IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CE-39 Feature-Model Synergy in Cry-Based Detection of Neonatal Asphyxia Using Hybrid Cepstral and Neural Approaches

N Sriraam [1], Aditi Anil Kulkarni [2], Smruthi Arun Kumar [3], Sumedha Tatti[4]

[1]*Center for Medical Electronics and Computational Intelligence, Dayananda Sagar College of Engineering, Bengaluru, India Department of Medical Electronics Engineering, Dayananda Sagar College of Engineering, Bengaluru, India Pin:560111; dr.sriraam.natarajan@ieee.org*

[2,3,4]*Department of Medical Electronics, Dayananda Sagar College of Engineering, Bengaluru, India Pin: 560111*

Abstract – *This research deals with the automated identification of neonatal asphyxia based on cry signals, where the performance of various cry signal feature extraction techniques and machine learning classifiers are compared. Four sets of hybrid features are used to assess their performance in extracting useful pathological features: Δ and ΔΔ Mel-Frequency Cepstral Coefficients (MFCC) with Mean Greenwood Cepstral Coefficients (µGWCC), Gammatone Cepstral Coefficients (GTCC) with µGWCC, Teager Energy Operator (TEO) features with µGWCC, and Wavelet Transform-based MFCC features with µGWCC. These features are used to train and test eight classifiers: Multilayer Perceptron (MLP), Support Vector Machine (SVM), Random Forest (RFM), Convolutional Neural Network (CNN), Deep Neural Network (DNN), Recurrent Neural Network (RNN), Bidirectional Long Short-Term Memory (BiLSTM), and Convolutional Recurrent Neural Network (CRNN). Among all tested configurations, it is found that the GTCC + µGWCC feature set with the MLP classifier produced the highest performance with an accuracy of 95.88% and AUC of 0.99 for asphyxia classification. The performance of CNNs is also found to be high with Δ + $Δ^2$ MFCC and TEO features, yielding up to 95.29% accuracy and AUC between 0.98 and 0.99. The results of this study show that it is possible to distinguish between asphyxiated cries and healthy cries by using enriched perceptual and temporal features. This study shows that cry-based acoustic analysis is a potential non-invasive tool for early screening of neonatal asphyxia*

Keywords- *Asphyxia, Mel-Frequency Cepstral Coefficients (MFCCs). Gammatone Cepstral Coefficients (GTCC), Teager Energy Operator (TEO), Wavelet-transformed MFCCs, Convolutional Neural Network (CNN), Support Vector Machine (SVM), Bidirectional Long Short-Term Memory (BiLSTM)*

INTRODUCTION

A. Background

Infant cries are the first form of vocal communication, which carry important information concerning the physiological and neurological state of the newborn. While the main information contained in an infant's cries is related to their basic needs, the acoustic features of the cries also carry information concerning the presence of medical abnormalities. For instance, changes in the acoustic features of the cries have been related to neurological disorders, which indicates that the analysis of the cries can be an important tool for the assessment of the medical state of the newborn [1, 2].

One of the important medical conditions that have been related to the acoustic features of the cries is neonatal asphyxia, which is usually caused by hypoxic-ischemic encephalopathy due to the lack of oxygen at the time of birth. Neonatal asphyxia is one of the major factors that cause mortality and neurological disorders in the neonatal population. Research indicates that the lack of oxygen affects the brain, which controls the vocal system, resulting in changes in the acoustic features of the cries. As a result, the acoustic features of the cries have distinctive characteristics that can be used for the early detection of the presence of pathological conditions [3].

In recent times, with advancements in biomedical signal processing and machine learning, it has been possible to automatically analyze infant cry signals for diagnostic purposes. Various research has been conducted to explore the application of acoustic feature extraction techniques, such as cry classification and disease detection, by employing techniques such as deep learning [1], [4]. Comprehensive surveys have been conducted on this topic, which has highlighted the increasing trend of employing machine learning techniques to differentiate between normal and pathological cry signals, thereby establishing the potential of employing cry signal analysis systems for neonatal healthcare applications [2], [5].

