Intelligent Baby Cry Monitoring System Using IoT and Machine Learning for Deaf Parents

Lukas L. Moreira[1], Matheus T. Arce[1], Robson M. S. Nogueira[1], Thiago P. T. Costa[1],
Celso B. Carvalho[1], Waldir S.S. Júnior[1], Andrey R.R. Bessa[1], Diego A. Amoedo[1], Edma V. C. U. Mattos[1]
[1]Federal University of Amazonas and Center for R&D in Elec. and Inf. Tech. (UFAM/CETELI), AM-Brazil
Emails: {lukasmoreira, matheusarce, robsonnogueira, celso, waldir, andreybessa, diegoamoedo, edmamattos}@super.ufam.edu.br

Abstract— **This paper presents an intelligent baby monitoring system developed for deaf parents, leveraging the Internet of Things (IoT) and machine learning. The system captures infant cries via a microphone, processes the audio data on a microcontroller using Mel-Frequency Cepstral Coefficients (MFCC) and Long Short-Term Memory (LSTM) networks, and dispatches visual and tactile alerts to parents through a lamp and a vibrating bracelet. The proposed system achieved 100% accuracy on the tested dataset, integrating a lightweight and cost-effective architecture with advanced features suitable for embedded system applications. Compared to existing solutions, this system prioritizes accessibility and affordability, making it a practical option for parents with hearing impairments. This solution represents a significant advancement in accessibility, decreasing reliance on expensive devices and enhancing parental autonomy.**

Keywords— **Internet of Things, Machine Learning, LSTM (Long Short-Term Memory), MFCC (Mel-Frequency Cepstral Coefficients), Baby Monitoring, Deaf Parents, Accessibility**

I. Introduction

Technology has profoundly transformed people's lives over the years, enabling widespread connectivity with machines and internet-connected devices. This has significantly facilitated inclusion and become a crucial ally for individuals with disabilities, including those with hearing loss can be significantly impact people's routines and daily lives, and they often require assistance to carry out their daily activities. However, using technology, such as sensors, microcontrollers, and microphones, can provide greater autonomy and facilitate interaction with other people [1], [2].

Taking care of a newborn baby is not an easy task. Parents often use baby monitoring devices to keep an eye on their young children at night and when they have to leave for work. However, the alert mechanism of these devices is usually triggered by audio rather than visual information, which makes it difficult for deaf parents to perceive the sounds that the baby is making at the time [3], [4]. This technological gap can create a significant barrier for hearing-impaired parents, who need adapted solutions beyond traditional monitoring models based exclusively on audio.

Several technological solutions for deaf parents have been developed, including vibration devices, mobile applications, and IoT-based systems [1], [3], [5]. These approaches investigate using sensors and neural networks to capture sounds and other data. However, the technical and cost challenges still vary between solutions, depending on the contexts and conditions of use reported in the analyzed works. These issues point to the need for systems that are accessible, reliable, and integrated.

This paper aims to develop an intelligent baby monitoring system for deaf parents. Unlike previous approaches, the proposed system combines a lightweight, low-cost architecture with high accuracy in controlled experiments, making it suitable for implementation in embedded systems. The solution integrates modern machine learning techniques, specifically LSTM networks, and signal processing methods, such as Mel-Frequency Cepstral Coefficients (MFCC). These techniques extract relevant features from audio signals to identify a baby's cry, enabling the system to send real-time visual and tactile alerts to parents. This combination ensures high precision and efficiency, making the system robust for real-world applications and suitable for embedded systems.

II. Related Work

Technology has played a significant role in people's lives, becoming increasingly integrated into their daily routines and contributing to the execution of everyday activities. In recent years, there has been a significant increase in the number of individuals with special needs who require assistance. Technology offers them the means to live as normally as possible, for instance, enabling deaf parents to detect their baby's crying. Previous studies have presented diverse approaches, but there is room for advances in precision and integration with new technologies.

Paper [1] proposes a new algorithm to monitor children during sleep. Upon detecting any change in their environment, the smart monitoring system alerts parents by shaking a wearable bracelet or ringing their mobile phone, enabling deaf parents to monitor their children through periodic recording of sound and/or images.

