Smart Cradle System for Automated Baby Monitoring with Computer Vision

Fahim Ahmed
Department of Computer Science and Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
2022409@iub.edu.bd

Anshara Chowdhury
Department of Computer Science and Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
2021768@iub.edu.bd

Hasin Anzum Mehtaj
Department of Computer Science and Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
1830947@iub.edu.bd

SK Sadia Tasnim Elma
Department of Computer Science and Engineering
Independent University, Bangladesh
Dhaka, Bangladesh
1930921@iub.edu.bd

***Abstract*—Parents attending to newborns often suffer from sleep deprivation. Our smart cradle attempts to reduce parents' loads by attempting to put the baby to sleep without parental intervention. The system uses a combination of a motion sensor and a camera system to detect the baby waking up and employs an AI component to decide the appropriate lullabies to play and the intensity of the cradle's rocking motion depending on the baby's current mood. The system notifies the parents via WiFi only if it fails to put the baby to sleep within a certain time threshold, thereby allowing parents to sleep better without needing to attend to their newborn unless necessary. The AI component can be used to monitor for other signs of distress and notify the parents accordingly.**

***Index Terms*—baby monitor, computer vision, cry detection, cradle**

I. Introduction

The reason behind our idea was the endless complaints from parents of toddlers from all over the world, be it our close relative, a mentor, or a colleague, and posts on social media, about one thing, "Sleep". The recommended hours of sleep for an adult would be at least 7 hours [1], but for parents with toddlers, it's just a dream. To make it a reality, we have proposed this idea which will take care of both parties, the child and the parent. Our main goal is to create a device that will be very customizable according to the baby's needs. It will act as a part-time nanny to take care of a baby's immediate tantrums and the system will be built in such a way that the robot will only alert the parents if the baby's cry is a cry for help. Primary users of this system are going to be both parents as well as neonatal wards of healthcare facilities. The system will reduce the stress on caregivers and babies.

Parents of newborns lose an average of 6 months of sleep in the first 24 months of their parenthood [1]. A survey was carried out to understand the difficulties of new parents that they go through regularly, especially while the baby is asleep and wakes up suddenly for certain needs. The responses are supposed to be effective enough for us to understand how an automated cradle system would reduce the troubles of new parents so that they can maintain a better lifestyle at that point.

Out of 70 participants that were mostly around the age of 25-40, it was found that 72.7% of new parents are suffering from sleep deprivation, and it's also seen that more than half (54.5%) of the people ask for help from family and friends, which seems to be a hassle for both parties. All participants believed that it is very difficult to respond to the baby waking up in the middle of the night. The baby might wake up due to any possible reasons like hunger, defecation, etc., but the results show that 50% of the people delay in responding, which turns the situation into a risk-taking one. Also, most people (91%) believe that getting the baby back to sleep requires extremely hard work and is also time-consuming. Hence 72.7% of people voted yes to the fact that an automated cradle system would be of great help to them, and they would buy it if it were available in the market, which proves the necessity and market demand for the automated cradle.

II. Related Works

A. Cry Recognition and Classification

Infant express various emotions via vocalisations [2] and so it is no surprise that detection and classification of infant cry is a widely studied topic. Various traditional machine learning methods and novel neural network approaches have been employed to recognise and identify the type of cry from audio signals [3]. Convolution Neural Networks (CNN) have been used to great effect at identifying baby cries into four categories - pain, hunger, sleepiness, and wet diaper by Chang and Tsai (2019). The audio signal was converted into a spectrogram using Short Term Fourier Transform before being fed to a 2D CNN that detected if the signal was of a baby crying or not. Another 1D CNN was used to classify the signal into the four types. [4].

Audio-based cry detection has been proposed for use in smart cradle systems to detect infant distress in order to provide adequate automated care. Kumar et al. (2023) suggests such a system that uses an multi-functional android app that provides processing and monitoring capabilities [5].

