Unravelling Infant Communication Through Advanced Machine Learning for Enhanced Baby Care

Kalyani Phursule
23229268
Master of Science in Data Analytics
National College of Ireland

Abstract

Accurately knowing the underlying causes of infant cries is valuable for ensuring the well-being of babies and aid parents and caregivers in addressing their needs effectively. This paper aims to explores application of the advanced machine learning algorithms to classify infant cries into predefined distinct classes. To achieve this, multiple audio feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, Spectrograms are explored to capture relevant data from the audio signals. By employing models such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), CNN combined with XGBoost (ConvXGB), Support Vector Machine (SVM) and Random Forest (RF), the most efficient approach is targeted to be identified for the classification task. In this study thorough data preprocessing, feature extraction and model evaluation are performed to identify algorithm with highest performance in baby cry classification. This study is capable to providing valuable insights of baby behaviour and help parents catering baby's needs effectively.

Keywords: Infant cry classification, Audio features extraction, Machine learning, Audio analysis, CNN, XGBoost, RNN, SVM, RF

1. Introduction

Languages helps to communicate sentiments to any grown-up person, but baby cries are their primary mean of communication, to signal any discomfort or need. It is the biggest challenge to decipher infant's cry to parents and care givers especially in case of lack of experience. Even though the parents are very attentive, it is still difficult for them to understand the underlying reason of a baby's cry. It is crucial to understand and interpret baby cries for addressing their needs promptly, leading to improved infant wellbeing and parental responsiveness. Conventionally, parents and care-givers rely on personal experience and intuitions to interpret these cries, which could be subjective as well as inconsistent. Using machine learning algorithms, a promising solution can be offered with consistency and accuracy in cry classification. This research emphasises on application of advanced machine learning techniques to classify infant cries into denominators: 'hungry', 'burping', 'discomfort', 'belly-pain' and 'tired'. This approach can significantly reduce response time to address infant's needs, which will impactfully enhance baby care.

The significance of this study is to provide a robust approach to accurately classify baby cries for enhanced infant care, which will help caregivers to respond timely, reducing baby's distress. An objective cry classification model can provide a valuable support to new parents, which helps them to understand and

address their baby's needs and early detection of potential health issues. In addition to this, the study will contribute to the knowledge on infant behaviour and communication. Also, integration of the advanced cry classification with baby care technologies such as baby monitors can bring this technology one step closer to intelligent assistance in parental support.

The existing research explores a range of approaches to audio signal classification, involving feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCC) and Spectrograms, and different machine learning models like CNN and RNN. However, the work focusing on baby cries is limited, indicating the need for more investigation further.

This paper addresses following research questions:

- How do different audio features extraction techniques (for example, MFCCs, chroma features, spectrograms) impact the performance of infant cry classification models?
- How does hybrid Convolutional Neural Networks with XGBoost(ConvXGB) perform compared to Convolutional Neural Networks (CNN) , Recurrent Neural Networks (RNN) models , Support Vector Machine (SVM) and Random Forest?

To answer these questions, the proposed research involves extracting relevant audio signal features, training the machine learning models and evaluating their performance. In this study, to handle the classification problem of a dataset of labeled baby cries, development and comparison of ConvXGB, CNN, RNN, SVM and RF is demonstrated.

This study aims to contribute to the existing work through a rigorous comparative analysis of advanced machine techniques for infant cry classification, outlining the most effective feature extraction technique for this task. Moreover, the research will provide insights into the practical challenges involved in deploying these models within real-work environments, aiming to develop robust and accurate solutions. Ultimately, the study aims to enhance infant care by providing a reliable tool to caregivers to better understand baby's needs and improve responsiveness.

The document is organized as follows:

In **Section 2** (Literature Review), provides a comprehensive review of previous on audio classification, with emphasis on the challenges and techniques associated with infant cry analysis. This section provides a base for proposed research by detecting gaps and potential direction for improvements. **Section 3 (**Research Method and Specification) focuses on the methodological approach, and offers details information about data collection process, feature extraction methods, building machine learning models and evaluation metrices used to assess performance. **Section 4** (Results and Evaluation), presents the findings of the research, including performance of models and comparative analysis of various methods. **Section 5** (Conclusion and Future Work) , this is the final section, which summarizes the key findings, provides the implication of results and outlines recommendations for future research.

