

DCRNet: A DEEP LEARNING APPROACH FOR CLASSIFYING PRETERM INFANT CRY SIGNALS

Syed Hauider Abbas ^{1,*}, Sivasubramanian R ²

¹ Faculty, Department of Computer Science Engineering, Integral University, Lucknow, Uttar Pradesh, India

² Associate Professor, AIML, Malla Reddy University, Hyderabad, Telangana, India

*Corresponding e-mail: abbaspdcse@gmail.com

Abstract – Infant crying is a vital communication tool reflecting physiological and emotional states, especially for preterm newborns. Manual analysis of cry signals is subjective and limited in accuracy, necessitating the development of automated systems. This study introduces DCRNet, a novel Deep Convolutional Recurrent Neural Network designed for classifying preterm infant cry signals into five categories: "eair," "neh," "eh," "heh," and "owh." The cry signals of the subjected preterm infant babies in the data collection step in which a longer acquisition period is required. An integrated feature fusion matrix is used for the categorization phase to separate pathological crying created using the integrated features significant to multi-class frequency features retrieved by the Cepstral Coefficients with Bark-Frequency (BFCC), Mel-Frequency (MFCC), and Linear Prediction (LPCC) characteristics. Depending on these features, the sounds of the preterm infant cry are categorised using the DCR Net. The target cry signal has five distinct groups: "owh" for tiredness, "heh" for discomfort, "eh" for burping, "eair" for cramping, and "neh" for hunger. The effectiveness of the proposed DCRNet was evaluated using F1 score, accuracy, precision, recall, and specificity. The proposed DCRNet model achieved a classification accuracy of 97.27%, outperforming state-of-the-art models, including AlexNet, ResNet-101, and VGG-19 enhance the total accuracy value by 7.92%, 6.17%, 4.83%, and 3.49%, respectively.

Keywords – Convolutional Recurrent Neural Network, infant cry signals, deep learning, Bark-Frequency Cepstral Coefficient.

1. INTRODUCTION

Humans possess a unique ability to formally express their emotions through language, but infants rely on crying as their primary form of communication during their early months [1]. A baby's cry, originating from the central nervous system, serves as an essential indicator of their cognitive and physiological health [2]. This acoustic signal is a rich source of information, providing insights into an infant's needs and overall well-being. [3] Researchers have identified various cry types such as "Neh" for hunger or "Heh" for discomfort helping caregivers and medical experts interpret these signals [4]. However, the subjective nature of manual cry analysis highlights the need for automated

systems to accurately determine the reasons behind infant crying [5].

Recent advancements in ML and DL algorithms, such as Decision Tree (DT), Adaptive Boosting (AB), and Support Vector Machine (SVM), have facilitated the development of automated cry analysis systems [6, 7]. These models analyze cry signals to identify patterns that can reveal an infant's physiological or emotional state. [8] Feature extraction plays a crucial role in this process, focusing on capturing essential features of the cry signal. However, the abundance of features may introduce redundancy and dimensionality issues, making feature selection and dimensionality reduction essential for effective classification [9]. Statistical approaches, including skewness, kurtosis, mean, and standard deviation, aid in summarizing the extracted information into meaningful values [10]. Additionally, feature selection techniques like Kruskal-Wallis's analysis and Pearson Correlation help prioritize features with high discriminant strength while eliminating irrelevant ones, enhancing model performance and reducing computational complexity [11].

For audio-based linguistic feature extraction, techniques STFT are critical [12]. STFT enables time-frequency analysis by segmenting the audio signal into smaller windows, capturing its dynamic nature and frequency variations [13]. However, selecting optimal window size and overlap is essential to balance temporal and spectral resolution. Key acoustic properties like frequency, intensity, and timbre can then be analyzed, providing a comprehensive understanding of the audio signal and supporting accurate infant cry analysis. To overcome this problem, a novel DCRNet model is proposed for classifying the preterm infant cry signals. The major research contributions are mentioned below:

- The main goal of this strategy is to develop a DCRNet method Deep Convolutional Recurrent Neural Network (DCRNet) for accurate classification of preterm infant cry signals.

- Zero-Time Frame Windowing (ZTFW) and Discrete Wavelet Transform (DWT) to enhance cry signal quality by denoising.
- Fused MFCC, BFCC, and LPCC features to create a robust feature matrix for improved classification accuracy.
- Finally, the extracted features are subsequently fed into Convolutional Recurrent neural network that classify eair, eh, neh, heh, owh.
- The effectiveness of the proposed DCRNet was evaluated using F1 score, accuracy, precision, recall, and specificity.

The remainder of the work is organized accordingly. The literature review is summarized in Section 2. Section 3 discusses about the proposed model and Section 4 focuses at the performance of the proposed approach and compares it to other methods. Section 5 provides a final explanation of the conclusion and future scope.

2. LITERATURE SURVEY

In this section, the state-of-the-art in the domains that are pertinent to the work that is being presented, including data, deep learning, machine learning methods, and related studies. Related studies are listed in the following paragraphs, with an emphasis on those that make use of categorize infant baby cries.

In 2019 Kheddache, Y., & Tadj, C. [14] suggests developing a system that uses a probabilistic neural network classifier to distinguish between normal and abnormal crying. The database utilized comprises 3250 cry samples from healthy and pathologic babies, both full-term and preterm. The greatest results were 82% for correctly identifying full-term babies with a certain ailment and 88.71% for accurately determining the preterm newborns' health state.

