

Leakage-Free Biomedical Acoustic Signal Classification of Infant Cry for Reliable Neonatal Monitoring

Humam Abdullah Khudair^{1*}, Ziyad K. Farej², and Omar Younis Alani³

¹ Engineering Technical College – Mosul, Northern Technical University, Mosul, Iraq

² Northern Technical University, Mosul, Iraq

³ School of Science, Engineering and Environment, University of Salford, Manchester, United Kingdom

*e-mail: eng.humam@ntu.edu.iq

Abstract. Infant cry analysis is a non-invasive biomedical acoustic signal processing task that can support reliable neonatal monitoring. However, reported high classification accuracy may be affected by dataset imbalance and data leakage rather than true model generalization. This study investigates the effect of leakage-free evaluation and hybrid acoustic feature representation on infant cry classification under imbalanced data conditions. A file-level partitioning protocol was applied before augmentation, feature scaling, and model training to preserve independence between training and testing data. Hybrid acoustic features, including Mel Frequency Cepstral Coefficients (MFCC) statistics, root-mean-square (RMS) energy, zero-crossing rate, pitch-related descriptors, and spectral features, were extracted from eight infant cry categories. Four lightweight machine learning models were evaluated using accuracy, Macro-F1, and balanced accuracy. Under leakage-free evaluation, support vector machines (SVMs) achieved the most balanced performance with 86.41% accuracy, 0.7376 Macro-F1, and 68.43% balanced accuracy. In contrast, leaky evaluation inflated performance above 90% across models. These findings indicate that reliable infant cry classification depends on leakage-free evaluation, imbalance-aware metrics, and biomedical acoustic feature integration rather than model complexity alone.

Keywords: infant cry classification; biomedical acoustic signal processing; neonatal monitoring; leakage-free evaluation; hybrid acoustic features; support vector machine (SVM).

Paper #9636 received 17 Jun 2026; revised manuscript received 21 Jul 2026; accepted for publication 27 Jul 2026; published online 30 Sep 2026. doi: [10.18287/JBPE26.12.030305](https://doi.org/10.18287/JBPE26.12.030305).

1 Introduction

The cry classification of infants is a significant research area in biomedical acoustic signal processing since crying is an early non-invasive behavioral cue of the infant's needs and physiological discomfort. Cry signals can offer complementary data that can be used for neonatal monitoring, such as hunger, pain, fatigue, thermal stress, fear, loneliness, or overall distress. Unlike conventional physiological measures, cry analysis does not necessitate direct contact with the body and thus can become an extra layer of behavior sensing in smart neonatal care systems, and even be used for continuous monitoring [1].

Although there is potential for reliable classification of infant cries, it continues to be a challenge. Cry signals

are highly non-stationary and affected by the infant's physiological state, emotional state and recording environment [2]. This means that there can be acoustic similarities between different cry categories, which makes the classification problem more complicated than just labelling. Discriminative information is spread over multiple acoustic dimensions such as spectral structure, temporal irregularity, vocal intensity, pitch behavior and harmonic content. Hence, one feature type may not be enough to capture the entire structure of infant cry signals [3].

The other major problem is methodological reliability. High classification accuracy has been reported in several studies, but this could be due to data imbalance, inadequate evaluation control, or data leakage. Leakage happens when two samples, which are acoustically

similar or enhanced versions of the same sample, are in both the training and testing sets. In this situation, the model might learn the patterns of the recordings instead of the cry patterns. Furthermore, the accuracy of the overall performance may be misleading in imbalanced data sets because good performance on the majority class may mask poor performance on the minority classes. Therefore, for a reliable performance assessment it is crucial to have leakage-free evaluation and imbalance-sensitive metrics like Macro-F1 and balanced accuracy [4].

To overcome these challenges, this study aims to introduce a leakage-free infant cry classification framework using a combination of hybrid biomedical acoustic features and lightweight machine learning. To ensure independence of the test set, the original recordings are split at the file level prior to augmentation, normalization and model training [5]. A hybrid of feature representation is extracted to include complementary spectral, temporal, energy-based and pitch-related features of the cry signals. Then, four lightweight machine learning models are tested under the same experimental conditions with accuracy, Macro-F1 and balanced accuracy [6].

What is most important in this study is that the classification of infant cries is reliable and that the consistency between acoustic representation, leakage-free evaluation design, and model suitability are crucial. This offers a more accurate foundation for the incorporation of cry analysis into smart neonatal monitoring systems.

