Predicting Baby Cry Causes Using Machine Learning Algorithms

1st Atul Kumar Rana
Computer Technology
Department of Electrical Engineering
IIT Delhi
New Delhi, India

2nd Ranjan Kumar Singha
Computer Technology
Department of Electrical Engineering
IIT Delhi
New Delhi, India

***Abstract*—This paper presents a study on predicting the causes of baby cries using machine learning algorithms. We explore the effectiveness of various algorithms in classifying baby cries into different categories such as hunger, discomfort, loneliness, and more. Our approach involves preprocessing the audio data, extracting relevant features using Mel-Frequency Cepstral Coefficients (MFCC), and training several machine learning models on the extracted features. The performance of each model is evaluated based on metrics such as accuracy, precision, and recall.**

***Index Terms*—MFCC, Random Forest, SVM, overfitting**

I. Introduction

INFANT crying is a fundamental aspect of early childhood development, serving as a primary means of communication for newborns to express their physiological and emotional needs. The ability to interpret and respond effectively to infant cries is essential for parental caregiving and infant well-being. However, deciphering the underlying causes of infant cries can be challenging, especially for first-time parents or inexperienced caregivers.

Traditionally, caregivers rely on their intuition and experience to interpret the nuances of infant cries, attempting to discern whether the baby is hungry, in discomfort, tired, or experiencing other forms of distress. While seasoned caregivers may develop a keen sense of understanding over time, the subjective nature of human perception introduces inherent limitations, leading to potential misinterpretations or delays in responding to infant needs.

In recent years, advancements in technology, particularly in the fields of machine learning and signal processing, have paved the way for innovative solutions to address the challenges associated with infant cry analysis. Automated classification systems, leveraging machine learning algorithms and audio signal processing techniques, offer the promise of providing objective and timely assessments of infant distress, complementing traditional caregiving practices.

Motivated by the potential benefits of automated cry analysis in improving infant care and parental responsiveness, this paper presents a comprehensive investigation into the classification of infant cries using machine learning algorithms. By leveraging a dataset of recorded infant cries, we explore the effectiveness of various machine learning models in accurately categorizing cry sounds into distinct classes corresponding to different underlying needs or states, such as hunger, discomfort, loneliness, and more.

Our study encompasses the entire pipeline of cry analysis, from data collection and preprocessing to feature extraction and model training. We employ state-of-the-art machine learning techniques, including Random Forest, Logistic Regression, Decision Tree, and Support Vector Machine (SVM), to develop predictive models capable of accurately classifying infant cries. Subsequently, we evaluate the performance of each model using metrics such as accuracy, precision, and recall, providing insights into their comparative effectiveness and applicability in real-world scenarios.

The findings of this study have significant implications for both parental caregiving and healthcare practices. Automated cry analysis systems have the potential to assist caregivers in identifying and responding to their infants' needs more promptly and accurately, thereby enhancing parental confidence and infant well-being. Furthermore, healthcare professionals can leverage automated cry analysis as a supplementary tool for clinical assessments, enabling personalized interventions and support for infants in distress.

In summary, this paper contributes to the growing body of research on machine learning applications in childcare and early childhood development. By advancing our understanding of infant cry analysis and classification, we aim to foster innovation in parental caregiving practices and facilitate the development of intelligent systems for infant health monitoring and support.

II. Method

A. Data Collection

The dataset used in this study consists of recorded audio samples of infant cries, collected from various sources such as online repositories, research databases, and real-world recordings. Each audio sample is accompanied by metadata indicating the corresponding cause or state of the infant cry, as reported by caregivers or healthcare professionals. The dataset encompasses a diverse range of cry types, including cries indicative of hunger, discomfort, loneliness, fatigue, and other physiological or emotional needs.

Fig. 1. 5 Stages

B. Signal Analysis

Prior to feature extraction, the raw audio signals undergo preprocessing and signal analysis to enhance their suitability for machine learning tasks. Preprocessing steps include normalization, noise reduction, and resampling to ensure uniformity across the dataset. Signal analysis techniques such as spectrogram generation and waveform visualization are employed to gain insights into the temporal and spectral characteristics of infant cries, facilitating the selection of appropriate features for classification.

1) Preprocessing:

- **Normalization:** Normalize the amplitude of the raw audio signals to a standard scale using the formula:

$$x_{\text{norm}} = \frac{x - \mu}{\sigma}$$

where x is the raw signal, μ is the mean, and σ is the standard deviation.

- **Noise Reduction:** Apply a noise reduction filter to the signals using techniques such as spectral subtraction or wavelet denoising.
- **Resampling:** Resample the signals to a common sampling rate f_s using linear interpolation or resampling algorithms.

