DEEPCRAFT™ Ready Model for Snore Detection

Table of Contents

- Introduction
- Model Specification
  - Model Overview
  - Expected Performance
  - Operations
  - Model Tech Specs and Deployment
- Data Set
  - Positive Data
  - Negative Data
- Model Evaluation
  - Validation Set Results
  - Test Set Results
  - Test Results on 20 Participants' Recordings
  - On-Device Test Results
  - Summary
- Appendix I - Data Sources
- Appendix II - Detailed Test Results on 20 Participants' Recordings
- Appendix III - Memory Footprint Calculation

Introduction

In this document, we describe the DEEPCRAFT™ Ready Model for Snore Detection, an audio-based AI model developed by Imagimob, an Infineon Technologies company, that detects when a person is snoring. 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.

Please note that this model is not certified for medical use. It must not be used as a medical device or as a substitute for professional medical advice, diagnosis or treatment.

Model Specification

Model Overview

The DEEPCRAFT™ Ready Model for Snore Detection is designed to detect snoring when the user is sleeping, in different kinds of typical sleeping environments.

Expected Performance

The aim of this model is to detect snoring while the user is sleeping. The model does not detect every single snore; a snoring period counts as detected if at least one snore within it is detected.

The model is expected to detect snoring in 92% of complete test-set files at distances of up to 2 meters. In the separate recording-based evaluation, five two-minute segments were randomly selected from each participant's full sleep recording, and at least one segment per participant contained snoring. A participant was considered successfully classified when the model detected at least one snore in every selected segment that contained snoring. By this criterion, the model successfully classified 16 of 20 participants (80%). The model is expected to detect more snoring over a prolonged period and to be robust against typical sleeping-environment backgrounds such as TV sounds and fan or air-conditioning noise. A snoring event, compared to its background, needs to have a Signal To Noise Ratio (SNR) of at least 10 decibels (dB) in order to be detected.

See the Test Set Results and Test Results on 20 Participants' Recordings sections for details about these evaluations, with detailed results provided in Appendix II - Detailed Test Results on 20 Participants' Recordings.

Operations

The model detects any snoring in its vicinity without distinguishing the person who is snoring, which means that it may detect the snoring of someone else nearby.

If the person is further away than 2 meters or there is an obstacle between the person and the device, performance may degrade.

In conditions where the SNR of a snoring event is too low compared to its background, the model may not be able to detect it. For instance, loud background noise from a TV, fan, or nearby conversation can be louder than the snoring sound and make detection harder.

The model does not detect every single snore, especially at the beginning of a snoring period. Loud talking and loud music, for example from a TV or speaker, may trigger false positives. It is recommended to use this model to detect snoring periods over a longer timespan rather than to count every single snore.

Model Tech Specs and Deployment

The DEEPCRAFT™ Ready Model for Snore Detection processes sound data with the following characteristics:

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

Target board	Model version	library - Flash (KiB)	library - RAM (KiB)	ml-middleware - Flash (KiB)	ml-middleware - RAM (KiB)	tflite-micro - Flash (KiB)	tflite-micro - RAM (KiB)
Any	ANSI C99	53.07	25.98	N/A	N/A	N/A	N/A
PSOC™ 6	float32	61.71	108.21	3.14	4.52	259.69	0.09

Target board	Model version	library - Flash (KiB)	library - RAM (KiB)	ml-middleware - Flash (KiB)	ml-middleware - RAM (KiB)	tflite-micro - Flash (KiB)	tflite-micro - RAM (KiB)
PSOC™ 6	int8x8	N/A	N/A	N/A	N/A	N/A	N/A
PSOC™ Edge	float32	99.58	108.21	4.27	4.62	264.32	0.12
PSOC™ Edge	int8x8	N/A	N/A	N/A	N/A	N/A	N/A

Memory-footprint calculation details are provided in Appendix III - Memory Footprint Calculation.

