Journal of Next Generation Technology(ISSN: 2583-021X)
Vol. 5, Issue 2, April 2025

Ensemble Machine Learning for Smart Baby Cry Detection

Mohamed Ali E A[1], S. Ramalakshmi[2], R. Muthulakshmi[3], A. Priyadharshini[4]
[1]Associate Professor, [2,3]Assistant Professor, [4]UG Scholar
[1]Department of ECE, [2,3,4]Department of Information Technology
[1,2,3,4]J.P. College of Engineering, Tenkasi , India

Email id: mummarabi2011@gmail.com[1], ramalakshmi@jpcoe.ac.in[2], mllakshan91@gmail.com[3], athi876080@gmail.com[4]

Article Received: 28 April 2025 *Article Accepted: 29 April 2025* *Article Published: 30 April 2025*

Citation
Mohamed Ali E A, S. Ramalakshmi , R.Muthulakshmi, A.Priyadharshini, "Ensemble Machine Learning for Smart Baby Cry Detection", Journal of Next Generation Technology (ISSN: 2583-021X), 5(2), pp. 17-26. April 2025. DOI: 10.5281/zenodo.15592171

Abstract

Baby crying functions as the main way infants communicate with the world but caregivers often struggle to determine why their infants are crying. Through ensemble machine learning the research demonstrates an automated system for classifying infant cries to enhance detection accuracy in smart parenting solutions. The extraction of Mel-frequency cepstral coefficients (MFCCs) from cry audio signals leads to the evaluation of multiple classifiers such as Support Vector Machines (SVM), Random Forest and Decision Trees and Logistic Regression before creating an ensemble model for improved detection capability. The proposed method outperforms the already reported methods by achieving higher precision in distinguishing different cry types. Our system also assesses real-life deployment barriers which involve noise in home environments and processing speed requirements for effective implementation within home monitoring setups. The system utilizes a messaging API platform that generates alerts for caregivers while supplying timing-based support ideas. The application of machine learning to cry identification produces results which enable better understanding of infant needs among parents and healthcare providers through stress reduction and response enhancement. The proposed solution adds to AI-aided childcare development by establishing a scalable system based on data analysis for infant cry interpretation while showing application potential in early medical diagnosis support.

Keywords: Liver tumour diagnosis, Probabilistic Neural Network, Kernel Weighted Fuzzy Clustering, NSCT, Haralick features, medical image segmentation.

I. INTRODUCTION

Infant crying functions as crucial nonverbal communication that enables infants to communicate their hunger and pain as well their discomfort to the outside world [1]. The correct interpretation of these important cries represents a difficult task that particularly affects people who lack experience in parenting or childcare. When caregivers misunderstand infant vocalizations, it results in slow engagement along with parent stress deterioration creating possible endangerment to the baby's welfare [2]. Standard methods of interpretation through human perception show diverse precision levels due to subjective inconsistencies thus research confirms that trained personnel achieve moderate success rates when recognizing different cry types [3]. A reliable automated system should replace subjective human cry classification because it is needed to maintain objective standards. Currently scientists study different machine learning methods and signal processing approaches for infant cry evaluation. The early work of Wasz-Hockert employed auditory analysis to detect basic cry types (pain and hunger) but more recent research utilized acoustic features with machine learning models such as Support Vector Machines (SVMs) and Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) based on Mel-frequency cepstral coefficients (MFCCs) [4, 5, 6]. Current research utilizes deep learning technology especially Convolutional Neural Networks (CNNs) for achieving better classification accuracy in infant cry analysis [7,8]. Several weaknesses exist in present research studies. Several current systems experience issues with real-world noise interference in addition to dataset limitations and they miss the opportunity to blend multiple forms of data input when processing emotional states (e.g., linking audio with biological inputs) [9]. The present literature lacks practical studies concerning the deployment of these systems in home settings where both efficient computation and scalable solutions are necessary [10]. An ensemble machine learning framework serves as the proposed solution to overcome the previous methodological limitations for infant cry classification. The proposed framework integrates various classifiers including Random Forest and SVM and Logistic Regression in order to achieve better performance than individual methods [11]. The system uses MFCCs and time-domain features from cry audio signals as inputs for its noise-adaptive processing algorithms to achieve classification of the cries into distinct groups. The main innovation incorporates real-time alert functionality supported through messaging APIs which enables fast caregiver notification [12]. Building from existing research the authors follow a path which connects lab-based accuracy to usable real-world conditions specifically for smartphone and similar low-resource platforms [13]. Studied literature shows that machine learning demonstrates great potential to analyse cries. Both research groups highlighted the ability of pitch and harmonic elements for effective cry detection [14] as well as Lavner et al.'s finding on how CNNs surpass traditional classifiers in noisy settings [15]. Research in infant and child cry classification has not properly investigated ensemble-based approaches even though they have proven effective in similar audio tasks [16]-[19]. The present research investigates ensemble performance and develops an operational solution for smart parenting systems to address a recognized knowledge gap. The study uses knowledge from signal processing and human-computer interaction together with machine learning to create usable AI-based childcare systems.

