IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 42, No. 2, April, 2025, pp. 1085-1096
Journal homepage: http://iieta.org/journals/ts

MSDP-GAPCNN: MODWT Based Statistical Distribution Patterns and Global Average Pooling Convolutional Neural Network for Improved Infant Cry Classification

Jayasree T., Blessy S.*

Department of Electronics and Communication Engineering, Government College of Engineering, Bodinayakanur 625582, India

Corresponding Author Email: gabiblessy@gmail.com

https://doi.org/10.18280/ts.420240

Received: 24 July 2024
Revised: 16 January 2025
Accepted: 8 April 2025
Available online: 30 April 2025

Keywords:
Wavelet, infant cry signal, deep learning, feature, statistical patterns

ABSTRACT

In this paper, classification of infant cry signals based on Maximal Overlap Discrete Wavelet Transform (MODWT) based statistical distribution patterns and Global Average Pooling Convolutional Neural Networks (MSDP-GAPCNN) is presented. First, the raw audio infant cry signals are transformed into a set of coefficients using MODWT. The statistical features of the signals are derived by finding the statical features such as mean, median, variance, energy, entropy, skewness and kurtosis from the MODWT coefficients at various decomposition levels. Then statistical distribution patterns are obtained from all the statistical features. Each statistical pattern is found to be unique and different. These patterns are fed as input to the Global Average Pooling Convolutional Neural Network (GAPCNN) for classifying the infant cry signal into different types. The performance of the proposed methodology is estimated using donate-a-cry corpus and Neo-cry datasets. The experimental results obtained are compared with other state-of-art methods. The comparison results reveal that, the MSDP-GAPCNN using statistical distribution patterns outperformed better and produced improved performances in classifying infant cry signals compared to the other methods.

1. INTRODUCTION

Infants communicate by means of crying and hence well experienced training is required for distinguishing infant cry types. Classification and interpretation of infant cries is one of the important challenging issues faced by many caregivers, parents, and pediatricians. Each type of cry comprises numerous auditory characteristics, and there is a pattern for each kind of infant cry signal [1]. The main research areas in infant cry involve infant cry signal processing, feature extraction, and classification. Signal preprocessing is vital to extract significant time-frequency domain features from the audio infant cry signals [2]. Mel-frequency cepstral coefficients (MFCC) and spectrograms are used for the analysis and obtaining features from the infant cry signals [3]. Discrete Wavelet Transform (DWT) and Wavelet Packet Transform (WPT) allows to maintain multiresolution signal decomposition in different types of coefficients and preserves key signal information [4]. In all these methods, the raw infant cry signals are converted into a set of coefficients and relevant features are extracted from the coefficients and they are fed to different classifiers, namely Support Vector Machine (SVM) and K-Nearest Neighborhood (KNN) classifiers. But such types of extracted features are usually in one-dimensional form, i.e., a set of fixed values. But it is very difficult to extract such 1D features from the non-stationary time-varying signals [5]. That is, some of the features might be lost while applying feature extraction algorithms. More accurate information about the time and frequency of the signal can be attained if the extracted features are in two-dimensional form, i.e., in the form of images. Moreover, the SVM, K-NN, and other Artificial Neural Network classifiers require 1D features as the input. Also, it requires a large data size, which leads to overfitting problems in the network models.

The research objectives of the proposed model are:

•To develop statistical distribution patterns from infant cry signals: It is necessary to extract robust statistical distribution patterns from the time-varying infant cry signals. This is achieved by applying MODWT, which decomposes the cry signals into coefficients at various decomposition levels. From these coefficients, key statistical features such as mean, median, variance, energy, entropy, skewness, and kurtosis are derived.

•To transform 1D statistical features into 2D representation for classification: To address the limitations of traditional 1D feature extraction techniques by transforming the statistical features into 2D patterns (i.e., images) to preserve both time and frequency information that is often lost in one-dimensional representations. This 2D transformation enhances the classification process by capturing richer, more complex data patterns.

•To employ GAPCNN for classification: Another key objective is to utilize GAPCNN for classifying the generated 2D statistical patterns. GAPCNN has been chosen because it is well-suited for handling image-like data and reduces overfitting in infant cry classification.

•To compare the proposed methodology with existing techniques: The final objective is to evaluate the key performance metrics of the proposed MSDP-GAPCNN methodology on benchmark datasets, specifically the donate-a-cry corpus and Neo-cry datasets, to demonstrate the superior classification performance.

Existing methods, such as SVM and KNN, require manual feature extraction, which can result in overfitting, especially when the dataset is limited or noisy [6]. This issue is particularly relevant in the context of infant cry classification, where small dataset sizes are common [7]. As shown in several studies, overfitting occurs when a model learns to memorize the training data, reducing its capability to generalize to new, hidden data [8, 9]. In contrast, the proposed GAPCNN model reduces the risk of overfitting by utilizing global average pooling, which minimizes the amount of parameters and improves generalization, even with smaller datasets.

Pre-trained networks, such as Convolutional Neural Networks (CNNs) that are commonly applied in image classification, may not be efficient when applied to infant cry signal classification due to the domain-specific nature of the data [10, 11]. These networks typically need large amounts of labeled data to fine-tune effectively, and they may not be optimized for the type of audio features in infant cries [12, 13]. As discussed, transferring a pre-trained model to an innovative task without adequate retraining can lead to inefficiencies and suboptimal performance [14, 15]. Our approach, which directly classifies the 2D statistical patterns extracted from infant cry signals using MODWT, avoids the need for pre-trained networks and is optimized specifically for this task, resulting in a more efficient and specialized model.

Research gaps and novel contributions of the proposed model:

Existing methods commonly rely on 1D features such as MFCCs or statistical measures extracted directly from infant cry signals. While these methods provide useful information, they are inadequate for non-stationary, time-varying signals like infant cries. Such 1D representations often fail to capture the fine time-frequency variations of the signal, leading to the potential loss of important discriminative features.

