DEEPCRAFT™ Ready Model for Baby Cry Detection

Introduction

In this document, we describe the DEEPCRAFT™ Ready Model for Baby Cry Detection, an audio-based AI model developed by Imagimob, an Infineon Technologies company, that detects when there is a baby or young child crying. 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

imagimob

Model Specification

Model Overview

The DEEPCRAFT™ Ready Model for Baby Cry Detection is designed to detect crying in babies and young children between 0 to 4 years old. This model can be used in a wearable device or smart baby product to alert the parents of a crying or active baby. The model is designed and tested to work up to 5 meters after which the performance will start to deteriorate.

Model Tech Specs

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

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

Model Deployment on a PSOC™ 6 Board

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

- RAM: 27KB
- FLASH: 26.5KB

And its inference time is about 127ms 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 516 ms.

Data Properties

The DEEPCRAFT™ Ready Model has been built using various positive and negative sounds. The positive sounds are child cries 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 consist of different types of baby cry in different backgrounds. And these cries are from different genders and different ages.

Some of the data used was guaranteed to be from different genders, different origins and ages between zero to four. We randomly chose 20 children and tested on their crying and non-crying sounds to make sure that the model generalised to different kinds of child sounds.

Included Negative Data

To ensure that the model does not trigger false positives, the model has been built using sound recordings belonging to the following non-baby cry or negative sounds that from indoor and outdoor categories:

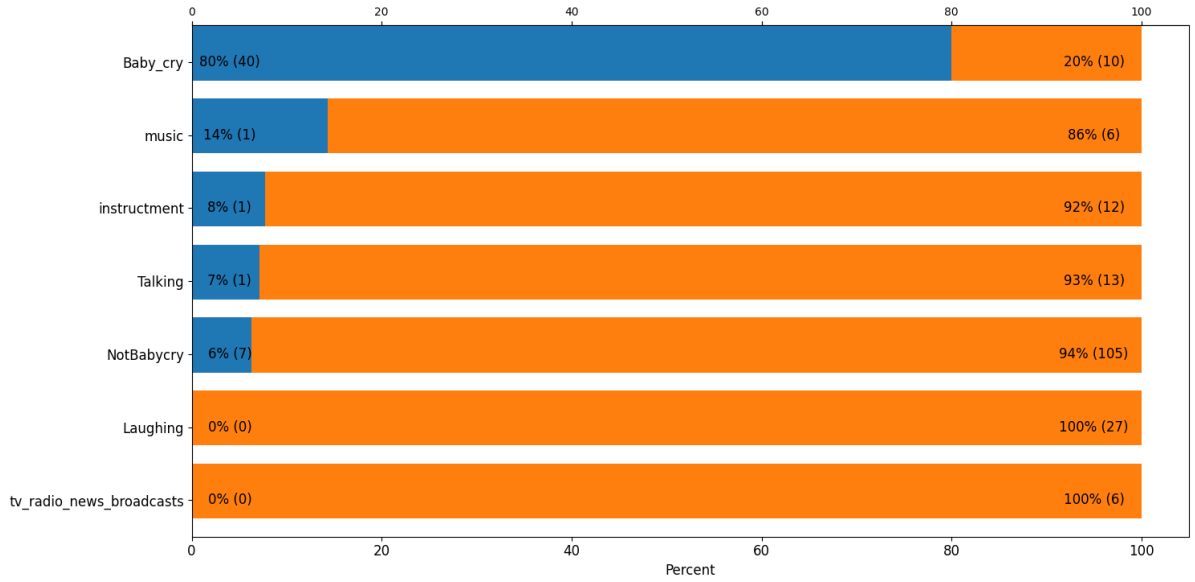
Negative sounds

- Cat sounds
- Door sounds
- Generic kitchen sounds
- Blender
- Dishes, cups, cutlery sounds
- Alarms
- Dish washing
- Other child sounds : babbling, laughing, talking and whining.
- Alarm
- Dog sounds
- Adult talking
- Dishwasher
- Electrical shaver
- Vacuum cleaner
- Frying food
- Glass break
- Different instruments
- Music
- TV/ radio
- Adult laughing
- Shower
- Firework
- Microwave
- Hammering
- Running water
- Toothbrush
- Washing machine
- Traffic
- Drilling
- Sawing
- Hairdryer
- Water boiler
- Computer
- Bottle Opening