Paper [2] developed a baby monitoring system utilizing Raspberry Pi, IoT sensors, and a Convolutional Neural Network (CNN). Data is processed and analyzed by the CNN to accurately detect abnormal patterns and behaviors in the baby's activities. The Raspberry Pi serves as the processing hub for this IoT-based baby monitoring system, capable of Wi-Fi communication. The system achieved 97% accuracy and 96% precision.

Paper [5] proposes an IoT device built using a lightweight Recurrent Neural Network (RNN) architecture that monitors a baby's vital signs, safety, and environmental conditions in real time. The baby is monitored using humidity (DHT22) and temperature (DS18b120) sensors, and a microphone module is incorporated to detect the baby's crying. The data is sent to the Raspberry Pi 4 microcontroller, which is integrated with a Wi-Fi module and alerts the parents.

Paper [3] developed a smart tool (smart bracelet) and a unique application for smartphones to alert deaf parents when their children call them by vibration after identifying the nature of the call and matching it with a list of words and situations that can be managed and modified in the application by the parents.

Paper [4] proposes an IoT-based alarm system that warns non-hearing users to value their safety as its highest priority. The technology uses multiple vibration frequencies to notify the user, and smart wearable devices such as smartwatches display live images of the external environment while a camera is recording them.

Paper [6] developed an intelligent baby monitoring system that detects infant crying and movements within the crib and monitors room temperature. This system utilizes a Raspberry Pi 3 B+ board, a Pi camera, and sound and temperature sensors to gather information about the baby. For intelligent monitoring, it leverages Convolutional Neural Networks (CNN) to identify and interpret the baby's status in its crib. Results from 500 iterations showed 99.9% accuracy during training and 93.2% during validation. This work was tested in controlled scenarios, similar to our proposal, underscoring the importance of expanding tests to more varied environments.

Paper [7] proposes an Internet of Things (IoT)-based baby monitoring system designed for innovative cribs, implemented to meet infant needs and provide information to parents. The crib includes an MP3 player for soothing music, a temperature detector, and a bedwetting sensor embedded in the ESP32 WROOM platform (microcontroller). High-speed internet connectivity facilitates the seamless use of the IoT platform, and any detected abnormality concerning the baby will be reported to parents via SMS using GSM.

Paper [8] has developed a highly efficient IoT-based baby monitoring system for real-time monitoring connected to the crib. The measured parameters about the baby's health, such as temperature, heart rate, and humidity in the crib, will be displayed on the mobile application. If the recorded readings show any abnormality, necessary actions such as controlling the temperature, turning the fan on or off, setting the crib movement, and playing music for the baby will be taken, and the caregiver, along with the parents, will receive an alert message.

Paper [9] proposes an IoT-based intelligent system that functions as a baby crib monitoring system. A Raspberry Pi B+ module provides overall hardware control. A MIC capacitor (microphone) is implemented for infant crying detection, a PIR motion sensor is designed to identify the baby's movement, and a Pi camera captures the baby's activity, with a display showing the video output of the sleeping baby.

To facilitate the analysis of main approaches and highlight the distinctions of the proposed system, Table I summarizes related works, detailing their precision and technologies used.

The analysis of related work shows that the proposed system more directly addresses the needs of deaf parents by combining high accuracy with a lightweight and accessible architecture. While previous approaches, such as [6], are limited by being tested only in controlled scenarios, the current proposal shares this challenge, underscoring the need to validate the system in environments with varying noise levels and a greater diversity of input data. Although the Rock Pi has a higher initial cost than the Raspberry Pi 3, its choice is justified by the performance required to run machine learning models directly on the board without additional hardware or cloud processing. This renders the system scalable and accessible in the long term, particularly when compared to commercial devices such as smartwatches or advanced camera-based systems.

III. Proposed System

A. Architecture of the proposed system

In Fig. 1, we present the architecture of the proposed system. The architecture is composed of three stages: (i) Data Acquisition, (ii) Data Processing, and (iii) Alerts.