•Challenges in classifying non-stationary signals: Infant cries exhibit non-stationary characteristics with high variability across different cry types and individuals. Traditional feature extraction techniques like DWT or WPT often struggle to preserve critical time and frequency information simultaneously, resulting in suboptimal feature representations for classification.

•Limitations of conventional classifiers: Classifiers such as SVM and KNN are commonly employed in infant cry classification. These methods require manually engineered features, which may not fully exploit the underlying patterns in complex data.

•Overfitting challenges in neural network models: While Artificial Neural Networks have shown promise, traditional models often require large datasets to achieve generalization and avoid overfitting, which poses a challenge for training deep learning models effectively. This gap necessitates a novel approach that can handle datasets while maintaining robust classification performance.

•Absence of a two-dimensional feature representation: Few studies explore transforming infant cry signals into two-dimensional representations, such as images, which can better capture the time-frequency characteristics of the signal. This gap indicates an opportunity to improve classification accuracy by introducing advanced neural network architectures designed for 2D data, such as CNNs.

•Lack of statistical pattern-based analysis in infant cry classification: Existing works have not explored the potential of statistical distribution patterns derived from wavelet coefficients for distinguishing between different cry types. These patterns could provide a richer and more discriminative feature set, enabling more accurate classification.

The novel contributions of the proposed model are:

•Introduced statistical patterns derived from MODWT coefficients, capturing key features such as mean, variance, entropy, and kurtosis, to represent infant cry signals comprehensively.

•Transformed 1D statistical features into 2D distribution patterns, preserving critical time-frequency information for improved classification accuracy.

•Utilized a GAPCNN to classify the 2D patterns, addressing overfitting challenges due to the presence of Global Average Pooling (GAP), which helps in the stabilization of validation accuracy [16] thus reducing the overall computation time of the CNN model. The incorporation of the GAP layer into the base CNN model calculates the average output of each feature map from the preceding layer, preparing the model for final classification. Unlike Max Pooling, the GAP layer has no trainable parameters, which helps reduce data complexity. This inclusion significantly improves the model's generalization ability and enhances overall computational efficiency.

•Demonstrated the effectiveness of the methodology on the donate-a-cry corpus and neo-cry datasets, ensuring robustness across diverse cry types.

•Achieved superior performance compared to traditional methods (e.g., SVM, KNN) and existing deep learning approaches, validating the innovation and effectiveness of the proposed approach.

The innovation in statistical patterns observed in infant cry signals and the necessity of GAPCNN are given as follows:

Unlike traditional feature extraction methods that rely on fixed-value, one-dimensional features like MFCC and spectrograms, this approach extracts statistical distribution patterns from MODWT coefficients. It preserves critical time-frequency characteristics, offering a more comprehensive signal representation. The generated patterns provide a unique two-dimensional visualization of statistical properties like mean, variance, entropy, etc., making them more discriminative and suitable for classification. Traditional classifiers like SVM and KNN require manual feature selection and struggle to handle the complexity and variability of non-stationary infant cry signals effectively. GAPCNN, on the other hand, efficiently processes the 2D statistical patterns and addresses overfitting challenges by reducing the number of parameters through global average pooling, which is a common constraint in infant cry classification tasks.

The proposed method in advancement over existing approaches is due to the following criteria, most existing methods focus on 1D features or require extensive pre-processing steps, which may lead to the loss of important signal information. The proposed methodology directly addresses these gaps by introducing MODWT to create rich statistical patterns and using GAPCNN to analyse them. The combination of these techniques enables a significant improvement in classification accuracy, as demonstrated through our comparative results on benchmark datasets.

2. RELATED WORKS

Several signal processing techniques such as MFCCs, spectrograms [17, 18], DWT, and DWPT are employed for the extraction of important characteristics from the infant cry signals [5]. Different machine learning techniques, such as K-NN, SVM, and CNN, are used for the classification of infant cry signals [6]. Recent studies show that the classification of infant cry signals using deep neural networks is utilized to retrieve spatial features in infant cry signal spectrograms [9]. Dewi et al. [3] employed MFCC and KNN for the classification of infant cry signals. Gujral et al. [12] used CNN with a transfer learning approach to classify raw infant cry signals and attained an accuracy of 79%. Franti et al. [9] employed a spectrogram and CNN for classifying five classes of infant cry signals and achieved an accuracy of 89%. Anders et al. [19] investigated the CNN model and Short Time Fourier Transform-based spectrogram images for recognizing an infant's cry. Ji et al. [20] used MFCCs to detect changes in the baby's cry signal. In recent years, different types of CNN models have been used for the classification of infant cry signals. Cohen et al. [21] and Ting et al. [14] suggested MFCC and CNN to classify different types of infant cry. Lahmiri et al. [15] employed deep feedforward neural networks (DFNN) and Long-short term Memory (LSTM) Networks for the discrimination of infant cries and attained an accuracy of 85%. Ozseven [11] adopted handcrafted features and 1D CNN model for the classification of infant cries. The authors investigated that the 1D CNN provided less performance. Mala and Darandale [22] derived statistical feature vectors from the infant cry signals and used 1D CNN for classification purposes.

The pretrained networks are also used for classifying infant cries, but the authors investigated that less performance is attained for such types of networks because they are not optimized for specific applications, and also these networks require large size data [11]. These problems can be overcome by designing a new network model based on Global Average Pooling CNN (GAPCNN) algorithms, which is proposed in this paper. The input to the proposed CNN model is the statistical distribution patterns, which are derived by applying MODWT to the infant cry signals.

Traditional techniques such as WT and WPT have been widely used for time-frequency analysis due to their ability to capture non-stationary characteristics of signals [12]. However, their fixed resolution and limited adaptability make them less effective in distinguishing subtle acoustic variations present in different cry types. Short-Time Fourier Transform (STFT) and spectrogram-based approaches have also gained prominence for visualizing time-frequency information. While these methods offer improved temporal and frequency localization, their fixed window sizes can result in suboptimal feature extraction for varying temporal patterns. Mel-spectrograms have emerged as a preferred feature representation in acoustic analysis, as they mimic the human auditory system by focusing on perceptual frequency bands. Despite their effectiveness, Mel-spectrograms rely heavily on the capability of downstream models to extract meaningful features [6].