For on-device testing and inference time calculation, the model uses generic C99 code containing the float32 model implementation and its preprocessing. It does not require a separate ML runtime. The inference time is about 23 ms when running on a PSOC™ 6 (model CY8CKIT-062S2-43012) overclocked to 150 MHz and mounting an IoT Sense expansion kit (model CY8CKIT-028-SENSE). The model outputs a prediction every 60 ms.

Data Set

The DEEPCRAFT™ Ready Model has been built using various positive and negative sounds. The positive sounds are different kinds of snoring from different individuals occurring in different indoor environments. The negative data represents different kinds of sounds that could happen indoors. The sounds are listed in the next sections.

Positive Data

The positive data consists of different types of snoring from individuals around the world, at different distances and ages. Each audio file contains one or more snoring events.

We also augmented the data with different background noises, such as white noise, and simulated different distances. Background sounds present in the recordings include, for example:

- TV sounds
- Fan/air conditioning
- Ambient sounds

Negative Data

Many different negative sounds were used in building and evaluating the model to ensure that it performs well in different environments. The following is a representative, non-exhaustive list of prominent negative sound categories validated against:

- Adult laughing
- Adult talking

- Alarms
- Baby sounds
- Blender
- Cat
- Clapping
- Crowd
- Dish washing
- Dishes, cups, cutlery
- Dishwasher
- Dog
- Door
- Electrical shaver
- Fan
- Frying food
- Generic kitchen
- Glass break
- Music
- Pig
- Running water
- Sneezing
- Traffic
- TV/radio
- Vacuum cleaner

Model Evaluation

In this section we present the evaluation of the model. The model was evaluated in four ways, each described in one of the following subsections:

1. Validation set results: performance on the validation set used during model development.
2. Test set results: performance on a separate test set not used during training.
3. Test results on 20 participants' recordings: performance on sleep recordings from 20 participants.
4. On-device test results: real-time performance with the model deployed on a PSOC™ 6 board.

We evaluate model performance at two levels: file level and sliding-window level. At file level, each sound file counts as one result: a positive file is classified as detected if the model triggers at least once anywhere in the file, and a negative file is classified as a false positive if the model triggers at least once anywhere in the file. At sliding-window level, each model input window and resulting prediction counts as one result.