Automated *Smart Cradle System with Emotion Recognition* [6] has existed in the market for some time now. However, uplifting effectiveness has been a constant attempt, hence the innovations. The main goal is to monitor babies and notify parents when needed, and the system comprises features like emotion recognition, automatic swinging of the cradle, sensing the wetness of the baby's bed, monitoring the presence of the baby in the cradle, and detecting the crying voice of a baby. The features accumulate to send emails to the parent and were made using Raspberry Pi 3, wet sensor, PIR sensor, sound sensor, cry pattern circuit, and camera.

In *Design of a Smart Baby Cradle Using Blynk and Local Customer Priorities* [7], IoT-based automation of the baby cradle is used to allow parents and housewives during housework to take care of the baby with minimal attention. The smart cradle uses the Blynk IoT platform to display the baby's status, including crying, wetness, and temperature. Sensors are used to detect crying, moisture, and body temperature.

In *Intelligent, Automated, and Web Application-Based Cradle Monitoring System* [8], The cradle with microphone, wet sensor, motor, camera, motor driver, and the controller sends signals to the motor, updates the status on the web portal, and sends alerts to parents. The study showed how to monitor the cradle with less human involvement, giving parents and other caregivers access to the cradle through a web gateway.

B. Sleeping Infant State Recognition

Computer vision methods have been utilised to detect the status of infants sleeping in a crib. Hussain (2019) has proposed the use of an easy-to-compute motion detection mechanism based on combining frame difference and background substitution methods to detect abnormal breath behaviour [9]. Naz et al. (2021) have used convolution neural networks to detect whether a baby was sleeping in a hazardous position [10]. Khan (2021) takes this a step further by proposing computer vision techniques to detect baby waking up via frequent movement detection and eye aspect ratio calculation. The proposed baby monitor system detects baby sleeping posture, the covering of face with blanket, the removal of blanket, along with the infant's sleep status. [11].

C. Facial Expression Set

In this article, it was introduced that the *Child's Affective Facial Expression (CAFE) Set* [12]sets a new stimulus set for the study of emotional development. To explore the developmental interpretation of these emotions, researchers recently began to emphasize the significance of having kid examples of the many emotional expressions. Even though the CAFE set only consists of seven ostensibly fundamental emotions, the set's inherent diversity will enable researchers to spot faces that are reminiscent of more subtly expressed forms or faces that combine other emotional expressions, as it includes 1192 images. CAFE includes two subsets of faces: Subset A, which only contains highly stereotypical examples of the various facial expressions, and Subset B, which solely consists of faces.

With the development of neuroscience technologies, clinical research has augmented emotion perception. Many people's faces were photographed for getting the Child Emotional Faces Picture Set [13] to make use of the stimulus to tabulate and gather results to increase reliability and validity. The pictures were categorized into 7 forms. A dataset has been created then with two conditions: direct and averted gaze. To check for validity, 20 raters were appointed and asked to understand the emotion and rate the intensity and representativeness. The limitation set was to appoint adult raters from places where they used to work with children. The results tabulated successfully ended up providing a dataset that will be used for manufacturing relevant bots. According to The *Perception of Facial Expressions in Newborns* [14], the newborns' facial movements did not alter with the various simulated expressions, and observers were unable to determine the modeled expression by looking at the infants' features, hence these authors could not uncover evidence for selective imitation of emotional facial expressions. Importantly, this was not a result of the technique's sensitivity as there was strong evidence that the tongue protrusion measurements matched the model. Three tests have been conducted to distinguish between neutral, fearful, and happy facial expressions. The results were compared using two different parametric tests. In this study, evidence has been found that people prefer to look at cheerful faces for longer periods than at scared ones.

In The Child Emotion Facial Expression Set: A Database for Emotion Recognition in Children [15], photos and videos of facial expressions of children from 4 to 6 were used as a database to understand the expression 7 universal emotions and neutral expressions. The paper is used to understand what methods can be used to judge the expression. They set up two rounds of assessment, with different judges and different parameters to grasp a universal agreement on 7 universal facial expressions [15].