1) Data Acquisition: This step consists of acquiring data using an INMP441 microphone that captures the baby's crying audio and saves it in the internal memory of the ESP32 microcontroller in .wav format. The microcontroller is a web server capable of making the audio file available for download through an access link. Any connected device that has access link can download the file. The Rock Pi downloads this data, which is used as input for training the model. We present the audio capture prototype in Fig. 2.

2) Data Processing, Training and Testing: The Mel Frequency Cepstral Coefficients (MFCCs) extraction operation is applied to better represent the audio characteristics for training purposes, which represents information on a logarithmic scale like human auditory perception. With this operation, the lower frequencies are highlighted to the detriment of the higher frequencies, causing the learning model to focus on the most perceptible audio characteristics. The choice of MFCC is especially relevant, as it reduces the dimensionality of the data by about 12 times compared to the original audio, in addition to eliminating noise and irrelevant variations, optimizing the processing and efficiency of the model.

Thus, a set of MFCCs will be generated for each audio file. These sets are used as samples for both training and testing. It is important to note that randomization of these coefficients is not applied here so that each set of MFCCs uniquely represents each audio signal.

We used a combination of LSTM (Long Short-Term Memory) networks and a dense layer at the output as the architecture to build the classification model. The LSTM network is especially suitable for handling temporal dependencies of MFCCs, maintaining long-term memory while discarding irrelevant information, considering the entire context. The adopted methodology allows modeling the temporal transitions present in audio patterns, ideal for differentiating complex sounds, such as baby crying, from other noises.

The database is made up of two classes. One class comes from audios containing baby crying and another class is extracted from audios with sounds other than baby crying (cars, horns, dogs, cats, hiccups, among others). The database is divided into training and testing sets, where the learning algorithms are applied in a defined number of epochs.

The obtained model is then loaded onto the Rock Pi board, where a new set of crying data is applied to verify the final classification accuracy. Future improvements to the model may include the use of compression techniques, such as quantization or pruning, to reduce power consumption and optimize computational performance in embedded systems. This data is obtained by Rock Pi from a new acquisition made by the ESP32/Microphone module, and saved on the web server.

3) Alerts: When the baby's crying is detected, an alert is sent to the lamp and the bracelet. In Figs. 3 and 4, we present the Lamp Activation Prototype and the Bracelet Prototype. These devices were chosen for their simplicity and affordable cost, as well as their ability to integrate with the proposed system. Future expansions of the proposal may investigate multimodal alerts, such as notifications in mobile applications.

TABLE I: Comparison of IoT-based Baby Cry Monitoring Systems.

Paper	Precision	Technology Used	Limitations
[1]	Not informed	IoT, Wearable Bracelet, Monitoring Algorithm	Dependence on specific devices, such as wristbands or cell phones compatible. Limitation of multimodal functionalities.
[2]	97%	IoT, Raspberry Pi, CNN, Mobile App	Image dependence (limitations in low light). Requires constant Wi-Fi connectivity.
[3]	Not informed	IoT, Smart Bracelet, Android App	Restricted to vibration for alerts. Exclusive to Android. Low accuracy due to simplicity of the system.
[4]	Not informed	IoT, Wearable Devices, Live Images	High dependence on constant connectivity. It does not mention the detection of specific sound patterns.
[5]	Not informed	IoT, RNN, Raspberry Pi 4, Humidity Sensors, Temperature and Pulse	Restricted to specific sensors. Dependence on expensive hardware. (Raspberry Pi 4). Does not address noise or sound interference.
[6]	93.2%	IoT, Raspberry Pi 3 B+, CNN, Sound and Temperature Sensors	Tested in controlled scenarios.
[7]	Not informed	IoT, Smart Crib, ESP32, GSM	Restricted to the crib environment, without multimodal integration.
[8]	Not informed	IoT, Mobile App, Motion Eye OS, Environmental Sensors	Greater focus on environmental and physiological parameters. High complexity due to multiple sensors.
[9]	Not informed	IoT, Raspberry Pi B+, MIC Capacitor, PIR Sensor	Restricted to the crib environment. Dependence on additional devices such as sensors and cameras.
Proposed Method	100% (in the tested data)	IoT, LSTM, MFCC, Rock Pi	Requires validation in more diverse environments.