The development of reliable systems for automatically analyzing infant cry signals has great potential to improve healthcare conditions by enabling early detection and timely intervention for conditions such as asphyxia, which affect neonates.

B. Database

The current research uses the Baby Chillanto Cry Database, which is a collection of cry sounds from infants developed for the purpose of conducting research in the field of biomedical signal processing. The dataset used for the current research includes audio samples from full-term newborns, which have been categorized into two different groups: healthy and newborn asphyxia. The audio files have been stored as WAV files with a standard sampling rate, which makes the data suitable for conducting computational research. The dataset has been developed for the purpose of conducting research on the diagnostic potential of cry sounds, which helps researchers explore the possibilities of non-invasive diagnosis of physiological and neurological abnormalities among newborns. The dataset also includes metadata regarding the diagnosis and conditions, which makes it suitable for conducting research in the field of health informatics.

C. Feature Extraction Methods

To effectively characterize the acoustic properties of infant cries and distinguish between healthy and asphyxiated conditions, several hybrid feature sets are extracted. Each set incorporates µGWCC (mean Greenwood Function Cepstral Coefficients), which provide a biologically inspired frequency representation aligned with the tonotopic organization of the human cochlea. The following combinations are evaluated:

i. ΔMFCC + Δ²MFCC + µGWCC

This set of features is a combination of first-order and second-order temporal derivative features of Mel Frequency Cepstral Coefficients (MFCCs) and Mean Greenwood Function Cepstral Coefficients (µGWCC). The features ΔMFCC and Δ²MFCC are temporal derivative features of MFCCs and are efficient in describing changes in vocal characteristics during a cry signal, which could be indicative of neurological distress. The addition of µGWCC makes it a more biologically relevant set of features as it represents frequency information in terms of human auditory perception.

ii. GTCC + µGWCC

Gammatone Frequency Cepstral Coefficients (GTCCs), obtained by using Gammatone filter banks, have been used to represent the frequency selectivity of the human auditory system in a more accurate manner. The GTCCs represent the perceptually relevant features that can be used for analysing the cry signal. The feature set obtained by combining GTCCs with µGWCC provides a rich representation of both auditory physiology and perceptual features that can be used to detect anomalies in the cry signal.

iii. TEO + µGWCC

The Teager Energy Operator (TEO) is used to obtain features that represent the instantaneous non-linear changes in energy for the cry signal. TEO features can be used to represent the energetic changes in the signal that can be associated with stress/pathology. The feature set obtained by combining TEO with µGWCC provides a dual representation of the signal.

iv. Wavelet MFCC + µGWCC

Wavelet-based MFCCs are extracted by applying discrete wavelet transforms before Cepstral analysis. This provides a multiresolution analysis of the temporal and spectral information contained in the signal. This is because it is able to capture subtle and broad acoustic changes, which are important in analysing complex patterns in cries of infants. The feature set is further enhanced by including µGWCC, which is able to map features as per human cochlear frequency mapping and is sensitive to abnormal cry features of asphyxia.

D. Classification Models

To evaluate the performance of the extracted cry features, several machine learning and deep learning models were employed, each offering distinct strengths in handling structured and temporal data.

i. **Multi-Layer Perceptron (MLP)**

MLP is a feedforward neural network with multiple fully connected layers and nonlinear activations. It was used as a baseline deep learning model due to its simplicity and effectiveness with structured features like MFCCs and GWCCs.

ii. Support Vector Machine (SVM)

SVM is a supervised classifier that identifies an optimal hyperplane to separate classes with maximum margin. It performs well in high-dimensional spaces and was applied to feature sets combining Cepstral and perceptual descriptors.

iii. Random Forest Model (RFM)

Random Forest is an ensemble method based on multiple decision trees. Its ability to handle non-linear feature interactions and resist overfitting made it suitable for classifying hybrid cry features.

iv. Convolutional Neural Network (CNN)

CNNs are used to detect spatial patterns in structured inputs. Here, they analysed time-frequency representations such as spectrograms, capturing local acoustic features linked to distress.

v. Recurrent Neural Network (RNN)