D. NICU

In *UWB Baby Monitor* [16], we see that our system has room for more innovations and can be improved in the future for use of NICU by including more advanced medical help with addition of features like respiratory and heart rate monitors. The Ultrawideband and Ultrashort Impulse Baby Monitor system can be used both at home and hospital to monitor and help with Sudden Infant Death Syndrome.

In Effect of Home Monitoring on a High-Risk Population, we see how home monitoring systems can be helpful. The paper is mostly a rebuttal paper that tells how important it is to have a diagnostic tool like a baby monitoring system and how it can potentially be a life saving addition to medical management systems [17].

In NICU, are filled with low birth weight(VLBW) infants and to assess the quality of nicus, baby monitors combined indicators are used to judge the quality of NICUs in California [18].

In Accuracy of Pulse Oximetry-Based Home Baby Monitors [19], two mainstream smartphone-integrated consumer baby monitors that use pulse oximetry tested against one another to see the accuracy and how the Pulse Oximetry-Based Home Baby Monitors can be improved more. We also see that method can be used to test the accuracy. We see that in mainstream baby monitoring systems there is an absence of FDA and many times that monitor provides wrong information like some claim to detect hypoxemia but never actually does and also sometimes displayed falsely low pulse rates. The paper also stated that for medical emergencies relying on this will hamper the medical decision so while making a baby monitor there should be a sense of responsibility as they can be used to detect medical conditions.

E. Infant Cry analysis and detection

For *Cry detection in Vehicles* [20], extraction and processing was used to differentiate crying audios from background noises like vehicle horns, passer-bys. Using K-NN algorithm, the audio was recognized as 'cry' and 'no cry'. After several experiment using real-time data and pre-proccessed dataset, this algorithm has proven to be robust and has high detection rate.

In search of more ways to detect baby crying detection and analysis, this paper [21] was summarized. As Crying is one kind of Natural language that is not given much thought about, this paper mentions and categorizes different emotions for crying: hunger, sleep, attention, discomfort, diaper. LPC, LPCC, BFCC, MFCC is used for this experiment.

Using *Dunstan Baby Language*, in this paper [22], several baby cry classification experiments were performed on a database of about 40 healthy babies. For automatic classification, two methods employed that were successful in speaker and language recognition: the GMM-UBM (Gaussian Mixture Model - Universal Background Model) and the i-vectors modeling methods.

We know how useful cry analysis is, and this paper [23] just takes it up a notch. They diffuse some structured solutions along with detection of crying which intercepts with the project we are working on.

III. Methodology

A Passive Infrared (PIR) motion sensor is used to detect the baby's motion. An Arduino microcontroller interprets motion above a set threshold to signal distress and triggers a video camera. Processing the video feed occurs on an Android smartphone using a low-power ARM chip. Computer vision is used to analyze the video feed to recognize the current emotional state of the baby in the cradle.

Convolution Neural Networks (CNN) is the current state of the art for computer vision applications, with a variety of different high-performance architectures available. Due to the low processing power available to mobile CPUs, MobileNetV2 is used in this project since it provides an excellent balance between detection accuracy and lightweight computational requirements [24]. Transfer learning is used to train the pre-trained model on a dataset consisting of baby faces that range between the classes: "distressed", "neutral", and "happy". 6

For each of the states, a playlist of lullabies is maintained that is best fit to soothe the baby towards falling asleep. The system continuously monitors the baby while awake and dynamically adjusts the music to the baby's current emotional state.

Failing to put the baby to sleep within a set threshold period, the system pushes notifications to a list of caretakers' cell phones via WiFi.

The system is reset either by a responding caretaker or if the baby falls asleep.