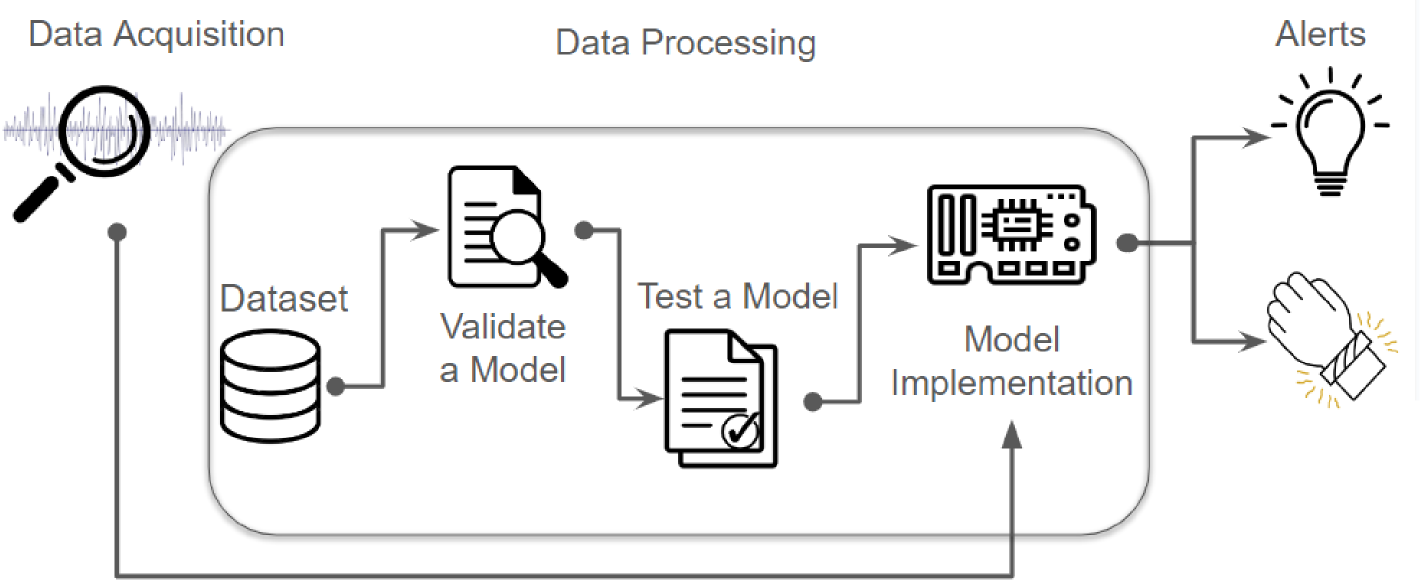

Fig. 1: Diagram of the architecture of the proposed system.

IV. Experimental procedures

1) Sensors: Initially, the KY-037 microphone was used to capture the audio. This microphone has a digital output and an analog output. The digital output provides a high or low logic signal depending on whether the detected signal exceeds a previously set limit. The analog output, in turn, provides an analog signal proportional to the received sound pressure, which would make it ideal for playback. For storage and, later, training and testing purposes, ESP analog-to-digital converters would have to be used, which, due to their low sampling rate and resolution, are not suitable for capturing high-fidelity audio. In this sense, the INMP441 digital microphone was adopted. This microphone uses MEMS (micro-electro-mechanical systems) digital technology, which are miniaturized electromechanical devices that capture physical signals and convert them to digital data that can be processed by microcontrollers via I²C or I²S. In ESP32, for example, I2S communication is used, with a sampling rate of 16 kHz, resolution of 16 bits, and ADPCM (Adaptive Differential Pulse Code Modulation) encoding for audio compression and bandwidth reduction. All audio captures take a fixed time of 5 s. For each MFCC application window, 13 coefficients are calculated.

2) Storage: When the capture process starts, the audio samples are sent to the ESP32 SRAM and, from there, to a web server until the 5s audio capture is finished. When the capture is finished, the audio recording is available on the server and the memory used in the acquisition is freed so that new acquisitions can be performed.