2 Literature Review

Cry analysis of infants has become a biomedical acoustic signal processing problem which seeks to extract clinically relevant behavior from raw vocalizations. Previous studies typically used a two-stage pipeline: acoustic signal representation and classification using machine learning. Mel-frequency cepstral coefficients (MFCCs) are still popular handcrafted descriptors due to their ability to give a compact and perceptually meaningful representation of the spectral envelope of infant cry signals. But infant crying is very non-stationary and is not completely described by the spectral envelope information [7]. Other cry characteristics that can vary between distress cries and non-distress cries include temporal instability, intensity of cries, pitch of cries, harmonic structure, and frequency distribution [8]. Therefore, recent studies have turned to hybrid acoustic representations, which include the MFCCs along with root-mean-square (RMS) energy, zero-crossing rate, pitch-related descriptors, spectral centroid, spectral bandwidth, spectral contrast, chroma and Mel-spectrogram statistics. This multi-domain representation is necessary because the discriminative structure of infant cry categories is not always determined by a single acoustic cue, but is more often the result of the interaction of complementary acoustic cues [9].

Infant cry classification has been a focus point of machine learning. Structured acoustic descriptors have been shown to be useful for classification, for example

by classical models like support vector machine (SVM), k-nearest neighbors (KNN), Random Forest or boosting based classifiers, especially when data is limited. The nonlinear decision boundaries of SVM are particularly important, as they can model complex relationships in high dimensional hybrid feature spaces [10]. Meanwhile, deep learning methods such as convolutional neural networks (CNNs), long short-term memory (LSTM), CNN-LSTM, and spectrogram-based models have been used to directly learn time-frequency patterns from the transformed acoustic data. While these models can enhance the representational ability, they strongly rely on the size of the data set, class balance, recording conditions, and evaluation protocol. Thus, complexity of the model does not necessarily imply generalization between small, imbalanced biomedical acoustic datasets [11].

One of the major drawbacks in literature is that high accuracy could be due to methodological artifacts, not necessarily to recognition ability. Leakage may happen if augmented or acoustically correlated samples are used in both the training and testing sets, which results in models learning the patterns of the recording rather than the cry characteristics. Moreover, the overall accuracy may be deceiving in cases of severe class imbalance as good performance on majority class can hide poor performance on minority classes. Hence, to achieve credible evaluation, it is essential to use leakage-free partitioning, training-only augmented data, and imbalance sensitive measures like Macro-F1 and balanced accuracy [12].

In terms of systems, infant cry analysis is best utilized as a non-contact behavioral sensing layer on smart neonatal monitoring platforms. It should be used in conjunction with the physiological and environmental monitoring and not as a substitute for clinical diagnosis [13]. Therefore, models which are interpretable and edge-deployable are of crucial importance for practical applications in the neonatal space, where the computational cost is low, the ability to make local inferences is required, and the models must be robust to real acoustic conditions [14]. This inspires the current research, which focuses on leakage-free evaluation, hybrid acoustic feature fusion and lightweight machine learning for robust infant cry classification.

3 Materials and Methods

3.1 Dataset Description and Preprocessing

The dataset for this study is comprised of 512 original infant cries, divided into eight categories: 1) belly pain, 2) burping, 3) cold/hot, 4) discomfort, 5) hungry, 6) lonely, 7) scared, and 8) tired. The data set is very skewed, with the hungry class being the majority. This imbalance is due to the practical challenge of infant cry classification and thus, evaluation metrics need to take into account class-wise behavior [15]. The distribution of the original recordings across the eight cry categories is presented in Table 1.

Table 1 Distribution of infant cry samples across categories.

Class Label	Number of Files	Percentage (%)
Belly Pain	16	3.13%
Burping	18	3.52%
Cold/Hot	7	1.37%
Discomfort	30	5.86%
Hungry	382	74.61%
Lonely	11	2.15%
Scared	20	3.91%
Tired	28	5.47%
Total	512	100%

3.2 Acoustic Feature Extraction

The feature extraction was designed to be a multi domain representation of the audio, not just a Mel Frequency Cepstral Coefficients (MFCC). The spectral envelope was measured in MFCC statistics, the RMS energy was used for vocal intensity and respiratory effort, the zero-crossing rate (ZCR) was used for irregularity of the vocal

excitation, and pitch-related descriptors for laryngeal tension and vocal stability. Other features, such as chroma, spectral contrast, spectral centroid, spectral bandwidth and Mel-spectrogram statistics, were related to harmonic structure and frequency-energy distribution [16]. This hybrid representation combines complementary spectral, temporal, energy-based and pitch related information to minimize the confusion of acoustically identical cry classes. The audio recordings were resampled to 22,050 Hz and cut to a maximum length of 5 s. For each recording, 13 MFCC coefficients were extracted, and the mean and standard deviation of each coefficient were calculated. Chroma and spectral contrast were computed with the Librosa default parameters and presented as average values. The temporal averages of the spectral centroid, spectral bandwidth, zero crossing rate and RMS energy were used. Global mean and standard deviation were calculated for mel-spectrograms obtained from 64 Mel frequency bands, and added as additional acoustic descriptors. The Librosa Python library was used for all the procedures for feature extraction. The extracted acoustic features and their interpretations are summarized in Table 2. Representative acoustic representations of an infant cry signal, including the waveform, Mel-spectrogram, and MFCCs, are shown in Fig. 1.

Table 2 Extracted acoustic features and their physiological interpretation.

