

YAMNET-based Transfer Learning for Baby Noise Classification and Poop Detection

B. McGraw^{1,2}

¹ Department of Overfitting Analytics, Cranberry-Lemon University Medical School, Pittsburgh, PA, USA

² Recent new dad

Abstract

I am two weeks into being a new dad. After many sleepless nights, we have discovered one item we failed to add to the baby registry: an adequate audio-based classification model to determine what our baby needs from us at four in the morning. I may be biased by I'm confident Hannah is the cutest baby in the world, but this does not mean she's easy to take care of at night. She wakes up every 2-3 hours and there are only a few things a new parent must take care of; wet/dirty diapers, hunger, burping, attention, pain (medical), discomfort (too cold/hot/spit up everywhere). Every diagnostic step will wake the baby more and it is ideal to feed them last so that they may ride their milk drunken state back to bed. Too much burping or an unnecessary diaper check will require exponentially more rocking to sleep. If the diagnostic step can be done using an audio classification model, valuable parental sleep time could be saved. In this paper we will develop and analyze a YAMNet based transfer learning model to make early classifications of baby needs based on crying audio clips and vibrationally analyzing baby diaper to classify farts and poops. We found that most baby cries sound too similar to be analyzed with state-of-the-art CNNs, but poops were significantly different than farts due to diaper aftershocks measured on the Richter scale.

Keywords: Audio Classification, New Borns, CNN, YAMNet, oh my God I need Sleep, Poopy Diapers, Baby Cry Classification, Parenting, Feeding Optimization, Baby Comfort Diagnostics

1. Introduction

The world of infant parenting has already been transformed in the 20th century [1]; formula provides an alternative to breast feeding, disposable diapers, baby monitors, electronic bassinets, mass produced baby gear sold at a discount at goodwill, and an endless stream of trash tv for the parents while breastfeeding. Despite all of the innovations, parents still must wake up multiple times throughout the night to feed, sooth, diaper, and burp the baby without any prior knowledge what they need until they stop crying [2].

When you are a new parent, you want to savor every moment; they are only infants once. This is why I have held my infant daughter under constant video and audio surveillance to capture every moment [3]. Little did I know this would create the data set necessary to fine tune a transfer learning model. Despite having collected every single noise, cry, fart, poop, and nipple suck sound through the baby monitor, 437 feature associated samples are not enough to train an audio classification model from scratch [4-5]. This is where a

transfer learning method is incredibly useful. As shown in figure 1, the previously trained YAMNet audio classification model may be fine tuned with a small dataset of Hannah audio to create a useful classification model to determine why she is crying.

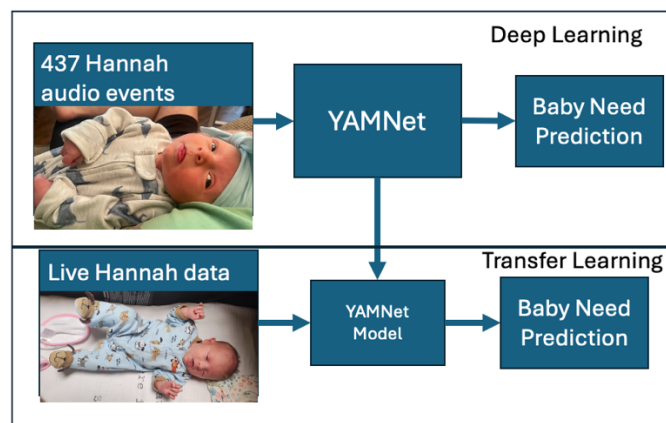


Figure 1: Baby noise transfer learning model process

2. Background

First the 300-600Hz baby cries are transformed into the frequency domain using a Short Form Discrete Fourier Transform using the equation below. Be sure to use a baby monitor which can sample more than 2kHz to capture the highest recorded baby screams over 1kHz when in extreme distress [6]. Otherwise, any signal above the Nyquist rate will add misleading aliasing effects.

$$X(m, k) = \sum_{n=0}^{N-1} x[n]w[n-m]e^{-j2\pi kn/N}$$

It is not recommended to band pass and down mix the time signal to make up for a cheaper baby monitor or the frequency signal will miss low 20Hz frequency farts [7]. Next, a spectrogram is taken of the baby cry frequency signal $X(m, k)$ to determine how much assistance they need.

$$|X(m, k)|^2$$

Then the Mel scale is applied to each frequency bin using:

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

One log scale isn't good enough [8], so we'll log it again with:

$$S_{log} = \log(S + \epsilon)$$

Then a Mobile Net Convolution Neural Network (CNN) is applied by convolving classification function K with the time signal x with the equation.

$$y(i, j) = \sum_m \sum_n K(m, n)x(i-m, j-n)$$

Finally, the signal is embedded into a 1024-dimension output layer that is soft-maxed into the final classification layer. But we don't need to do any of this ourselves because that's all included in a handful of python calls in the YAMNet library. We just wanted to prove that we know how this works.