3) Training and Testing: Labeling consisted of naming the audios, and data segmentation consisted of separating the data into 200 audios of baby crying and 1,400 audios of non-baby crying (audios of cars, horns, dogs, cats, among others). The dataset was divided into 80% for training and 20% for testing

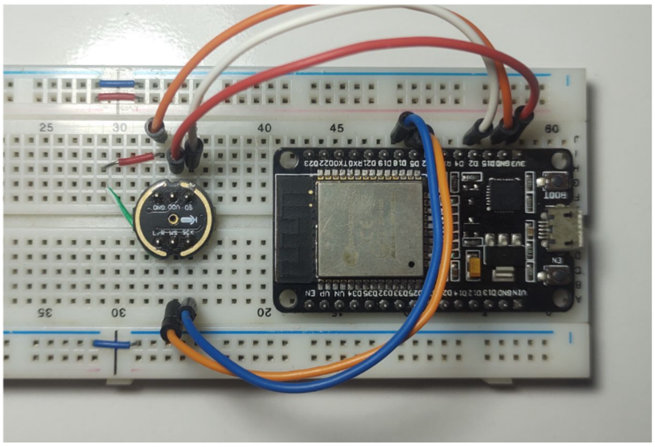

Fig. 2: Audio Capture Prototype.

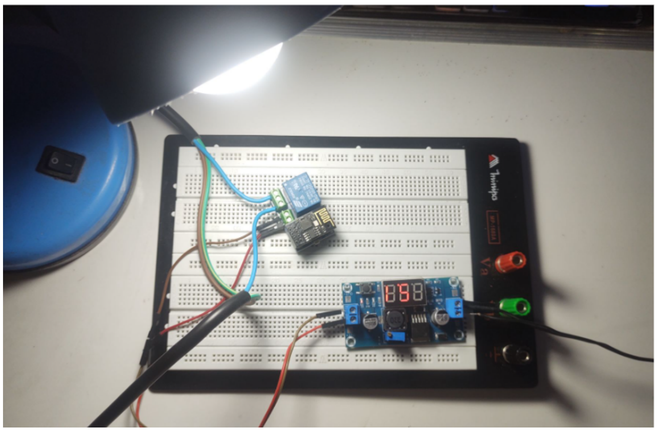

Fig. 3: Lamp Activation Prototype.

to verify the viability of the model. This dataset was chosen to represent real scenarios faced by deaf parents, where various types of noises and environmental sounds can coexist with a baby's crying. This representativeness was essential to ensure that the model could differentiate crying from other sounds.

The baby crying and non-crying data were obtained from both the YAMNet database and audio collected on the internet. To ensure consistency and standardization, all audios were adjusted to have a fixed duration of 5 seconds through scripts developed specifically for this purpose.

The experiments were performed using audio recordings of babies crying and unrelated sounds (such as cars, horns, and dogs, among others), to evaluate the model's performance in correctly identifying crying patterns. To this end, the tests sought to simulate common conditions, varying the characteristics of the collection environment and including typical noises from domestic environments. However, in future work, more experiments can be conducted in different scenarios to evaluate the robustness of the system in practical situations.

The LSTM network has 32 neurons. The learning algorithm used was sequential, with Adam optimization, binary cross-entropy error function and accuracy metric. The trained model was loaded into Rock Pi, where its viability was tested on a new dataset to verify the final accuracy.

4) Actuators: After classification in Rock Pi, where the presence of a baby crying is detected, two signals are sent, one goes to the relay and lamp prototype, simulating the turning on of the light in the parents' room, and the other, via Wi-Fi, to the ESP32 module and vibration sensor, simulating the bracelets that can be installed on the guardians.