In 2020 Bashiri, A., & Hosseinkhani, R. [15] uses a genetic algorithm, LPC (Linear Predictive Coding). In this investigation, a database of 2268 infant screams from Baby Chillanto were employed. Equal to the number of features chosen, there was the same number of input neurons. Methods for categorizing audio signals are being developed, and they show promise for use in a variety of contexts, such as analyzing baby cries.

In 2021 Sutanto, E., et al., [16] devised to used computerized audio evaluation and information mining to identify the distinguishing characteristics of premature and fully developed newborn cries. This technique for interpreting infant cry language is presented that uses spectrograms of windowed signals to extract invariant features using desired features. 98.34% of all crying impressions were successfully identified, based on the classification of five categories of crying impression according to different features.

In 2021 Ashwini, K., et al., [17] introduced the main differences in newborn sounds indicating pain, hunger, and fatigue. When kernel techniques based on evaluating linear, polynomial, and radial basis function (RBF) SVM, it is

demonstrated that SVM-RBF offers the best accuracy of any kernel-based newborn cry categorization system, with an accuracy rate of 88.89%.

In 2022 Cabon, S., et al., [18] predicated on the first distinction between silent and audible occurrences, feature extraction from the resulting audio clips, and classification of sobbing and non-crying. Replaying the automatically chosen cries produced an accuracy of 92.2%. Also investigated was the effect of mistakes on the fundamental frequency characterization.

In 2022 Khalilzad, Z., et al., [19] proposed to recognize employing Support Vector Machine (SVM) and Multilayer Perceptron (MLP) machine learning (ML) techniques to treat septic neonates with Neonatal Respiratory Distress Syndrome (RDS). Therefore, the 95.3% accuracy achieved in the classification of two combined diagnostic categories combining sepsis and RDS indicated the encouraging possibility for more research with larger datasets and more pathology groups.

In 2023 Abbaskhah, A., et al., [20] introduced newborn cry data set from Dunstan is pre-processed in this work using the feature vector technique, which includes the features energy (one feature) and MFCC (19 features). The analysis of all three classifiers reveals classifier performs the best. Results indicate that when model complexity rises, performance will also rise. The suggested CNN structure achieves an accuracy of 92.1% under ideal conditions for five classes of newborn screams.

In the literature review, above existing techniques had developed using various DL and ML approaches to infant cry signals. However, existing methods for infant cry signal classification often suffer from limited accuracy due to inadequate feature extraction and preprocessing techniques. Additionally, many models fail to effectively handle the variability and noise inherent in real-world cry signal datasets. In this research, DCRNet method was proposed for classifying the infant baby cries.

3. PROPOSED METHODOLOGY

In order to classify the preterm newborn cry signal, a unique Deep Convolutional Recurrent Neural Network (DCRNet) was developed for classifying preterm newborn cry signals. To ensure accuracy, the cry signals undergo dual pre-processing using ZTFW and DWT filters to prevent erroneous activity analysis. Frequency features, including BFCC, MFCC, and LPCC, are combined with prosodic characteristics to detect pathological crying effectively. These features enable the DCRNet to classify the cry patterns into five groups: "eair," "he," "ne," "e," and "ow," representing different needs or conditions. The overall methodology is illustrated in Figure 1.