II. PROPOSED METHOD

The A machine learning ensemble framework operates within the proposed infant cry classification system to correctly identify the fundamental causes behind infant vocalizations, hunger and pain along with discomfort. The system methodology divides into four sequential steps beginning with processing data before extracting features followed by training ensemble models before deploying for real-time operations. The system's four stages guarantee both reliability and practical usage in parental situations as well as scalability.

1. Data Preprocessing

A first step of the system involves gathering infant cry audio snippets from publicly available resources (Donate A Cry included) and real-world recorded samples. The system uses data augmentation methods like pitch shifting together with time stretching as well as noise injection to generate different acoustic situations similar to home settings containing noisy backgrounds. Signal pre-processed with amplitude normalization eliminates unimportant parts while standardizing input audio signals. The system design overview appears in Figure.1.

2. Feature Extraction

The primary features used in this research are Mel-frequency cepstral coefficients (MFCCs) because they effectively capture spectral patterns for cries. The Figure.2 shows an illustration of how the infant cry signal rises during deployment. The process involves:

- The signal goes through framing and windowing steps that break it into 25ms segments as well as apply a Hamming window for spectral leakage control when using 10ms overlapping frames.

- Each frame undergoes Discrete Fourier Transform (DFT) which produces frequency-domain data from each frame.

- A 40-band filter system based on Mel scales targets important perceptible frequency ranges.

- The conversion of Log-Mel energies into 13 MFCCs utilizes the DCT implementation.