Feature Category	Feature Name	Description	Physiological Interpretation
Spectral	MFCC (Mean, Std)	Compact representation of spectral envelope	Captures vocal tract characteristics and overall acoustic structure of the cry
Temporal	Zero-Crossing Rate (ZCR)	Rate of sign changes in the signal	Reflects excitation irregularity and instability in vocal fold vibration
Energy-based	RMS Energy	Root Mean Square of signal amplitude	Indicates respiratory effort and vocal intensity during crying
Pitch-related	Fundamental Frequency (Pitch Mean)	Average pitch over the segment	Represents laryngeal tension and emotional intensity (e.g., pain vs. discomfort)
Pitch-related	Pitch Variability	Variation of pitch across time	Indicates instability or modulation in vocal production
Spectral	Spectral Centroid	Center of mass of the spectrum	Reflects perceived brightness or sharpness of the cry
Spectral	Spectral Bandwidth	Spread of frequencies around the centroid	Indicates distribution of acoustic energy across frequencies
Spectral	Spectral Contrast	Difference between spectral peaks and valleys	Captures harmonic structure and resonance differences
Harmonic	Chroma Features	Distribution of energy across pitch classes	Represents harmonic content and tonal structure of vocalization
Spectral	Mel-Spectrogram (Statistics)	Time–frequency energy representation	Encodes global acoustic patterns and temporal evolution of cry signals
Dynamic (Temporal)	Δ MFCC (Delta)	First-order derivative of MFCC	Captures temporal transitions in spectral structure
Dynamic (Temporal)	Δ^2 MFCC (Delta-Delta)	Second-order derivative of MFCC	Models acceleration and dynamic variation in cry patterns

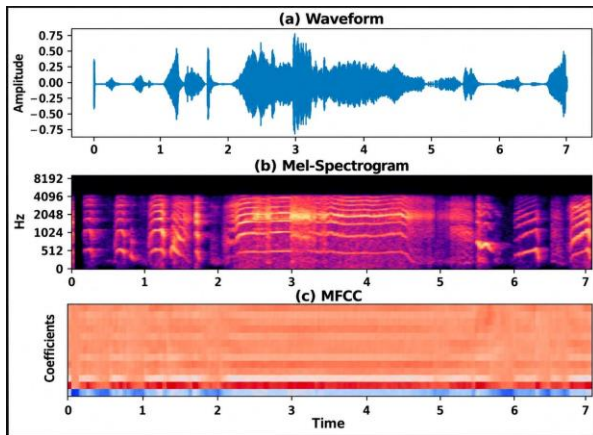


Fig. 1 Acoustic representations of infant cry signals: (a) waveform, (b) Mel-spectrogram, (c) MFCC.

3.3 Evaluation Protocol and Model Comparison

The four lightweight classifiers were tested under the same preprocessing and feature environment: radial basis function (RBF)-SVM, Random Forest, Gradient Boosting and multilayer perceptron (MLP). For severe imbalance, three metrics were used to assess performance: accuracy, Macro-F1, and balanced accuracy. The same models were also tested using a leaky protocol to investigate methodological sensitivity. These results were only used as a comparative measure of performance inflation and were not considered to be valid deployment performance [17]. The leakage-free evaluation workflow is illustrated in Fig. 2. The classifier hyperparameters are summarized in Table 3.

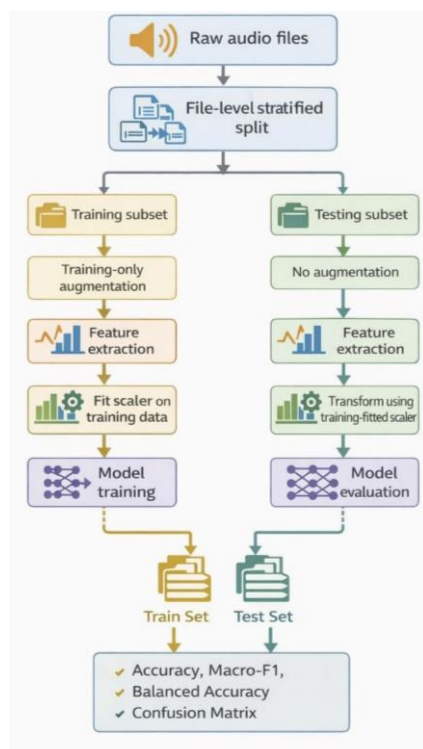


Fig. 2 Leakage-free evaluation pipeline.

Table 3 Hyperparameter Settings of the Evaluated Classifiers.

Model	Hyperparameter Settings
SVM (RBF)	Kernel = RBF, C = 5, Gamma = scale
Random Forest	Number of trees = 250, Max depth = None (unrestricted), Random state = 42
Gradient Boosting	Number of estimators = 200, Learning rate = 0.1, Random state = 42
MLP	Hidden layers = (256, 128), Dropout = (0.35, 0.25), Optimizer = Adam, Loss = Categorical Cross-Entropy, Batch size = 16, Epochs = 100, Early Stopping (patience = 10, restore_best_weights = True)

The hyperparameter settings listed in Table 3 were fixed throughout all experiments to ensure fair comparison among the evaluated classifiers. The experiments were done in Python with the following libraries: Scikit-learn, TensorFlow (Keras), Librosa, NumPy, Pandas, Matplotlib and Seaborn.