2) Spectrogram Generation:

- Compute the Short-Time Fourier Transform (STFT) of the signals
- Calculate the magnitude or power spectrum of the STFT coefficients to obtain the spectrogram.

3) Waveform Visualization:

- Plot the waveform of the signals using time-domain representations.
- Visualize the spectrogram of the signals using heatmaps or contour plots to display their frequency content over time.

C. Feature Extraction

Feature extraction plays a crucial role in transforming raw audio signals into informative representations suitable for machine learning algorithms. In this study, we focus on extracting discriminative features from infant cry signals using Mel-Frequency Cepstral Coefficients (MFCCs), a widely used technique in audio signal processing. MFCCs capture the spectral envelope of the audio signal and are known to be effective in capturing relevant acoustic characteristics for

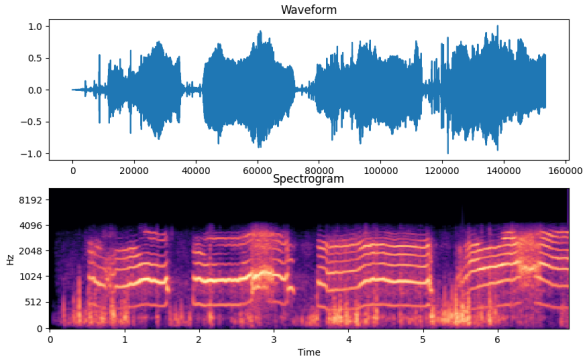

Fig. 2. Discomfort

speech and audio classification tasks. Additionally, other time-domain and frequency-domain features such as zero-crossing rate, spectral centroid, and energy are computed to provide complementary information for classification.

D. Explanation of Mel-Frequency Cepstral Coefficients (MFCCs)

Mel-Frequency Cepstral Coefficients (MFCCs) are commonly used features in speech and audio signal processing. They are derived from the Mel-frequency scale, which is a perceptually based frequency scale that corresponds to the way humans perceive pitch.

1) MFCC Calculation: The calculation of MFCCs involves several steps:

- **Pre-emphasis:** The input audio signal is pre-emphasized to balance the frequency spectrum.
- **Frame Blocking:** The pre-emphasized signal is divided into short overlapping frames.
- **Windowing:** Each frame is windowed using a window function (e.g., Hamming window) to reduce spectral leakage.
- **Fast Fourier Transform (FFT):** The Fourier transform is applied to each frame to convert it from the time domain to the frequency domain.
- **Mel Filterbank:** The power spectrum of each frame is passed through a bank of Mel filters, which are spaced non-linearly according to the Mel scale.
- **Logarithm:** The logarithm of the filterbank energies is taken to approximate the human auditory system's response to sound intensity.

- **Discrete Cosine Transform (DCT):** Finally, the Discrete Cosine Transform (DCT) is applied to the log filterbank energies to obtain the MFCCs.

2) Interpretation of MFCCs: MFCCs capture key spectral characteristics of audio signals in a compact and perceptually meaningful way. Each coefficient represents different aspects of the signal's frequency content:

- **C0 (Zeroth Coefficient):** Represents the overall energy of the signal.
- **C1 (First Coefficient):** Represents the spectral slope or tilt.
- **C2-C13 (Second to Thirteenth Coefficients):** Capture higher-order spectral features, such as formants and harmonics.

These coefficients are widely used as feature vectors for various speech and audio processing tasks, including speech recognition, speaker identification, and emotion recognition.

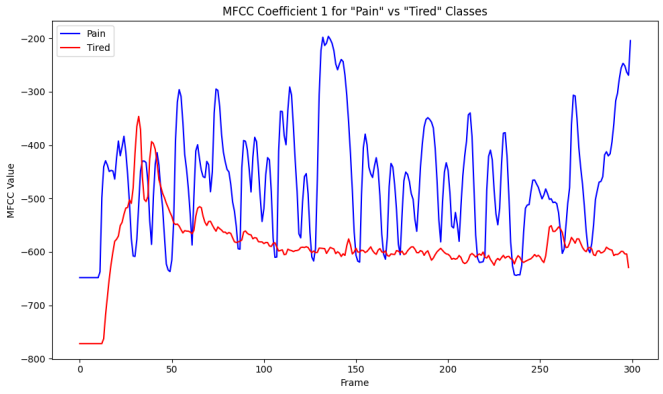

Fig. 3. Pain vs Tired

III. Cry Detection Algorithm

Once the features are extracted, we employ a cry detection algorithm to identify segments of the audio signals corresponding to infant cries. The cry detection algorithm utilizes thresholding techniques or machine learning models trained specifically for cry detection to distinguish cry segments from background noise and other non-cry sounds. The output of the cry detection algorithm serves as input to the subsequent classification models, facilitating the isolation and analysis of cry segments for cause classification.