3. Methodology

Two methods will be used with YAMNet, one classifying baby monitor recordings and another classifying vibrational recordings from her diaper to detect the difference between a poop and a fart.

3.1 Crying Classification

As shown in the table below, each audio clip was associated with the baby problem that needed solving. Hunger, attention, and wet diapers were the largest problems for Hannah though they weren't the only issues in her first two weeks of life. Each audio clip was stored as a .wav file inside a folder labeled with each baby problem.

Baby Problem	Number of Samples
Wet Diaper	83
Dirty Diaper	16
Hungry	144
Too Hot/Cold	24
Needs Attention	84
Wants me to lift her up into the air really fast which Hannah really likes for some reason	12
Unknown	74
Total	437

Table 1: Baby cry audio samples across two weeks of infancy

The .wav files were then down sampled to 20kHz and ingested into multiple K-fold cross validating training and testing datasets for the YAMNet classifier. The resulting model was then applied to live audio data to continuously output a probability of a baby problem when a cry is detected.

3.2 Fart vs Poop Classification

A Raspberry Pi was attached to an infant swaddler diaper and variance lengths of wires to connect to eight different pressure readings across Hannah's poopspllosion zone. One of the poop detector diapers can be seen in figure 2.



Figure 2: Poop Detection Diaper

There are multiple reasons poop-fart differentiation required special monitoring and a dedicated raspberry pi for processing. Firstly, from a distant sensor, a poop and fart will largely sound the exact same [9]. Additionally, with a more limited audio data set from a poop induced cry, it may be difficult to accurately classify from the baby monitor alone. Secondly and more importantly, a baby poop diagnosis type 2 error is associated with higher material risk. Given a false negative during a full on poopspllosion event, containment and baby clean up procedures exponentially increases with time before making a correct diagnosis [10].

4. Results and Discussion

A few days of baby collection were used to test the new live baby cry classification algorithm and the resulting detection probabilities were recorded and averaged into the table below. As expected, the algorithm performed much better for the baby problems which had more data. The poop detection diaper performed extremely well by detecting all but one out of eleven poops. That eleventh poop snuck up on everyone. Further analysis suggested that a poop has a much stronger after shock affect than a fart does because of the multiple splattering effects that was easily differentiated by the eight different diaper pressure sensors [11].

Baby Problem	Detection probability
Wet Diaper	0.189
Dirty Diaper	0.037
Hungry	0.330
Too Hot/Cold	0.055
Needs Attention	0.193
Wants me to lift her up into the air really fast which Hannah really likes for some reason	0.027
Unknown	0.169
Poop Detection Diaper	
Poop	0.91

Table 2: Baby cry audio samples across two weeks of

Unfortunately, the added discomfort, poking, and duct taping of the poop detection diapers added to the total time spent taking care of Hannah in the middle of the night. It took her on average 23.2 minutes longer to settle down between feedings with that raspberry pi sticking between her legs. Even when swaddled, she would also move around disconnecting many of the pressure sensors. Hannah additionally soiled many of the Pis which created additional work re-programming and constructing each diaper. It was not ideal for a disposable garment. At the current manufacturing cost, it would be cheaper to hire a full-time nanny.

The cry classification algorithm however did not appear to add much value. Closer analysis of the results discovered why.

Data visualization blow revealed something peculiar when analyzing the class detection probability outputs. As shown in figure 3, the baby problem probability rarely changed over time as a result of the algorithm.