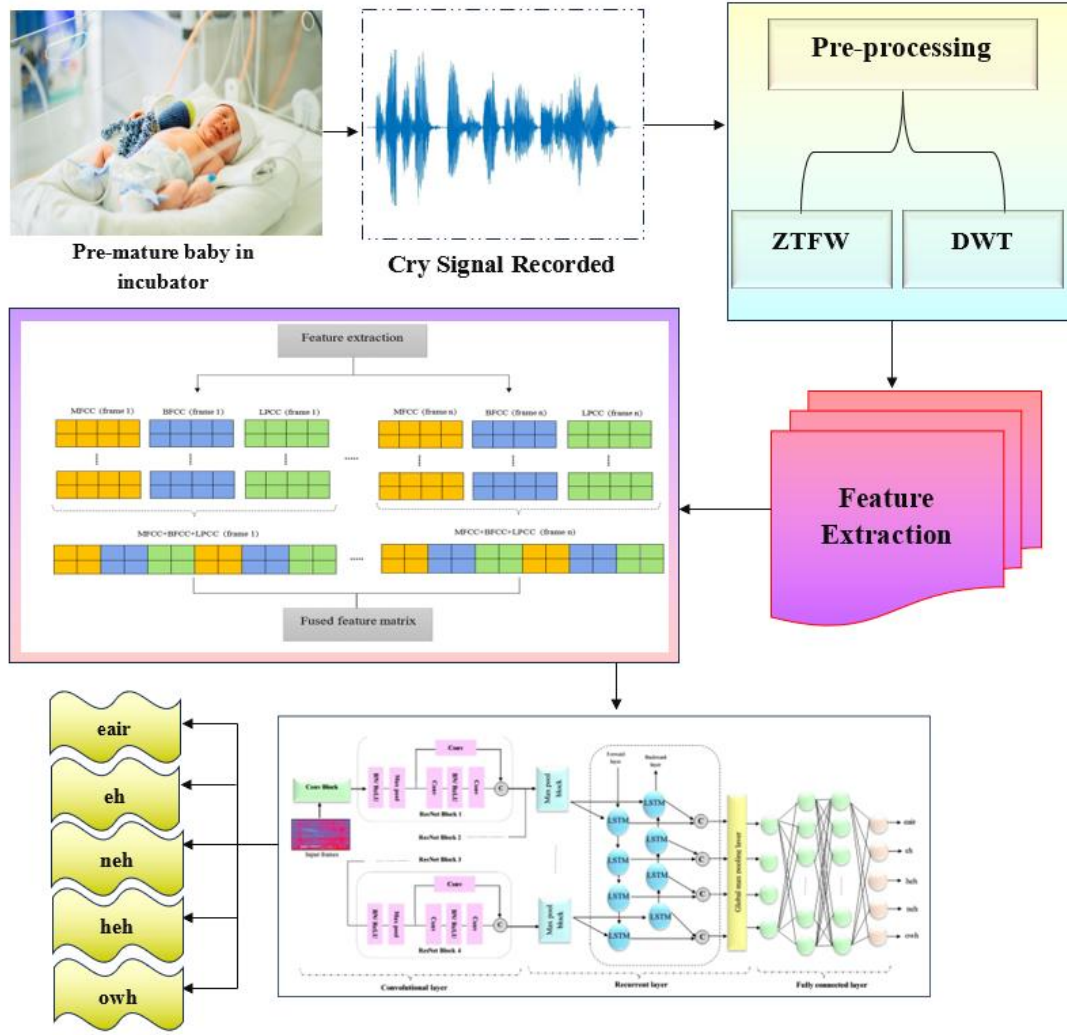


Figure 1. The proposed XXX method

3.1. Dataset Description

A complete description of the infants who were submitted to the study is given in table 3.1. The dataset came from nearby clinics' NICUs and surrounding colonies. Both at home and at the clinic, the baby's activity levels are constantly tracked. Both parents and hospital employees have given their approval for the data collecting. The database contains twenty-seven acoustic signal streams, twelve from different neonates and the remainder from related births. For sixteen preterm male newborns and eleven premature boy infants in this tape, distinct voices may be heard. In the hospital setting, the cry signals are captured for Twenty seconds. Each of the twenty-seven children who were captured as one hundred and thirty-five samples had five distinct sounds.

3.2 Pre-processing

Denoising, this is an essential step in signal processing that converts the input cry signal into a signal that is important for production. The first cry signal is cleaned up of noisy artefacts including ecological noise, speech intrusion, and others. Frame windowing and signal denoising are the first two phases of the pre-processing of the acquired cry signals.

3.2.1 Zero-time frame windowing (ZTFW)

The spectral features of the cry signal are tracked and preprocessed from short sample durations using zero-time frame windowing (ZTFW). This technique lengthens the brief testing period in order to assess an electrode by a window coefficient that mimics the frequency response of a zero-frequency resonator. The calculation for the window operator was given by,

$$\sigma(n) = \begin{cases} 0, & n = 0 \\ \frac{1}{4\sin^2(\frac{n\pi}{2w_l})}, & n = 1, 2, \dots, w_l - 1 \end{cases} \quad (1)$$

In the equation (1), n represents the total inputs, w_l denotes the window length. The actual window value has been set to zero as $\sigma(0)$ in order to reduce the divided by zero errors and possess average spectrum window value of zero with no further modifications of spectral peaks. σ represents the window function that eliminates rapidity of signal prevailing in time domain opposed to remaining window functions.

3.2.2 Discrete wavelet transforms (DWT)

The Discrete Wavelet Transform (DWT) is used in preprocessing infant cry signals by decomposing them into

sub-band wavelet components that represent temporal variations within specific frequency ranges. The DWT parameters A and B, representing discrete dilation and translation values, are often logarithmically discretized to relate these factors. Signal reconstruction is achieved by selecting relevant wavelet coefficients in the upper half-plane, avoiding the need to analyze every coefficient computationally.

$$\phi(n, m) = \frac{1}{A_0^n} \phi\left(\frac{t - B_0 m A_0^n}{A_0^n}\right) \quad (2)$$

The equation (2) is obtained by considering $B=B_0 m A_0^n$ and $A=A_0^n$, accordingly. n, m represent the integers to handle dilation and translation in constant time gap T. The stable dilation phase is represented by A_0 that remains higher than

one. The term B_0 represents the position parameter greater than one.

3.3 Feature extraction

Multi-class feature extraction is a crucial step in removing pertinent characteristics from a single signal. A few multi-class feature extraction methods that can parametrically characterize the cry signal for identification include BFCC, LPCC, and MFCC. The fused feature matrix, as seen in figure 2, combines LPCC, BFCC, and MFCC characteristics. These characteristics stand in for the high-level newborn cry data that is used to categories the various premature infant cry signals.