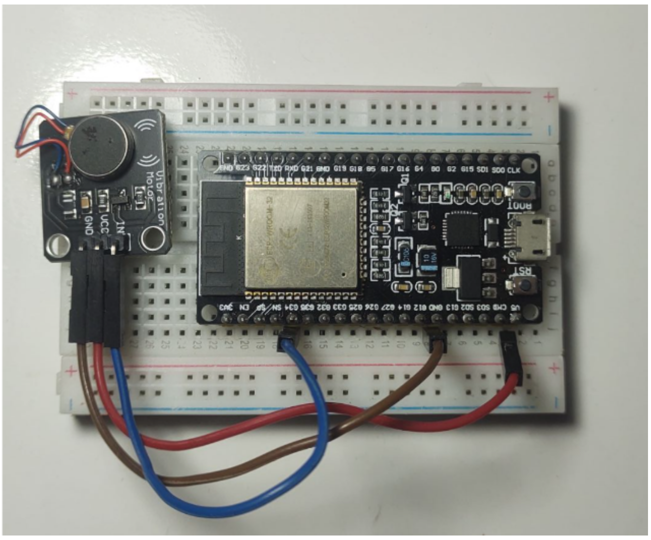

Fig. 4: Bracelet Prototype.

V. Experimental results

To compare the effectiveness of the model obtained in this work, two more models were trained: the first one used YAMNet, and the other one was obtained by transfer learning from YAMNet, where the labeling performed is binary, considering the crying and non-crying classes. We call this second model YAMNet Transfer. Table II presents the results obtained.

From the results obtained, it is possible to verify the effectiveness of the method developed for the classification of baby crying, achieving an accuracy of 100% for the 40 test audios. In addition to the accuracy, additional metrics were calculated to evaluate the model's performance in greater detail. The recall was 100%, indicating that almost all crying audios were correctly identified. The precision, which evaluates the proportion of correct predictions concerning the total of positive predictions, was 100%. The F1-score, which harmonizes recall and precision, reached 100%, demonstrating a robust balance between the metrics. These results reinforce the reliability of the model for the proposed task.

An important point to highlight is that YAMNet presented a very low prediction time compared to our model and YAMNet Transfer, despite the lower accuracy (77.5%). In this sense, it can be noted that the prediction time of our method can be improved in future work, since 25s is a high value, considering the test audio of 5s. Among the possible solutions, the application of model compression techniques, such as quantization and pruning, stands out, which can significantly reduce the prediction time. Another future approach is to replace the LSTM architecture with lighter neural networks, such as GRU or optimized versions of CNNs, without substantially compromising the model accuracy.

However, it is worth noting the accuracy of the method, which achieved an accuracy of 100% with the advantage of being much lighter than YAMNet and YAMNet Transfer, making it suitable for porting to embedded systems. Future analyses should investigate the impact of real-world conditions, including high noise levels or environments with

TABLE II: Comparison of IoT-based Baby Cry Monitoring Systems.

Model	**Testing Audios**	**Positive**	**Negative**	**Prediction Time (s)**	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**F1-score (%)**
YAMNet	40	31	9	1.31	77.5	100	77.5	88.75
YAMNet Transfer	40	40	0	25.85	100	100	100	100
Proposed Model	40	40	0	25.85	100	100	100	100

multiple sound sources, to validate the robustness of the model in different scenarios.

VI. Research Limitations

The baby monitoring system represents a significant advance in independence and peace of mind for deaf parents. However, like all technology, it can have some limitations that may affect its application, such as the specific environment in which it is used, interference from other electronic devices present, various nearby sounds, internet connection stability, variation in crying intensity, and the detection of false positives and false negatives.

Devices such as televisions and radios can interfere with the perception of a baby's crying, as can external sounds such as the noise of a blender in the kitchen near the baby's room, a vacuum cleaner running around the house, people talking, a car passing by on the street, or even someone ringing the doorbell. The system also depends on Wi-Fi connectivity for data transmission, making it vulnerable in places with unstable or no connection. Another limitation is that the baby's crying data currently needs to be stored first before being analyzed; ideally, this capture would be in real-time. These limitations highlight important areas for improving the system and guiding future research and development efforts.