RNNs are designed for sequential data and use feedback connections to model temporal dependencies. They were employed to track changes in cry features across time.

vi. Deep Neural Network (DNN)

DNNs extend MLPs with additional hidden layers, offering greater capacity for modelling complex patterns. They were used to learn non-linear relationships across high-dimensional acoustic features.

vii. Bidirectional LSTM (BiLSTM)

BiLSTM networks process input sequences in both forward and backward directions, capturing full temporal context. This made them effective for modelling the time structure of infant cries.

viii. **Convolutional Recurrent Neural Network (CRNN)**

CRNNs integrate CNN and RNN architectures, enabling joint learning of spatial and temporal features. They were well-suited for cry analysis by capturing both spectral content and its evolution.

It may be observed that the proposed hybrid feature extraction framework is quite efficient in extracting the complex acoustic characteristics associated with the cries of the infant. In the proposed framework, the incorporation of the feature set involving the μGWCC ensures a biologically meaningful representation that is consistent with the human perception of sound. The other feature extraction techniques used in the proposed framework, i.e., MFCC derivatives, GTCC, TEO, and wavelet transforms, ensure the enhanced sensitivity of the feature extraction mechanism to the subtle pathological changes.

LITERATURE REVIE

Table.1- Existing Literature

Title & Authors	Dataset	Features	Models	Results
A Multistage Heterogeneous Stacking Ensemble Model for Augmented Infant Cry Classification – Joshi et al. [1]	Infant cry dataset (multiple cry types)	MFCC spectrogram features	CNN variants (VGG16, YOLOv4), NuSVC, RF, XGBoost, AdaBoost	93.7%
CQT-Based Cepstral Features for Classification of Normal vs Pathological Infant Cry – Patil et al. [2]	Baby Chillanto, DA-IICT dataset	CQCC, MFCC, LFCC	GMM, SVM	99.8%, 98.24%

Infant Cry Signal Diagnostic System Using Deep Learning and Fused Features – Zayed et al. [3]	Infant cry dataset (RDS, sepsis, healthy)	GFCC, Harmonic Ratio, Spectrogram	RF, SVM, DNN	97.5%
On the Use of Long-Term Features in a Newborn Cry Diagnostic System – Salehian Matikolaie et al. [4]	Infant cry dataset for RDS detection	MFCC, Melody, Rhythm	SVM	F-score
Using CCA-Fused Cepstral Features in a Deep Learning-Based Cry Diagnostic System – Khalilzad et al. [5]	Inspiratory and expiratory cry datasets	MFCC, GFCC, CCA-fused cepstral features	SVM, LSTM	99.86%, 99.44%

There are some studies on the automatic analysis of infants' cries for determining infants' needs and diagnosing pathological states. Joshi et al. [6] presented a multistage heterogeneous stacking ensemble model for classification of infants' cries using different features such as MFCC-based spectrogram features and different ensemble learning algorithms. The accuracy of 93.7% is obtained in this study. Patil et al. [7] presented a set of features known as Constant Q Cepstral Coefficients for classifying normal and pathological cries of infants using Baby Chillanto dataset and obtained classification accuracies of 99.8% and 98.24% using GMM and SVM classifiers, respectively. Zayed et al. [8] presented a deep learning-based diagnostic tool using features such as GFCC, harmonic ratio, and spectrogram for diagnosing neonatal states such as respiratory distress syndrome and sepsis with 97.5% accuracy. Salehian Matikolaie et al. [9] used both short-term and long-term features like MFCC, melody, and rhythm to diagnose newborn baby cry using an SVM classifier. Khalilzad et al. [10] proposed a system for cry detection using a combination of MFCC and GFCC features with the help of canonical correlation analysis. In this system, SVM and LSTM classifiers have been used to achieve F-scores of 99.86% and 99.44%, respectively. Furthermore, Hashemi et al. [11] proposed a survey that covers various datasets, feature extraction techniques, and machine learning approaches for cry detection and classification.

Overall, these studies show the effectiveness of sophisticated techniques for extracting acoustic features and the use of machine and deep learning techniques for the classification of infant cry signals and the early diagnosis of various conditions.