4 Results and Discussion

4.1 Data Characteristics and Evaluation Conditions

The dataset also bears a high level of imbalance in the classes, as hungry takes an overbearing size in the sample space. Such an imbalance directly affects the behavior of the classifier since models do not need to be overly accurate in order to obtain relatively high overall accuracy by promoting majority categories. This is why the results that are being reported are mostly interpreted using Macro-F1 and Balanced Accuracy as opposed to accuracy [18]. In order to reduce the class imbalance, the data augmentation was performed only on the training subset following the file-level partitioning. One of the following augmentation methods was applied to each of the songs randomly: additive Gaussian noise, time stretching ($0.75\text{--}0.95\times$ or $1.05\text{--}1.25\times$), pitch shifting (-3 to -1 or $+1$ to $+3$ semitones), or no transformation. The number of training samples per class for minority classes was increased to 180 and the testing set was not augmented. The class-wise distribution of the training and testing sets is presented in Table 4.

4.2 Leakage-Free Performance

The leakage-free protocol yielded the most balanced performance for SVM with accuracy of 86.41%, Macro-F1 of 0.7376 and balanced accuracy of 68.43%.

Table 4 Dataset distribution under the leakage-free protocol.

Cry Category	Original Files	Training Files (Before Augmentation)	Training Samples (After Augmentation)	Testing Files
Belly Pain	16	13	180	3
Burping	18	14	180	4
Cold–Hot	7	6	180	1
Discomfort	30	24	180	6
Hungry	382	305	305	77
Lonely	11	9	180	2
Scared	20	16	180	4
Tired	28	22	180	6
Total	512	409	1565	103

Table 5 Model performance under leakage-free evaluation.

Model	Accuracy (%)	Macro-F1	Balanced Accuracy (%)
SVM (RBF)	86.41	0.7376	68.43
Random Forest	86.41	0.7068	65.46
MLP	81.55	0.6289	65.69
Gradient Boosting	84.47	0.5649	53.68

Random Forest had the same accuracy, but lower Macro-F1 and balanced accuracy scores mean that it is less sensitive to the minority classes. The performance of the evaluated classifiers is summarized in Table 5. The results of MLP and Gradient Boosting indicated a lower class-balanced reliability, thus illustrating that overall accuracy can overestimate the model's performance in the extreme case of class imbalance. Overall, SVM provided the best trade-off between accuracy and class-wise robustness.

4.3 Effect of Data Leakage

All models exhibited significant overperformance when using the leaky protocol. The results showed that MLP obtained the highest accuracy of 94.83%, with the Macro-F1 and balanced accuracy of 0.9483 and 95.02%, respectively, while SVM gave the accuracy of 91.79% and the balanced accuracy of 92.10%. This improvement was observed across models, and was primarily due to the evaluation protocol, and not to the model architecture. The results show that data leakage can greatly overestimate the reliability of classification, and leakage-free evaluation is a prerequisite for meaningful biomedical acoustic evaluation. The model-wise performance comparison is shown in Fig. 3. The effect of data leakage on classification performance is

illustrated in Fig. 4. The detailed results under the leaky evaluation protocol are reported in Table 6.

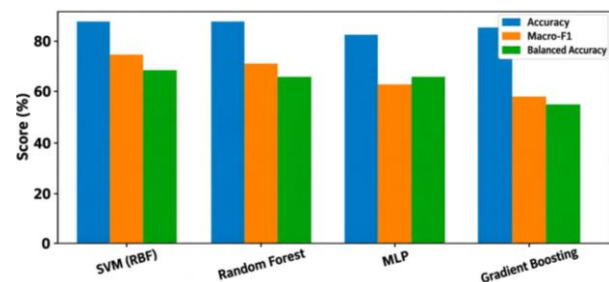


Fig. 3 Model performance comparison.

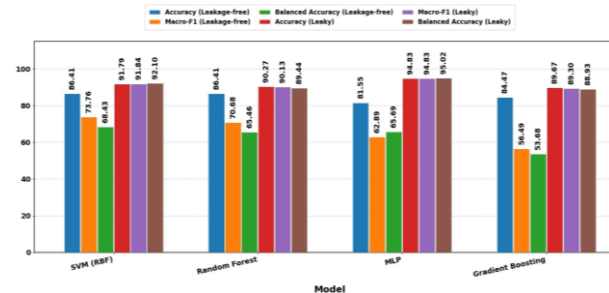


Fig. 4 Effect of data leakage on performance.

Table 6 Model performance under leaky evaluation.