2. Literature Review

A. Feature Extraction for Baby Cry Classification

Liu et al. (2019) focused on analysing infant cries employing advanced audio analysis techniques to extract features from audio signal including Linear Predictive Coding (LPC), Mel Frequency Cepstral Coefficients (MFCC) and others. The experiment implemented machine learning and compressed sensing to classify these signals, the results indicate that the methods can accurately recognize and classify infant cries. This work highlighted the importance of feature extraction setting a foundation for later research in this area. Building on this, Ekinci and Küçükkülahlı (2023), introduced data augmentation and utilized audio feature Mel Frequency Cepstral Coefficients (MFCC) to improve the accuracy of cry class determination. Their approach highlights that data handling also significant along with audio features to enhance the classification model outcomes. This evolution towards sophisticated techniques starting from basic classification shows ongoing advancement in the field, indicating enhanced approach to interpret infant cries.

B. Role of Machine Learning in analysing Infant Cry

Numerous studies have explored classification of audio using machine learning models. Schmuck et al. (2019) explored the application of Support Vector Machines (SVM) and Random Forest (RF) for classifying infant cries into different classes. The study used Mel Frequency Cepstral Coefficients (MFCC) as features and resulted in moderate success, indicating models limited performance of SVM compared to RF to capture complex pattern in audio data. In contrast, Kulkarni et al. (2021) employed Random Forests, which demonstrated better performance due to the model's robustness against overfitting.

C. Advanced Deep Learning Approaches

The potential of deep learning models in cry classification is explored by Ozseven (2023), showcasing that these advanced models outperform traditional methods by leveraging large datasets. The shift towards these deep learning models demonstrates broader capabilities where the model can learn to recognize cry patterns without intervention, leading more accurate results. Conversely, Tusty et al. (2020) focused on research where Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are combined to temporal features of baby cries. Their analysis highlights the importance of model architecture attaining high classification accuracy. Both of these studies illustrate the diverse approaches with deep learning applications, signifying the choice of model can impact he effectiveness of classification task.

D. Innovating Solutions: Genetic Algorithms and Ensemble Modelling

Bashiri and Hosseinkhani (2020) introduced genetic algorithms in along with artificial neural networks, targeting to optimize neural network parameters for improved model accuracy in classification. This approach brings a new perspective to enhance performance in cry classification using hybrid methodologies. On the other hand, Cohen et al. (2020) employed ensemble classification techniques, indicating that combining multiple classifiers can effectively reduce incorrect classifications. This method is called as ensemble classification, demonstrates the significant impact of combining different approaches. Both studies underlines the growing trend of using hybrid methods in machine learning to achieve better outcomes, suggesting new possibilities to interpret infant cries with the help of these models.

E. Comparative Analysis

Dewi et al. (2019), analysed baby classification utilizing audio features MFCC and Linear Predictive Coding (LPC) with machine learning models Support Vector Machine (SVM) and K-Nearest Neighbors (KNN). The research highlights SVM as best performing model and the importance of feature selection and large dataset can improve model performance further. On contrary, Pradhan et al. (2022), proposed a system to classify infant cries, using traditional machine learning models K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) as well as deep learning models like ResNet and EfficientNet to classify the cry types based on audio signal features like Mel-Frequency Cepstral Coefficients (MFCC) and spectrograms. The comparative study demonstrated that the deep learning model ResNet outperformed other models achieving 72.6% accuracy.

F. Recent Developments in Cry Classification

Latest research has introduced several innovative techniques for cry classification. For instance Matikolaie et al. (2022) has created an automated newborn cry diagnostic system using machine learning, which focuses on the practical application of these technologies in real-world. Building upon this Joshi et al. (2022), who demonstrates a multistage heterogeneous stacking ensemble model which significantly improved classification performance. The integration of these advanced techniques highlight the rapid advancement in cry classification techniques. Furthermore, the work of Lahmiri et al. (2021) employed deep feedforward neural networks using cepstrum analysis with to diagnose infant cries. This approach demonstrated the effectiveness of integrating traditional signal processing techniques with advanced machine learning methods.