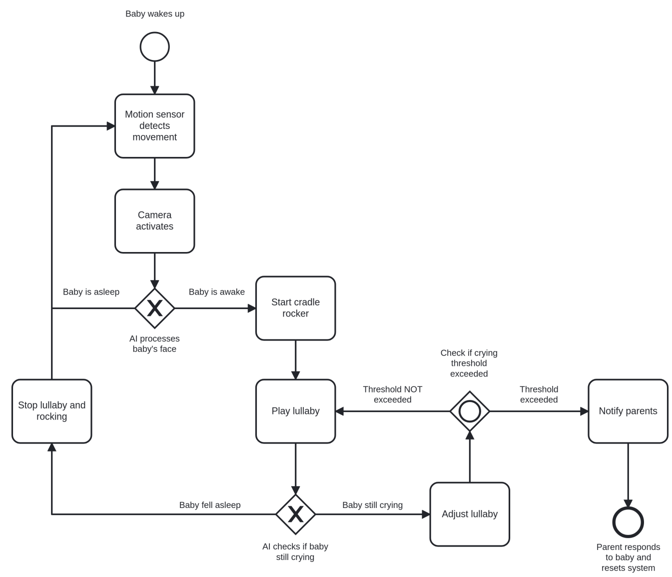

Fig. 1. Event Loop of Smart Cradle

IV. System Design

A. Hardware

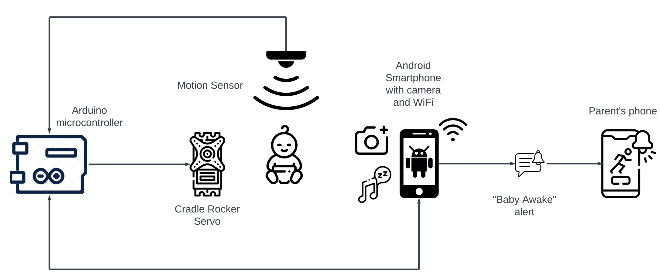

Fig. 2. Hardware Components of Smart Cradle

The scaled-down prototype consists of readily available hobbyist-grade components. An Arduino Uno microcontroller is connected to an HC-SR501 PIR motion sensor, an SG90 mini servo motor for rocking the cradle, and an Android Smartphone.

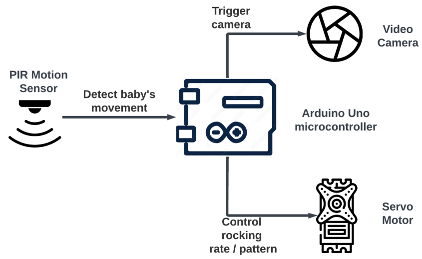

Fig. 3. Micro-controller responsibilities

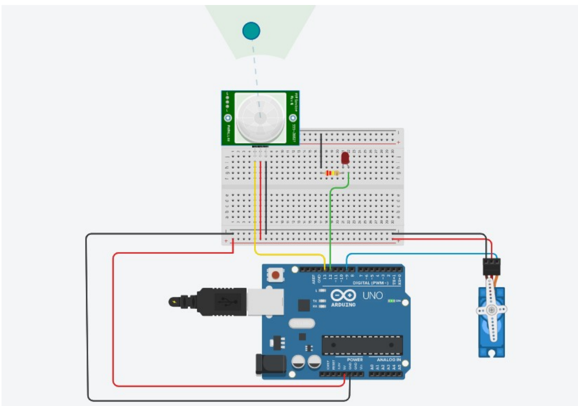

Fig. 4. Prototype of Smart Cradle

The smartphone is an extremely flexible, versatile platform that is already available at hand and performs the task of multiple standalone components - saving both prototyping cost and complexity. It performs the role of a standalone camera unit, music player, speaker, Internet connectivity/WiFi module, and computer vision processor. Communication between the microcontroller and the smartphone occurs via USB.; which also provides DC power to the Arduino and the servo motor.

B. Software

The system has two separate Android applications.