For instance, pre-trained networks like AlexNet and VGG-16 have been explored in cry classification tasks due to their success in general image recognition problems [19]. However, their inherent architectures, designed for visual features, may not efficiently capture the intricate temporal and frequency patterns of infant cries, resulting in suboptimal performance when applied directly [23]. Ozseven [11] employed pre-trained networks for evaluation and attained ~92% classification performance. Mala and Darandale [22] also investigated and attained ~95%. To address these limitations, approaches utilizing wavelet-based methods, such as the MODWT, have demonstrated promise due to their ability to preserve temporal resolution across decomposition levels [24]. By coupling MODWT with advanced deep learning frameworks, such as GAPCNN, as proposed in this study, it is possible to extract unique statistical distribution patterns that better capture the variability in infant cry signals. This combination not only overcomes the limitations of traditional methods but also provides a more robust representation for accurate classification, outperforming state-of-the-art techniques as demonstrated through extensive experiments on datasets like the donate-a-cry corpus and Neo-cry.

3. PROPOSED METHODOLOGY

In the proposed infant cry classification system, first, the audio infant cry signals are applied with MODWT, which produces a set of coefficients called approximations and details. Then, statistical features and statistical distribution patterns are derived from the detailed coefficients and they are given as input to the GAPCNN, which classifies the infant cry signals into different types. The flow diagram of the proposed infant cry classification system is shown in Figure 1.

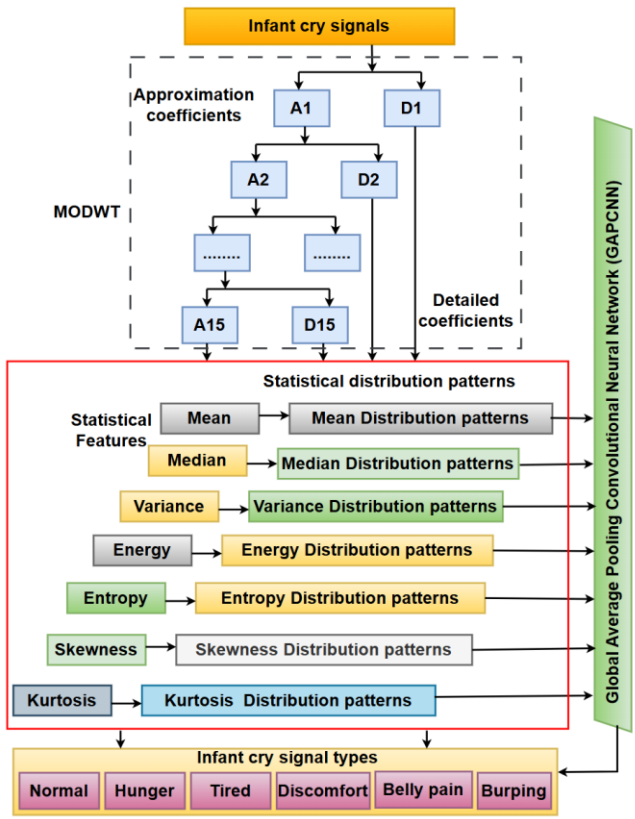

Figure 1. Proposed framework for infant cry signal classification system

3.1 MODWT coefficients

The Maximal Overlap Discrete Wavelet Transform (MODWT) is an undecimated Discrete Wavelet Transform (DWT) that is mainly used for analysing infant cry signals at different scales [25]. The MODWT offers significant advantages over DWT and WPT, particularly in the infant cry signal analysis. Its ability to decompose signals into multiple frequency sub-bands without down-sampling ensures that critical time-domain resolution is preserved, retaining transient features in infant cry signals that might otherwise be lost during decomposition. Infant cry signals often involve large-sized recorded audio, which can be challenging to process directly. The MODWT-based approach effectively addresses this issue by transforming large recorded signals into significantly reduced feature vectors. Unlike DWT and WPT, which require signal lengths to be integral multiples of two, MODWT can handle signals with variable lengths.

Moreover, MODWT facilitates frequency-specific feature extraction by capturing both low- and high-frequency components with greater precision compared to conventional methods like MFCC, DWT, and WPT, which lack such detailed frequency-specific resolution.

Additionally, the wavelet and scaling coefficients in MODWT-based Multi-Resolution Analysis (MRA) are associated with zero-phase filters. This property results in more accurate and efficient wavelet estimators compared to DWT and WPT. Finally, MODWT decomposition ensures that the length of the coefficients remains consistent across all decomposition levels due to the absence of decimation operations. This consistency preserves more information at all levels, enhancing the overall analysis and classification of infant cry signals. Compared to DWT, MODWT is shift-invariant and redundant, making it more robust for time-frequency analysis, especially when analyzing the non-stationary infant cry signals. Moreover, MODWT does not downsample the signal during decomposition, thus preserving the temporal resolution across scales. This characteristic is crucial for identifying subtle features in the cry signals that might otherwise be lost during downsampling. Similarly, while WPT offers full decomposition of both approximation and detail coefficients, it tends to increase computational complexity significantly. MODWT achieves this by providing a redundant representation without the computational burden of WPT [26].