G. Research Niche and Expected contribution

Analysis the existing work illuminates the significance of advancements in baby cry classification with the help of machine learning techniques. Research highlights the significant roles of feature extraction, model architecture and innovative approaches in improving classification accuracy. Although diverse methodologies have been explored, there is an opportunity to develop hybrid models combining feature extraction with advanced machine learning models. This research fills the gap by providing a details comparison of multiple feature techniques Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, Spectrograms combined with machine learning techniques Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), CNN combined with XGBoost (ConvXGB), Support Vector Machine (SVM) and Random Forest (RF) for infant cry classification. Ultimately contribution of the through comparative study is to provide most effective feature extraction technique as well as the machine learning model, which will help to improve baby care and fostering healthier development for infants.

3. Research Method & Specification

In this section the methodological approach is described to classify infant cries using various feature and advanced machine learning algorithms. The CRISP-DM (Cross-Industry Standard Process for Data Mining) framework is utilized in the research for a structed and systematic approach to address the research question.

The aim of this study is to develop a approach using relevant feature extraction using machine learning system capable of identifying infant cries into categories like hunger, burping, discomfort, belly pain and tiredness. Multiple audio feature extraction methods such as Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, and Spectrograms are employed to capture relevant data from the audio signals. By implementing models like Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), CNN combined with XGBoost (ConvXGB), Support Vector Machine (SVM) and Random Forest (RF), the project targets the most effective approach for this classification task.

The Key Components of the solution are as follows [Figure 1]:

1. Feature Extraction: To capture distinct characteristics of audio signals methods used are Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, and Spectrograms
2. Model Development: To determine most accurate classifier model various machine learning models are trained and optimized.

3. Evaluation: Build machine learning models are compare based on performance metrices: accuracy, precision, recall and F1 score.
4. Deployment: Best-performing model is deployed as a user-friendly application for real-time cry classification.

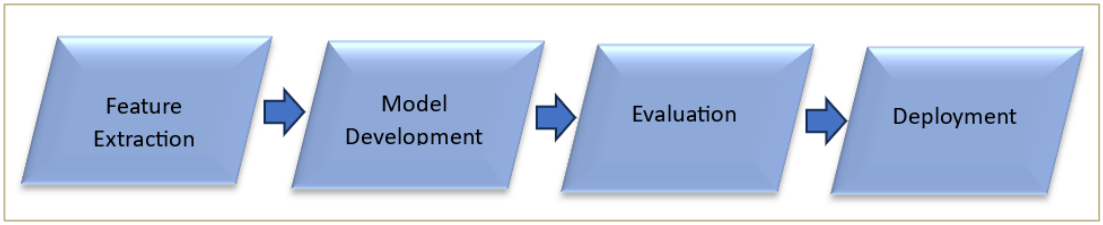

Figure 1 : Key Component of the solution

The project is designed in various phases [Figure 2], each having specific tasks to ensure system progress to achieve the research objectives.

