Infant Health Risk Detection Using Deep Learning-Based Cry Analysis for Classifying Normal Infant Cry Signal

Sameena Bano1, Ayodeji Olalekan Salau2

1 Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia.

2 Department of Electrical/Electronics and Computer Engineering Afe Babalola University, Ado-Ekiti, Nigeria.

Email ID

[1] * drsameenakhayum@gmail.com

[2] ayodejisalau98@gmail.com

Abstract: Infant cry signals carry critical diagnostic information regarding a newborn's physical state, discomfort, and health risks. While parents rely on intuition, manual identification of specific cry reasons is highly subjective and error-prone. The paper proposes an automated framework for infant health risk detection using deep learning-based cry signal analysis. This paper focus on the classification of normal infant's cry signal into 5 different classes: Neh for Hungry, Eh for burp or Pain, Owh for Sleepy, Eairh for Internal pain or Tummy pain and Heh for discomfort. By applying DSP (Digital Signal Processing), we extract a bonded acoustic feature containing MFCCs (Mel-Frequency Cepstral Coefficients) and pitch (F0) parameters. A convolutional neural network (CNN) output is match with traditional K-Nearest Neighbors (KNN) and support vector machine (SVM) models. The paper proposes a CNN model, which attains good accuracy in multi-class cry classification. The trial evaluation for the classification is conducted on the Kaggle Infant Cry Audio Corpus and the ICSD dataset using Python with TensorFlow, PyTorch. The experimental findings indicate that the combination of MFCC features with pitch improves the classification accuracy and offering a reliable tool for the caregivers and the pediatricians.

Keywords: Infant Cry Analysis; Speech Signal Processing; Mel-Frequency Cepstral Coefficients (MFCCs); Convolutional Neural Networks; Deep Learning; Neonatal Diagnostics.

Introduction

Infant crying is the primary non-verbal communication channel for newborns to signal their physiological needs, emotional states, and physical discomfort. Recently, infant cry analysis has gained greater prominence due to advancements in digital signal processing (DSP), acoustic pattern recognition, and soft computing. Automated cry analysis has changed the diagnostic capabilities of physicians to correctly evaluate newborns.

Classifying infant cries, colloquially referred to as "baby talk," supports a mother's natural intuition by providing an objective classification of her baby's immediate needs (such as hunger, sleepiness, or pain). This classification also helps pediatricians detect early-stage health anomalies or risks. However, manual observation of infant crying is highly subjective, and caregivers struggle to distinguish between different normal infant cries consistently [1].

This paper proposes an automated system that extracts pitch and MFCC features to classify normal infant cries into functional categories, laying the foundation for early health risk detection.

Related work

Research on automated infant cry analysis has evolved from traditional manual detection approach to deep spatial-temporal neural networks approach. A timeline summary of relevant methodologies is provided in Table I.

Table 1. Literature Survey on Automated Infant Cry Classification

Year	Method	Features Used	Dataset / Focus	Key Findings
2022	KNN, SVM	Pitch, MFCC	Normal vs. Abnormal Cry	Traditional ML methods are more effective to small data set
2023	KNN	MFCC	Cry classification (basic categories)	Simple and fast, but less robust to acoustic noise
2025	KNN + Feature Fusion	Pitch + MFCC	Lightweight systems	Useful for real-time, low-cost microcontrollers.
2026	CNN + Attention	MFCC + Pitch + Spectrogram	Smart healthcare (abnormal cry)	Better feature focus and classification
2019	Hybrid DL (CNN + Transformer)	MFCC + Pitch + Temporal	Advanced infant cry analysis	SOTA performance, optimized for real-

Key Contribution

To build a reliable dataset for model training, the following criteria were established:

- Audio recordings of infant cries must have clear sound quality and minimal ambient noise.
- Cry samples must be labeled as normal (non-pathological) under the five standard categories.
- Audio signals must be suitable for feature extraction (specifically containing trackable pitch contours and MFCC structures).
- Infant age must fall within the range of newborn to 24 months.

Method, Experiments and Results

The pipeline of the proposed infant cry classification frame-work is structured into preprocessing, acoustic feature extraction, and target categorization.

- Preprocessing Step

  An audio signals from the Kaggle-hosted dataset of Infant Cry and ICSD-hosted dataset are converted to 16 kHz monophonic wav format. The Silent or ambiguous segments are neglected, and the filter is applied to get the high-frequency cry formant.

- Feature Extraction Step

  MFCCs (Mel-Frequency Cepstral Coefficients) Step: MFCC captures the structural characteristics of infant's vocal-tract and reflect the changes in acoustic resonant-cavities in associated with the different emotional/physical conditions.

