

Advancements in Infant Cry Classifications: A Literature Review on Methods and Systems

Athira R Kurup¹, Dr. B Ben Sujitha², and Deva Mahilan³

¹PhD research scholar, Computer Applications, Noorul Islam Centre for Higher Education, Kumaracoil

² Professor, Department of Computer Science and Engineering, Noorul Islam Centre for Higher Education, Kumaracoil

³BTECH UG Student, Department of Computer Science and Engineering, TOMS college of Engineering

Abstract—The use of spoken language, a complex instrument developed over thousands of years, is one of the essential elements of human communication. Unfortunately, throughout the first several years of life, we still lack this ability to communicate; instead, we cry to express pain, hunger, and other emotions. The literature review highlights the importance of pitch, intensity, and length in classification, the impact of medical issues and prematurity on cry features, and foundational work on cry acoustics. There are still problems with noise reduction, dataset standardization, and practical implementation because of individual differences in kid screams and ambient unpredictability. This literature review looks at the current state of research in infant cry classification with a focus on how the technology can be used for the identification of early developmental diseases, emotional state, and medical diagnosis. Advanced machine learning algorithms, deep learning algorithms, and audio analysis methods have been developed in recent years which enable an automated system to distinguish between screams that are made due to pain, discomfort, hunger or any other emotional or physical state.

Index Terms—Adaptive models, Annotated datasets, Automatic recognition, Cry analysis, Cry classification, Databases, Deep learning, Emotion detection, Ethical considerations, Feature fusion, Infant cry recognition, Lightweight algorithms, Machine learning, Multimodal approaches, Neural networks, Pain detection, Real-world applications, Signal processing, Transfer learning, Wavelet-based methods.

I. INTRODUCTION

In recent years, academics have become increasingly interested in the identification and analysis of newborn cries. Many academics are investigating a wide range of creative methods to successfully recognize,

classify, and interpret these screams. Knowing why babies scream is important because it helps medical workers identify emotional states of distress and possible health problems that may impact babies. This literature review summarizes the latest developments in this fascinating field and identifies important discoveries, research methodologies, and challenges faced by scientists. We can improve our ability to respond to the demands of newborns in a way that will ultimately lead to better care and better results for their wellbeing. This growing body of research could help parents better understand infant communication.

II. AUTOMATIC CLASSIFICATION OF INFANT CRIES

A lot of research on newborn cry analysis focuses on classifying cry signals into classes, such as hunger or discomfort. For cry categorization, several traditional methods have been employed, including decision trees, Support Vector machines (SVMs), and Artificial Neural Networks (ANNs). Due to the lack of labeled data, cry categorization is inherently problematic. This study [2] addresses this issue by introducing a hierarchical classification strategy leveraging data imbalance. The study suggests a multi-stage classification technique to increase classification systems' accuracy.