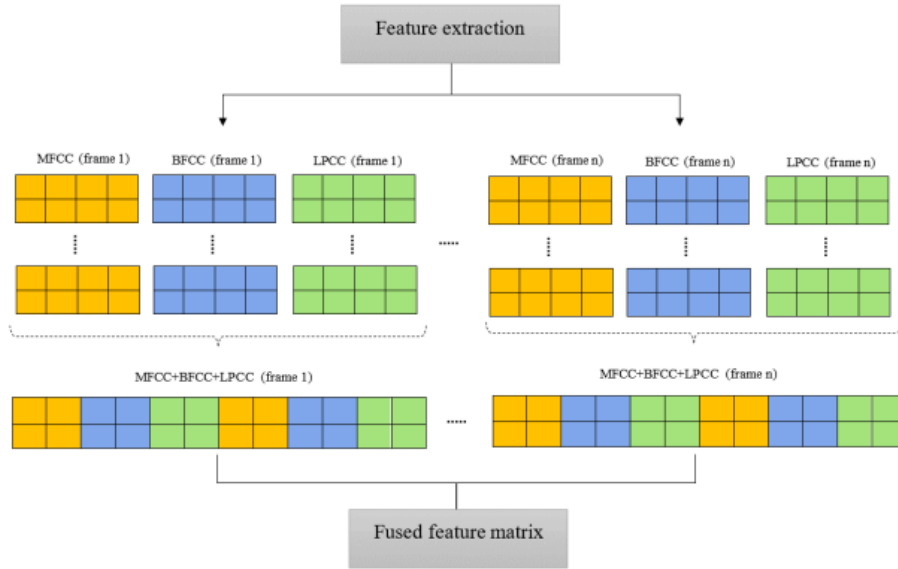


Figure 2. Proposed feature extraction phase

3.3.1 Mel-Frequency cepstral coefficient

MFCC is a commonly employed feature extraction technique. Mel-frequency parameters are used to transform the sound waves into vector impulses. Less than 1000 Hz, the MFCC method functions similarly to human hearing. Signal frames split out by 15–40 milliseconds due to frame obstruction. In Hanning windows, frame blocking removes aliasing problems. The windowing procedure for the baby's cry signal is depicted in equation (3).

$$w_H(t) = \frac{1}{2} (1 - \cos(2f\pi)) \quad (4)$$

In the equation (3) the Hanning window variable is denoted by $w_H(t)$, the total samples of input in a sequential frame are represented by t and frequency in terms of hertz is denoted by f . The evaluations from window have been altered as frequency domain through applying Fourier transform. A filter bank of frequency domain is applied to these signals to obtain Mel frequencies. The derived expression for Mel frequency is given by,

$$f_{mel} = 2595 \log_{10}\left(\frac{f}{700} + 1\right) \quad (4)$$

MFCC makes use of mel-scale bank filters. The resulting broader bandwidth is produced by high-frequency filters. Reducing higher order factors by aggregating these to lower-level factors can drastically minimize the computing cost of cry signals.

3.3.2 Bark Frequency cepstral coefficient

To obtain features, BFCC may also be used to analyze a cry signal. The BFCC uses power spectrum distortion, similar to the MFCC, to alter how loudness is perceived. Two terms separate the BFCC technique from the MFCC approach. Frequency may be changed into bark level using equation (5) below.

$$f_{bark} = 3.5 \arctan\left(\frac{f}{7500}\right)^2 + 13 \arctan(0.00075f) \quad (5)$$

where f is expressed in hertz and f_{bark} stands for the frequency of the bark. Eighteen filters are mapped to calculate bark frequency. The nonlinear connection among observed loudness and signal strength must be taken into consideration is utilized to reduce the cry signal by applying the arctan logarithmic scaling to compress the cry signal, just like in MFCC, and then decorrelated utilizing DCT.

3.3.3 Linear Prediction Cepstral Coefficient

In speech and signal processing, digital voice signals spectral shape is reduced by the use of linear predictive modelling. In signal processing, the LPC approach is the most often employed one. Similarly to LPC, LPCC is used in cry signal processing to create sample points from a speech waveform by utilizing intensity as the vertical axis as well as time as the axis that runs horizontally.

$$f_c = l_c + \sum_{k=1}^{c-1} \frac{k}{c} * l_{c-k} f_c \quad (6)$$

Where f_c represents the value being computed, l_c represents a base or initial value for f_c , Σ Denotes the summation, l_{c-k} indicates a value at a shifted index relative

to l_c and f_c it appears the summation implying recursive or iterative computation. The frequency domain cepstral evaluation is hard to isolate high pitched audio signals from input filters.

3.4 Classification

Recurrent neural networks (RNN) are also good at learning temporal sequences and capturing long-term relationships. In order to conduct categorization characteristics retrieved from the cry signal on a multi-class show, this study integrates CNN and RNN. The RNN is used to gather long-term dependencies, while embedded local properties are retrieved by the CNN.

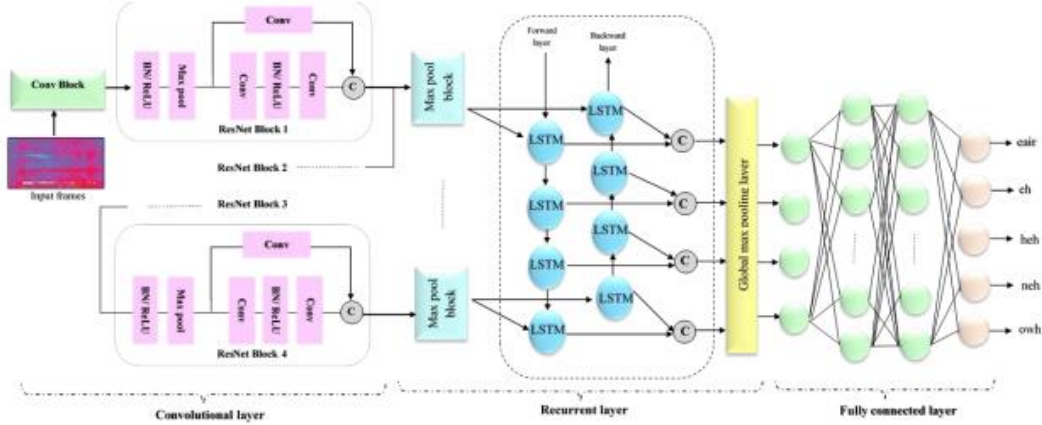
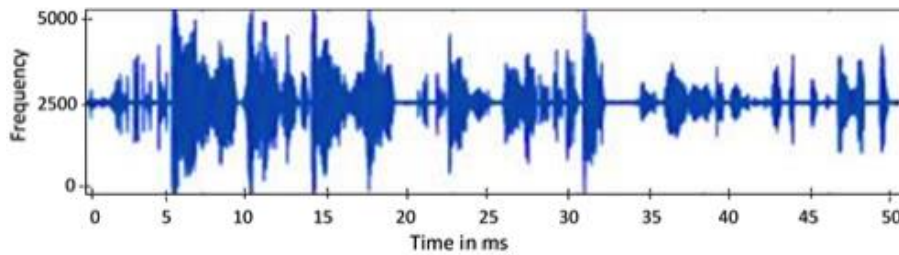