- **Pitch (F0) Tracking Step:** Fundamental frequency (F0) of infant's cry is extracted using the shorttime autocorrelation. The variations in pitch will help in distinguishing the different cry types. The hunger cry generally exhibiting a rising-falling-frequency, while the pain cry tends to show high-frequency-spikes.

Target Classification Categories are

- **Neh for Hungry:** It is characterized by a distinct number of sounds that are present at the start of the cry, which is associated with the infant sucking impulse.
- **Eh for Burp-me or Pain:** It is characterized by a chesty and constrained sound, which indicate the presence of the trapped air.
- **Owh for Sleepy:** It is characterized by a yawning-like acoustic pattern with an open-mouth vocal structure.
- **Eairh for Internal Tummy Pain:** It is characterized by a deep, strained, low-pitched sound accompanied by abdominal muscle contractions.

A. **Heh foe Discomfort :** It is characterized by a high-frequency and a shallow cry that may indicate irritation caused by cold, wetness, or skin discomfort.

B. Pitch Analysis for "Neh" (Hungry)

During evaluation, the segmentation of the "Neh: Hungry" cry was measured across three distinct signals to evaluate tracking error rates, as shown in Table II.

TABLE 2. Pitch Segmentation Verification for "Neh" (Hungry) Cry Signals

Segment Metric	Sample 1	Sample 2	Sample 3	Mean
Correct Detection	0.63	0.90	0.90	0.810
Tracking Error	0.37	0.10	0.10	0.190
Correct (Alternate)	0.68	0.56	0.58	0.607
Error (Alternate)	0.32	0.44	0.42	0.393

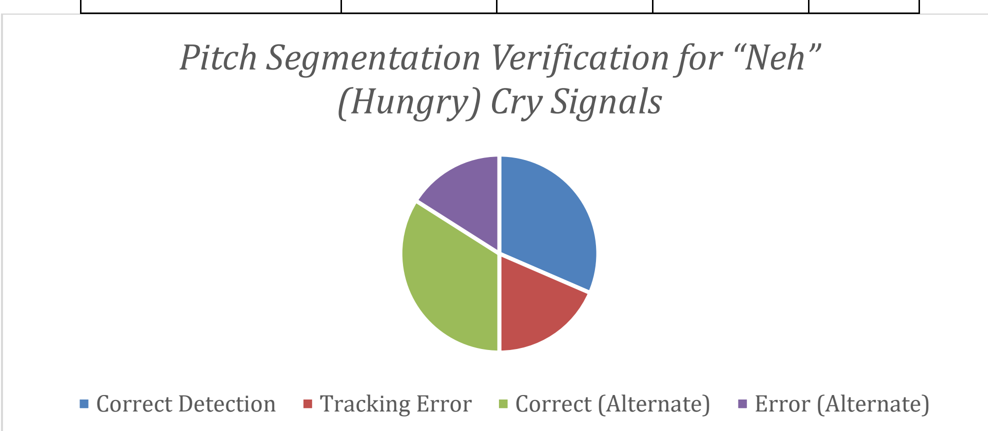


Figure 1. Pitch Segmentation Verification for "Neh" (Hungry) Cry Signals

C. Classification Performance

The CNN model outperforms baseline models (KNN, SVM) in overall accuracy and classification robustness:

- **CNN Model Accuracy:** High-accuracy multi-class categorization achieved compared to standard

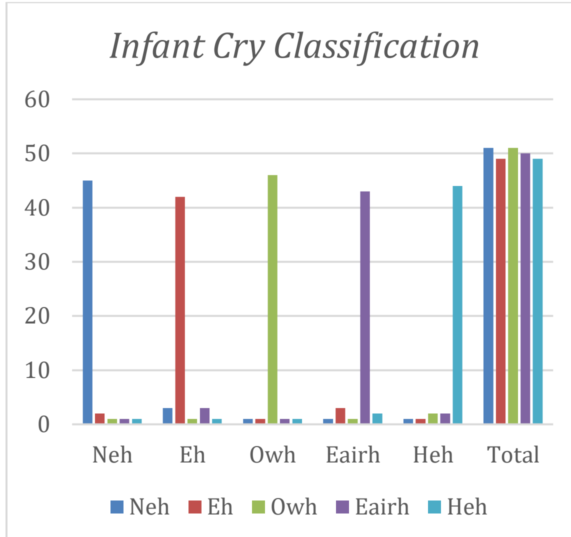
ML techniques.

- **Feature Combination Efficacy:** Combining MFCC with Pitch features yields significantly higher classification precision compared to using MFCCs alone.