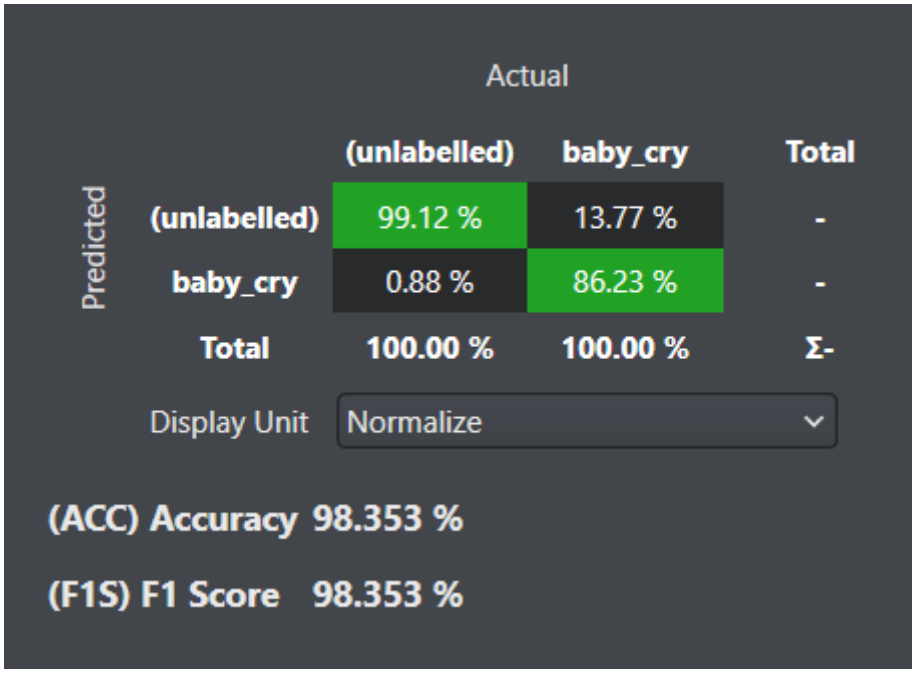
Testing

In this section different kinds of tests will be carried out to show the model reacting to positive and negative sounds with the performance documented.

Validation Set Results



The plot above shows the predictions of validation set on the file level. As we can see from the plot, the model predicts 80% of the baby crying files correctly. And there are some false triggers on the negative data. The 7 false positives among NotBabycry are angry cat, electrical shaver and siren sounds.