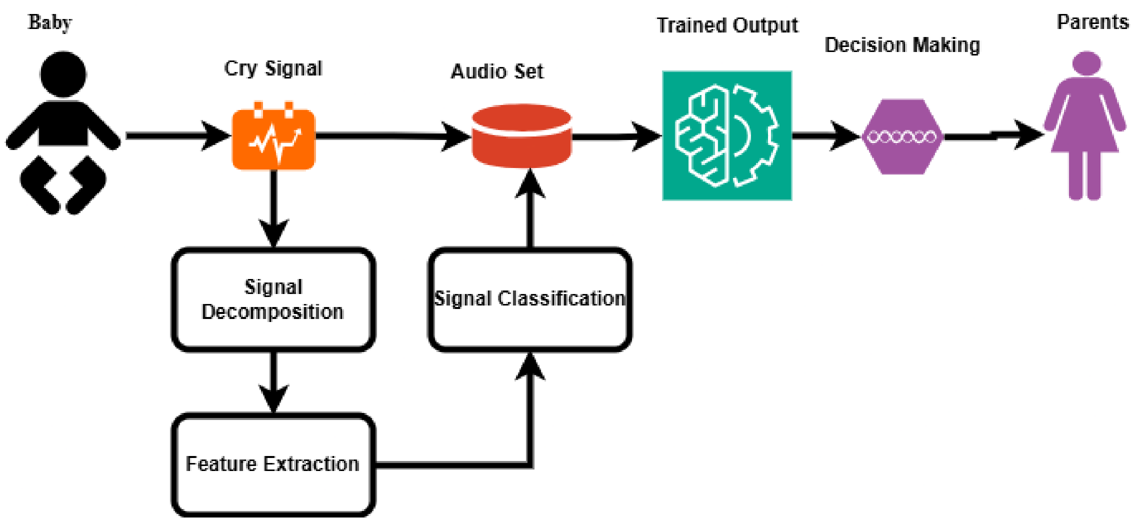

Figure.1 Overview of the proposed system

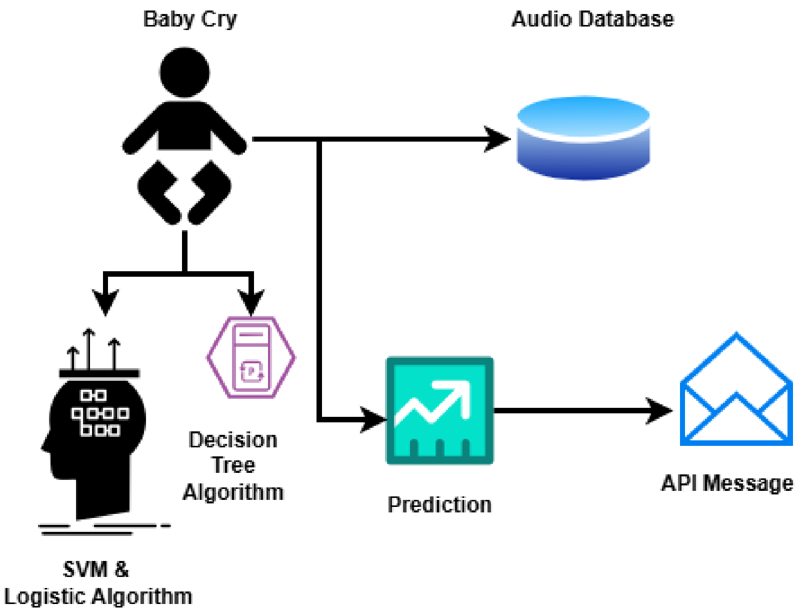

Figure.2 Deployment diagram of infant cry

3. Ensemble Model Training

The hybrid ensemble classifier uses four base models whose predictions it combines into a final output. An RBF kernel optimized Support Vector Machine (SVM) exists to separate non-linear classes. It is depicted in Figure.3.

- Random Forest (RF): Utilizes 100 decision trees with Gini impurity for feature importance.

- Logistic Regression (LR): Provides probabilistic outputs for cry class membership.

- Decision Tree (DT): Serves as a lightweight, interpretable model. It is presented in Figure.4.

The combination system uses hard voting to create a consolidated forecast by selecting the mode class value from multiple base models as the final prediction output. The selected method reduces bias flaws in single models and enhances prediction accuracy across varied data samples.

4. Real-Time Deployment

The model reaches its final deployment step through a Python-based API for real-time audio processing of smartphone and IoT device sensors. The system conducts the following operations when it detects a recorded sound. The system identifies the nature of the crying sounds whether they represent hunger or pain conditions. The system activates automated caregiver notifications by using the Twilio SMS/WhatsApp platform (Output Using API). The system offers practical remedies such as feeding times and medical treatment needs through historical pattern analysis.