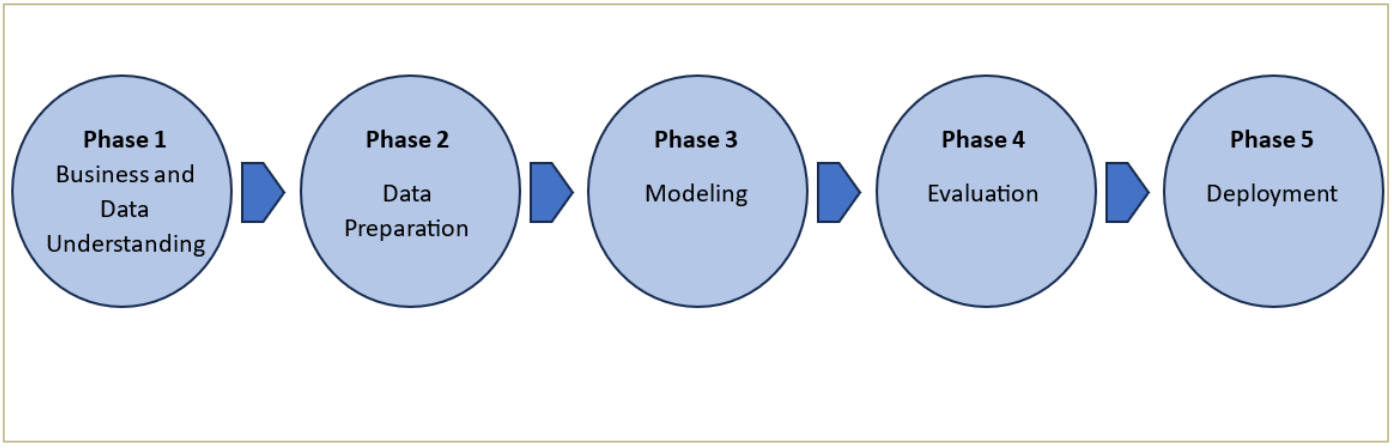

Figure 2 : Project Phases using CRISP-DM

Phase 1 : Business and Data Understanding

Understand the problem statement and define project objectives in this phase

Actives :

- Perform thorough literature review on baby cry classification and related methodologies for feature extraction and machine learning algorithms.
- Define the research questions and draw the project scope
- Identify datasets by exploring available data repositories

Phase 2 : Data Preparation

Gather and preprocess data, which is suitable for analysis

Actives :

- Data collection : Collect pre labelled infant cry audio recordings from open data repositories.
- Data Cleaning: Clean audio by removing noise, handling missing values and standardizing audio formats
- Feature Extraction : Employ feature extraction techniques : MFCC, Chroma and Spectrogram

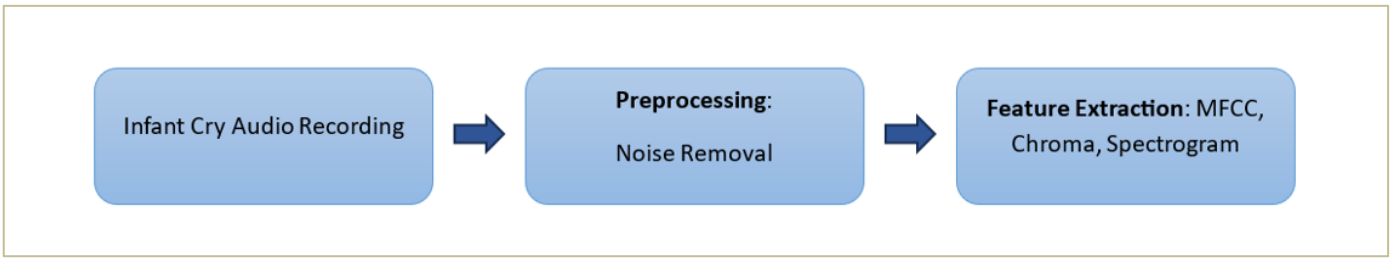

Figure 3 : Data Preparation

Phase 3: Model Building

Develop machine learning models to cry classification and optimise further

Actives :

- **Selecting Model :** Apply Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), CNN combined with XGBoost (ConvXGB), Support Vector Machine (SVM) and Random Forest (RF) models
- **Training and Optimization :** Create training , validation and test datasets splits. Optimize models using hyperparameter tuning and cross validation.

Phase 4: Evaluation

Model evaluation is done based on performance metrices and efficient model is identified

Actives :

- **Performance Metrices :** Model assessment using accuracy, precision, recall and F1 score
- **Comparison and Analysis :** Selection of best approach based on comparison of model results.
- **Confusion Matrix Analysis :** Identify misclassification using confusion matrices

Phase 5:

- **Deployment**

  Deploy identified best preforming model into real-time user-friendly application for infant-cry classification.

- **Documentation and Presentation**

  Compile the research findings into a report and prepare presentation to communicate the obtained insights and results.