METHOLOGY

A. Data Acquisition

The dataset used in this study is the Baby Chillanto Cry Database, which contains labeled audio samples of full-term babies, divided into normal and asphyxiated categories. For consistency, the audio samples must be using a sampling rate of 16 kHz with clear signal quality. The quality of the samples will be checked, and they will be labeled as normal (0) or asphyxiated (1).

B. Preprocessing

Each cry sound in its WAV format was run through a standard preprocessing procedure to make the sound clearer before the extraction of the features:

1. Pre-emphasis filtering was done to increase the sound level at higher frequency ranges to make them more audible.
2. Overlapping frames of 25 ms with a step size of 10 ms were created to capture the sound in terms of time.
3. A hamming window was applied to the frames to avoid spectral leakage.
4. Each frame was converted to the frequency domain using the FFT.
5. Each sound was log-compressed to make it more suitable to human hearing.
6. Standardization was done to bring all the sound features to a standard normal distribution.

To verify preprocessing steps and assist in pattern interpretation, spectrograms (via STFT) and scalograms (via CWT) were generated for selected recordings.

C. Feature Extraction

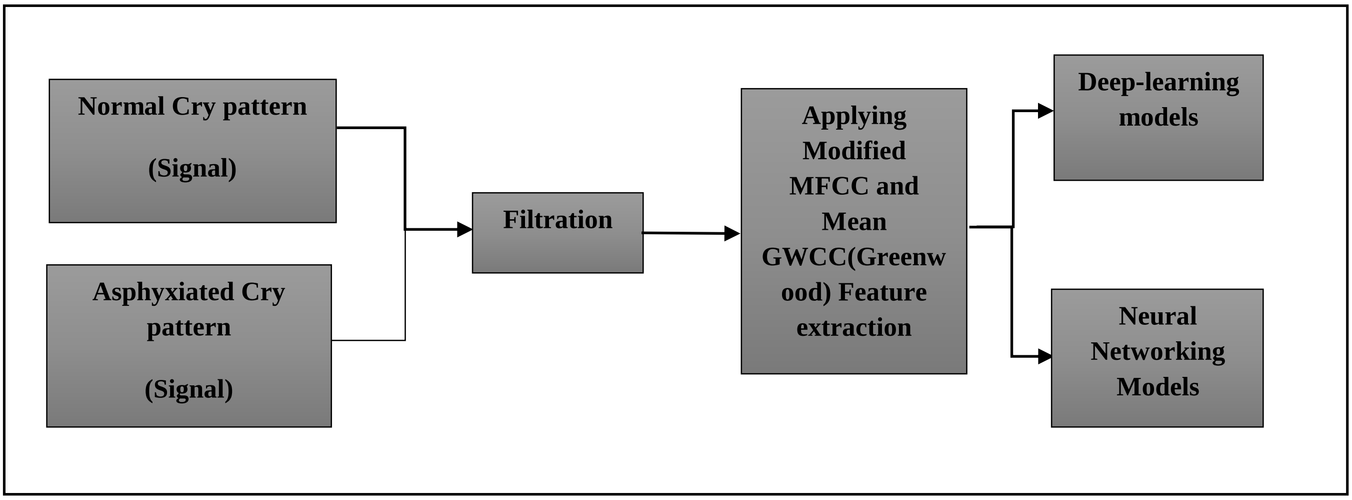

Fig.1: Block diagram of proposed methodology

Four different composite feature sets were identified to represent various acoustic as well as physiological characteristics of the cry signal. They are as follows:

i. **Δ + Δ² MFCC + µGWCC**: The feature set includes the first and second derivatives of Mel Frequency Cepstral Coefficients as well as Mean Greenwood Cepstral Coefficients.
ii. **GTCC + µGWCC**: Gammatone Cepstral Coefficients were used to model the auditory filter response. In addition to this, µGWCC provides cochlear-based frequency mapping.
iii. **TEO + µGWCC**: The feature set includes the application of the Teager Energy Operator to highlight non-linear energy changes. In addition to this, µGWCC provides spectral characteristics of the signal.
iv. **Wavelet MFCC + µGWCC**: In this feature set, Mel Frequency Cepstral Coefficients are applied to the wavelet-transformed signal to provide multiple resolution analysis. In addition to this, µGWCC provides biological characteristics.