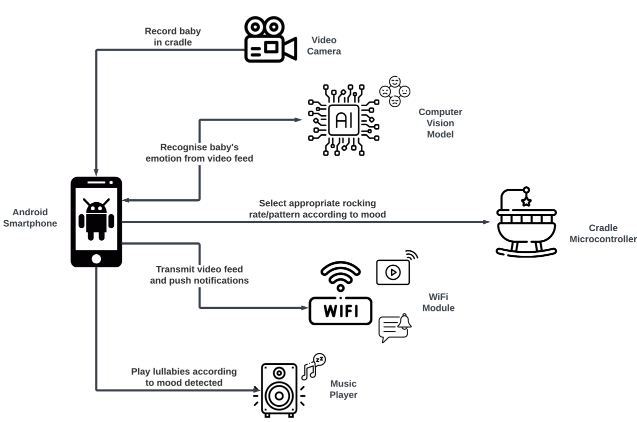

Fig. 5. Android smartphone responsibilities

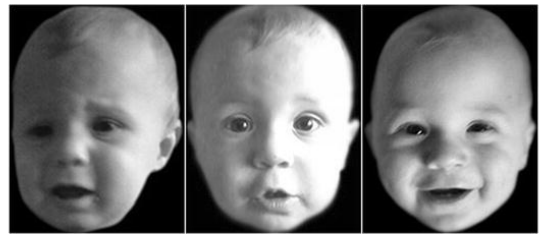

Fig. 6. Samples from the City Infant Faces Database

TABLE I
ACCURACY PER CLASS

Class	Accuracy	# Samples
happy	0.89	9
neutral	0.71	7
negative	0.56	9

1) The caretaker-side application (i.e. the baby monitor app) does the following:
   - receives alerts via push-notifications
   - receives and views video feed from the cradle on demand.
2) The cradle-side applications are responsible for:
   - recording and transmitting video from cradle
   - running inferences on the video feed
   - playing appropriate lullabies based on inference
   - pushing alert notifications to baby monitor app via WiFi

V. Results

The City Infant Faces Database - comprising 60 photographs of positive ("happy") infant faces, 54 photographs of negative ("distressed") infant faces, and 40 photographs of neutral infant face [25] - is used to train the AI model used in the prototype.

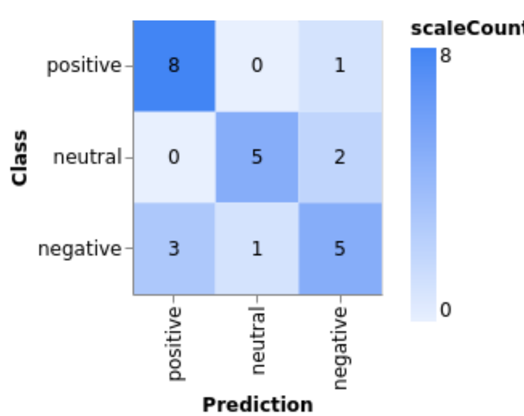

Fig. 7. Confusion Matrix

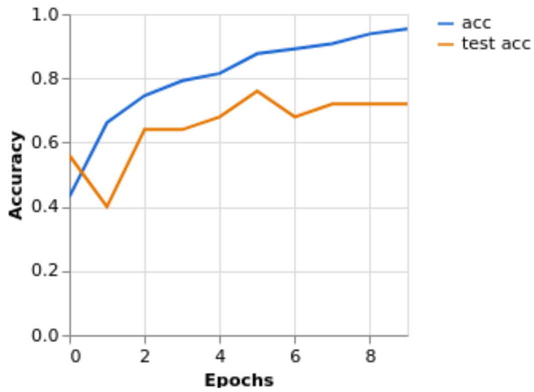

Fig. 8. Accuracy per epoch

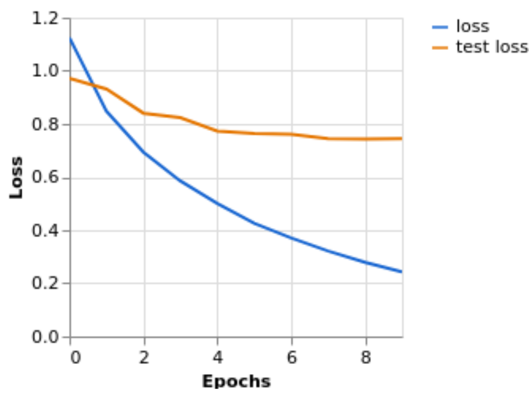

Fig. 9. Loss per epoch

VI. Discussion and Conclusion

A. Training Limitations

Transfer learning allows for training computer vision models despite limited data availability. The dataset used to train this model is much too small even then. To improve system reliability, a larger dataset must be used to train the model to yield better detection accuracy.