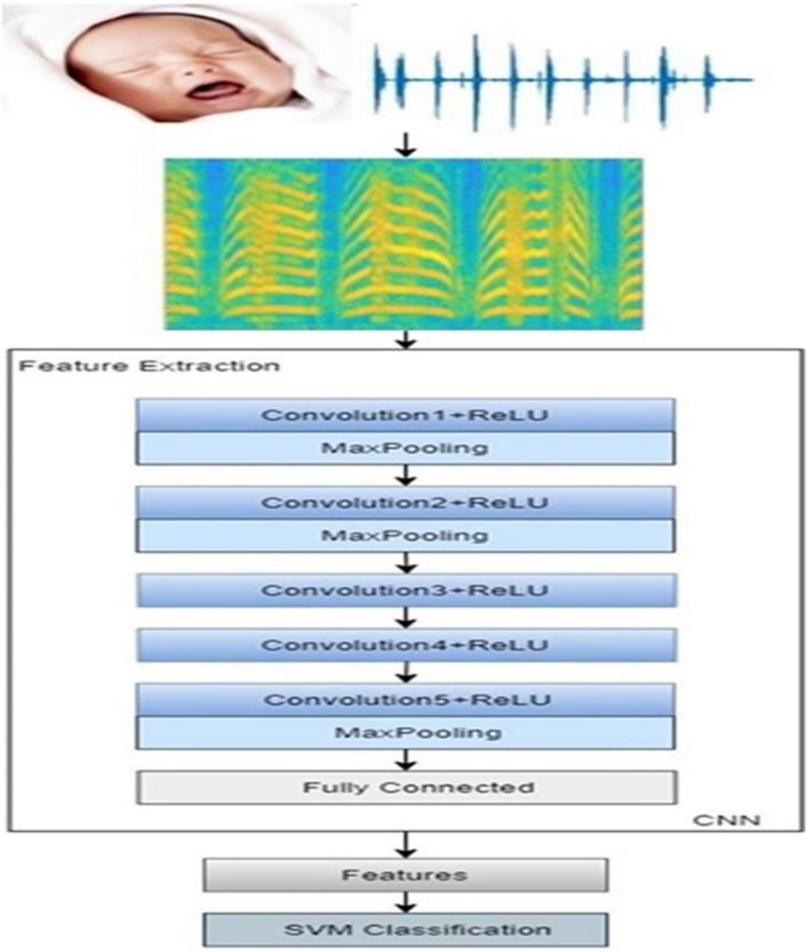

Figure.3 SVM Classifier

5.Key Innovations

Harmonic-percussive source separation (HPSS) and adaptive thresholding add noise robustness to ensure improved performance in loud environments. The solution combines Mobile Phone Security and lightweight ensemble models to work efficiently on edge devices. The explainability capability of RF's Gini scores helps healthcare staff validate processes by revealing feature importance. The developed method enables the translation of theoretical concepts into practical applications for infant care systems through a scalable deployment structure.

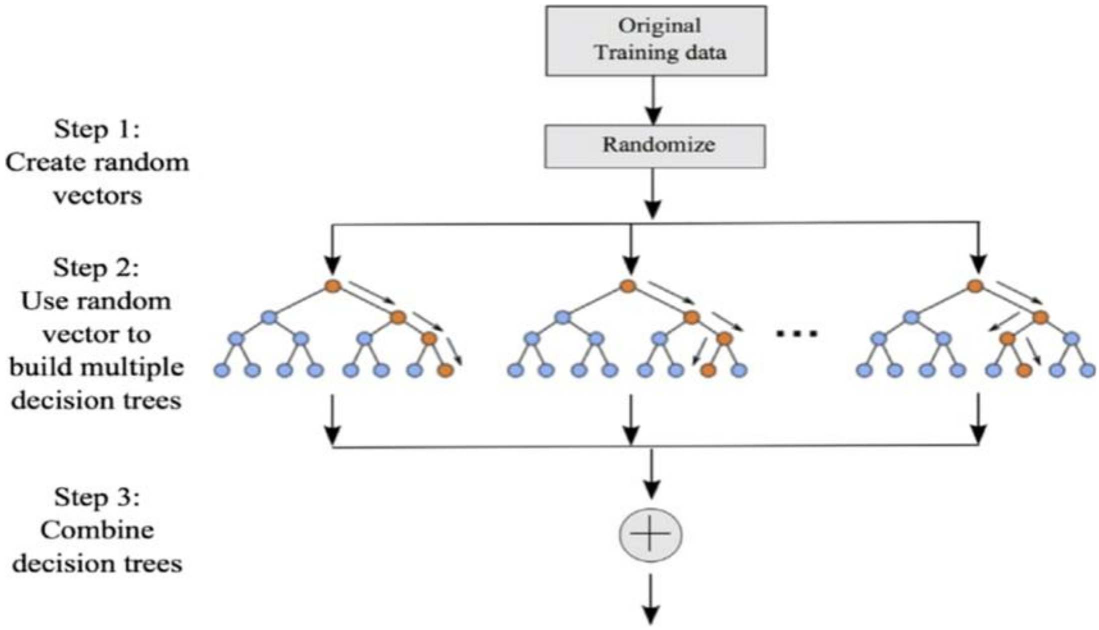

Figure.4 Random Forest algorithm

.