Model	Accuracy (%)	Macro-F1	Balanced Accuracy (%)
SVM (RBF)	91.79	0.9184	92.10
Random Forest	90.27	0.9013	89.44
MLP	94.83	0.9483	95.02
Gradient Boosting	89.67	0.8930	88.93

4.4 Structure and Confusion Patterns of Errors

The analysis of confusion matrix also explains the behavior of models. The leakage-free SVM results depict high diagonal concentration of dominant classes, which are an indication that the model represents the major structure of the dataset. Nonetheless, there are still systematic cases of misclassification between acoustically similar classes, especially between classes which have overlapping energy and pitch properties.

The misclassification errors are not arbitrary and mean something underlying in the feature space. As an illustration, the similarity between the forms of discomfort and tiredness can be explained by their comparable RMS distribution and limited differences in pitch variation, whereas the overlap between the scared and pain-related conditions can be the excessively high pitch in both cases.

Unlike SVM, this model is not as separable with respect to classes, which are represented by the higher the off-diagonal values of the model. This demonstrates the significance of nonlinear separation processes in the determination of the finer acoustic variations. The confusion matrices of the SVM and Gradient Boosting models are presented in Figs. 5 and 6, respectively.

4.5 Multidimensional Feature Space Analysis

To facilitate qualitative interpretation of the multidimensional acoustic feature representation, selected combinations of representative acoustic descriptors (e.g., RMS energy, zero-crossing rate, and pitch-related features) were visualized. These visualizations provide an intuitive representation of the feature distribution rather than a formal dimensionality reduction technique such as principal component analysis (PCA), *t*-distributed stochastic neighbor embedding (*t*-SNE), or uniform manifold approximation and projection (UMAP).

When individual features are viewed separately, it is demonstrated in the visualization that they do not offer clear separation. Rather, the class clusters are just found when features are studied together in a multidimensional

space. Intersection between clusters is a positive indication that the classification problem is inherently nonlinear. This fact is the direct cause of the high performance of SVM with an RBF kernel which can capture smooth surfaces with the underlying geometry of the feature space. Conversely, simple partitioning strategies are not effective in models that use the simplistic strategies. The pairwise distributions of the selected acoustic features are illustrated in Fig. 7.

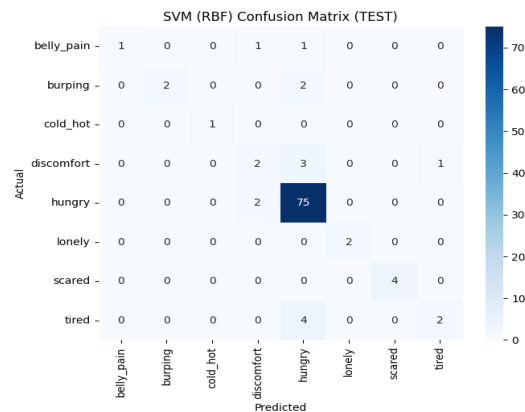


Fig. 5 Confusion matrix of the SVM model.

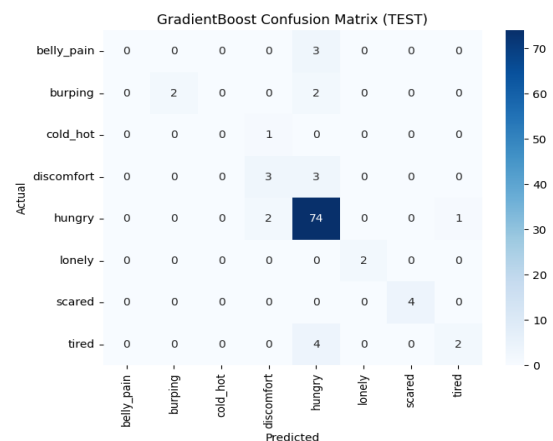


Fig. 6 Confusion matrix of the Gradient Boosting model.