B. Future Scope

For ease of prototyping, only three classes of emotions have been chosen for the model to detect. An actual system will of course not be subject to such a limitation. Computer vision can be used to detect a much wider and finer range of emotions and states. Parents may even be warned of the baby in risky positions.

The goal of this project was to build a base to expand upon and we believe we have succeeded in that regard. The android app developed can easily be integrated with a more advanced model.

Acknowledgement

We are expressing our heartfelt gratitude to our instructor, Dr. Mohammad Shidujaman, for his constant support and assistance throughout the process. His proficiency, perceptions, and approval were vital in allowing us to complete the work. This project has been a part of the course work of Instrumentation and Measurement (CSE441), and hence we are also grateful to the Computer Science and Engineering Department of Independent University, Bangladesh (IUB) for all the support and for arranging the required sessions for us, which allowed us to conduct our study and complete our work.

References

[1] "New Parents Have 6 Months Sleep Deficit During First 24 Months Of Baby's Life," New Parents Have 6 Months Sleep Deficit During First 24 Months Of Baby's Life. [Online]. Available: https://www.medicalnewstoday.com/articles/195821

[2] O. Wasz-Höckert, T. J. Partanen, V. Vuorenkoski, K. Michelsson, E. Valanne, The identification of some specific meanings in infant vocalization. Experientia. 20(3), 154 (1964).

[3] Ji, C., Mudiyanselage, T.B., Gao, Y. et al. A review of infant cry analysis and classification. J AUDIO SPEECH MUSIC PROC. 2021, 8 (2021). https://doi.org/10.1186/s13636-021-00197-5

[4] C.-Y. Chang and L.-Y. Tsai, "A CNN-Based Method for Infant Cry Detection and Recognition", Mar. 15, 2019. https://link.springer.com/chapter/10.1007/978-3-030-15035-8_76

[5] V. S. Kumar et al., "Internet of Things-Based Patient Cradle System with an Android App for Baby Monitoring with Machine Learning", Jul. 16, 2022. https://www.hindawi.com/journals/wcmc/2022/1140789/ (accessed Mar. 12, 2023).

[6] A. A. A. M. S. S. Suresh, Maheen, "10.1063/5.0020709.7." Default Digital Object Group, vol. 7, no. 7, pp. 3788-3793, 2020, doi: 10.1063/5.0020709.7

[7] A. V. Karanjkar and R. Kumawat, "Design of a Smart Baby Cradle Using Blynk and Local Customer Priorities." SAMRIDDHI: A Journal of Physical Sciences, Engineering, and Technology, vol. 14, no. 2, pp. 159-165, 2022, doi: 10.18090/samriddhi.v14i02.5.

[8] P. J. Nair and V. Ravi, "Intelligent, Automated, and Web Application-Based Cradle Monitoring System." Handbook of Smart Materials, Technologies, and Devices, pp. 1-14, 2021, doi: 10.1007/978-3-030-58675-1_166-1

[9] T. Hussain; K. Muhammad; S. Khan; A. Ullah; M. Y. Lee; S. W. Baik, "Intelligent Baby Behavior Monitoring using Embedded Vision in IoT for Smart Healthcare Centers", Journal of Artificial Intelligence and Systems, 1, 110–124, 2019, https://doi.org/10.33969/AIS.2019.11007

[10] T. Naz, R. Shukla, and K. Tiwari, "Affordable ML Based Collaborative Approach for Baby Monitoring", Asian Journal of Research in Computer Science, Nov. 24, 2021. [Online]. Available: https://doi.org/10.9734/ajrcos/2021/

[11] T. Khan, "An Intelligent Baby Monitor with Automatic Sleeping Posture Detection and Notification", MDPI, Jun. 18, 2021. https://www.mdpi.com/2673-2688/2/2/18