Figure 3. Deep Convolutional Recurrent Neural Network (DCRNet) architecture

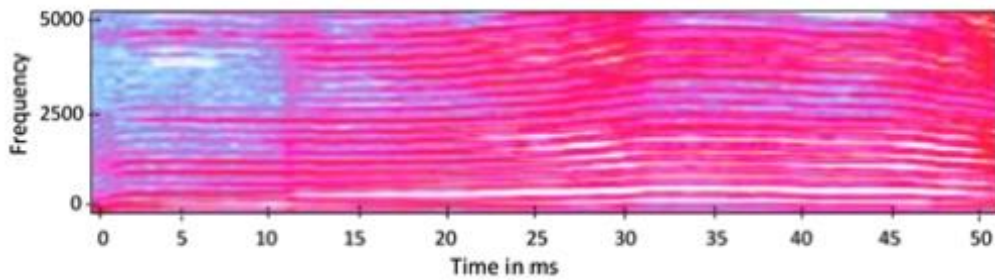
The architecture consists of three layers: ResNet block with Convolutional layer, Bi-LSTM with Recurrent layer, and fully linked layer. The ResNet block combines cry signal detection task characteristics, identify premature infant cry signals. Primary or secondary systems combine convolution block data to produce the final loss function

4. RESULTS AND DISCUSSIONS

This part evaluates the suggested DCRNet efficacy utilizing multi-class features using Matlab-2019b. the cry signals have been trimmed be assembled into suitable frames for further analysis.



(a)



(b)

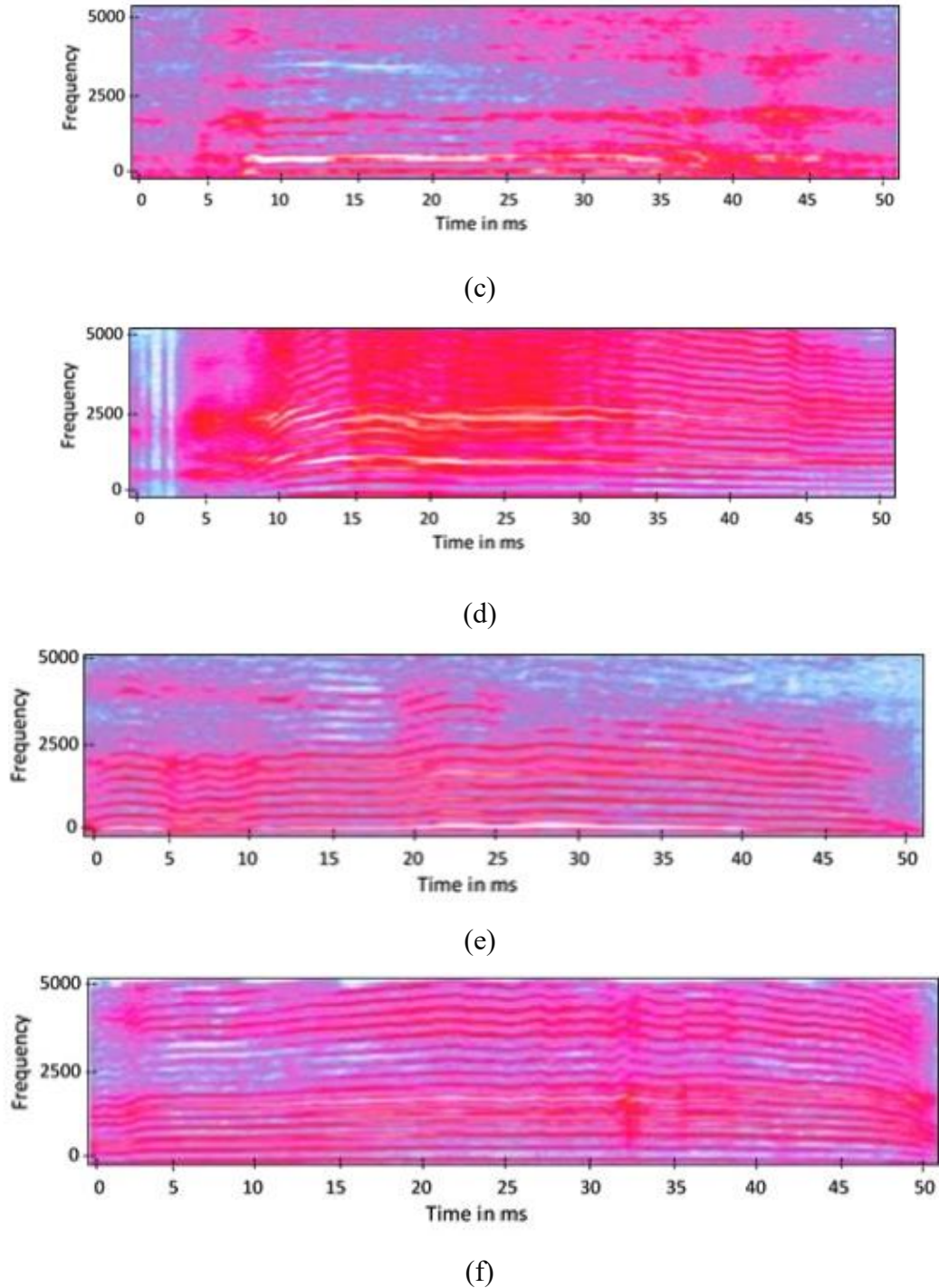


Figure 4. Performance outcomes of spectrogram data (a) input signal (b) eair (c) eh (d) heh (E) neh and (F) owh of pre-term baby cry