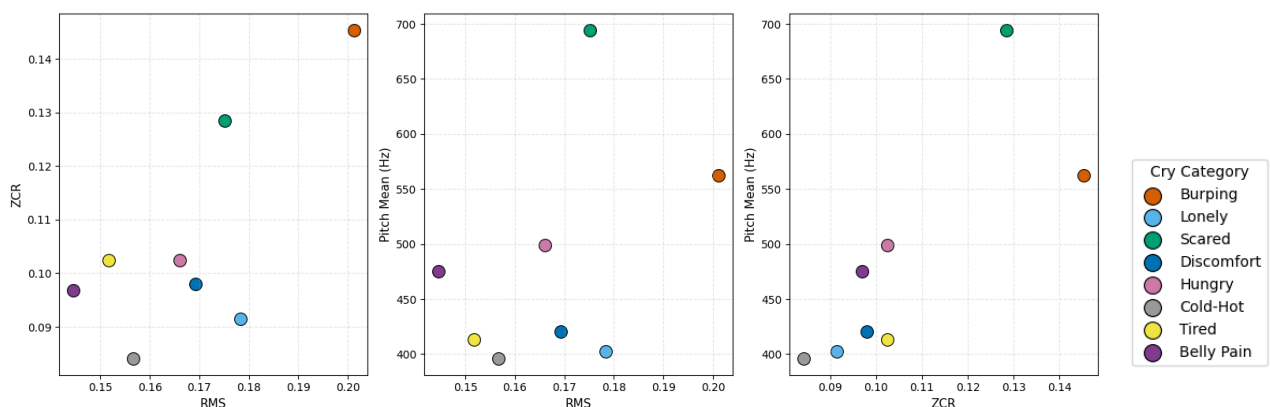


Fig. 7 Acoustic feature space visualization. (a) RMS vs ZCR, (b) RMS vs Pitch Mean, and (c) ZCR vs Pitch Mean.

4.6 Interpretation of Cry Classification As a Feature

The classification behaviour observed indicates that the recognition of infant crying is more related to the interaction of a multitude of acoustic features than to the complexity of the model. The RMS energy represents the intensity of the voice and the effort used to produce it, the ZCR measures the irregularity of the voice's excitation, and the pitch-related descriptors show the tension and stability of the voice. Variations in pitch, temporal irregularity and energy thus combine to produce differences between the cry categories, not just one of these. The results are consistent with the use of hybrid acoustic feature integration to be a more reliable representation of distinguishing overlapping infant cry patterns. Class-wise acoustic feature characteristics are summarized in Table 7.

4.7 Integration into Smart Neonatal Monitoring Systems

For easier implementation, the proposed classification system can be embedded in a smart neonatal monitoring system as illustrated in Fig. 8. The system integrates environmental sensors, physiological sensors and an acoustic sensing unit based on the microphone. The features extracted and processed on the edge device (Raspberry Pi) are then fed into the SVM model for real time inference. The summarized outputs can then be sent remotely via an IEEE 802.11ac wireless network for scalable multi-incubator monitoring [19]. This architecture is not designed to use cry classification as a diagnostic tool, but rather as a behavioral inference layer, working in conjunction with physiological and environmental monitoring, and to help raise the alert of the caregiver [20].

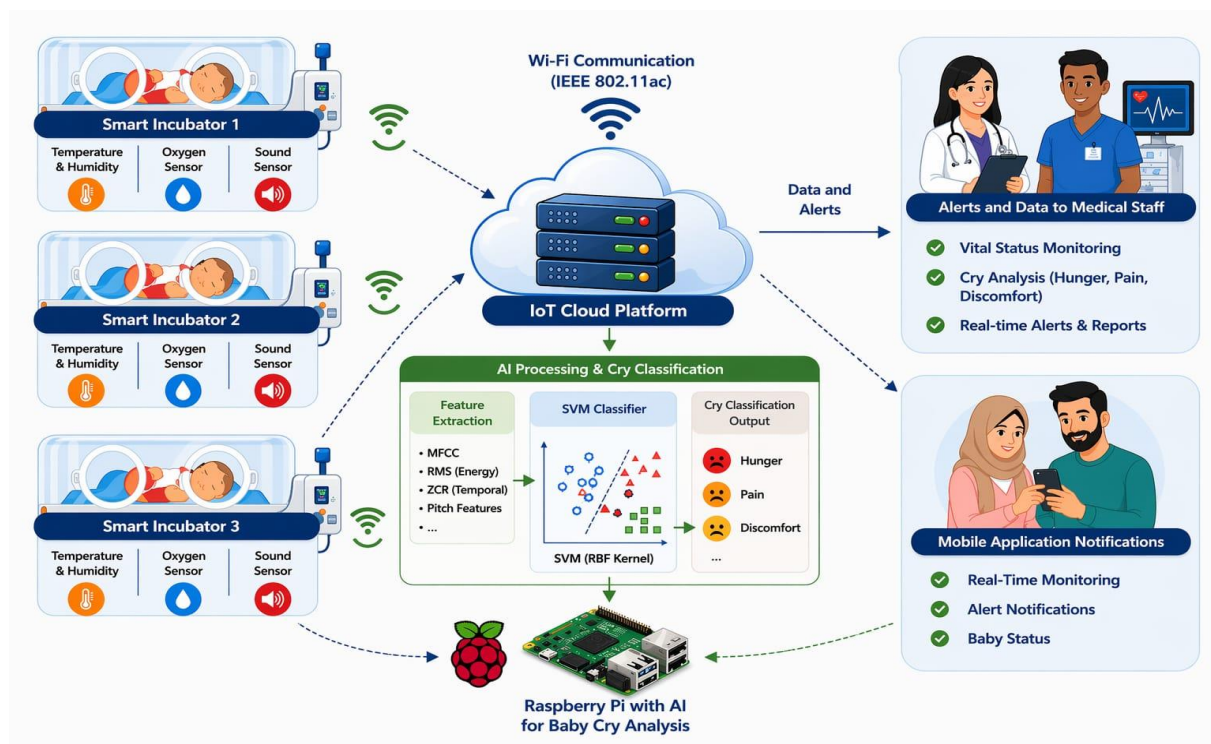


Fig. 8 IoT-based neonatal monitoring architecture.