Gantt Chart : Below chat illustrates project timeline

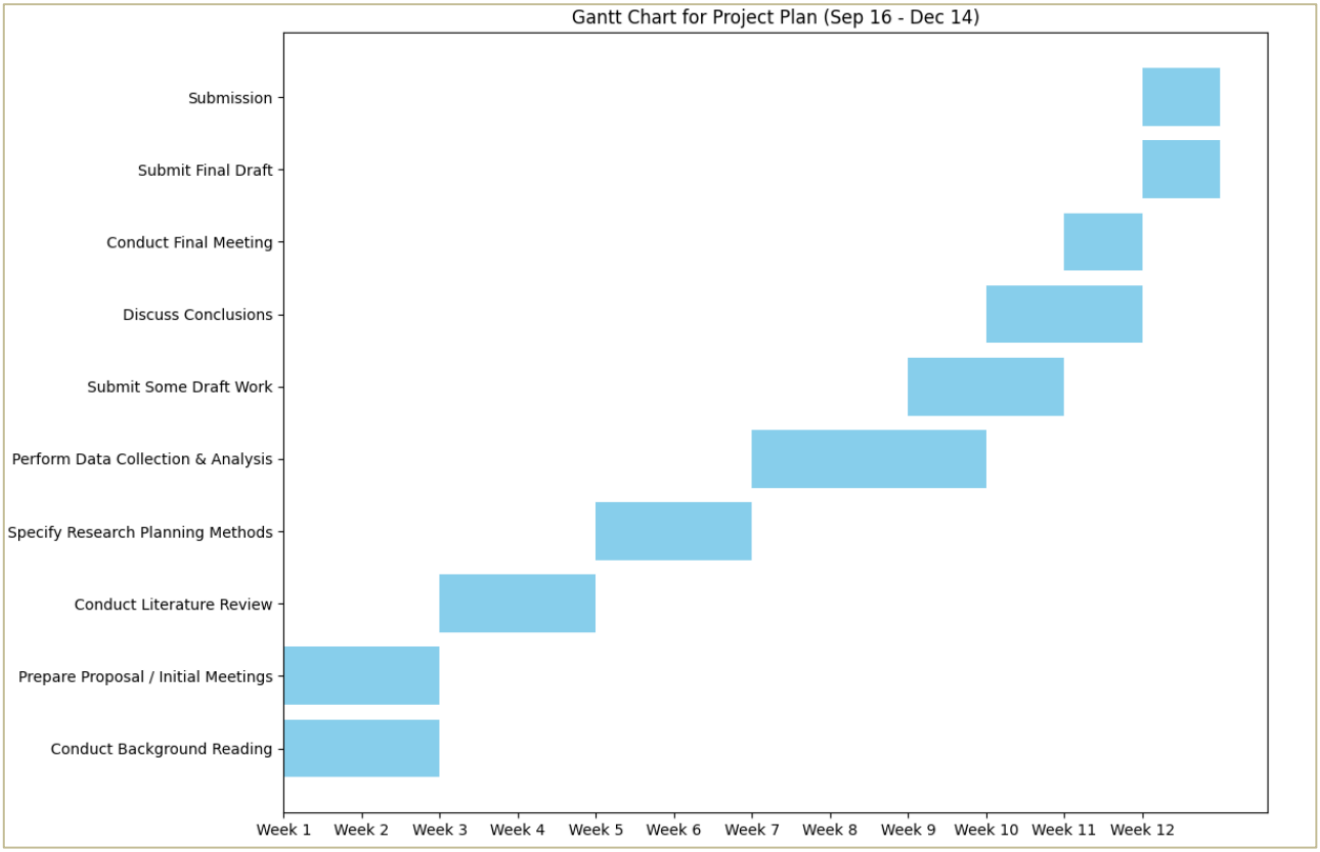

Figure 4 Gantt Chart

Tools used in the project:

- **Programming Technique : Python**
- **Libraries :**
  - **Librosa :** Used for audio signal processing and feature extraction
  - **Numpy/Pandas :** Data manipulation and analysis
  - **Tensorflow/Keras :** For implementing deep learning models

  - **Scikit-learn :** For building SVM, Random Forest
  - **Matplotlib/Seaborn :** Data Visualization
- **Development Environment :** Jupyter Notebook
- **Version Control :** Github for code management

Test Data :

- **Dataset :** Pre labelled audio recordings of baby cries is sourced from public repositories
- **Data Augmentation :** Data augmentation techniques like pitch shifting and time stretching will be applied to enhance dataset diversity and robustness.

