INFANT NEEDS CLASSIFICATION BASED ON CRIES USING MACHINE LEARNING TECHNIQUES

Team 1

Abstract

Cry is the sole indication that infants use to indicate their demands. The Smart Baby Monitoring System employs Cry as a sign for monitoring as well. Cries are the only sign that infants use. After a while, a baby will have several wants, including the need to eat, the need to be comfortable, the need to burp, the need for noise, the need for laughter, the need for solitude, the need for fatigue, and the need for their belly to hurt. This study proposes a deep learning-based approach to recognizing newborn cries. We make use of characteristics such as the melfrequency cepstral coefficients (MFCC), the short time energy (STE), the short time zero crossing rate (STZCR), the short time energy acceleration, the spectral centroid, the spectral bandwidth, the spectral rolloff, and the spectral flatness. Here, we collected infant sounds of varying durations, each of which was 500 milliseconds. Eighty percent of the data collected was used for training, and the remaining twenty percent was used for testing. In this step, we train the model with a multi class support vector machine, often known as an SVM. The classification of baby sounds is accomplished with an accuracy score of 84 percent by using this model. In recent years, a considerable amount of research attention has been focused on various aspects related to the processing of natural languages. However, because it is not a language that is easily understood, the baby's cry, which is the primary form of communication for newborns, has not yet been properly examined. This is due to the fact that a baby's scream is their major method of communication in their early years. The recognition and analysis of a baby's cries are not only technically achievable, but they also have the potential to have substantial repercussions for the area of medicine as well as for society as a whole. This is due to the fact that an infant's cries convey information about their health, and both experienced parents and medical professionals are capable of understanding these cries to some extent. As part of our investigation, we are collecting and analyzing the spectral and temporal properties of the sound signals that are generated by the cries of newborns.

Keywords—SVM (Support Vector Machine), Baby cry, deep learning, scipy, audio classification, SKlearn.

Introduction

The baby phone has been the ideal companion for parents to monitor their infant children for several decades, despite all of the restrictions that we are aware of. The rapidly accelerating development of pervasive computing technology, in particular for wireless sensor networks and connected things, encourages the investigation for developing context aware systems such as a new generation baby-phone or any other replacement device that will have cognitive capabilities on which parents can rely. These kinds of technologies will pave the way for a number of applications that make use of ambient intelligence (AmI), which has the potential to dramatically improve the overall quality of life and wellness of infants as well as their parents. The global scope of the research for innovative AmI applications for infants poses several challenges to develop smart systems that can monitor, for example, sleep phases, crying, motion, or monitor vital signs to detect symptoms of infections or monitor some contextual parameters of infants' local living environment such as temperature, humidity, noise, air quality, luminosity, and so on. These smart systems need to be able to monitor for example sleep phases, crying, motion, or monitor vital signs to detect symptoms of infections. However, one of the most essential and difficult problems to solve is the identification and analysis of the complex events that characterize an infant's cry. This can be done by taking into consideration audio, visual, and/or physiological inputs. Because of this, the primary research problem that is discussed in this article is how to create an automatic cognitive process that will provide a more nuanced study and interpretation of the infant's cries. Cry is a universal language used by infants from their earliest days as a biological alarm system to express their basic needs such as food or react to pain or discomfort; and with experience, parents learn how to recognize some patterns of the cry and infer baby's needs. Cry is a universal language used by infants from their earliest days as a biological alarm system to express their basic needs such as food or react to pain or discomfort. The majority of the time, a baby will cry because they are uncomfortable. According to the data sheet on the world population that was issued by the Population Reference Bureau in 2015, this issue affects 25 percent of the 145 million newborns around the world on a regular basis. It has been asserted in [11] that there is a considerable association between a baby's cries and the particular requirements that he has at the time. Cry is the final step of the expression of need, when a newborn is disturbed after having tried to reduce the discomfort by himself, as the authors found more recently in [14]. It is a series of automatic responses that includes phonetic sounds, motions, back-arching, knees flexion, and so on [5]. Several precry signals are created before weeping begins. This article suggests the use of a machine learning mechanism as the central component of an overall ambient intelligence system in order to continually monitor newborn behavior in order to identify and alleviate pain. We propose a comprehensive approach to machine learning that incorporates low-level audio feature selection methods from labeled infant pre-cry recordings, such as spectral descriptors and Mel frequency cepstral coefficients, as well as high-level features that characterize the envelop of the crying. These low-level and high-level features can be extracted from recordings of infants before they cry. After the phase of selecting the features, various machine learning algorithms are utilized in order to carry out the classification.