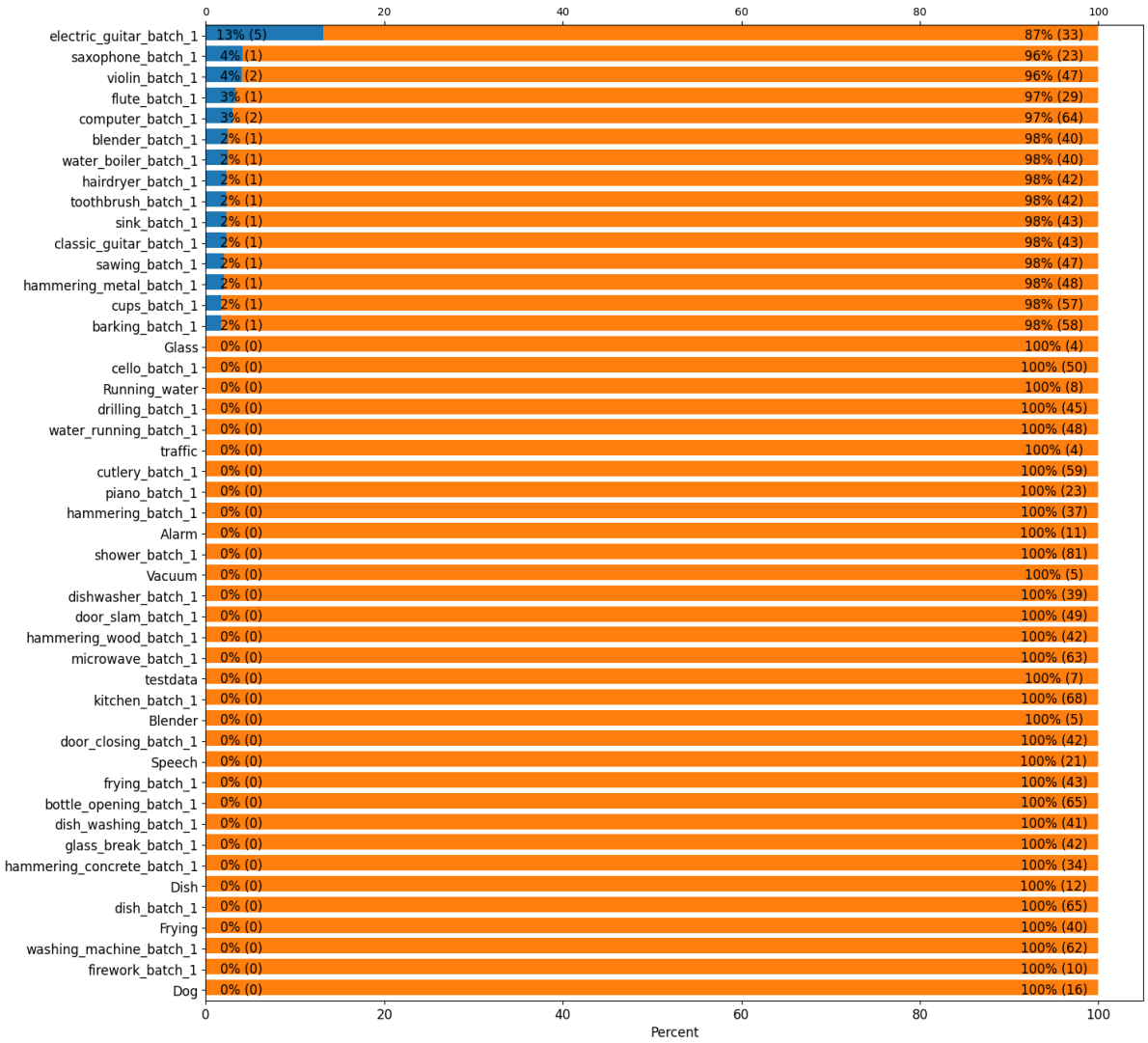


The picture above is the confusion matrix of validation set from our studio.

The meaning of the percentages/values is as follows:
Top Left Value (True Negatives): actual negative/NotBabycry data predicted as negative/ NotBabycry data
Bottom Left Value (False Positives): actual negative/NotBabycry data predicted as positives/ Baby_cry data
Top Right Value (False Negatives): actual positive/Baby_cry data predicted as negative/non-childcry data
Bottom Right Value (True Positives): actual positive/Baby_cry data predicted as positive/ Baby_cry data

Here we can see that the model correctly caught 99% of these timesteps as true negatives among the non-baby crying data. For the baby crying sound, it was able to catch around 86%.

Test set



This plot shows the prediction result of different kinds of negative sounds in the test set.

Note for this plot:

- This child cry model has the most problem with electric guitar sound even though we already trained the model with such sound
- We only see a few FPs when using Confidence level 85% and consecutive number three. We can consider reducing the confidence level depending on the needs.