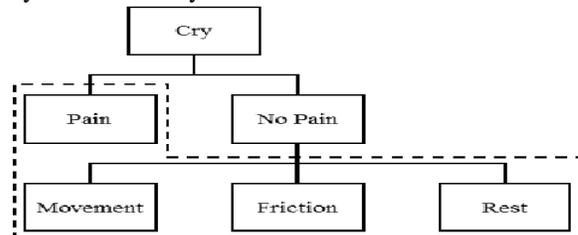


Figure 1 Flat Classification Method

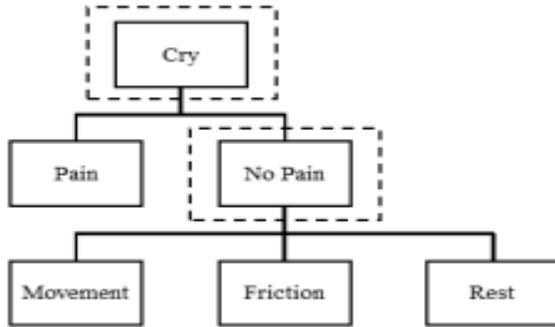


Figure 2 Local Classifier per parent node approach Pain and no-pain samples, which can again be further subdivided into three subclasses, make up the natural hierarchy of our dataset. Based on [2], we have chosen two primary methods for hierarchical organization. The first is the flat classification, which is the most straightforward approach to hierarchical classification problems and is shown in Fig. 1. This approach typically just predicts classes in the leaf nodes, ignoring the class hierarchy entirely. In essence, this turns our landscape into a typical multi-class classification problem. The other method, depicted in Fig. 2, was the local classifier per parent nodes technique, in which a multi-class classifier is trained to differentiate between each parent node in the class hierarchy and its offspring.

Furthermore, [21] looks at various algorithms for analyzing cries and notes that deep learning methods are now more popular than traditional machine learning. These methods are useful in real-world applications because they help manage larger and more complex information effectively. Automatically classifying newborn cries is typically a pattern recognition problem. It is comparable to the two primary steps of the Automatic Speech Recognition (ASR) process, which are pattern classification and signal processing (Fig. 3). The newborn baby's raw cry signal is fed into the automatic system. Normalization, cleaning, and filtering are the initial preprocessing steps in the signal processing step. The significant feature vectors are then obtained by analyzing the smoothed signal using appropriate feature extraction algorithms. Using appropriate feature selection or feature reduction techniques, the high-dimensionality retrieved features are reduced into lower-dimensional space. After that, the reduced feature vector goes through a classification phase in which it is compared to the computer's knowledge to get a result.

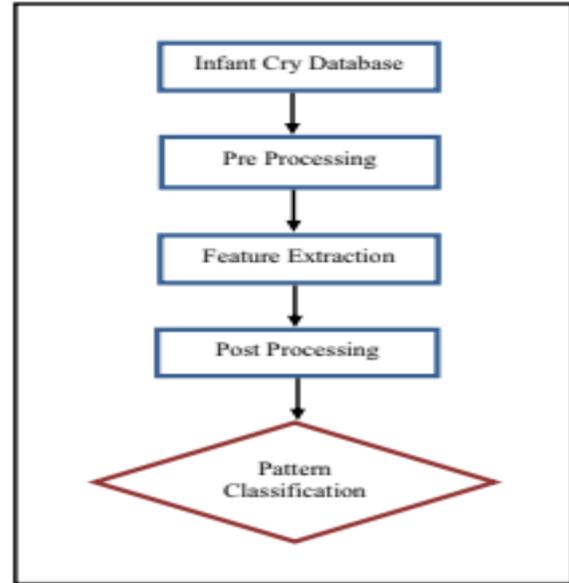


Figure 3 Automatic infant cry classification

III.SIGNAL PROCESSING METHODS FOR INFANT CRY ANALYSIS

Traditional signal processing techniques are critical to cry analysis because of their efficiency and dependability. Over time, they have survived and are still pretty useful for the acoustic characteristics of cry signals. For instance, using spectrogram analysis, researchers could gather valuable information from the signals for various applications. It can be said that combining machine learning and signal processing methods has proved the strength and promise of this combination, as presented in [6]. It is very noticeable in the context of cry categorization tasks where techniques such as time-frequency representations are commonly used as neural network inputs to improve the precision and accuracy of classification.

Currently, there are four types of images related to audio signal in literatures. They are spectrogram, mel-spectrograms, waveforms, and prosodic feature images. A spectrogram is an image that is a time-frequency representation of an audio clip. It is known that spectrogram has a strong ability to present the signal and include both acoustic and prosodic information. Spectrograms can be extracted through framing and windowing, FFT, and calculating the log of the filtered spectrum steps, which are illustrated in Figure 4. Feeding spectrograms into classifiers can solve the problem of different cry signals having different durations. Instead of using zero padding to