Each of these feature sets was arranged in a structured form that is easily readable by classification models as a vector or a 2D/sequence matrix.

D. Classification

To classify the cry signals into *normal* or *asphyxiated*, a range of supervised models were employed, covering both traditional and deep learning approaches:

- **Machine Learning Models:**

i. **Support Vector Machine (SVM):** This algorithm works fine with high-dimensional spaces.
ii. **Random Forest (RFM)**: This is an ensemble method that performs well in avoiding overfitting.
iii. **Multi-Layer Perceptron (MLP)**: This is a basic neural network that works fine with structured features.

- **Deep Learning Models:**

i. **Convolutional Neural Network (CNN):** This network is generally used to handle 2D features to find patterns.
ii. **Deep Neural Network (DNN):** This is an extension of the MLP model with deeper architecture.
iii. **Recurrent Neural Network (RNN):** This model is generally used to handle the sequential patterns between time steps.
iv. **Bidirectional LSTM (BiLSTM):** This model is generally used to find patterns in time series in both directions.
v. **Convolutional Recurrent Neural Network (CRNN):** This model combines the spectral features with the time series patterns.

Machine learning models received **flattened feature vectors**, while deep learning architectures used **2D matrices** or **sequential inputs**, depending on model design.

RESULTS AND DISCUSSIONS

This study assessed the effectiveness of various models for binary classification of infant cry signals, aiming to differentiate between healthy and asphyxiated neonates. The evaluation focused on feature combinations involving Greenwood Cepstral Coefficients (GWCC) and Gammatone Frequency Cepstral Coefficients (GFCC). Model performance was analyzed using key metrics: classification accuracy, area under the ROC curve (AUC), and average precision derived from precision-recall analysis.

A. Performance Comparison: Machine Learning vs. Deep Learning

All models demonstrated strong potential in identifying cry patterns associated with neonatal asphyxia, though distinct differences were observed between the machine learning and deep learning paradigms.

i. Machine Learning Models

Within the group of machine learning classifiers, the Multi-Layer Perceptron (MLP) emerged as the top performer, consistently achieving classification accuracies above 95%. The Random Forest Model (RFM) also delivered solid results, demonstrating strong precision and generalization ability across various feature sets. Support Vector Machine (SVM), though reliable, yielded slightly lower accuracy and AUC scores in comparison to MLP and RFM. These models proved to be effective when paired with well-structured features and maintained a clear advantage in terms of lower computational overhead, making them suitable for constrained environments.

In contrast, deep learning models exhibited superior overall performance, particularly in capturing complex acoustic patterns. The Convolutional Neural Network (CNN) achieved the highest accuracy among all models, likely due to its ability to extract localized spectral features from time-frequency representations. Similarly, Deep Neural Networks (DNNs) demonstrated strong performance in learning non-linear relationships within high-dimensional feature spaces. Hybrid models such as the Convolutional Recurrent Neural Network (CRNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks also performed well, effectively capturing both spatial and temporal patterns in the cry signals.The standard Recurrent Neural Network (RNN), however, showed relatively modest performance, likely due to its limitations in modelling long-term dependencies and susceptibility to vanishing gradient issues.

Table.2 Tabulated Results

Sl.No	Features	Model	AUC	Classification Accuracy (%)
1	$\Delta + \Delta^2$ MFCC + µGWCC	MLP	0.99	93.53%
		SVM	0.94	87.65%
		RFM	0.98	89.41%
		CNN	0.98	95.29%
		DNN	0.98	92.35%
		RNN	0.98	91.76%
		BiLSTM	0.96	87.65%
		CRNN	0.98	90.59%
2	GTCC + µGWCC	MLP	0.99	95.88%
		SVM	0.95	87.65%
		RFM	0.98	92.94%
		CNN	0.98	92.94%
		DNN	0.98	92.94%
		RNN	0.98	93.93%
		BiLSTM	0.98	92.35%