Employing strategies for multi-class feature extraction, periodic features are retrieved from the spectrograms that are produced from the original wavelet signals and a DCRNet is used to classify the signal. Accuracy, specificity, precision, F1 score, and recall have all been employed to estimate the analysis of the test samples. A, B, C, D, E, and F.

The Proposed DCRNet multi-class characteristics for categorizing the preterm newborn cry as five classes—"eair," "eh," "heh," "neh," and "owh"—are depicted in Figure 4. Since babies cry more when they are in pain, on the discomfort spectrum, the spectrogram's blue lines are

dense. The little red and blue bars indicate a healthy cry. When a baby is sick or suffocating the spectrogram bands get twisted and the red and blue bands become thicker. This study demonstrates that the suggested method accurately predicts the baby's acoustic cry signal.

4.1 Performance analysis

S Through the use of system measurements including specificity, F1 score, precision, accuracy and recall, the suggested DCRNet's competency was evaluated.

$$Specificity = \frac{t_{neg}}{t_{pos} + f_{pos}} \quad (7)$$

$$Accuracy = \frac{t_{pos} + t_{neg}}{total\ samples} \quad (8)$$

$$Precision = \frac{t_{pos}}{f_{pos} + t_{pos}} \quad (9)$$

$$f_1 - score = 2 \left(\frac{recall \cdot precision}{recall + precision} \right) \quad (10)$$

$$recall = \frac{t_{pos}}{f_{neg} + t_{pos}} \quad (11)$$

Where f_{pos} and f_{neg} specify false positives and false negatives of the input cry signals, and t_{pos} and t_{neg} specify true positives and negatives of the input pictures, respectively. In table 1, the suggested DCRNet suitability for categorizing the various causes of cry signals is shown, and figure 5 shows the classification results visually.

Table 1. Performance Analysis of proposed model

Class es	Precisio n	Specifici ty	Reca ll	Accura cy	F1 scor e
cair	96.24	95.12	96.25	98.41	96.19
neh	97.28	96.24	96.08	95.13	96.52
eh	96.12	94.27	95.05	97.25	95.48
owh	95.22	97.26	95.18	94.32	96.02
heh	97.05	95.15	96.12	94.32	96.02

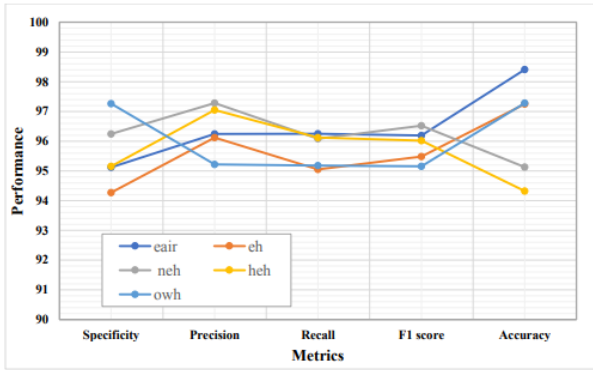


Figure 5. Illustration of proposed DCRNet performance

The table displays the DCRNet classification accuracy for several cry signal kinds, including "cair," "eh," "neh," "heh," and "owh." The precision, recall, specificity, accuracy, and f1 score are used to assess the effectiveness of the DCRNet. The suggested DCRNet achieves 97.27% overall accuracy. With an overall recall of 96.67%, precision of 95.67%, and F1 score of 96.05%, specificity of 95.19%, respectively, the recommended DCRNet performs well.