achieve same length of feature vectors, normalization is applied in the process of spectrogram generation, which produces the same size images without changing the original signal. A mel-spectrogram is another visual representation of a signal indicating how people hear sound by converting the Y-axis to mel-scale. Mel-spectrogram can relatively represent human sound perception characteristics, which presents the linear distribution under the 1000Hz and the logarithm growth above the 1000Hz on a logarithmic scale rather than a linear scale. People are more sensitive to lower frequency sound and the difference between high frequency sound is not as easy to distinguish as the ones between lower frequency sound. We hear frequencies on a logarithmic scale rather than a linear scale. In mel-spectrogram, each unit is judged by listeners to be equal in pitch distance from the next. Waveforms, representing the pattern of sound pressure amplitude in the time domain, are not as effective as spectrograms in deep learning models for infant cry

classification. We extract spectrograms, waveforms, prosodic lines images, and mel-spectrograms from infant cry signals and feed them into our neural network models for disease identification and cry type classification.

To capture information in both temporal and frequency, the study [7] presents wavelet transforms. This reduces computing costs without leading to a loss of precision in classification, which is a desired approach for real-time cry detection on the limited resources. Wavelet scattering coefficients (WSC) is indeed a feature extraction method from signals or images and is very useful for applications like pattern recognition, regression, or classification. This multi-scale, hierarchical method is founded on wavelet transforms and provides a way of identifying both local and global patterns in data. WSC combines the principle of scattering, which involves modifying the signal and then iteratively applying operations to capture pertinent patterns, with the power of wavelets,

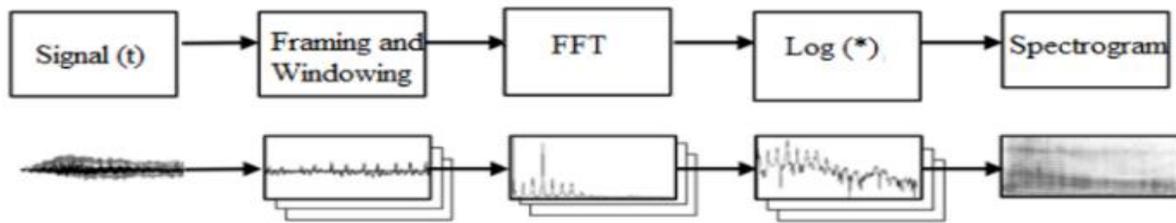


Figure 4 The flowchart of spectrogram generation

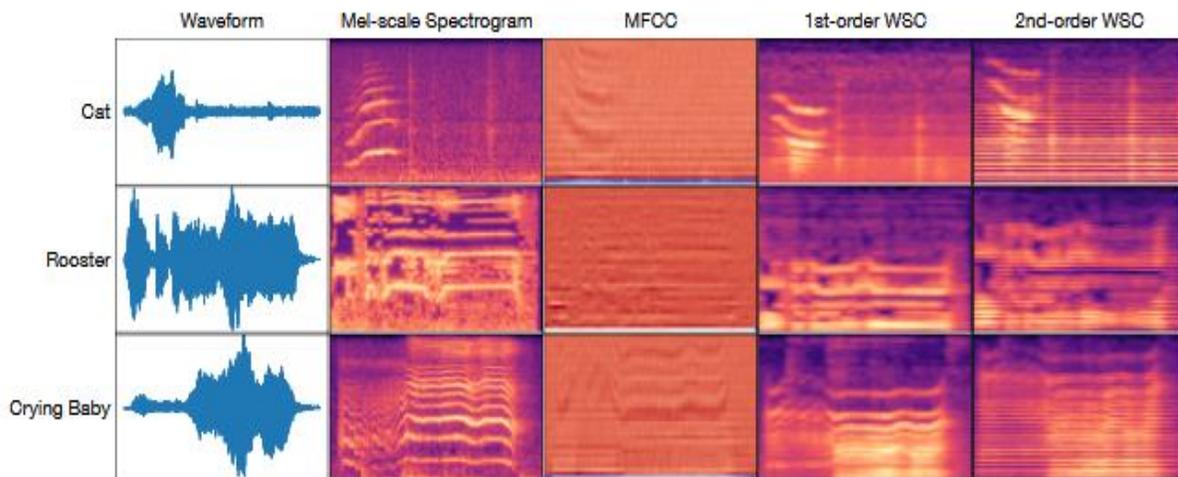


Figure 5 The first column displays the waveform, followed by the Mel-scale spectrogram in the second column, MFCC in the third column, first-order WSC in the fourth column, and second-order WSC in the fifth column. Each row corresponds to a specific sound class: the first represents a cat, the second represents a rooster, and the third showcases a crying baby.

which are functions used to shape signals or images in a way that shows multi-scale features.