Table 7 Acoustic feature characteristics across cry categories.

Cry Category	RMS (After)	ZCR (After)	Pitch Mean (Hz)	Pitch Variability (Hz)	Dominant Acoustic Profile	Physiological / Behavioral Interpretation
Burping	0.201178	0.145341	562.609645	164.036446	High energy, high irregularity, moderately high pitch, very high variability	Explosive airflow bursts with unstable excitation; energy–excitation dominated vocalization
Loneliness	0.178320	0.091464	402.334851	54.401323	Relatively low irregularity, low-to-moderate pitch, very low variability	Sustained and stable phonation with limited emotional escalation; low-activation cry

Table 7 (Cont.)

Scared	0.175196	0.128498	694.229246	125.675044	High irregularity, very high pitch, high variability	Emotionally aroused and unstable vocal excitation; fear-related high-arousal cry
Discomfort	0.169246	0.097998	420.154448	116.828038	Relatively high energy, moderate irregularity, moderate pitch, moderate variability	Sustained physical irritation with controlled but effortful phonation; energy-dominant discomfort cry
Hungry	0.166051	0.102475	499.155349	143.861676	Moderate energy, moderate irregularity, moderately high pitch, high variability	Rhythmic and repetitive distress signaling with balanced multidimensional activation
Cold–Hot	0.156627	0.084085	395.957094	82.902765	Low irregularity, low pitch, low variability	Intermittent and relatively weak vocal effort with limited activation; low-energy cry
Tired	0.151715	0.102363	413.109759	89.635140	Low energy, low-to-moderate irregularity, low-to-moderate pitch, low variability	Reduced respiratory support and diminished arousal; fatigue-dominated cry
Belly Pain	0.144520	0.096835	475.233171	84.366574	Low-to-moderate energy, moderate irregularity, moderate-to-high pitch, low variability	Sustained laryngeal tension with limited modulation; pain-related cry dominated by controlled pitch tension

5 Key Insight

The findings indicate that infant cry classification depends on feature interaction, leakage-free evaluation, and model suitability rather than model selection alone for reliable biomedical monitoring.

6 Conclusion

This paper has looked at infant cry classification in terms of methodological reliability as opposed to raw benchmarks performance. The findings reveal that the effect of acoustic representation, imbalance among classes, and testing procedures are leading to highly influential classification results. In a leakage free environment, SVM performed best in terms of balanced performance relative to the tested models, whereas, Random Forest performed equally well in terms of accuracy but not in terms of reliability by class. Limited data were more sensitive to neural performance than to Gradient Boosting, and the balanced behavior of leaky evaluation was the poorest. Increased performance of over 90 percent, in leaky conditions, was not a measure

of greater generalization but less experiment independence. This shows that the effectiveness of evaluation design can overshadow the seemingly successful models. As an applied contribution, the results endorse lightweight, interpretable acoustic model usage as a part of smart neonatal monitoring systems. Nevertheless, they too indicate that cry-only inference is constrained in extreme imbalance and acoustic overlap. It should then be a focus of future development to use a more diverse data, more subject-independent evaluation, and multimodal integration with physiological sensing in order to implement it in intelligent neonatal care settings.

7 Limitations and Future Work

7.1 Limitations

The interpretation of the current findings has a number of limitations. To begin with, the dataset is extremely skewed, and augmentation is not able to replace authentic acoustic diversity in minority classes. Second, the feature representation can be considered statistical

and potentially underrepresents longer temporal dependencies in the sequence of cry. Third, the experiments were done on rather controlled recordings but in the actual incubator settings there is a lot of acoustic interference. Fourth, it could not ensure strict subject-level independence in spite of leakage being managed at the file level. Lastly, the research was narrowed down to the classification of cry only, and is not yet multimodal in terms of physiological context. No formal statistical significance testing (such as McNemar's test or bootstrap confidence intervals) was conducted to compare the classifiers evaluated under the leakage-free evaluation protocol, although the SVM classifier yielded the highest Macro-F1 score and balanced accuracy. So, the results of the observed performance differences should be taken as an empirical difference and not statistically significant. Future work will consider the use of formal statistical hypothesis testing for providing more evidence in favor of model comparison.

7.2 Future Work

In the future all the research must emphasize data-centric enhancement of a system rather than architectural magnification. Balanced and more diverse data sets should be sought especially on underrepresented distresses. Where possible, subject-independent partitioning must be imposed. Statistical frameworks with time-frequency representations can be more effective at capturing global and dynamic structure in a hybrid way. Fusion with physiological measurements, including temperature, SpO₂ and heart rate, is of particular interest to Multimodal fusion as it allows to reduce the similarity of acoustically similar conditions when deploying the smart incubator. Moreover, lighter edge-conscious executions are also to be considered in order to enable real-time inference of embedded neonatal applications.