Validation Set Results

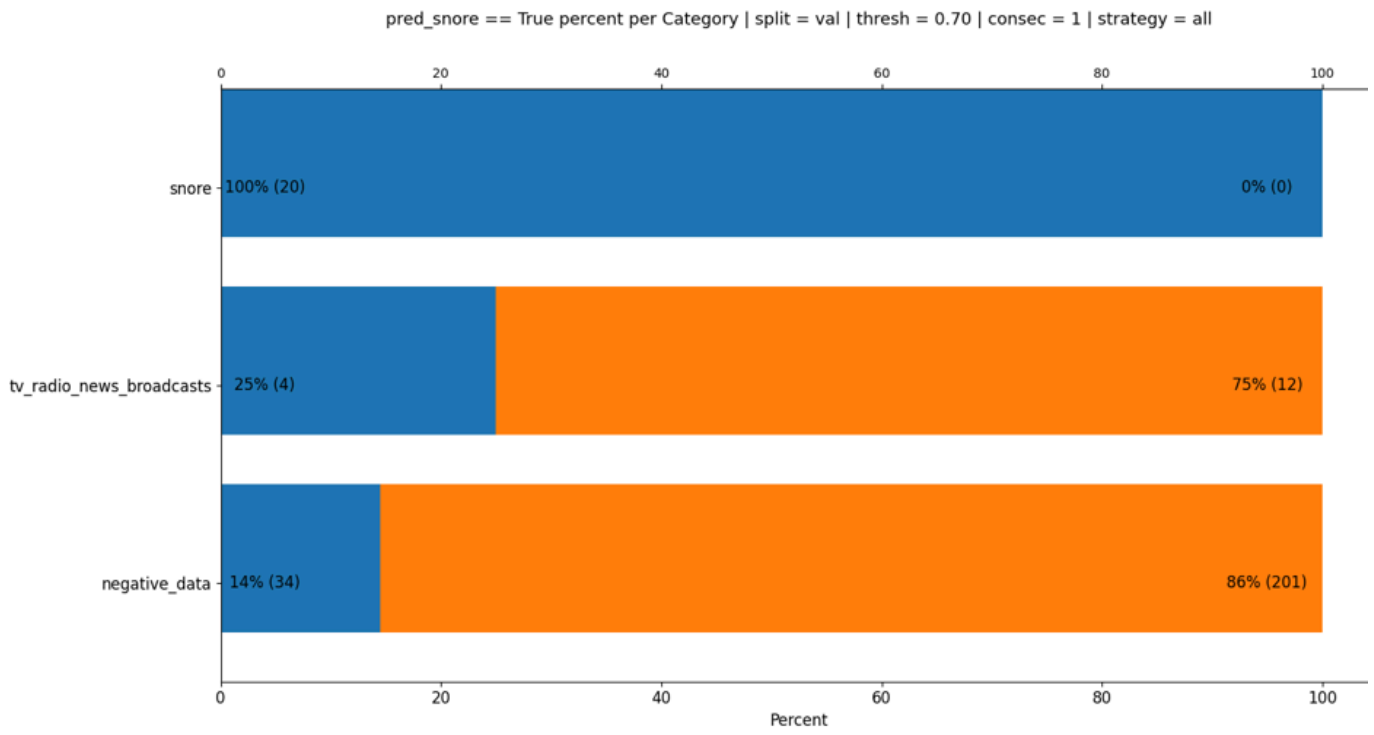


Figure 1: Validation set predictions per category at file level.

The plot above shows the predictions on the validation set at file level.

As we can see from the plot, the model predicts all the snore files correctly. However, there are some false positives on the negative data, including TV sounds. The 34 false positives among the negative data are mainly alarms, angry cat noises, electrical shaver and loud people talking sounds.

The confusion matrix below summarises the performance of the model on the validation set, as reported by DEEPCRAFT™ Studio. The percentages are normalised within each actual class column, so the diagonal values are recall values for the corresponding class:

- Top Left Value (Negative Recall / True-Negative Rate): actual negative data predicted as negative
- Bottom Left Value (False-Positive Rate): actual negative data predicted as positive
- Top Right Value (False-Negative Rate): actual positive data predicted as negative
- Bottom Right Value (Positive Recall / True-Positive Rate): actual positive data predicted as positive

Confusion Matrix (Validation Set)	**Actual negative**	**Actual snore**
Predicted negative	97.04 %	4.42 %
Predicted snore	2.96 %	95.58 %

Table 1: Confusion matrix of the validation set from DEEPCRAFT™ Studio.

This confusion matrix uses the sliding-window-level results. The model correctly classified about 97% of the sliding-window results in the negative data as true negatives. For the snoring data, it correctly detected about

96% of the sliding-window results.

Test Set Results

In addition to the validation set, the model was evaluated offline on a separate test set that was not used during training. The test set contains both snoring recordings and negative home sounds.