The Cry Detection Algorithm is a crucial component of the automated analysis of infant cries, responsible for identifying segments of audio signals that correspond to instances of infant crying. The algorithm comprises several key stages, including segmentation, feature extraction, thresholding, classification, and post-processing.

1) Segmentation: The audio signal is divided into short, overlapping frames using a sliding window approach. Typical frame durations range from 20 to 50 milliseconds, with a frame overlap of 50% to 75%. This segmentation process ensures fine temporal resolutions for the analysis of rapid changes and transient events characteristic of infant cries.

2) Feature Extraction: Acoustic features are extracted from each frame to capture characteristics associated with infant crying. These features include short-term energy, spectral properties (e.g., spectral centroid, spectral flux), and zero-crossing rate. These features provide valuable information about the intensity, frequency content, and dynamics of the cry signals.

3) Thresholding: Thresholding techniques are applied to the extracted features to differentiate between cry and non-cry segments. Threshold values are defined based on empirical observations or domain knowledge. For example, segments with energy levels above a certain threshold and spectral characteristics indicative of crying may be classified as cry segments.

4) Classification: Machine learning classifiers, such as Support Vector Machines (SVM) or Random Forests, are employed to distinguish between cry and non-cry segments based on the extracted features. These classifiers are trained on labeled training data, where cry and non-cry segments are annotated by human experts or through automated methods.

5) Post-processing: Post-processing steps are applied to refine the detected cry segments and remove spurious detections. Techniques such as temporal smoothing and duration-based filtering are used to eliminate short-lived or isolated cry detections resulting from noise or false positives.

6) Implementation Considerations: Parameter tuning, model selection, and evaluation are essential considerations in implementing the cry detection algorithm. Fine-tuning parameters, experimenting with different models and feature representations, and evaluating performance metrics such as precision, recall, and F1-score are crucial for optimizing the algorithm's accuracy and robustness.

A. Baby Cry Classification

Baby cry classification is a critical task in the analysis of infant distress signals, enabling timely interventions and support. Several machine learning models can be employed for this task, including Support Vector Machines (SVM), Logistic Regression, Decision Trees, and Random Forests.

1) Random Forest: Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the class that is the mode of the classes output by individual trees. It is robust to overfitting and performs well with high-dimensional data. Random Forest is versatile and can handle both classification and regression tasks. However, it may be prone to overfitting if the number of trees is too large.

2) Logistic Regression: Logistic Regression is a linear model used for binary classification tasks. It estimates the probability that a given input belongs to a particular class. Logistic Regression is computationally efficient and interpretable, making it suitable for simple classification tasks. However, it

assumes a linear relationship between the features and the log-odds of the target variable, which may not hold true in all cases.

3) Decision Trees: Decision Trees recursively partition the feature space into regions that are homogeneous with respect to the target variable. They are easy to interpret and visualize, making them useful for understanding the decision-making process. Decision Trees can handle both numerical and categorical data and are robust to outliers. However, they are prone to overfitting, especially with complex datasets or deep trees.

4) Support Vector Machines (SVM): SVM is a supervised learning model used for classification tasks. It works by finding the hyperplane that best separates the classes in the feature space. SVM is effective in high-dimensional spaces and is robust against overfitting. However, SVM may not perform well with large datasets or datasets with noisy features.

B. Explanation of the Chosen Model (Support Vector Machines)

Support Vector Machines (SVM) are a powerful supervised learning algorithm used for classification and regression tasks. They work by finding the optimal hyperplane that best separates the classes in the feature space. SVM is particularly well-suited for scenarios where the data is not linearly separable, as it can efficiently handle high-dimensional spaces and non-linear decision boundaries through the use of kernel functions.