Motivation

Infants scream for a number of reasons, including pain, discomfort, and general discontent. Therefore, the most crucial part of creating useful applications for ambient intelligence to enhance the lives of newborns and their parents is the automatic detection of infant cry patterns. In this research, we describe a complete method for machine learning that involves the creation of reliable datasets based on newborn cries and the identification of pertinent sound characteristics. The experimental outcomes are promising for enhancing newborn monitoring in more realistic environments. About twenty-five percent of infants experience pain, and the ability of the proposed system to detect and automatically analyze pain signals is what makes it so novel. Selecting low-level audio characteristics from labeled recordings of newborns just before they cry, and high-level components that describe the envelop of the crying sound, are both necessary steps in the machine learning process. Following an up-front phase of feature selection, the categorization is executed by ensemble learning techniques. Additionally crucial to the suggested method is the utilization of pre-crying signals to enhance recognition accuracy. Results on a real-world dataset demonstrate that this method improves the precision of the learning stage. Additionally, the suggested method heavily relies on pre-crying signals to boost identification quality. This discovery paves the way for the development of smart baby monitors that can foresee their users' needs.

Objectives

1. Clipping Audio Files
2. Feature Extraction from Audio files
3. Data Preparation
4. Model Construction
5. Model Training
6. Model Analysis

Related Work

Instinctively and spontaneously, infants scream to communicate distressing feelings like hunger, pain, indigestion, discomfort, etc. Research on the causes of baby crying has been ongoing for quite some time. In the 1940s, the first books on the topic were published [1, 2]. Several active research groups have been founded since then, particularly, the Scandinavian and the Hungarian study group. These discussions centered on crying's role in child development and on its potential pathological causes. In fact, writers in [3] have investigated newborn crying in an effort to diagnose, for instance, the signs of severe diseases through study of qualitative aspects of the crying. Despite this excitement about the potential of studying newborn screams in early illness identification, scientists did not limit themselves to this area. It has also been of some interest to try to classify the causes of crying. Since the seventies [11], numerous research have focused specifically on the notion that we can discriminate between cry kinds. Recent advances in automatic signal processing and machine learning have opened the possibility of automating the treatment of infant crying, the diagnosis of diseases, and most importantly, the detection and monitoring of situations due to infant discomfort, whereas earlier studies relied on the manual inspection of experts in the field. Older research primarily shifted around the newborn cry's basic frequency. Since then, several signal processing and voice recognition-based methods have been adapted for use in identifying the origins of children's distress. Mel frequency cepstum coefficients were derived from a real-world database of newborn cries and used in concert with a feed-forward neural network to perform the classification step of the study [4]. With the help of the support vector machines technique, the authors of [13] were able to make use of a collection of temporal and spectral features. Good results have been found in these research for identifying newborns' emotional states from their screams. Furthermore, authors in [14] have succeeded in finding crucial contextual elements that provide fascinating viewpoint to better distinguish the newborn scream, complementing automatic cry analysis utilizing signal fundamental features. The authors found that babies use universal phonetic sounds prior to crying, which can be categorized as follows: (1) Eairh, which stands for flatulence or the accumulation of gas in the alimentary canal; (2) Eh, which stands for eructation, burping, i.e. the release of gas from the digestive system through the mouth; (3) Heh, which indicates discomfort (cold, hot, wet diaper, change position, etc.); (4) Neh, which indicates that baby is hungry; and In our lab, we use this categorization to guide our studies.