Disclosures

All authors declare that there is no conflict of interest in this paper.

References

1. S. Sirithemmontree, N. Katchamat, and S. Nuampa, "The application of machine learning for infant cries classification and pathological cries detection: A systematic review," *Science Progress* 109(1), 00368504251410776 (2026).
2. T. Ozcan, H. Gungor, "Baby Cry Classification Using Structure-Tuned Artificial Neural Networks with Data Augmentation and MFCC Features," *Applied Sciences* 15(5), 2648 (2025).
3. G. Aggarwal, K. Jhajharia, J. Izhar, M. Kumar, and L. Abualigah, "A Machine Learning Approach to Classify Biomedical Acoustic Features for Baby Cries," *Journal of Voice* 39(6), 1446–1455 (2025).
4. M. Mekhfioui, W. Fadel, F. E. Hammouch, O. Laayati, M. Bouchouirbat, N. El Bazi, A. Satif, T. Boujiha, and A. Chebak, "Development of a Baby Cry Identification System Using a Raspberry Pi-Based Embedded System and Machine Learning," *Technologies* 13(4), 130 (2025).
5. M. Hammoud, M. N. Getahun, A. Baldycheva, and A. Somov, "Machine learning-based infant crying interpretation," *Frontiers in Artificial Intelligence* 7, 1337356 (2024).
6. K. Rezaee, H. G. Zadeh, L. Qi, H. Rabiee, and M. R. Khosravi, "Can You Understand Why I Am Crying? A Decision-making System for Classifying Infants' Cry Languages Based on DeepSVM Model," *ACM Transactions on Asian and Low-Resource Language Information Processing* 23(1), 1–17 (2024).
7. X. Qiao, S. Jiao, H. Li, G. Liu, X. Gao, and Z. Li, "Infant cry classification using an efficient graph structure and attention-based model," *Kuwait Journal of Science* 51(3), 100221 (2024).
8. F. S. Matikolaie, C. Tadj, "Machine Learning-Based Cry Diagnostic System for Identifying Septic Newborns," *Journal of Voice* 38(4), 963.e1-963.e14 (2024).
9. F. Li, C. Cui, and Y. Hu, "Classification of Infant Crying Sounds Using SE-ResNet-Transformer," *Sensors* 24(20), 6575 (2024).
10. K. Zhang, H.-N. Ting, and Y.-M. Choo, "Baby cry recognition based on WOA-VMD and an improved Dempster-Shafer evidence theory," *Computer Methods and Programs in Biomedicine* 245, 108043 (2024).
11. H. A. Patil, A. Kachhi, and A. T. Patil, "CQT-Based Cepstral Features for Classification of Normal vs. Pathological Infant Cry," *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 32, 4713–4726 (2024).
12. Y. Zayed, A. Hasasneh, and C. Tadj, "Infant Cry Signal Diagnostic System Using Deep Learning and Fused Features," *Diagnostics* 13(12), 2107 (2023).
13. T. Ozseven, "Infant cry classification by using different deep neural network models and hand-crafted features," *Biomedical Signal Processing and Control* 83, 104648 (2023).
14. Z. Wang, Y. Cai, X. Wang, S. Wu, Y. Cao, F. Xu, and M. Huang, "The significance of an infant's cry: a narrative review of physiological, pathological, and analytical perspectives," *Frontiers in Pediatrics* 13, 1558951 (2025).
15. M. Ahmed, "Baby Cry Dataset," Kaggle, 2023 (accessed 11 November 2025). <https://www.kaggle.com/datasets/mennaahmed23/baby-cry-dataset>.
16. A. Abbaskhah, H. Sedighi, and H. Marvi, "Infant cry classification by MFCC feature extraction with MLP and CNN structures," *Biomedical Signal Processing and Control* 86, 105261 (2023).

17. M. Ashraf, F. Abid, I. U. Din, J. Rasheed, M. Yesiltepe, S. F. Yeo, and M. T. Ersoy, “[A Hybrid CNN and RNN Variant Model for Music Classification](#),” *Applied Sciences* 13(3), 1476 (2023).
18. Y.-C. Liang, I. Wijaya, M.-T. Yang, J. R. Cuevas Juarez, and H.-T. Chang, “[Deep Learning for Infant Cry Recognition](#),” *International Journal of Environmental Research and Public Health* 19(10), 6311 (2022).
19. Z. K. Farej and H. Y. Al-hayaly, “[Accuracy Evaluation of Healthcare Monitoring System Based on ESP32 Microcontroller with IoT](#),” in 2023 International Conference on Engineering, Science and Advanced Technology (ICESAT) (IEEE, 2023), pp. 90–94.
20. Z. K. Farej, M. A. Hassan, “[Analysis and Performance Evaluation of the Static and Dynamic Channel Bonding for the IEEE 802.11ac WLAN](#),” *International Journal of Interactive Mobile Technologies* 17(13), 168–185 (2023).