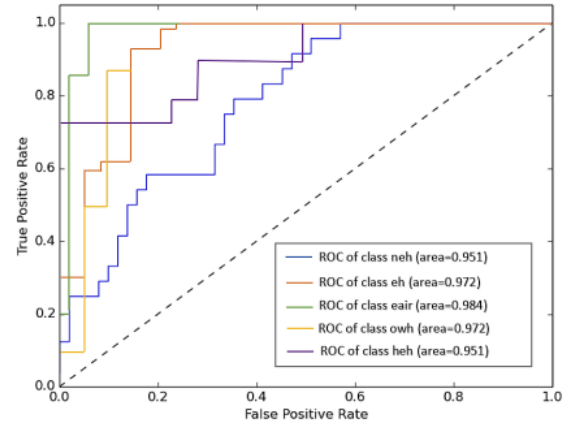


Figure 6. Proposed DCRNet ROC

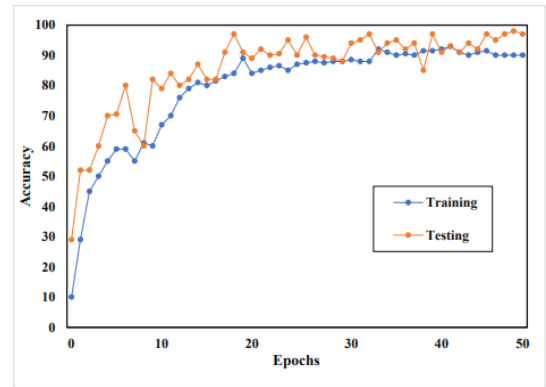


Figure 7. Proposed DCRNet accuracy Curve

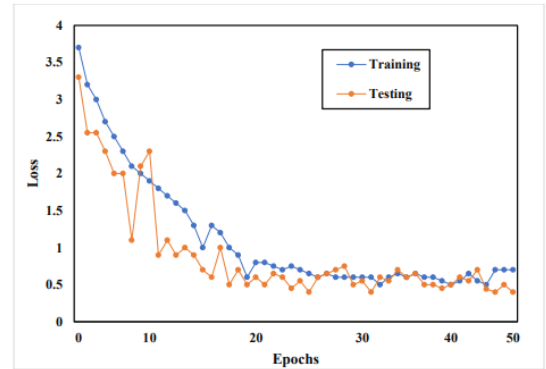


Figure 8. Proposed DCRNet loss curve

The collected database provides higher AUC, and the calculated ROC corresponding to different varieties of premature infant cry has been illustrated in figure 6. For "cair," DCRNet produces AUC values of 0.98, whereas for "eh" and "owh," it generates values of 0.97. For "heh," 0.95 is used, while for "neh," 0.95 may be calculated using variables such as t_{pos} rate on the y-axis and f_{pos} rate on the x-axis.

Figure 7 indicates the accuracy curve, where the y-axis corresponds for accuracy and x-axis corresponds for epochs.

respectively. As the quantity of epochs rises, the DCRNet accuracy gets better. Figure 8 illustrates the epochs and loss that indicate how DCRNet model's loss decreases as the epochs grow. The suggested DCRNet achieves a high level of accuracy for the classification of various cry signal types. The goal of this study was to determine how many training epochs are required to get a high level of testing accuracy. The DCRNet model achieved its classification accuracy of 97.27% at 50 epochs with minimum error rate using the obtained values.

4.2 Comparative analysis

Each Deep Learning network's capacity was evaluated to corroborate the results of the suggested DCRNet. AlexNet, ResNet-101, LeNet, and VGG-19 were the four deep learning classifiers used to evaluate the proposed DCRNet. The suggested DCRNet achieves an accuracy of 97.27%, surpassing the accuracy of the conventional DL networks. A number of parameters, including the DL network's f1 score, recall, accuracy, precision, and specificity, were used to estimate performance.

Table 2. Proposed method comparative assessment with existing models

Class es	Precis ion	Specificity	Reca ll	Accurac y	F1 scor e
LeNet	84.4	85.5	84.3	88.5	85.2
Alex Net	86.7	88.1	85.6	90.2	87.1
ResNet	92.1	93.4	93.5	93.8	93.5
VGG	88.2	89.7	87.7	91.5	88.4
DCR Net	95.6	95.4	96.8	97.2	96.4

By achieving the greatest accuracy limit during the categorization stage of preterm newborn cry, table 3.5 compares several DL techniques based on the particular performance factors. In addition, the suggested DCRNet achieves more accuracy than the conventional networks. By using the acquired dataset, the suggested DCRNet improves the whole accuracy span by 7.92%, 6.17%, 3.49% and 4.83%, better than LeNet, AlexNet, ResNet-101, and VGG-19, respectively

Table 3. Comparative evaluation of proposed and existing models in terms of accuracy

Authors	Model	Accuracy
Ji, C., et al. (2021) [1]	Optimized deep learning model	96.74%
Naithani, G., et al. (2018) [2]	Hidden Markov model	80.7%
Matikolaie, F. S., & Tadj, C. (2022) [4]	MFCC-SVM	86.0%

Vincent, P.D.R., et al. (2021) [13]	SVM-RBF	88.9%
Proposed	DCRNet	97.27%

The time frame of the experiment having evaluation images from the dataset compiled throughout evaluation stage was taken into consideration to assess the efficacy of various strategies. By achieving the necessary classification accuracy, previous methodologies were compared in terms of performance characteristics. The overall accuracy is improved by DCRNet more than SVM-RBF [Vincent, P.D.R., et al. (2021)], MFCC-SVM [Matikolaie, F. S., & Tadj, C. (2022)], Optimized deep learning model [Ji, C., et al. (2021)], and Hidden Markov model [Naithani, G., et al. (2018)] by 8.61%, 11.58%, 0.54%, and 17.03%, respectively.

5. CONCLUSION

In this research, newborn cry signals were analyzed and predicted, and it was discovered that spectrograms are excellent tools for categorizing sounds. The cry signals of the subjected preterm infant babies in the data collection step in which a longer acquisition period is required. An integrated feature fusion matrix is used for the categorization phase to separate pathological crying created using the integrated features significant to multi-class frequency features retrieved by the Cepstral Coefficients with Bark-Frequency (BFCC), Mel-Frequency (MFCC), and Linear Prediction (LPCC) characteristics. Depending on these features, the sounds of the preterm infant cry are categorised using the DCR Net. The target cry signal has five distinct groups: "owh" for tiredness, "heh" for discomfort, "eh" for burping, "eair" for cramping, and "neh" for hunger. The proposed DCRNet model achieved a classification accuracy of 97.27%. In comparison to LeNet, AlexNet, ResNet-101 and VGG-19, the suggested DCR Net enhances its total accuracy value by 7.92%, 6.17%, 4.83%, and 3.49%, respectively. Additionally, the DCR Net outperforms the MFCC-SVM, SVM-RBF, deep learning model optimization and the Hidden Markov model in terms of overall accuracy range by 11.58%, 8.61%, 0.54%, and 17.03%, respectively

CONFLICTS OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

FUNDING STATEMENT

Not applicable.