		CRNN	0.98	93.53%
3	TEO + μGWCC	MLP	0.99	94.71%
		SVM	0.97	92.35%
		RFM	0.98	91.76%
		CNN	0.99	95.29%
		DNN	0.99	94.71%
		RNN	0.93	84.12%
		BiLSTM	0.96	90.00%
		CRNN	0.97	91.18%
4	Wavelet MFCC + μGWCC	MLP	0.99	94.71%
		SVM	0.95	87.06%
		RFM	0.98	90.00%
		CNN	0.98	93.53%
		DNN	0.97	92.35%
		RNN	0.96	89.41%
		BiLSTM	0.96	89.41%
		CRNN	0.97	88.82%

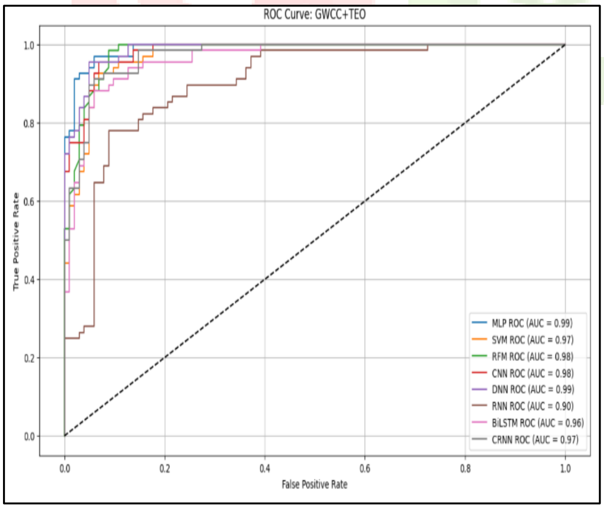

Fig.2- ROC plot of GWCC+ TEO

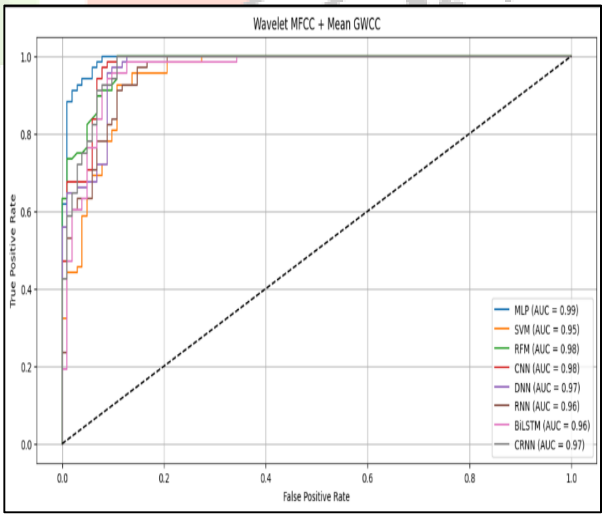

Fig.3- ROC plot of Wavelet MFCC+ Mean GWCC

B. Practical Considerations and Model Suitability

While deep learning models achieved higher classification metrics, their training complexity and computational demands may limit deployment in low-resource clinical settings. In contrast, simpler models like MLP and RFM offered a favorable balance between performance and efficiency, showing potential for use in mobile applications or point-of-care screening systems.

In summary, the results suggest that deep learning models are ideal when ample computational resources and large training datasets are available, whereas machine learning approaches remain viable for lightweight, real-time applications where interpretability and speed are prioritized.

CONCLUSION

The current study aims to provide a comprehensive framework for neonatal asphyxia detection with a set of hybrid cepstral features and various machine and deep learning models. In all evaluated models and configurations, the best-performing model was GTCC + μGWCC with an MLP model that showed 95.88% accuracy and an AUC of 0.99. Other models with a CNN showed robust and consistent results with a maximum accuracy of 95.29%. Most significantly, all proposed models with a set of hybrid features showed an AUC value between 0.97 and 0.99.