In MODWT, the input signal is decomposed into low and high frequency components, producing a set of coefficients called approximations and details, which are represented mathematically as follows:

$$A_{j,n} = \sum_{l=0}^{L-1} h_{l,n} X_{n-l} mod\ M \tag{1}$$

$$D_{j,n} = \sum_{l=0}^{L-1} g_{l,n} X_{n-l} mod\ M \tag{2}$$

where $A_i, i \rightarrow$ approximation coefficients, $D_i, i \rightarrow$ detailed coefficients, $h_{l,n} \rightarrow$ low pass filter, $g_{l,n} \rightarrow$ high pass filter, $X_{n-l}\ n = 0,1,2,\ M-1$, $M \rightarrow$ signal length, $mod\ M \rightarrow$ circular filtering. If the input signal is decomposed into 15 – levels, the coefficients obtained are given in Eqs. (3)-(6).

$$First\ level\ coefficients, \begin{bmatrix} A_1 \\ D_1 \end{bmatrix} = \begin{bmatrix} A_{(1,n)} \\ D_{(1,n)} \end{bmatrix} \tag{3}$$

$$Second\ level\ cofficients, \begin{bmatrix} A_2 \\ D_2 \end{bmatrix} = \begin{bmatrix} A_{(2,n)} \\ D_{(2,n)} \end{bmatrix} \tag{4}$$

$$Third\ level\ cofficients, \begin{bmatrix} A_2 \\ D_2 \end{bmatrix} = \begin{bmatrix} A_{(2,n)} \\ D_{(2,n)} \end{bmatrix} \tag{5}$$

$$Fifteenth\ level\ coefficients, \begin{bmatrix} A_{15} \\ D_{15} \end{bmatrix} = \begin{bmatrix} A_{(15,n)} \\ D_{(15,n)} \end{bmatrix} \tag{6}$$

3.2 Statistical distribution patterns

The statistical distribution patterns like 1) Mean 2) Median 3) Variance 4) Energy 5) Entropy 6) Skewness 7) Kurtosis distribution patterns are derived from the MODWT detailed coefficients of the infant cry signals using Eqs. (7)-(13). These patterns are obtained by first finding the statistical features for the detailed coefficients, $[D_1, D_2, \dots\dots\dots, D_{15}]$.

$$\text{Mean= } Mean\ [D_1, D_2, \dots\dots\dots, D_{15}] \tag{7}$$

$$\text{Median=} Med[D_1, D_2, \dots\dots\dots, D_{15}] \tag{8}$$

$$\text{Variance=} Var[D_1, D_2, \dots\dots\dots, D_{15}] \tag{9}$$

$$\text{Energy=} Energy[D_1, D_2, \dots\dots\dots, D_{15}] \tag{10}$$

$$\text{Entropy=} En[D_1, D_2, \dots\dots\dots, D_{15}] \tag{11}$$

$$\text{Skewness=} skew[D_1, D_2, \dots\dots\dots, D_{15}] \tag{12}$$

$$\text{Kurtosis=} Kur[D_1, D_2, \dots\dots\dots, D_{15}] \tag{13}$$

The selection of 15 decomposition levels was based on the signal characteristics and extensive experimental analysis. Infant cry signals exhibit complex time-frequency structures that require a sufficient number of levels to decompose both high-frequency transients and low-frequency patterns effectively. MODWT allows for deeper decomposition without losing temporal resolution. Experimental trials were conducted by varying the decomposition levels, and the classification accuracy was analysed for each configuration. The results demonstrated that performance improved progressively with increasing levels, stabilizing at 15 levels. Beyond this point, no significant gains in accuracy were observed, and computational complexity increased. Therefore, 15 levels were chosen as the optimal balance between preserving signal information and ensuring efficient processing. The statistical features were selected for their ability to effectively capture the distribution patterns in infant cry signals was validated through experiments, resulting in improved classification performance.

The significance of the statistical features in enhancing the infant cry classification systems performance are, the mean provides information about the general amplitude of the signal. It is useful for distinguishing between relatively stable and more fluctuating cry types. For example, hunger cries tend to have lower mean values, while distress cries, such as pain, show higher mean values [27]. As a robust measure of central tendency, the median reduces the influence of outliers and is especially effective for noisy or non-stationary signals particularly for cries with irregular characteristics, such as discomfort. Variance measures the spread of the signal, with higher variance indicating more dynamic and unpredictable cries. Lower variance, in contrast, corresponds to more stable cries like hunger, making variance crucial for differentiating

these types. Energy quantifies the total magnitude of the signal, differentiating between intense cries and less intense ones. This feature is pivotal for classifying cries based on urgency and intensity.

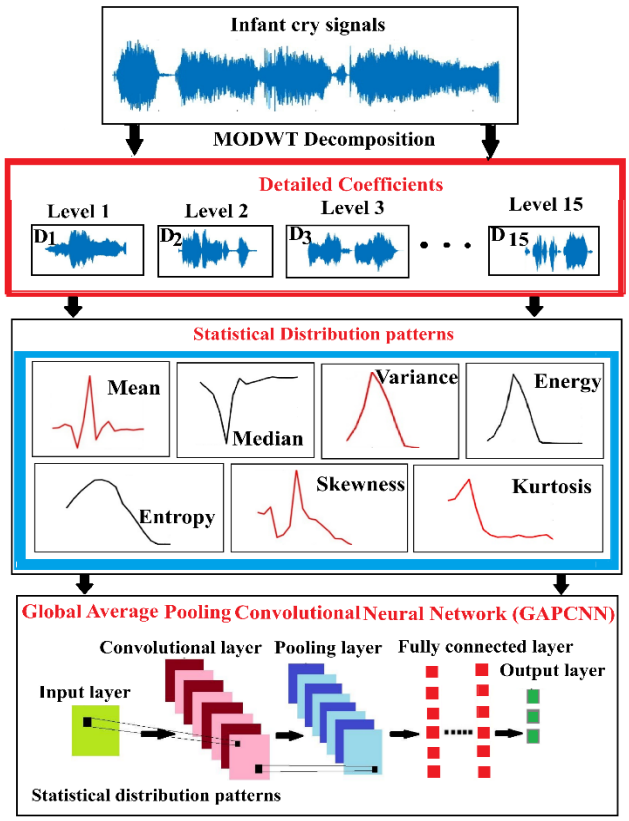

Figure 2. Schematic flow diagram representing MODWT decomposition, pattern generation and classification

Entropy captures the unpredictability of the signal. High entropy corresponds to chaotic, urgent cries, while low entropy reflects more structured, predictable cries. Skewness measures the asymmetry of the cry signal's distribution. Cries such as discomfort or pain exhibit higher skewness due to sudden onset, while hunger cries tend to be more symmetric, aiding in their distinction. Kurtosis reflects the peak of the signal distribution. Higher kurtosis indicates more abrupt and sharp cry patterns, typically seen in pain-related cries, whereas lower kurtosis indicates smoother, less intense cries, such as those associated with hunger. The schematic flow diagram for MODWT decomposition, pattern generation, and classification are shown in Figure 2. All the statistical patterns generated are in the form of images, which are given as input to the GAPCNN for feature extraction/classification.