Proposed Framework

When babies scream, they are expressing a wide range of feelings and demands, including those related to their health and development. Because infants can't talk yet, the smart baby monitoring system relies on their cries as a means of communication. A baby will eventually have many demands, including those related to hunger, discomfort, burping, noise, laughing, silence, stomach pain, and exhaustion. In this work, we introduce a deep learning-based approach to recognizing infant cries. An SVM multiclass classifier is used in the system to improve accuracy. Baby Sound Recognition (BSR) and Baby Cry Recognition (BCR) are two examples of the multi-model architecture we employ (BCR). The BSR can identify the caliber of the baby's sonic output. If the BSR determines that the sound is a cry, the BCR uses the information to determine the specific sort of cry being made. Both the BSR and the BCR are support vector machine-trained models. The characteristics are derived from infant sounds in.wav or.ogg format. From the sound file, we get the mel-frequency cepstral coefficients (MFCC), short-time energy (STE), short-time zero crossing rate (STZCR), short-time energy acceleration (STEA), spectral centroid (SC), spectral bandwidth (BB), spectral rolloff (SR), and spectral flatness (SF).

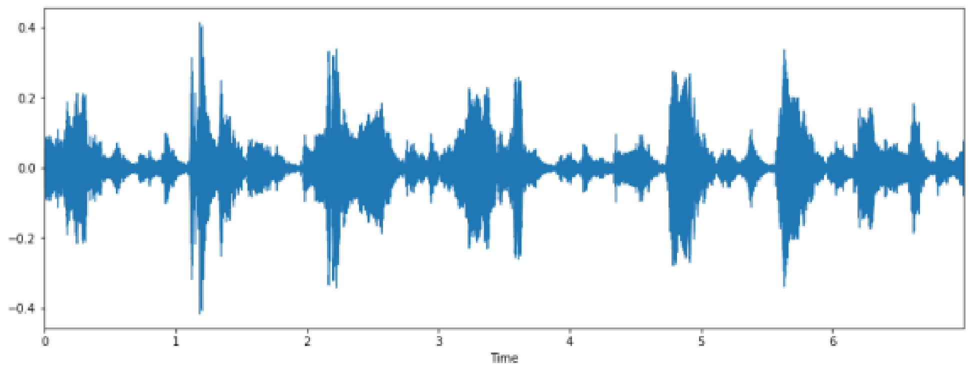

fig 1. 500MS audio Clip

The data comes from a wide range of medical centers, and each video was cut into 500 MS segments. When it comes to BSR, there are four different types, whereas BCR has five. Crying, Noise, Laughter, and Silence are all categories used by the BSR. Pain in the abdomen, bloating, belching, hunger, and exhaustion are all symptoms that may be sorted using the BCR. The audio files are parsed for features, and the data is saved in a CSV file for further use in the training process. Read on to see the specs.

The characteristics of each audio file are extracted with the help of the Librosa python library. Due to the vast size of the dataset, computing the characteristics of each audio file and saving the audio file as a CSV file might take up to 8 minutes. Here, we build a pipeline for training the model, the next step after data preparation. In order to simulate the BSR and BCR, two distinct datasets are used during training. In contrast to BCR, which is exclusively learned using audio of newborns' cries, BSR is trained utilizing various sorts of noises from the baby and its environment. Eighty percent of the data is utilized for training, while the remaining twenty percent is used for testing. Before being sent to the SVM classifier, the data undergoes standardization in the pipeline.