IV. EMOTION AND PAIN DETECTION IN INFANT CRIES

An essential feature of cry analysis is the ability to distinguish between different states of emotion or disease conditions, such as pain. Spectral entropy features are utilized in [11] for the discrimination between normal and pain-induced cries. In this research study, the authors opined that multi-band analysis can be employed to classify the distinctive spectral features of pain-induced cries. As Fig. 6 normal crying features consist of rising-falling pitch contour, ascending-descending melody and high intensity seen from the spectrum. In pathological infant's crying spectral intensity will be lower than normal, rapid pitch shifts, weak phonation and silences during the crying as Fig. 7.

The infant's Automatic Cry Recognition (ACR) process is in generally, a pattern recognition problem similar to Automatic Speech Recognition (ASR). The system takes the wave of infant's cry as the system input, and at the end obtains the kind of cry. Generally, the process of ACR is done in two steps. The first step is known as signal processing, whereas the second is called pattern classification. First, we make an aim on feature extraction and for that task the cry signal is normalized and cleaned up from all noise then analyzed for the extraction of most important characteristics that we called "acoustical analysis". The set of such characteristics obtained could be considered as a vector which may stand for a pattern for the next step. In second step, the set of all vectors is used to train the classifier. Later on, a set of unknown feature vector, the test patterns, is compared with the knowledge that the computer has to measure the classification output efficiency. Fig. 8 shows the different stages of the described recognition process.