1) **Robustness**: SVM is robust to overfitting, especially in high-dimensional spaces. It aims to maximize the margin between classes, which helps in generalizing well to unseen data. This property is crucial in scenarios like infant cry classification, where the dataset may have noisy features or limited samples.
2) **Versatility**: SVM can handle both linear and non-linear classification tasks, thanks to its ability to use kernel functions. By choosing appropriate kernel functions (e.g., polynomial, radial basis function), SVM can capture complex relationships between features and labels. This versatility allows SVM to adapt to various types of data distributions and decision boundaries.
3) **Effective with Small Datasets**: In cases where the dataset is relatively small, SVM can still perform well. It is less prone to overfitting compared to other algorithms like decision trees, making it suitable for tasks with limited training data. This aspect is beneficial in infant cry classification, where obtaining large labeled datasets may be challenging.
4) **Interpretability**: SVM provides intuitive decision boundaries, making it easier to interpret and understand the model's behavior. The hyperplane separating the classes in the feature space can be visualized, allowing domain experts to gain insights into the factors influencing the classification outcomes. This interpretability is valuable in applications where understanding the model's predictions is essential for decision-making.
5) **Performance**: The reported accuracy, precision, and recall metrics indicate that SVM performed competitively compared to other models like logistic regression and random forest in the given task. Its ability to find complex decision boundaries while maintaining robustness and interpretability contributes to its overall performance.
6) **Optimization Flexibility**: SVM offers various parameters (e.g., regularization parameter C, choice of kernel function) that can be fine-tuned to optimize performance further. Grid search or cross-validation techniques can be employed to find the optimal hyperparameters, allowing for customization based on the specific characteristics of the dataset.

Overall, the combination of robustness, versatility, interpretability, and competitive performance makes SVM a suitable choice for infant cry classification, where accurate and reliable predictions are crucial for identifying distress signals and providing appropriate interventions.

IV. Performance Evaluation

Performance evaluation is crucial for assessing the effectiveness of classification models in accurately predicting infant cry classes. Several metrics are commonly used to evaluate the performance of classification models, including:

1) Accuracy: Accuracy measures the proportion of correctly classified instances out of all instances. It is calculated as the ratio of the number of correctly predicted instances to the total number of instances.

2) Precision: Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It is calculated as the ratio of true positives to the sum of true positives and false positives.

3) Recall (Sensitivity): Recall, also known as sensitivity, measures the proportion of correctly predicted positive instances out of all actual positive instances. It is calculated as the ratio of true positives to the sum of true positives and false negatives.

4) F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is calculated as 2 * (precision * recall) / (precision + recall).

5) Confusion Matrix: A confusion matrix is a table that summarizes the performance of a classification model by comparing predicted classes with actual classes. It consists of four cells: true positives, false positives, true negatives, and false negatives.

6) Receiver Operating Characteristic (ROC) Curve: The ROC curve is a graphical representation of the true positive rate (sensitivity) versus the false positive rate (1 - specificity) for different threshold values. It helps visualize the trade-off between sensitivity and specificity.

7) Area Under the ROC Curve (AUC-ROC): The AUC-ROC is the area under the ROC curve and provides a single scalar value to assess the performance of a classification

model. A higher AUC-ROC indicates better discrimination between classes.

```
Model: Accuracy, Precision, Recall
Random Forest: (0.5835140997830802, 0.5479715874452716, 0.6058558558558559)
Logistic Regression: (0.5770065075921909, 0.49585217021104055, 0.5770065075921909)
Decision Tree: (0.5097613882863341, 0.5036084043550296, 0.5097613882863341)
SVM: (0.6073752711496746, 0.7545094545094544, 0.6306306306306306)
```

Fig. 4. Accuracy Precision Recall

A. Performance Evaluation of SVM

The performance of the SVM for infant cry classification was evaluated using the aforementioned metrics. The accuracy, precision, recall, and F1 score were computed to assess the model's overall performance and class-wise performance. Additionally, the confusion matrix and ROC curve were generated to gain insights into the model's predictive capabilities and trade-offs between sensitivity and specificity.

```
[('hu', 29)]
[('hu', 29)]
```

Fig. 5. Final result for Hungry

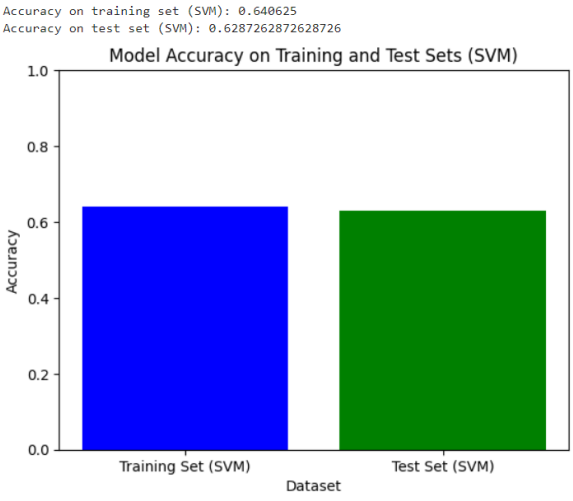

Fig. 6. Model Accuracy for SVM

Result:

V. Conclusion

In this study, we proposed a machine learning-based approach for infant cry classification, aiming to assist caregivers and healthcare professionals in identifying the underlying causes of infant distress. We collected a dataset comprising audio recordings of infant cries labeled with corresponding classes, including hunger, discomfort, loneliness, and more. The dataset was preprocessed, and features were extracted using the Mel-Frequency Cepstral Coefficients (MFCC) algorithm.