3.3 GAPCNN architecture

The proposed GAPCNN architecture consists of five convolutional layers (CL1 to CL5). Rectilinear (ReLu) activation layers (RL1 to RL5) and max pooling layers (PL1 to PL5) are placed after each convolution layer [27]. Batch normalization layers (BL1 to BL5) are placed after each max pooling layer [28] as shown in Figure 3. Additionally, the Global Average Pooling layer (GAPL) is placed after the last batch normalization layer. This layer is followed by a fully connected layer (FL1), which is connected in series with the flatten layer (FL). Finally, softmax layer (SL) and classification layers are connected as the terminal layers. Several convolutional layers with the rectified linear unit (ReLU)) activation function in one row creates different models for extracting different features.

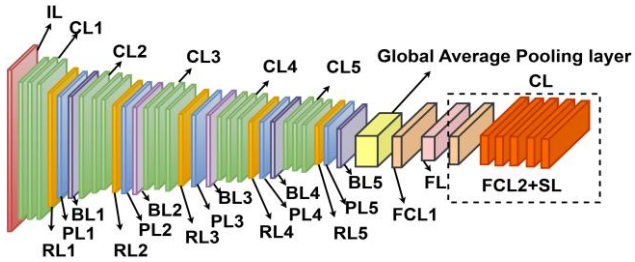

Figure 3. Architecture of GAPCNN

The fully connected layers with ReLU activation function and dropout are located alternatively, which is used for estimating the class of each data using Softmax activation function [29].

The details of each layer, filter size, and activation are given in Table 1. The pseudocode for the feature extraction and classification module is given elaborately.

Table 1. Details of the layers, filter size, activations and learnable parameters [30] used for the design of the proposed GAPCNN

Names	Filter Details	Activations	Learnable Parameters
Input layer (IL)	128 × 128 × 3	128 × 128 × 3	Learnable parameters not used
Convolution layer (CL1)	32 filters, 3× 3	128 × 128 × 32	w, (3 × 3 × 3 × 32), *bias* (1 × 1 × 32)
ReLU activation layer (RL1)	No filter used	64 × 64 × 32	learnable parameters not used
Max pooling layer (PL1)	Max, 2× 2	64 × 64 × 32	64 × 64 × 8
Batch normalization layer (BL1)	No filter used	32 × 32 × 16	offset 1 × 1 × 32, scale 1 × 1 × 32
Convolution layer (CL2)	64 filters, 3× 3	64 × 64 × 64	w, (3×3×3×64), bias 1×1×64
ReLU activation layer (RL2)	No filter used	32 × 32 × 64	Learnable parameters not used
Max pooling layer (PL2)	Max, 2× 2, 2	32 × 32 × 64	Learnable parameters not used
Batch normalization layer (BL2)	No filter used	32 × 32 × 64	offset 1 × 1 × 64, scale 1 × 1 × 64
Convolutional layer (CL3)	128 filters, 3× 3	32 × 32 × 128	w, (3 × 3 × 3 × 128), bias 1 × 1 × 128
ReLU activation layer (RL3)	No filter used	16 × 16 × 128	Learnable parameters not used
Max pooling layer (PL3)	Max, 2× 2, 2	16 × 16 × 128	Learnable parameters not used
Batch normalization layer (BL3)	No filter used	16 × 16 × 128	offset 1 × 1 × 128, scale 1 × 1 × 128
Convolutional layer (CL4)	256 filters, 3× 3	16 × 16 × 256	w, (3 × 3 × 3 × 256), bias 1 × 1 × 256
ReLU activation layer (RL4)	No filter used	8 × 8 × 256	Learnable parameters not used
Max pooling layer (PL4)	Max, 2× 2, 2	8 × 8 × 256	Learnable parameters not used
Batch normalization layer (BL4)	No filter used	8 × 8 × 256	offset 1 × 1 × 128, scale 1 × 1 × 128

Convolutional layer (CL5)	256 filters, 3× 3	8 × 8 × 512	w, (3 × 3 × 3 × 512), bias 1 × 1 × 512
ReLU activation layer (RL5)	No filter used	8 × 8 × 512	Learnable parameters not used
Max pooling layer (PL5)	Max, 2× 2, 2	4 × 4 × 512	Learnable parameters not used
Batch normalization layer (BL5)	No filter used	4 × 4 × 512	offset 1 × 1 × 64, scale 1 × 1 × 64
Global average pooling layer (GAPL)	-	1 × 1 × 10	Learnable parameters not used
Fully connected layer (FCL1)	No filter used	1 × 1 × 32	w, (10 × 2048) *b*ias, 10 × 1
Flatten layer (FL)	-	524288	Learnable parameters not used
Fully connected layer (FCL2)	No filter used	1 × 1 × 5	w, (10 × 2048) *b*ias, 5× 1
Softmax layer (SL)	No filter used	1 × 1 × 5	Learnable parameters not used
Classification layer (CL)	5 outputs	-	Learnable parameters not used