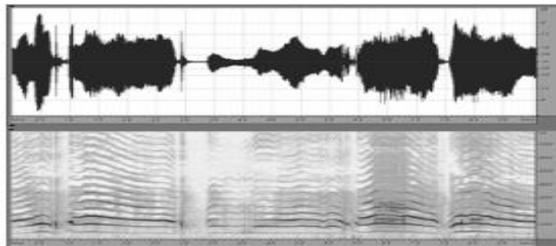


Figure 6 Waveform and spectrum of normal cry

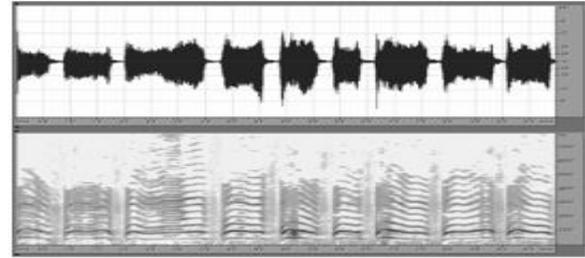


Figure 7 Waveform and spectrum of pathological cry

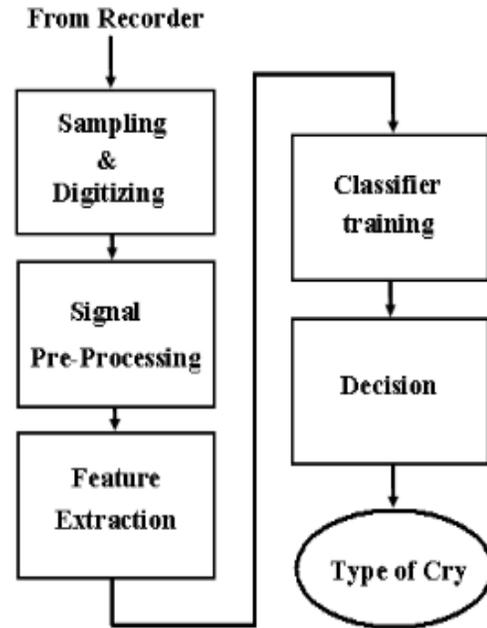


Figure 8 Automatic Cry Recognition (ACR) process

V. CHALLENGES AND FUTURE DIRECTION

1. **DATA SCARCITY:** The lack of diverse and annotated datasets remains one of the main barriers, even though valuable databases such as [16] and [17] are available. It is highlighted that international cooperation could be necessary for building up large-scale databases in papers like [15].
2. **MODEL GENERATION:** Many models work well in the laboratory setting but fail when applied to real-world scenarios because ambient noise and cry characteristics change significantly. More resilient and adaptable models are needed, according to [6] and [9].
3. **MULTIMODAL INTEGRATION:** A promising direction is to combine cry analysis with facial expression recognition, as seen in [13] and [14].

However, processing multimodal data efficiently is very complex, as reviewed in [22].

4. **ETHICAL CONSIDERATIONS:** The use of infant cry data does raise several issues concerning the ethics of privacy and consent. Data collection and analysis should then have proper frameworks on ethics. [21], [23].

VI. CONCLUSION

Rapid advancements in deep learning, signal processing, and multimodal techniques are propelling research in infant cry analysis. Despite the notable advancements, issues with data accessibility, generalization, and ethical considerations still exist. In conclusion, to create efficient and scalable solutions, future research should concentrate on developing different datasets, improving model robustness, and resolving ethical issues. The accuracy of cry detection systems has significantly increased due to the incorporation of deep learning models and sophisticated signal processing techniques, allowing for real-time monitoring and diagnostic assistance. There are still difficulties, nevertheless, especially when managing sizable, noisy, and varied datasets. For cry detection systems to be used in everyday settings, future research should concentrate on enhancing system scalability, real-time processing, and resilience.

REFERENCES

- [1] A self-training automatic infant-cry detector.
- [2] Automatic_Classification_of_Infants_Cry_Using_Data_Balancing_and_Hierarchical_Classification_Techniques.
- [3] Baby Cry Detection: Deep Learning and Classical Approaches.
- [4] Baby_Cry_Recognition_by_BCRNet_Using_Transfer_Learning_and_Deep_Feature_Fusion.
- [5] Infant Cry Signal Diagnostic System Using Deep Learning.
- [6] Infant Cry Signal Processing, Analysis, and Classification with Artificial Neural Networks.
- [7] Infant_Cry_Detection_with_Lightweight_Wavelet_Scat.
- [8] Convolutional Neural Networks for Audio-Based Continuous Infant Cry Monitoring at Home 2021.
- [9] Contributions to speech processing and ambient sound.
- [10] Evolutionary-Neural System to Classify Infant Cry Units for Pathologies 2008.
- [11] A System for Detecting Infants with Pain from Normal Infants Based on MultiBand Spectral Entropy by Infant's Cry Analysis.
- [12] Automatic Recognition Methods Supporting Pain 2019.
- [13] Emotion_Detection_From_Infant_Facial_Expressions.
- [14] A_Systematic_Review_on_Multimodal_Emotion_Recognition_Building_Blocks_Current_State_Applications_and_Challenges.
- [15] A Database of Neonatal Facial 2023.
- [16] FENP_A_Database_of_Neonatal_Facial_Expression_for_Pain_Analysis.
- [17] diagnostics-12-02802.
- [18] entropy-24-01194-v2.
- [19] A Comprehensive Literature Review on Children's 2021.
- [20] A review of infant cry analysis and classification.
- [21] Survey on automatic infant cry analysis and classification 2018.
- [22] Infant+Cry+Language+Analysis+and+Recognition+_An+Experimental+Approach.
- [23] Validating a model to detect infant crying from naturalistic audio.
- [24] A review of infant cry signal processing techniques.
- [25] Survey on classification methods for infant cries.
- [26] Real-world infant cry detection systems.
- [27] Scalable cry analysis using self-training methods.
- [28] Evolutionary neural systems in infant cry monitoring.