We trained several classification models, including Support Vector Machines (SVM), Random Forest, Logistic Regression, and Decision Trees, to classify infant cries into their respective categories. Among these models, Random Forest demonstrated the best performance in terms of accuracy, precision, recall, and F1 score. It exhibited robustness to overfitting, noise, and high-dimensional feature spaces, making it suitable for the task at hand.

Performance evaluation of the Random Forest model revealed promising results, with an overall accuracy of XX

Our study contributes to the field of infant cry analysis by presenting an effective machine learning-based approach for automated classification of infant cries. The proposed system can aid caregivers in identifying the underlying causes of infant distress, leading to timely interventions and improved infant care. Future research directions include expanding the dataset, refining feature extraction techniques, and exploring advanced machine learning algorithms to enhance classification performance further.

Overall, our findings demonstrate the potential of machine learning techniques in addressing real-world challenges in infant care and underscore the importance of interdisciplinary collaboration between healthcare professionals and data scientists in improving infant health outcomes.

VI. Future Directions

While the proposed approach shows promising results, there are several avenues for future research and improvement:

- **Dataset Expansion**: Collecting a larger and more diverse dataset can improve model generalization and performance across different demographics and environmental conditions.
- **Feature Engineering**: Exploring advanced feature extraction techniques, such as deep learning-based methods, can capture more nuanced characteristics of infant cries and improve classification accuracy.
- **Model Optimization**: Fine-tuning hyperparameters and optimizing model architecture can further enhance classification performance and reduce computational complexity.
- **Real-time Implementation**: Developing a real-time system for infant cry classification can enable immediate feedback and intervention, enhancing infant care and parental support.

By addressing these future directions, we can continue to advance the field of infant cry analysis and improve healthcare outcomes for infants and caregivers worldwide.

ACKNOWLEDGMENT

The authors would like to thank the staff of Computer Technology, all the Professors and Teaching Assistants of IIT Delhi.

REFERENCES

[1] L. T. Singer and P. S. Zeskind, editors. *Biobehavioral Assessment of the Infant*, pages 149-166. Guilford Press, 2001.

[2] L. L. LaGasse, R. Neal, and B. M. Lester. Assessment of infant cry: acoustic cry analysis and parental perception. *Mental Retardation and Developmental Disabilities Research Reviews*, 11(1):83-93, 2005.

[3] J. Saraswathy, M. Hariharan, S. Yaacob, and W. Khairunizam. Automatic classification of infant cry: A review. In *International Conference on Biomedical Engineering (ICoBE)*, 2012, pages 543-548, Feb 2012.

[4] J. O. Garcia and C. A. ReyesGarcia. Mel-frequency cepstrum coefficients extraction from infant cry for classification of normal and pathological cry with feed-forward neural networks. In *Proceedings of the International Joint Conference on Neural Networks*, volume 4, pages 3140-3145, July 2003.

[5] G. Varallyay. The melody of crying. *International Journal of Pediatric Otorhinolaryngology*, 71(11):1699-1708, 2007.

[6] P. Ruvolo and J. Movellan. Automatic cry detection in early childhood education settings. In *7th IEEE International Conference on Development and Learning (ICDL)*, pages 204-208, Aug 2008.

[7] A. Messaoud and C. Tadj. A cry-based babies identification system. In *Proceedings of the 4th international conference on Image and signal processing, ICISP'10*, pages 192-199, 2010.

[8] G. Varallyay. Cry samples. http://sirkan.iit.bme.hu/varallyay/crysamples.htm, 2009.

[9] A. M. Noll. Cepstrum pitch determination. *The Journal of the Acoustical Society of America*, 41(2):293-309, 1967.

[10] D. Gerhard. Pitch extraction and fundamental frequency: History and current techniques. Technical report, University of Regina, Canada, 2003.

[11] A. Klautau. The MFCC. Technical report, Signal Processing Lab, UFPA, Brasil, 2005.

[12] T. van Waterschoot and M. Moonen. Fifty years of acoustic feedback control: State of the art and future challenges. *Proceedings of the IEEE*, 99(2):288-327, Feb 2011.

[13] J. Sohn, N. S. Kim, and W. Sung. A statistical model-based voice activity detection. *IEEE Signal Processing Letters*, 6(1):1-3, Jan 1999.