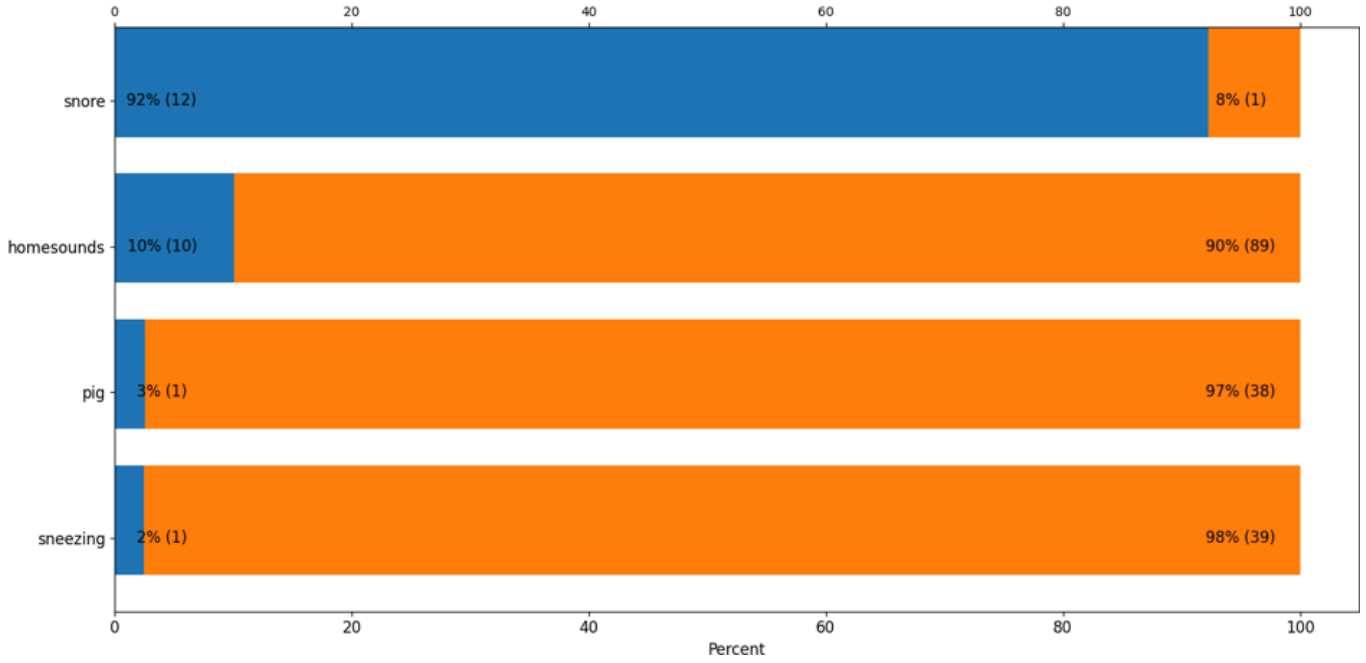


Figure 2: Test set predictions per category at file level.

As we can see from the predictions on the test set, the model detects snoring in 92% of complete snore files, where a file counts as detected if at least one snore in it is detected. The 10 false positives among home sounds are mainly sounds of a crowded room of people having conversations.

Test Results on 20 Participants' Recordings

The model was evaluated offline on sleep recordings from 20 participants. The detailed results are shown in Appendix II. Using the participant-level criterion defined in Expected Performance, the model successfully classified 16 of 20 participants (80%). The model behaved well with noisy backgrounds but underperformed for a few participants.

On-Device Test Results

We performed on-device testing with 4 people in total, recorded during the night while they were sleeping. For these tests, the model was deployed on a PSOC™ 6 (model CY8CKIT-062S2-43012) mounting an IoT Sense expansion kit (model CY8CKIT-028-SENSE) and used a 70% confidence threshold as the only post-processing. Two of them snored during the recorded time; the remaining two did not snore, providing sleeping sounds without snoring as negative data. Post-processing used a confidence threshold only. The testing results are shown below.

We can see that the model detects snoring well and does not react to normal sleeping data. However, there are false positives on loud music and TV sounds.

Subject	Duration	Sounds Tested	Snoring Detection	False Positives
Subject 1	Around 8 hours (live)	Snoring, people talking, partially TV background	Snoring detected well	High number of false positives on people talking with TV background
Subject 2	Around 6 hours (live)	Mainly quiet sleeping sounds without any snoring	No snoring present (negative test)	Zero
Subject 3	7 hours (board recording)	Sleeping sounds without snoring, door opening and closing a few times, possibly some very silent breathing	No snoring present (negative test)	Three
Subject 4	8 hours (board recording)	Sleeping sounds with snoring, partially loud TV sound	Snoring detected well, though not every snore is detected, especially at the beginning of the snoring period	Zero; the model does not react to the TV sounds

Table 2: On-device test results.