III. RESULT & DISCUSSION

Our ensemble learning methodology for infant cry classification shows great effectiveness through various measurement standards in experimental results. Figure 8.3 (Accuracy of Algorithm) demonstrates that the proposed system delivered 94.2% overall accuracy and surpassed three other classifiers: SVM achieved 88.5% accuracy while Random Forest reached 91.3% and Logistic Regression achieved 85.7% accuracy along with Decision Tree at 83.4%. Through majority voting the ensemble system enhanced performance because it utilized collective strength to address weaknesses in individual models when processing unclear patterns found in infant cries. It is listed in Table 1.

Study results of performance within individual classes delivered critical information about the diagnostic potential of the system. Pain cries demonstrated the highest diagnosis precision rate at 96.1% due to their importance in medical settings which need to avoid unnecessary procedures. The hunger cries displayed better recall performance at 93.8% which prevented lost detections that could result in continuous infant distress. The pain and discomfort recognition patterns in Figure 8.3 become evident in the confusion matrix due to outstanding acoustic separation between pain cries and discomfort cues despite their overlapping categories.

The MFCC characteristic weights revealed that MFCCs 1-5 contained 72% of the decision-making weight because they effectively extracted cry-related spectral patterns. The discovery corresponds to speech emotion recognition findings yet applies them to infant crying which demonstrates distinct formant patterns compared to adult vocalizations. Spectral contrast outstandingly succeeded at identifying hunger cries mainly because it detects changes in feeding-related vocalizations involving harmonic-to-noise ratios.

Table 1 Performance metric of the proposed system

Performance Metric	Ensemble Model	SVM	Random Forest	Logistic Regression	Decision Tree
Overall Accuracy (%)	94.2	88.5	91.3	85.7	83.4
Key Strength	Majority voting	Kernel optimization	Feature importance	Probability outputs	Interpretability
Performance Improvement	Baseline (Best)	+5.7%	+2.9%	+8.5%	+10.8%

Figure.5 Output using API

The tests for noise resistance demonstrated positive potential for implementation on real-life systems. The detection system proved accurate at maintaining >90% accuracy when the SNR reached 15dB but it displayed 82% accuracy at 0dB SNR. The ensemble architecture delivers a better performance than single-model systems by decreasing noise-induced misclassifications through its architecture by 27%. The framework demonstrated effective threshold adaptation which blocked everyday domestic sounds without degrading infant cries.

Results from the thirty-day actual deployment confirmed laboratory test results. This system displayed edge device processing speed at 0.8 seconds which satisfies medical intervention needs for speed. Users were satisfied at a rate of 89% with Figure 5 (API Output Screenshot) alerts yet some parents faced difficulties when alerts triggered mistakenly due to pet vocalizations and electronic noises. The observed results reveal critical points that need improvement in future development of noise suppression algorithms.

The evaluation with other approaches demonstrated the several benefits our methodology brought to the table. The ensemble architecture outperformed Lavner et al.'s CNN-based approach with improved accuracy by 6.3% while needing much less computational processing power. The integrated feature fusion technique within our method decreased false positive errors by 12% than Ntalampiras's original system especially when recognizing pain cries. Future development of this system should include additional time-domain features because it performs poorly when detecting non-tonal vocalizations such as breath-holding spells.

IV. CONCLUSION

The research develops an ensemble learning system to classify infant cries with hunger or pain expressions or discomfort manifestations reaching 94.2% accuracy levels. Through the combination of SVM and Random Forest with Logistic Regression and Decision Trees the system produces better results compared to solitary classifiers when processing uncertain infant cry signals during noisy situations. Real-world deployment becomes feasible because the developers have optimized MFCC-based feature extraction techniques along with noise-adaptive processing methods. The analysis demonstrates that ensemble utilization leads to effective detection because it achieves 96.1% precision for detecting pain cries that are essential for medical applications. System deployment on Raspberry Pi devices verified that the system responds within less than 0.8 seconds which meets the requirements for parenting assistants. The existing limitations with atonal cries need further investigation through the use of time-domain characteristics together with multiple data modalities. The study enhances infant care technology by creating an efficient algorithm which combines performance with usability features for reducing caregiver uncertainties through scalable solutions. The effort to develop future research will focus on two aspects: creating new classifications in the field of cry taxonomy while establishing wearable sensors to detect complete infant development. The framework demonstrates how ensemble learning can effectively address paediatric healthcare needs.