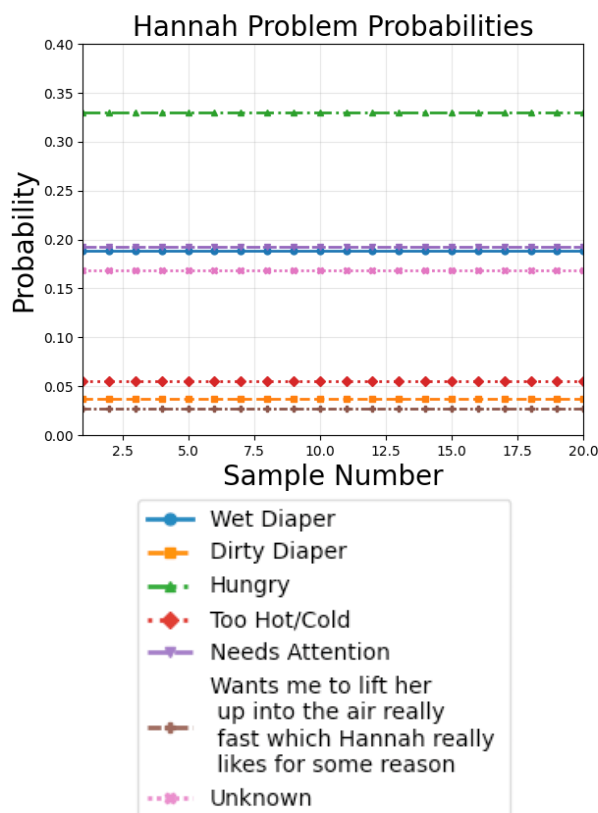


Figure 3: Cry probability across twenty detections

We back tracked these results to table 1 of this paper and realized that the entire YAMNet detection probability could be expressed by the equation below.

$$P(d) = \text{Number of Recorded Events} / \text{Total Events}$$

The detection probabilities did change subtly so we confirmed that the algorithm was working as intended but rarely deviated from the flat prior of baby problem frequency. As far as YAMNet can determine, all of Hannah's cries sound the exact same in 1024-dimensional feature space.

5. Conclusion

Despite the advances in Machine Learning and live audio processing technology, we must still diagnose and manage Hannah's needs manually just like our ancestors thousands of years ago. One could suppose that mathematics is still incapable of matching the innate nature of humans to take care of their young. One would be wrong because we are about as good as YAMNet at predicting Hannah's problems [12].

Thankfully, Hannah is so precious and cute that we don't mind taking care of her even if it's tough work.

References

- [1] Dr. Gerber P. 1998, *The Socioeconomic Impact of Purchasing 97% of Baby Equipment Second-Hand :: Proceedings of the Society for Frugal Parenthood*
- [2] McSnuggles G. and H.G. Wobblesworth 2018 *Multi-Armed Bandit Approaches to Midnight Infant Troubleshooting :: IEEE Transactions on Parental Sleep Deprivation*
- [3] Bottles McGee 2015 *Big Data Begins at Home: Opportunities in Continuous Baby Surveillance :: Journal of Consumer Baby Informatics*
- [4] B. McGraw 2026 *Audio and Video of Hannah's Infancy: Week 1 Data package*
- [5] B. McGraw 2026 *Audio and Video of Hannah's Infancy: Week 2 Data package*
- [6] O'Scilloscope M. 2015 *A Demon on the Baby Monitor: Extreme Spectral Characteristics of Infants Experiencing Existential Distress :: Proceedings of the International Conference on Baby Sampling*
- [7] Brian Flatulus the IV and John N. McToot 2020 *Towards a Unified Theory of Burps, Grunts, and Farts :: Journal of Acoustic Parenting Systems*
- [8] Kenny Loggins and Wyatt the Log Whisperer Johnson 2015 *Recursive Logarithmic Transformations for Data Scientists on the Go Who Refuse to Normalize :: Journal of Aggressive Data Compression*
- [9] Brian Flatulus the IV and John N. McToot 2022 *On the Classification of Ambiguous Rearward Acoustic Emissions :: Journal of Acoustic Parenting Systems*
- [10] Crumbly T. 2011 *A Cost-Benefit Analysis of Early Poopspllosion Detection Systems and the Escalation of Wriggling in a Dirty :: International Review of Diaper Economics*
- [11] Tremblay, G. and Patel H 2024 *Aftershocks, Reverberations, and Residual Momentum: A Study of Post-Poop Dynamics :: Journal of Diaper Dynamics*
- [12] B. McGraw 2026 *Comparative Performance of Exhausted Parents and Convolutional Neural Networks in Cry Based Infant Need Classification :: Journal of Computational Childcare*