More robust machine learning techniques can be incorporated to reduce false positives and false negatives, such as ensemble models or multimodal systems that combine audio with other inputs, such as movement data or the infant's respiratory rate. These improvements can increase accuracy and reduce the need to rely solely on audio features. Additionally, experiments are planned in real-world scenarios to mitigate noise interference, including domestic environments with different noise levels. These tests will allow us to assess the impact of different sounds and adjust the model to improve its robustness in practical conditions.

Regarding the dependence on Wi-Fi connectivity, alternative solutions such as temporary local storage and asynchronous data transmission when the connection is reestablished can make the system more resilient in locations with limited connectivity. Another possibility is to investigate the use of local communication protocols, such as Bluetooth, for environments where Wi-Fi is not viable. Although these limitations exist, the proposed system already presents promising results and demonstrates the potential to be adapted and improved to serve a wider audience in different scenarios.

VII. Conclusion

At the end of the study, the work demonstrated the effectiveness of IoT technology in creating a baby monitoring system for parents with disabilities, using an INMP441 microphone connected to a Rock Pi S microcontroller. By developing and testing a model that recognizes crying patterns, we achieved 100% accuracy in our final tests. This accuracy, combined with the system's lightweight architecture, contributes to the state of the art by offering a solution for deaf parents, overcoming challenges of existing systems such as high cost and dependence on specific hardware.

This system addresses the specific needs of deaf parents and paves the way for future health and safety monitoring applications using audio recognition. Future work could investigate the integration of new sensors, such as accelerometers or motion detectors, to enrich the multimodal analysis of infant behavior. In addition, model compression techniques, such as quantization or pruning, could be investigated to optimize real-time processing, reducing energy consumption and inference time. Other possibilities include validating the system in more varied scenarios, such as environments with different noise levels or challenging acoustic conditions, to increase its robustness in practical applications.

Acknowledgment

The authors would like to thank the Center of Electronic and Information Technology Research and Development (CETELI/UFAM) for the financial and institutional support to the authors of the article.

References

[1] N. M. Bahbouh, A. B. Alkhodre, A. A. A. Sen, A. Namoun, and S. S. Albouq, "A cost effective iot-based system for monitoring baby incidents by deaf parents," in *2019 International Conference on Advances in the Emerging Computing Technologies (AECT)*, 2020, pp. 1–6.

[2] S. Kumaran, C. Manoj Kumar, D. Yashwanth, and B. Deepak, "Cnn-powered baby monitoring system using internet of things sensors," in *2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS)*, 2023, pp. 938–943.

[3] A. A. Abi Sen, A. A. S Aljohani, N. M. Bahbouh, and O. Alhaboob, "Designing a smart bracelet based on arduino for deaf parents to interact with their children," in *2021 8th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2021, pp. 380–384.

[4] S. J. V, H. J. A, D. E. X. J, P. I, and T. Thiyagu, "Iot based wireless alert system for individuals with impaired hearing," in *2024 3rd International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, 2024, pp. 662–666.

[5] N. J. Kumar, H. R, A. N. Kumar, G. T, G. S, and T. A, "A lightweight (recurrent neural network) rnn architecture for iot based baby monitoring system to provide real time alerts and notification to parents or care takers," in *2023 Intelligent Computing and Control for Engineering and Business Systems (ICCEBS)*, 2023, pp. 1–6.

[6] R. Cheggou, S. S. h. mohand, O. Annad, and E. h. Khoumeri, "An intelligent baby monitoring system based on raspberry pi, iot sensors and convolutional neural network," in *2020 IEEE 21st International Conference on Information Reuse and Integration for Data Science (IRI)*, 2020, pp. 365–371.

[7] V. P. Hotur, A. T. N. R, A. P, C. R, and A. B. B, "Internet of things-based baby monitoring system for smart cradle," in *2021 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C)*, 2021, pp. 265–270.

[8] N. L. Pratap, K. Anuroop, P. N. Devi, A. Sandeep, and S. Nalajala, "Iot based smart cradle for baby monitoring system," in *2021 6th International Conference on Inventive Computation Technologies (ICICT)*, 2021, pp. 1298–1303.

[9] N. Saude and P. H. Vardhini, "Iot based smart baby cradle system using raspberry pi b+," in *2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC)*, 2020, pp. 273–278.