a. Pseudocode for Feature Extraction Module

1. Input: Input the training images of mean, median, variance, energy, entropy, skewness, and kurtosis distribution pattern
   Fix filter size, $(h_w^{j1} * h_h^{j1})$, $(h_{aw}^{j1} * h_{ah}^{j1})$, set, number of filters, M^{j1}, strides, M^{j1} at j^{th} stage of $1^{st} layer$ and max. pool layer
2. Output: Extract features from statistical distribution pattern images of infant cry signals.
3. Training: Train input images at input and feature extraction layer
   For each iteration, j ranges from $(1, R_t)$,
4. At convolution layer, perform convolution operation
   $$Cf^{jn} = [Cf_1^{j1}, Cf_2^{j1}, Cf_3^{j1}, Cf_4^{j1}, \ldots\ldots\ldots Cf_N^{jl}]$$
5. At Max pooling layer
   Max pooling on Cf^{jn},generated feature maps
   $$PM^{j1} = [PM_1^{j1}, PM_2^{j1}, PM_3^{j1}, PM_4^{j1}, \ldots\ldots\ldots PM_N^{j1}]$$
6. Batch Normalization Layer
   Perform batch normalization on feature maps PM_{ma}^{i1} ,
   $$N^{j1} = [N_{1(norm)}^{j1}, N_{2(norm)}^{j1}, \ldots\ldots\ldots N_{N(norm)}^{j1}]$$
7. At ReLU layer, perform normalization on feature maps $BN_{ma(norm)}^{i1}$ using
   $$f(x)^{j^1} = [f(x)_1^{j_1}, f(x)_2^{j_1}, \ldots\ldots\ldots\ldots\ldots f(x)_N^{j_1}]$$
8. Assign $f(z)^{j^1} = I^{(j+1)_1}$, j=j+1
   end for; Return $f(x)^{j^1}$; end

b. Classification module

1. Input: N training samples with $f(x)^{j^1}$ feature maps of on mean, median, variance, energy, entropy, skewness, and kurtosis distribution pattern images
2. Output: classes
3. Train N samples with feature maps $f(x)^{j^1}$
4. In fully connected layer net compute input using $z = W^T.f(x)^{j^1} + B$

4. RESULTS AND DISCUSSIONS

The proposed methodology is experimented using MATLAB 2021b software. The performance is evaluated by calculating metrics such as Accuracy (ACC), Precision (PRE), Recall (REC), and F-Score (FSC) [31].

4.1 Datasets and MODWT coefficients

We have used two datasets such the donate-a-cry corpus dataset (DS1) and the neo-cry dataset (DS2), for the evaluation of the network model. The donate-a-cry corpus dataset (DS1) consists of audio files containing 2500 infant cry signal samples, including normal, discomfort, hunger, tired, belly pain, and burp. The neo-cry dataset (DS2) is our own recorded dataset containing 2500 samples [32]. This real-time infant cry dataset is created by recording audio signals from the infants born in the Government hospital, Tiruchendur, which is in the Southern part of Tamil Nadu state in India. The audio signals are recorded using an ICD-PX470 voice recorder and then stored in the form of wave files. The signals were sampled at the rate of 8000 Hz with the 16-bit sample resolution. Figure 4 shows six different types of infant cry signals [22].

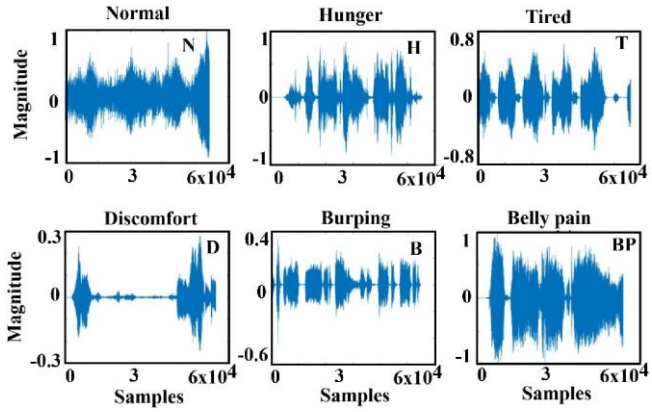

Figure 4. Different types of infant cry signals
Normal (N), Hunger (H), Tired (T), Discomfort (D), Burping and Belly pain (BP)

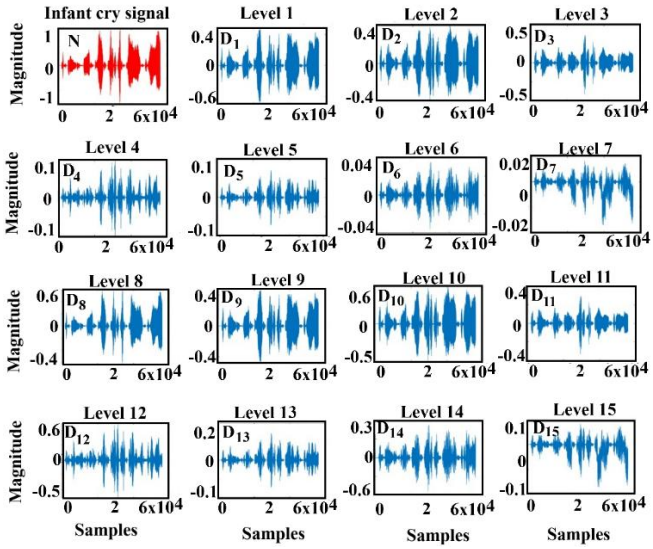

Figure 5. Infant cry signal and its corresponding detailed coefficients

Experiments were conducted in order to obtain the statistical distribution patterns for all the infant cry signals by considering 15 levels of MODWT decomposition. The MODWT-based detailed coefficients for all 15 levels of a particular infant cry signal are shown in Figure 5. Similarly, the results of the distribution patterns showing 15 levels are

shown in Figures 6-13. It is observed that the precise characteristics of the infant cry signals, ranging from lower to higher frequency levels, can be retrieved for better time localization and a wide range of frequency band analysis [33].

4.2 Statistical distribution patterns

The statistical features such as mean, median, variance, energy, entropy, kurtosis, and skewness are computed for all the 15 MODWT detailed coefficients and plotted. This gives different statistical distribution patterns in the form of images. The details of the patterns generated are given in Table 2. The mean values of the MODWT detailed coefficients for all the 15 levels are calculated and then plotted, thus getting mean distribution pattern as shown in Figure 6. All the mean distribution patterns are different, as observed in Figure 6. Figure 7 shows the median distribution patterns for N, H, T, D, B, and BP. According to Figure 7, all the median distribution patterns, i.e., Med-N, Med-H, Med-T, Med-D, Med-B, and Med-BP, are different. The number of peaks in all the patterns is also different. The maximum and minimum values of each distribution pattern are also different [34]. The variance distribution patterns are unique and different and are shown in Figure 8. The maximum value of variance occurred at the 4th level, as illustrated in Figure 8.