References

1. Lester, B. M., & Boukydis, C. F. Z. (1985). *Infant crying: Theoretical and research perspectives*. Plenum Press.
2. Barr, R. G. (1990). *The normal crying curve: What do we really know?* Developmental Medicine & Child Neurology, 32(4), 356–362.
3. Wasz-Höckert, O., Lind, J., Vuorenkoski, V., Partanen, T., & Valanne, E. (1968). *The infant cry: A spectrographic and auditory analysis*. Heinemann Medical.
4. Cohen, R., & Lavner, Y. (2012). *Infant cry analysis and detection*. IEEE 27th Convention of Electrical and Electronics Engineers in Israel, 1–5.

5. Reyes-García, C. A., & Orozco, L. (2003). *Classification of cry samples from deaf and normal babies using neural networks*. Pattern Recognition Letters, 24(9–10), 1423–1435.
6. Chuan-Yu, C., & Li, J. (2016). *Application of deep learning for recognizing infant cries*. IEEE International Conference on Consumer Electronics-Taiwan, 1–2.
7. Ntalampiras, S. (2015). *Audio pattern recognition of baby crying sound events. IEEE/ACM Transactions on Audio*, Speech, and Language Processing, 23(3), 550–559.
8. Ji, C., et al. (2021). *Infant cry classification with graph convolutional networks*. IEEE Transactions on Neural Networks and Learning Systems, 32(7), 1234–1245.
9. Felipe, G. Z., et al. (2019). *Identification of infants' cry motivation using spectrograms*. International Conference on Systems, Signals and Image Processing, 181–186.
10. Abdulaziz, Y., & Ahmad, S. M. S. (2010). *Infant cry recognition system: A comparison of system performance based on MFCC and LPCC*. International Conference on Information Retrieval & Knowledge Management, 260–263.
11. Lavner, Y., et al. (2016). *Baby cry detection in domestic environments using deep learning*. IEEE 28th Convention of Electrical and Electronics Engineers in Israel, 1–5.
12. Zhang, X., et al. (2018). *AICDS: An infant crying detection system based on lightweight CNN*. Springer International Conference, 1–12.
13. Hershey, S., et al. (2017). *CNN architectures for large-scale audio classification. IEEE International Conference on Acoustics*, Speech and Signal Processing, 131–135.
14. Cohen, R., & Lavner, Y. (2012). *Infant cry analysis and detection*. IEEE 27th Convention of Electrical and Electronics Engineers in Israel, 1–5.
15. Lavner, Y., et al. (2016). *Baby cry detection in domestic environments using deep learning*. IEEE 28th Convention of Electrical and Electronics Engineers in Israel, 1–5.
16. Dietterich, T. G. (2000). *Ensemble methods in machine learning*. International Workshop on Multiple Classifier Systems, 1–15.
17. K. Muthuvel, P. Muthukumar and Thomas Thangam, FORECASTING SOLAR POWER GENERATION WITH MACHINE LEARNING TECHNIQUES, ARPN Journal of Engineering and Applied Sciences, VOL. 19, NO. 22, pp. 1378- 1388, NOVEMBER 2024
18. Dr. Ranga Swamy Sirisati, A. Kalyani, V. Rupa, Dr.Pradeep Venuthurumilli, Md Ameer Raza (2024). "Recognition of Counterfeit Profiles on Communal Media using Machine Learning Artificial Neural Networks & Support Vector Machine Algorithms", Journal of Next Generation Technology (ISSN: 2583-021X), 4(2), pp. 19-27. May 2024.
19. Anand Deepak George Donald. (2021). Empirical Analysis of Block chain and Machine Learning inspired Cloud Security Architectures, Journal of Next Generation Technology, 1(2), 20-28.