Summary

The evaluation shows that the model reliably detects snoring while remaining robust against most negative sounds:

- On the validation set, the model detected all snore files and correctly classified 97% of the negative timesteps.
- On the test set, the model detected snoring in 92% of complete snore files.
- On recordings from 20 participants, the model successfully classified 80% of participants under the criterion described in Expected Performance; it struggled with insect sounds in the background and low-SNR snoring/ heavy breathing.
- On-device, snoring was detected well; false positives occurred mainly on loud talking and loud music played from speakers.

The model does not detect every single snore, especially at the beginning of a snoring period. It is therefore recommended to consider a snoring period detected if at least one detection is triggered during the period.

Appendix I - Data Sources

The positive data has one or more snoring events per file, with sound file lengths ranging from 15 seconds and up. These files have been downloaded from the following sources:

- Freesound

The negative data has been downloaded from the following sources:

- Freesound
- DESED

The datasets are licensed under a combination of CC0 1.0, CC BY 3.0, and CC BY 4.0.

Appendix II - Detailed Test Results on 20 Participants' Recordings

Participant	Result
30	Does not detect short snoring, but detects all strong snoring
17	Detects all snoring at file level
34	2 FPs within 10 minutes of background noise without any snoring. Detects the snoring in this noisy background in another 10 minute file
35	Detects most snoring and detects all snoring at file level
19	Does not detect any snoring for this participant; cricket sounds in the background
6	Detects most snoring and detects all snoring at file level
7	Detects most snoring and detects all snoring at file level
12	Detects most snoring and detects all snoring at file level
13	Does not detect all of the snoring but performs well at file level
11	Does not detect any snoring for this participant, the same case as participant 19; cricket sounds in the background
1	Detects most snoring and detects all snoring at file level
14	Does not detect any snoring for this participant, the same case as participant 19; cricket sounds in the background
16	Very noisy background; does not detect all of the snoring but performs well at file level
15	Detects most snoring and detects all snoring at file level
48	Detects all snoring at file level
49	Detects all snoring at file level
40	Detects most snoring and detects all snoring at file level
41	Detects most snoring and detects all snoring at file level
42	Detects most snoring and detects all snoring at file level
3	Does not detect light snoring/ heavy breathing, 0 FS on the non-snore data containing sleep talking, phone ringing

Table 3: Detailed test results on 20 participants' recordings.

In summary, the model detected snoring for 16 of the 20 listed participants. The three participants without any detections (IDs 19, 11 and 14) all had cricket sounds in the background. The model also struggles with participant 3's snore, which has light snoring with low SNR. For participants with noisy backgrounds (IDs 34 and 16) the model still detected the snoring at file level.

Appendix III - Memory Footprint Calculation

All memory values in the deployment table are KiB, where 1 KiB = 1024 bytes. Memory usage was measured by building the audio deployment examples with the GCC_ARM 14.2.1 compiler and its Debug configuration flags, then analyzing the generated linker map, as the examples existed on September 4, 2026:

- PSOC™ Edge deploy-audio
- PSOC™ 6 deploy-audio

The calculation parses the `Linker script and memory map` portion of the GNU linker map. Each input section is attributed to the generated model object, TensorFlow Lite for Microcontrollers (`libtensorflow-microlite.a`), or ML Middleware (`libraries_shared/ml-middleware`). The size reported for each input section is then added to the Flash and/or RAM total for its component according to its enclosing output section.

For PSOC™ 6, Flash is the sum of `.text + .rodata + .data`, and RAM is the sum of `.data + .bss + .noinit`. For PSOC™ Edge, Flash is the sum of `.app_code_main + .app_code_itcm + .app_code_socmem + .ram_vectors + .cy_socmem_data + .data`, and RAM is the sum of `.app_code_itcm + .app_code_socmem + .ram_vectors + .cy_socmem_data + .data + .bss + .noinit + .cy_sharedmem`. Sections that are stored in Flash and executed or initialized in RAM are counted in both totals. Consequently, these results describe the linked deployment example rather than a platform-independent measurement of the model alone.

The PSOC™ Edge Flash figures are currently overestimated because of a known bug in the code examples. A fix is work in progress, so the PSOC™ Edge Flash values should be treated as provisional until the corrected examples are available.