Table 2. Details of various statistical distribution patterns and their notations

Infant Cry Signals	Mean	Median	Variance	Energy	Entropy	Skewness	Kurtosis
Normal (N)	Mean-N	Med-N	Var-N	Energy-N	En-N	skew-N	kur-N
Hungry	Mean-H	Med-H	Var-H	Energy-H	En-H	skew-H	kur-H
Tired	Mean-T	Med-T	Var-T	Energy-T	En-T	skew-T	kur-T
Discomfort	Mean-D	Med-D	Var-D	Energy-D	En-D	skew-D	kur-D
Burping	Mean-B	Med-B	Var-B	Energy-B	En-B	Var-B	kur-B
Belly pain	Mean-BP	Med-BP	Var-BP	Energy-P	En-BP	skew-BP	kur-BP

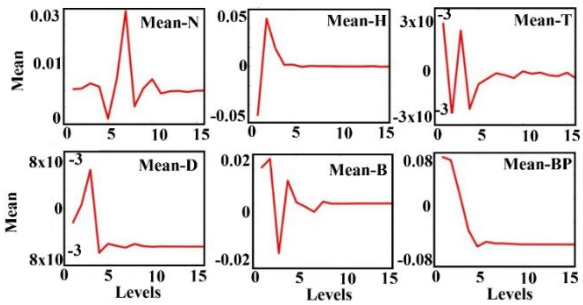

Figure 6. Mean distribution patterns (Mean-N, Mean-H, Mean-T, Mean-D, Mean-B, Mean-BP)

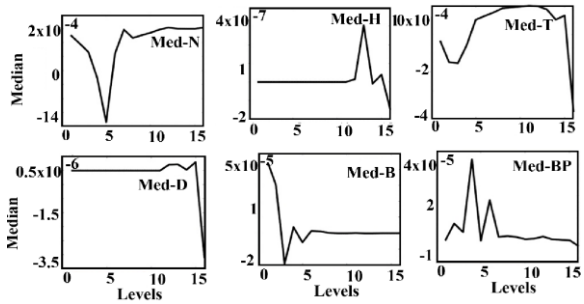

Figure 7. Median distribution patterns (Med-N, Med-H, Med-T, Med-D, Med-B and Med-BP)

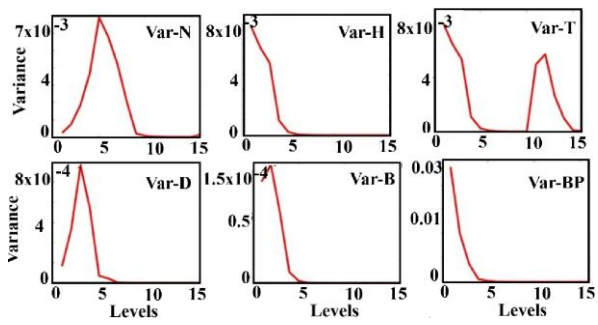

Figure 8. Variance distribution patterns (Var-N, Var-H, Var-T, Var-D, Var-B, Var-BP)

It is also noticed that, after the 5th level, the variance is approximately zero, as illustrated in var-H, var-D, var-B, and var-BP distribution patterns. Figure 9 shows energy distribution patterns (Energy-N, Energy-H, Energy-T, Energy-D, Energy-BP, Energy-B).

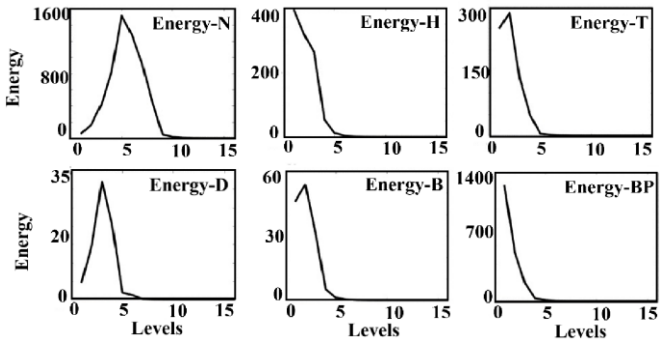

Figure 9. Energy distribution patterns

The energy distribution patterns provide very important information about signal energy and energy concentration of the signals at various levels. As shown in Figure 9, Energy-N provides the highest energy values.

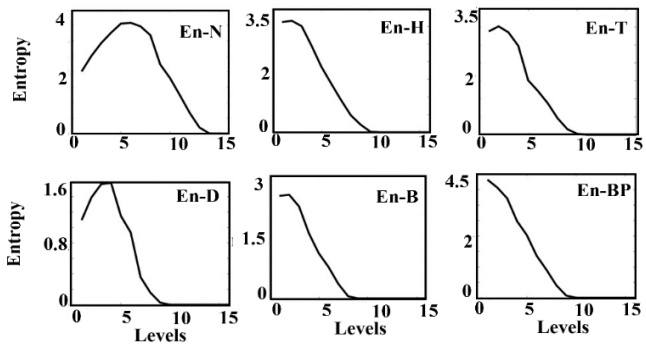

Figure 10. Entropy distribution pattern (En-N, En-H, En-T, En-D, En-B, En-BP)

It is also observed that there exists only one peak in all the

energy distribution patterns, and energy is maximum at the fifth level in most of the cases. As shown in Figure 10, the entropy distribution pattern is different from all other patterns. The Kurtosis distribution pattern for all types of infant cry signals is shown in Figure 11.