Results

Users can begin by segmenting their work into modules of development that have already been completed by validating a trained model. This will help them get started more quickly. This audio file has been cut down to a certain length for your listening pleasure. At least once, each and every audio file that is part of the training and testing will be subjected to both the training and the testing. The method that is being designed takes into account the requirements of the infant. In this instance, we are relying on methods that are associated with machine learning. A popular tool for data categorization is the support vector machine (SVM) multi class classifier. The findings that are generated by the support vectors and hyper planes are accurate, fair, and beneficial to the parents. Because BSR and BCR are serialized, they are machine agnostic, can be utilized everywhere, and are pleasant to work with on the backend. We also have the option of using Google Collab to test the models and examine the outcomes they provide.

GitHub Link: https://github.com/SUSHMASHETTY6/Final-Increment

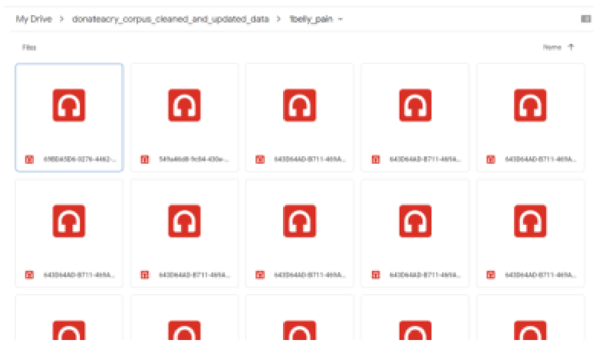

fig 5. Audio Files

The features are stored in a CSV file where it contains 13 MFCC and 8 other features


```
0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,label
0.05291110788863109,8.913723142024054,0.004821361250138543,0.02865391716018339,2501.8777665563616,3109.4224235140814,6401.756339725928,0.00
15091324457898736,-325.5334521838137,120.39443434211233,-3.440430942526252,8.02582873467579,-7.0414020530738135,-11.66649757022974,-2.93343
4770756346,9.193701831036925,-9.954863285846093,0.4768874564318091,6.466847981118129,3.9127593547224304,2.3267694460327992,1BabyCry
0.10710239631670533,2.0299377745198126,0.001479681412239152,0.057966247619647505,3561.2570708332173,2820.861974093493,7665.745370957801,0.0
006719750235788524,-341.0473344626243,65.95280760640796,-82.13343901372923,30.811138629257183,-6.66394954171495,9.112455949835462,-16.71965
3631306684,16.51218379552139,5.979892386716488,4.139631996613279,6.141187996336694,-11.977003729525713,4.221822021447004,1BabyCry
0.06189049811484919,2.567540432292458,0.002783061839227741,0.03349701011912925,2699.0619864966898,2884.0565966036165,6282.79913174304,0.000
479803973576054,-253.22978209893125,126.21316272117735,-58.12228441256116,6.0313675556186865,-10.75191825077819,-7.956529308072327,-4.77277
7227377985,23.16507293789255,3.6290392363483583,5.873406117211603,-3.2650048051546405,-3.2124801346557947,5.896133631872128,1BabyCry
0.10505184164733179,2.9990223179207343,0.0021857161203588204,0.05686209247722047,3733.927675222276,3022.3582234474,8314.164574572216,0.000
3062959585804492,-298.1215258258732,48.149684872444304,-86.7854493766613,18.64431477667477,-21.827170634868413,-0.8250496201242771,-23.5260
12193845407,26.538113876221928,4.303657051166598,-2.982080040387219,-2.036040991926254,-10.27878272918113,3.5055040141752047,1BabyCry
0.11039461281902552,3.074868691960276,0.0013271234438426566,0.059752569784372385,3919.590831970422,3629.525695912804,9415.630381289879,0.003
024633973836899,-338.8636222696679,88.77279940012558,-41.13351154015725,16.481831641315452,-16.899750097506864,19.21710163087352,-3.5387737
39465343,4.716053872330519,6.899653613899512,7.9418456932115555,-3.220335211005599,-1.927445752368097,1.5803328994297319,1BabyCry
0.08716330118909513,1.4593982406166393,0.0007122160217055244,0.04716788917255925,3540.190419743118,3636.650529175473,8945.721931191996,0.00
30335134360939264,-382.83353159947023,104.46992326027402,-21.385235496652363,19.46474484515855,-13.937386454613799,22.64674639087694,-2.059
0052123938887,4.944778810555679,3.194836123464038,7.671610011466254,2.3367650975954577,0.04535946380010574,-2.105160090902677,1BabyCry
0.15968043068445475,2.380803712937595,0.0011194112298685 7,0.08644850843476005,4060.1501521119458,3220.232139646644,8507.1640262471,0.005646
```


