely 0.6 triggers per hour.

Appendix I - Data Sources

The positive data have been downloaded from the following sources:

- Freesound - https://freesound.org/
- CoughVid - https://zenodo.org/record/7024894
- Coswara - https://github.com/iiscleap/Coswara-Data/tree/master

Whereas Freesound is a well established sound database website, CoughVid and Coswara are two crowdsourcing projects, including sound files with cough events of thousands of users around the world, aimed at developing AI models to detect if a person is sick with Covid-19 disease.

The negative data has been downloaded from the following sources:

- Freesound - https://freesound.org/
- DESED - https://project.inria.fr/desed/

Appendix II - Cough Sound Examples

Below we provide a list of sound examples for the two different cough types:

Short, single cough:
Man Coughing Sound Effect Part 2
AWWGHGHAGHAHH (Cough sound effect)
Types of Coughs in 60 Sec (2nd and 4th way of coughing)

Long, multiple cough:
Man Coughing Sound Effect
Types of Coughs in 60 Sec (1st and 3rd way of coughing)

Appendix III - Results of Model Testing on 25 People's Recordings

Person	10-15 cm	50-70 cm	1 m	2 m	False Positives	Actual Coughs
1	6			12	0	21
2			6	11	0	21
3		6		5	0	21
4		6		12	0	22
5		7		9	0	21
6			4	3	0	9
6 (tv playing)					0	
7 (sick)		4		5	0	9
8		10		11	0	22
9 (sick)	12			13	0	25
10		3		4	0	11
11			9	6	0	20
12 (sick)			3		0	5

13			4	3	0	7
14		10		9	0	20
14 (typing on laptop)					0	
14 (kitchen sounds, washing dishes)					0	
15 (sick)			12	15	0	30
16 (sick)			4	5	0	9
16 (dryer running)					0	
17	5			9	0	15
17 (playing piano, singing, sighing, whistling)					0	
18		4		5	0	9
19	5			7	0	12
20		1		1	0	9
21	2			2	0	7
22 (sick)			5		0	5
23 (sick)			7		0	10
24 (sick)			3		0	5
25	2			3	0	29
25 (music)					0	
25 (dish-washer)					0	
25 (tv, talking, laughing)					1 (inhaling/ laughing)	
TOTAL	32	51	57	150	1	374

The average True Positives rate is given by 290/374 = 0.775, so 77.5% of coughs have been detected in this test. Notice that the overall performance is good for all the people who participated except for person 20 and especially person 25.

Only 1 false positive has been detected on 442 seconds of non-cough data.