Evaluation Plan

The models evaluation will be done using the following performance metrices :

- Accuracy : Overall correctness of the model is measured
- Precision : Demonstrates the ration of true positive predictions among all positive predictions
- Recall : Shows model's ability to capture all relevant instances
- F1 Score : Harmonic mean of precision and recall, providing a balanced measure of model performance.

Experimental Evidence

- Confusion Matrix: Confusion matrices to be used to identify patterns of misclassification and improve model accuracy
- Cross-Validation: Utilize cross-validation to make sure model generalization and robustness
- Feature Importance: Evaluation of the importance of various features (MFCC, Chroma, Spectrogram) in contributing to model performance

Ethical Considerations

Ethical consideration includes :

- Data Privacy : Ensuring the data privacy of audio dataset by make use of anonymous data
- Informed Consent : Ensuring that the collected ata comply with ethical guidelines, including obtaining informed consent from parents or guardians
- Bias : Assessing the models for potential bias and mitigating the issue with proper actions

REFERENCES

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.

Ekinci, A., & Küçükkülahlı, E. (2023). Classification of Baby Cries Using Machine Learning Algorithms. Eastern Anatolian Journal of Science, 9(1), 16-26.

Schmuck, N., Michaelis, J., & Walther, S. (2019, August 28). A Pragmatic Approach for Infant Cry Analysis Using Support Vector Machine and Random Forest Classifiers. Springer Professional. https://link.springer.com/article/10.1007/s11277-024-11491-8

P. Kulkarni, S. Umarani, V. Diwan, V. Korde and P. P. Rege, Child Cry Classification - An Analysis of Features and Models, 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-7, doi: 10.1109/I2CT51068.2021.9418129

Turgut Ozseven, Infant cry classification by using different deep neural network models and hand-crafted features, Biomedical Signal Processing and Control, Volume 83, 2023, 104648, ISSN 1746-8094, https://doi.org/10.1016/j.bspc.2023.104648.

Tusty, N. M., Basaruddin, T., & Krisnadhi, A. (2020).). Infant cry classification using CNN – RNN. Journal of Physics Conference Series, 1528(1), 012019. https://doi.org/10.1088/1742-6596/1528/1/012019

Bashiri, A., & Hosseinkhani, R. (2020). Infant crying classification by using genetic algorithm and artificial neural network. ACTA MEDICA IRANICA. https://doi.org/10.18502/acta.v58i10.4916

Cohen, R., Ruinskiy, D., Zickfeld, J., Ijzerman, H., & Lavner, Y. (2019). Baby cry Detection: Deep learning and classical approaches. In Studies in computational intelligence (pp. 171–196). https://doi.org/10.1007/978-3-030-31764-5_7

S. P. Dewi, A. L. Prasasti, and B. Irawan, "Classification of Infant Behavioral Traits Using Acoustic Cry: An Empirical Study," in 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS), Bali, Indonesia, 2019, pp. 86-91, doi: 10.1109/IoTaIS47347.2019.8980401.

B. N. Pradhan, G. R. Shah, S. Ankalaki, K. Shorya, K. Venkatesh, and A. Kushwaha, "Baby Cry Decoder: A Boon for the Hearing Impaired Caregiver," in 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), Bengaluru, India, 2022, pp. 1-5, doi: 10.1109/MysuruCon55714.2022.9972437.

Matikolaie, F. S., Kheddache, Y., & Tadj, C. (2022). Automated newborn cry diagnostic system using machine learning approach. Biomedical Signal Processing and Control, 73, 103434.

Joshi, V. R., Srinivasan, K., Vincent, P. M. D. R., Rajinikanth, V., & Chang, C. Y. (2022). A multistage heterogeneous stacking ensemble model for augmented infant cry classification. Frontiers in Public Health, 10, 819865.

Lahmiri, S., Tadj, C., & Gargour, C. (2021). Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks. IEEE Instrumentation and Measurement Magazine, 24(2), 24-29.