Fig 6: Image after feature Extraction

Every single one of the models These objects have been serialized and may be employed by any monitoring system as a result of the fact that the BSR and BCR objects are pickled by making use of the joblib Python package. On addition, they are stored in the cloud, and APIs that make use of rest are used to gain access to them. The needs of the child being cared for are being taken into consideration by the approach that is currently being developed. On this particular scenario, we are putting our faith in strategies that belong to the field of machine learning. The support vector machine (SVM) multi class classifier is a technology that is commonly used for the classification of data. The results that are produced by the support vectors and the hyper planes are precise, objective, and helpful to the parents. Because BSR and BCR are encoded in a serialized manner, they are not dependent on the type of machine being used, they may be utilized everywhere, and they are advantageous to backend implementations. In addition to that, we can assess the models in terms of the outcomes that they produce by making use of Google Colab.

References

[1] R.G. Barr, N. Fairbrother, J. Pauwels, J. Green, M. Chen and R. Brant, "Maternal frustration emotional and behavioural

responses to prolonged infant crying", Infant Behavior and Development, vol. 37, no. 4, pp. 652-664, 2014.

[2] H.C. Lin and R. McFatter, "Empathy and distress: Two distinct but related emotions in response to infant crying", Infant

Behavior and Development, vol. 35, no. 4, pp. 887-897, December 2012.

[3]P. Dunstan, Calm the crying: The Secret Baby Language, pp. 16-17, 2012..

[4] H.L. Golub and M.J. Corwin, "Infant Cry: A Clue to Diagnosis", Pediatrics, vol. 2, no. 197, pp. 69, 1982.

[5] L.L. LaGasse, A.R. Neal and B.M. Lester, "Assessment of infant cry: acoustic cry analysis and parental perception",

Developmental Disabilities Research Reviews, vol. 11, no. 1, pp. 83-93, 2005.

[6] M. Uluba, H. Uzun and D.A. Senses, "A newborn with Cornelia de Lange syndrome: a case report", Cases J, vol. 1, pp.

329, 2008.

[7] H. Nakayama, "Emergence of amae crying in early infancy as a possible social communication tool between infants and

mothers", Infant Behavior and Development, vol. 40, pp. 122-130, 2015.

[8] H. Nakayama, "Development of infant crying behavior: A longitudinal case study", Infant behavior & development, vol.

33, no. 10, pp. 463-71, 2010.

[9] C. Parlak, B. Diri and F. Gürgen, "A cross-corpus experiment in speech emotion recognition", SLAM@ Interspeech, pp.

58-61, 2014.

[10]M. Chóliz, E. G. Fernández-Abascal and F. Martínez-Sánchez, "Infant Crying: Pattern of Weeping Recognition of Emotion

and Affective Reactions in Observers", The Spanish Journal of Psychology, vol. 15, no. 3, pp. 978-988, 2012.

GitHub Link: https://github.com/SUSHMASHETTY6/Final-Increment

[11] Ko Yi-Chun, A study on feature sets for emotion recognition from speech signals, 2015.

[12] B. Yang and M. Lugger, "Emotion recognition from speech signals using new harmony features", Signal Processing, vol.

90, no. 5, pp. 1415-1423, 2010.

[13] M.P. Robb and A. T. Cacace, "Estimation of formant frequencies in infant cry", Int. J. of Pediatric Otorhinolaryngology,

vol. 32, pp. 57-67, 1995.