Test Results on 20 People Recordings

We tested the model on recordings from 20 children, 10 females and 10 males. Half of these children were from Asia and the other half come from western countries. Each crying file is 20 seconds long and the not crying file is 10 seconds

long. The test result is listed at Appendix II. As you can see from the table, the model detects 95% of children's cries within 10 seconds. And the model performs differently on different types of crying. For example, the model reacts better on wailing and bawling compared to blubbering. For other child sounds, the model accuracy is 90%. We can see from the table that the model does not react to the child laughing and babbling but the model triggers on some child yelling and whining sounds.

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 one meter away from a child
2. Wait until the child starts crying
    a. You can set up a video recording to catch the moment without waiting in place
3. Prediction is generated:
    a. If using custom firmware: check whatever output you have created yourself
    b. If using hex file: check the terminal for prints

Test Results

We performed on-device testing with five different people. Three of the five tests contain child crying and the other two tests are non-child crying sounds. The testing results are shown below.

People	Testing Sounds	Testing Result
Person 1	Two hours of board recording, including home sounds, baby laughing, baby yelling, baby crying, music/tv background and people talking.	Detected all of the baby crying sessions, two False Positives (FPs) on baby yelling and two FPs on music/TV background.
Person 2	Eight hours of live testing including office sounds such as keyboard, online meetings, walking.	Zero FP even when people are loud talking from the speaker
Person 3	One hour live testing in the car, including people talking, baby fussing, baby crying (zero-one year old, male), traffic sound	Zero FP, one out of two baby crying sessions detected and each session lasts two-three minutes long.
Person 4	Less than 10 minutes in total. Including baby crying (zero-one years old, female), people talking	Zero FP, four out of four crying sessions detected and each session lasts around 0.5 - three minutes. However, for the three minute session, the model detects crying but not all of them.

Appendix I - Additional Details

The positive data has one or more child cry events per file, with sound file length ranging from two seconds up. These files have been downloaded from the following sources:

- Freesound - https://freesound.org/

Freesound is a well established sound database website.

The negative data has been downloaded from the following sources:

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

Appendix II - Details Test Results on 20 children's Recordings

ID	Gender	Origin	Age (in years)	Crying result	Non-crying result
58	female	China	1-2	3 out of 6	Laugh, 0FP
61	female	China	1-2	3 out of 4 detected	Babbling, 0FP
60	female	China	2-3	3 out of 4 detected	babbling, 0FP
63	female	China	3-4	1 out 4 detected (sounds like the mouth was covered)	Talking, 0FP
65	female	China	3-4	3 out 3 detected	talking,0FP
67	female	China	0-1	8 out of 9 detected	babbling,0 FP
39	female	Netherlands	1-2	2 out 5 detected (yelling, sounds like fake crying)	Laugh, 0FP
40	female	USA	0-1	6 out of 12 detected	0FP, laugh
42	female	USA	1-2	2 out of 5 detected	Laugh, 0FP

43	female	USA	3-4	3 out of 3 detected	Laugh, 0FP
4	male	China	1-2	5 out 7 detected	Yelling “biii” sound, 2FP
5	male	China	3-4	0 out 5 detected, (sound like month is covered, not break out crying)	Talking, 0FP
6	male	Philippines	1-2	1 out 5 detected	Talking, 0FP
7	male	China	0-1	3 out 3 detected	Babbling, 0FP
8	male	China	3-4	1 out of 5 (talking while crying)	Talking, 0FP
11	male	USA	2-3	1 out of 6 (not crying, more like whining	Laugh, 0FP
10	male	USA	2-3	3 out of 4	Laugh, 0FP
12	male	USA	0-1	4 out of 4	Babbling,0 FP
15	male	USA	2-3	2 out of 6	Laugh, 0FP
14	male	USA	2-3	2 out of 3	whining/talking, 1FP