When compared to previous literature on this topic, the current study and its proposed models show robust and consistent results with a higher accuracy compared to previous models. In all evaluated models and configurations, previous models showed an accuracy between 93.7% and ~99.8% by Joshi et al. and Patil et al., and ~99.86% by Khalilzad et al. Although these models showed slightly better accuracy compared to this study in their best-performing models, these models showed robust and consistent results with most models showing an accuracy between 90 and 96%.From a feature perspective, it is evident that most of the existing literature utilizes either MFCC, CQCC, or GFCC as individual features or in combination, whereas the proposed method utilizes μGWCC-enhanced hybrid features, thereby increasing the physiological relevance of the features extracted. The use of μGWCC in all feature sets guarantees that all feature sets are benefiting from cochlear-inspired frequency mapping, unlike the existing literature.

From a model evaluation perspective, it is evident that most of the existing literature compares results of 1-3 models at most, whereas the proposed method compares results of 8 different models, including both traditional models (SVM, RF) and state-of-the-art models (CNN, RNN, BiLSTM, CRNN). Moreover, it is evident from the results of the proposed method that a relatively simple model, such as the proposed MLP model, is capable of achieving results within 3-4% of the state-of-the-art results reported in the literature while requiring significantly lower computational complexity, making it more viable for deployment in resource-constrained environments. Moreover, unlike the prior works, the emphasis of the proposed study is on the evaluation of the model in terms of multiple criteria; the proposed study ensures a high value of AUC ≈ 0.99 for the top models, which shows the high discriminative ability of the model, thereby reducing the chance of false classification, a major requirement for medical screening systems.

In summary, it is evident that the proposed work is better than the prior works in terms of overall robustness, consistency, physiological relevance, practical applicability, and evaluation criteria, though the prior works achieve a higher peak accuracy of around 3-4%.

REFERENCES

1. *S. M. H. Hashemi, H. Kolivand, W. Khan, and T. Saba,"Infant cry analysis: A survey of datasets, features, and machine learning techniques," IEEE Transactions on Affective Computing, 2025.*
2. *S. Sirithepmontree, N. Katchamat, and S. Nuampa,"The application of machine learning for infant cries classification and pathological cries detection: A systematic review," 2026.*
3. *[3] M. Satar, C. Cengizler, S. Hamitoglu, and M. Ozdemir,"Investigation of relation between hypoxic-ischemic encephalopathy and spectral features of infant cry audio," 2024.Ioan Doroftei; Victor Grosu and Veaceslav Spinu; "Omnidirectional mobile robot- Design and Implimentation" from "Gh.Asachi" Technical university of lasi, Romania.*
4. *F. Li, C. Cui, and Y. Hu,"Classification of infant crying sounds using SE-ResNet-transformer," Sensors, vol. 24, 2024.*
5. *G. Owino and B. Shibwabo,"Advances in infant cry paralinguistic classification—Methods, implementation, and applications: Systematic review," 2025.*
6. *V. R. Joshi, K. Srinivasan, P. M. D. R. Vincent, V. Rajinikanth, and C.-Y. Chang, "A multistage heterogeneous stacking ensemble model for augmented infant cry classification," Frontiers in Public Health, vol. 10, 2022.*
7. *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, vol. 32, pp. 4713–4726, 2024.*
8. *Y. Zayed, A. Hasasneh, and C. Tadj, "Infant cry signal diagnostic system using deep learning and fused features," Diagnostics, vol. 13, no. 12, 2023*
9. *F. Salehian Matikolaie and C. Tadj, "On the use of long-term features in a newborn cry diagnostic system," Biomedical Signal Processing and Control, vol. 59, 2020.*
10. *Z. Khalilzad and C. Tadj, "Using CCA-fused cepstral features in a deep learning-based cry diagnostic system for detecting an ensemble of pathologies in newborns," Diagnostics, vol. 13, no. 5, 2023.*
11. *S. M. H. Hashemi, H. Kolivand, W. Khan, and T. Saba, "Infant cry analysis: A survey of datasets, features, and machine learning techniques," IEEE Transactions on Affective Computing, 2025.*