[14] R. Lavner, Y. Cohen, D. Ruinskiy and H. IJzerman, "Baby cry detection in a domestic environment using deep learning",

IEEE Int. Conf. on the Science of Electrical Engineering, pp. 1-5, 2016.

[15] M. Hariharan, S. Chee, L. Sazali and Y. Sazali, "Analysis of Infant Cry Through Weighted Linear Prediction Cepstral

Coefficients and Probabilistic Neural Network", Journal of medical systems, vol. 36, pp. 1309-15, 2010.

[16] J. Saraswathy, M. Hariharan, S. Yaacob and W. Khairunizam, "Automatic classification of infant cry: A review", Int.

Conf. on Biomedical Engineering (ICoBE), pp. 543-548, 2012.

GitHub Link: https://github.com/SUSHMASHETTY6/Final-Increment

[17] I-A Bănică, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Baby Cry Recognition in Real-World Conditions", Int.

Conf. on Telecommunications and Signal Proc., pp. 315-318, 2016.

[18] I.-A. Bănică, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Automatic Methods for Infant Cry Classification", Int.

Conf. on Communications (COMM), 2016.

[19] R. Cohen and L. Yizhar, "Infant cry analysis and detection", Conv. of Electrical & Electronics Engineers in Israel (IEEEI),

pp. 1-5, 2012..

[20] B. Schuller et al., "Deception Sincerity & Native Language", Interspeech Computational Paralinguistics Challenge, pp.

2001-2005, September 2016.

[21] B. Schuller, "The Interspeech 2013 Computational Paralinguistics Challenge: Social Signals Conflict Emotion Autism",

Proceedings of Interspeech, 2013.

GitHub Link: https://github.com/SUSHMASHETTY6/Final-Increment

[22] E. Frank, M. A. Hall and I. H. Witten, The WEKA Workbench. Online Appendix for “Data Mining: Practical Machine

Learning Tools and Techniques, Morgan Kaufmann, 2016.

[23] M.S. Rusu, S. Diaconescu, G. Sardescu and E. Bratila, "Database and system design for data collection of crying related

to infant's needs and diseases", Int. Conf. on Speech Technology and Human-Computer Dialogue SpeD 2015, 2015.

[24] F. Eyben, F. Weninger, F. Groβ and B. Schuller, "Recent developments in opensmile the Munich open-source multimedia

feature extractor", ACM Int. Conf. on Multimedia, pp. 835-838, 2013.

[25] Y. Oshrat, A. Lerner, A. Cohen and M. Avigal, The Fingerprints of Pain in Human Voice, Israel, 2014

[26] Patibandla, R. S. M., & VeeranJaneyulu, N. (2020). A simrank based ensemble method for resolving challenges of

partition clustering methods.

[27] Patibandla, R. L., Narayana, V. L., Gopi, A. P., & Rao, B. T. Recommender Systems for the Social Networking Context

GitHub Link: https://github.com/SUSHMASHETTY6/Final-Increment

for Collaborative Filtering and Content-Based Approaches. In Recommender Systems (pp. 121-137). CRC Press.

[28] R S M Lakshmi Patibandla, B. Tarakeswara Rao*, M. Ramakrishna Murty and V. Lakshman Narayana, "Predictive Time Series Analysis and Forecasting of COVID-19 Dataset", Recent Patents on Engineering 2021.

[29] Srinivas, V. S., Babu, B. S., & Patibandla, R. L. (2021). Mammographic Image Segmentation using Automatic Evolutionary Algorithms. Annals of the Romanian Society for Cell Biology, 4058-4066.

[30] Patibandla, R. L., Narayana, V. L., Gopi, A. P., & Rao, B. T. (2021). Comparative Study on Analysis of Medical Images Using Deep Learning Techniques. In Deep Learning for Biomedical Applications (pp. 329-345). CRC Press.

GitHub Link: https://github.com/SUSHMASHETTY6/Final-Increment