TABLE 3. CNN Matrix – Infant Cry Classification

	Neh	Eh	Owh	Eairh	Heh	Total
Neh	45	2	1	1	1	50
Eh	3	42	1	3	1	50
Owh	1	1	46	1	1	50
Eairh	1	3	1	43	2	50
Heh	1	1	2	2	44	50
Total	51	49	51	50	49	250

Figure 2. CNN Matrix – Infant Cry Classification



Overall Accuracy

$$Accuracy = \frac{45 + 42 + 46 + 43 + 44}{250} \times 100$$

$$Accuracy = 88\%$$

Discussions

Discussions should be carried out in the article for the readers to understand the meaning of your results and experimental findings

This system demonstrates that combining MFCC feature ex-traction with a convolutional neural network (CNN) provides an effective approach for normal infant cry classification. The spatial filter kernels in the CNN capture the transitions in Mel-spectrogram representations of the cry, matching the clinical labels (Neh, Eh, Owh, Eairh, Heh).

However, practical deployment in smart healthcare systems requires improvements in:

- **Dataset Quality:** Expanding clinical databases to avoid overfitting.
- **Noise Handling:** Enhancing denoising filters to cope with everyday domestic sounds.
- **Real-Time Processing:** Optimizing deep models using network pruning and quantization for mobile deployment.

Conclusions

1. Problem Statement / Motivation:
   Manual identification of infant cry causes is subjective and difficult. An automated system is needed for early health-risk detection.
2. Method Used:
   MFCC and pitch (F0) features were extracted and classified using CNN, with comparison to KNN

and SVM.

3. Key Findings:
   CNN performed better than KNN and SVM, and combining MFCC + pitch improved classification accuracy.
4. Limitations & Future Work:
   Noisy datasets and real-time deployment constraints remain. Future work should focus on larger clinical datasets, better noise handling, and model optimization for mobile use.

References

1. Y.-C. Liang, I. Wijaya, M.-T. Yang, J. R. C. Juarez and H.-T. Chang, "Deep learning for infant cry recognition", International Journal of Environmental Research and Public Health, vol. 19, no. 10, pp.6311,2022.
   https://doi.org/10.3390/ijerph19106311
2. Y. Zayed, A. Hasasneh and C. Tadj, "Infant cry signal diagnostic system using deep learning and fused features", Diagnostics, vol. 13, no. 12, pp. 2107, 2023. https://doi.org/10.3390/diagnostics13122107
3. D. M. Herlea, B. Iancu and E.-R. Ardelean, "A study of deep learning models for audio classification of infant crying in a baby monitoring system", Informatics, vol. 12, no. 2, pp. 50, 2025. https://doi.org/10.3390/informatics12020050
4. A. McCofie, D. Goldgof, J. Hausmann, P. R. Mouton, Y. Sun and M. I. Hossain, "A review of deep learning model approach for pain assessment in infant cry sounds", Machine Learning and Knowledge Extraction, vol. 8, no. 3, 2026. https://doi.org/10.3390/make8030076
5. X. Zhang, X. Zhang and L. Han, "An energy efficient Internet of Things network using restart artificial bee colony and wireless power transfer", IEEE Access, vol. 7, pp. 12686-12695, 2019.
   https://doi.org/10.1109/ACCESS.2019.2892798
6. X. Zhong, L. Zhang and Y. Wei, "Dynamic load-balancing vertical control for a large-scale software-defined Internet of Things", IEEE Access, vol. 7, pp. 140769-140780, 2019.
   https://doi.org/10.1109/ACCESS.2019.2943173
7. M. Malik, M. Dutta and J. Granjal, "A survey of key bootstrapping protocols based on public key cryptography in the Internet of Things", IEEE Access, vol. 7, pp. 27443-27464, 2019.
   https://doi.org/10.1109/ACCESS.2019.2900957
8. C. Ji, T. B. Mudiyanselage, Y. Gao and Y. Pan, "A review of infant cry analysis and classification", EURASIP Journal on Audio, Speech, and Music Processing, vol. 2021, no. 8, 2021. https://doi.org/10.1186/s13636-021-00197-5
9. A. Sharma, R. Kumar and S. Gupta, "AI ears for infant tears: Deep learning-based analysis of infant cries for early health diagnostics", SSRN, 2025. https://doi.org/10.2139/ssrn.5521038
10. "Deep learning systems for automatic diagnosis of infant cry signals", Chaos, Solitons & Fractals, vol.154,2022.
    https://doi.org/10.1016/j.chaos.2021.111694