ACKNOWLEDGEMENTS

The author would like to express his heartfelt gratitude to the supervisor for his guidance and unwavering support during this research for his guidance and support.

REFERENCES

- [1] C. Ji, X. Xiao, S. Basodi, Y. Pan, "Deep learning for asphyxiated infant cry classification based on acoustic features and weighted prosodic features", *In 2019 International Conference on Internet of Things (iThings) and IEEE Green*

- Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) IEEE.* pp. 1233-1240, 2019[[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [2] G. Naithani, J. Kivinummi, T. Virtanen, O. Tammela, M.J. Peltola, J.M. Leppänen, "Automatic segmentation of infant cry signals using Gaussian Mixture Models", *EURASIP J. Audio Speech Music Process.* vol. 1, pp. 1-14, 2018[[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [3] G. Anjali, S. Sanjeev, A. Mounika, G. Suhas, G.P. Reddy, and Y. Kshiraja. "Infant Cry Classification using Transfer Learning", In *TENCON 2022-2022 IEEE Region 10 Conference (TENCON)*, pp. 1-7, 2022[[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [4] F. S. Matikolaie, & C. Tadj, "Machine learning-based cry diagnostic system for identifying septic newborns", *Journal of Voice*, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [5] S. K. Singh, M. Sundararajan, & J. Yadav, "An Automatic Infant Cry Speech Recognition Using Artificial Neural Network", 2023[[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [6] A. Ekinci, & E. Küçükkülahli, "Classification of Baby Cries Using Machine Learning Algorithms", *Eastern Anatolian Journal of Science*, vol. 9, no. 1, pp. 16-26, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [7] J. Xie, "Baby Cry Detection Based on Audio Signals Using Deep Neural Networks (Doctoral dissertation, Master's thesis, Eindhoven University of Technology, Eindhoven, the Netherlands)", 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [8] J. Yan, G. Lu, X. Li, W. Zheng, C. Huang, Z. Cui, & H. Li, "FENP: a database of neonatal facial expression for pain analysis", *IEEE Transactions on Affective Computing*, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [9] S. Wang, J. Du, & Y. Wang, Baby Cry Recognition Based on Acoustic Segment Model. In *National Conference on Man-Machine Speech Communication* (pp. 16-29). Springer, Singapore, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [10] S. Vesangi, & S. R. Regatte, "A Novel Approach to Predict the Reason for Baby Cry using Machine Learning", *International Journal of Computer Science Trends and Technology (IJCTST)*, vol. 10, no. 3, pp. 108-112, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [11] J. Chunyan, M. Chen, L. Bin, & Y. Pan, "Infant cry classification with graph convolutional networks", In *2021 IEEE 6th International Conference on Computer and Communication Systems (ICCCS)*, pp. 322-327, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [12] V. Vaishnavi, P. S. Dhanaselvam, "Premature infant cry signal prediction and classification via dense convolution neural network". *J. Intell. Fuzzy Syst.* pp. 1-14, 2022[[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [13] P.D.R. Vincent, K. Srinivasan, C.Y. Chang, "Deep learning assisted premature infant cry classification via support vector machine models", *Public Health Front.* vol. 9, pp. 670352, 2021[[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- Biomedical signal processing and control, 50, 35-44, 2019. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [15] A. Bashiri, & R. Hosseinkhani, "Infant crying classification by using genetic algorithm and artificial neural network", *Acta Medica Iranica*, pp. 531-539, 2020. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [16] E. Sutanto, F. Fahmi, W. Shalannanda, & A. Aridarma, Cry recognition for infant incubator monitoring system based on internet of things using machine learning, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [17] K. Ashwini, P. D. R. Vincent, K. Srinivasan, & C. Y. Chang, "Deep learning assisted neonatal cry classification via support vector machine models", *Frontiers in Public Health*, vol. 9, 2021. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [18] S. Cabon, B. Met-Montot, F. Porée, O. Rosec, A. Simon, & G. Carrault, "Extraction of Premature Newborns' Spontaneous Cries in the Real Context of Neonatal Intensive Care Units". *Sensors*, vol. 22, no. 5, pp. 1823, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [19] Z. Khalilzad, Y. Kheddache, & C. Tadj, "An entropy-based architecture for detection of sepsis in newborn cry diagnostic systems", *Entropy*, vol. 24, no. 9, pp. 1194, 2022. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]
- [20] A. Abbaskhah, H. Sedighi, & H. Marvi, "Infant cry classification by MFCC feature extraction with MLP and CNN structures". *Biomedical Signal Processing and Control*, vol. 86, pp. 105261, 2023. [[CrossRef](#)] [[Google Scholar](#)] [[Publisher Link](#)]

AUTHORS



Syed Hauider Abbas is a faculty member at Integral University, Lucknow. He has holding a Ph.D. in Computer Science with over 15 years of teaching experience, 90+ research publications, 10+ edited book chapters, 8 authored books, a registered patent, and current postdoctoral research in cancer detection using machine learning, with expertise in AI, image processing, and convolutional neural networks



Sivasubramanian R having 17 years of teaching experience in Engineering Colleges. I completed my M.Tech 2008 and My doctorate 2023. I have published 7 Scopus indexed papers in reputed journals. My areas of interests are Cloud Computing, IoT, Grid Computing, Computer Networks, Data Structures

Arrived: 20.01.2025

Accepted: 26.02.2025