[12] V. LoBue and C. Thrasher, "The Child Affective Facial Expression (CAFE) set: validity and reliability from untrained adults." Frontiers in Psychology, vol. 5, 2015, doi: 10.3389/fpsyg.2014.01532

[13] Egger, H.L., Pine, D.S., Nelson, E., Leibenluft, E., Ernst, M., Towbin, K.E. and Angold, A. (2011), The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): a new set of children's facial emotion stimuli. Int. J. Methods Psychiatr. Res., 20: 145-156. https://doi.org/10.1002/mpr.343

[14] T. Farroni, E. Menon, S. Rigato, and M. H. Johnson, "The perception of facial expressions in newborns." European Journal of Developmental Psychology, vol. 4, no. 1, pp. 2-13, 2007, doi: 10.1080/17405620601046832

[15] Negrão, J., Osorio, A., Siciliano, R. F., Lederman, V. R. G., Kozasa, E. H., D'Antino, M. E. F., Tamborim, A., Luo, J., De Leucas, D. L. B., Camargo, P. M., Mograbi, D. C., Mecca, T. P., & Schwartzman, J. S. (2021). The Child Emotion Facial Expression Set: A Database for Emotion Recognition in Children. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.666245

[16] Ziganshin, E., Numerov, M. A., & Vygolov, S. A. (2010). UWB Baby Monitor. International Conference on Ultrawideband and Ultrashort Impulse Signals. https://doi.org/10.1109/uwbusis.2010.5609156

[17] Freed, G. L., Meny, R. G., Glomb, W. B., & Hageman, J. R. (2002). Effect of Home Monitoring on a High-Risk Population. Journal of Perinatology, 22(2), 165–167. https://doi.org/10.1038/sj.jp.7210662

[18] Profit, J., Kowalkowski, M. A., Zupancic, J. A., Pietz, K., Richardson, P. J., Draper, D., Hysong, S. J., Thomas, E. J., Petersen, L. A., & Gould, J. B. (2014b). Baby-MONITOR: A Composite Indicator of NICU Quality. Pediatrics, 134(1), 74–82. https://doi.org/10.1542/peds.2013-3552

[19] Bonafide, C. P., Localio, A. R., Ferro, D. F., Orenstein, E. W., Jamison, D. K., Lavanchy, C., & Foglia, E. E. (2018a). Accuracy of Pulse Oximetry-Based Home Baby Monitors. JAMA, 320(7), 717. https://doi.org/10.1001/jama.2018.9018

[20] Infant cry analysis and detection - R. Cohen and Y. Lavner, "Infant cry analysis and detection," 2012 IEEE 27th Convention of Electrical and Electronics Engineers in Israel, Eilat, Israel, 2012, pp. 1-5, doi: 10.1109/EEEI.2012.6376996.

[21] Infant cry language analysis and recognition: an experimental approach - L. Liu, W. Li, X. Wu and B. X. Zhou, "Infant cry language analysis and recognition: an experimental approach," in IEEE/CAA Journal of Automatica Sinica, vol. 6, no. 3, pp. 778-788, May 2019, doi: 10.1109/JAS.2019.1911435.

[22] Automatic methods for infant cry classification - I. -A. Bănică, H. Cucu, A. Buzo, D. Burileanu and C. Burileanu, "Automatic methods for infant cry classification," 2016 International Conference on Communications (COMM), Bucharest, Romania, 2016, pp. 51-54, doi: 10.1109/ICComm.2016.7528261

[23] Machine Learning Approach for Infant Cry Interpretation - A. Osmani, M. Hamidi and A. Chibani, "Machine Learning Approach for Infant Cry Interpretation," 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), Boston, MA, USA, 2017, pp. 182-186, doi: 10.1109/ICTAI.2017.00038

[24] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottlenecks", 2018, https://doi.org/10.48550/arXiv.1801.04381

[25] R. Webb, S. Ayers, A. Endress, "The City Infant Faces Database: A validated set of infant facial expressions", 2018, Behavioural reseach methods, 50(1), 151-159, https://pubmed.ncbi.nlm.nih.gov/28205132/