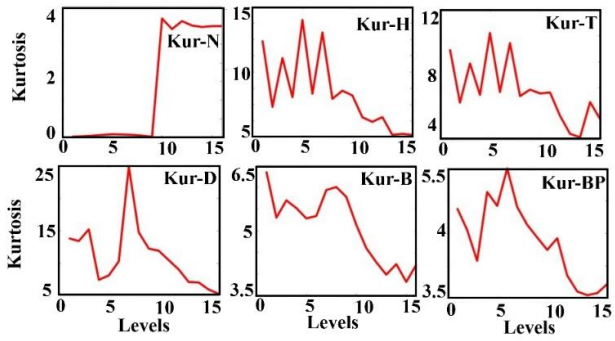

Figure 11. Kurtosis distribution patterns (Kurtosis-N, Kurtosis -H, Kurtosis -T, Kurtosis -D, Kurtosis -B, Kurtosis - BP)

It is observed that the kurtosis distribution pattern is different for each type of infant cry signal. According to Figure 11, the maximum kurtosis value is 25, i.e., for the discomfort infant cry signal (D), and the minimum kurtosis value is 4, i.e., for the normal cry signal (N). It is also inferred that the kurtosis value is maximum approximately at the 10th level.

The skewness distribution pattern shown in Figure 12 also varies for different infant cries. It is inferred that the number of peaks for each signal differs. The maximum peak value also differs in magnitude. The performance of the proposed methodology is evaluated for both DS1 and DS2 datasets. The network is trained and tested with 5-fold cross-validation. The performance measures such as Precision (PRE) [35], Recall (REC) [36], Accuracy (ACC) [37], Specificity (SP) [38] and F-Score (FS) [39] for all the distribution patterns (Mean, Median, Variance, Energy, Entropy) and the results are tabulated in Table 3. According to Table 3, the proposed GAPCNN produced high PRE, REC, ACC, SP, and FS for all the distribution types, i.e., in the range of 97 % to 99% for both the datasets. It is illustrated that, for all the distribution methods, the classification accuracy (ACC) obtained is more than 98%, and the highest accuracy obtained for DS1 is 99.96, i.e., for the entropy distribution pattern. The highest ACC for DS2 is 99.98%, i.e., for the kurtosis distribution pattern.

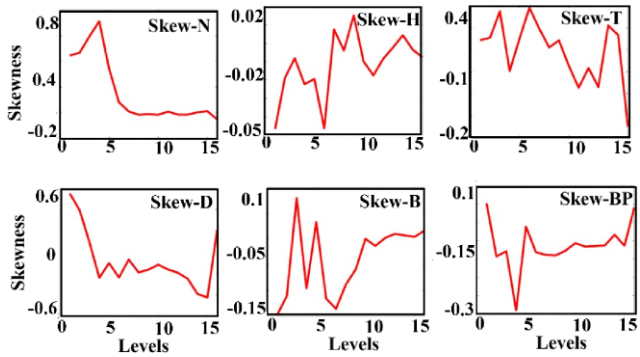

Figure 12. Skewness distribution patterns

Table 3. Performance results of GAPCNN using statistical distribution patterns (Mean, Median, Variance, Energy, Entropy)

Distribution Pattern Type	DS1 (Donat-a-Cry-Corpus Dataset)					DS2 (Neo-Cry Dataset)				
	PRE	REC	ACC	SP	FS	PRE	REC	ACC	SP	FS
Mean	98.15	98.76	98.95	97.20	97.85	98.35	97.40	98.17	98.34	98.11
Median	98.35	98.75	99.75	98.20	97.22	98.85	98.90	99.20	98.02	98.20
Variance	98.00	97.66	98.65	97.03	98.08	98.33	98.13	98.76	97.32	98.32
Energy	99.55	99.45	99.85	99.07	98.59	99.34	98.31	99.22	98.43	98.40
Entropy	99.65	98.54	99.96	99.34	98.21	99.09	98.87	99.11	98.20	98.32
Kurtosis	99.95	98.85	99.97	98.93	99.52	98.20	98.28	99.98	99.11	99.09
Skewness	98.55	98.46	99.22	99.10	99.09	99.70	99.75	99.85	98.90	98.89

Moreover, it is also examined that the highest REC obtained for DS1 is 99.45%, i.e., for the energy distribution pattern. The highest REC for DS2 is 99.75%, i.e., for the skewness distribution pattern. Similarly, the highest SP obtained for DS1 is 99.34%, i.e., for the entropy distribution pattern. The highest for DS2 is 99.11%, which is observed in the kurtosis distribution function. Similarly, the highest FS obtained for DS1 is 99.52%, i.e., for the kurtosis distribution pattern. The highest for DS2 is 99.09%, which is observed in the kurtosis distribution function. The experimental results reveal that, the proposed GAPCNN produced promising results for all the distribution pattern images for both the datasets. The GAPCNN is trained based on the stochastic gradient descent (SGD) method, and 5-fold cross-validation is performed to validate the performance of the model. Training and validation loss are measured after each epoch. The cross-entropy loss is the error function used for finding the network error. It is calculated by finding the sum of the average difference between the actual and the predicted probability distributions for predicting the output classes. Figure 13 shows the training accuracy, testing accuracy, training loss, and testing loss for GAPCNN using seven types of disturbance patterns (GAPCNN + Mean, GAPCNN + Median, GAPCNN + Variance, GAPCNN + Energy, GAPCNN + Entropy, GAPCNN + Skewness, GAPCNN + Kurtosis). Figure 13 shows that the training and testing accuracies of the GAPCNN with all the distribution patterns produced high accuracy rates, above 97% with the maximum of 99.89 %. It is inferred that mean + GAPCNN produced the accuracy of 98.15% and loss of 2% for DS1 and with the accuracy of 98.17% and loss of 1.83 % for DS2. Similarly, median+ GAPCNN produces the accuracy of 98.35% and loss of 1.65% for DS1 and with the accuracy of 99.20% and loss of 0.8% for DS2. It is also observed that, variance+ GAPCNN produced an accuracy of 98.00% and loss of 2% for DS1 and with an accuracy of 98.76% and loss of 1.24% for DS2. It is worth noting that energy+ GAPCNN produces the accuracy of 99.85% and a loss of 0.15% for DS1 and an accuracy of 99.22% and loss of 0.78% for DS2. Notably, entropy+ GAPCNN produce the accuracy of 99.96% and loss of 0.15% for DS1 and with the accuracy of 99.22% and loss of 0.78% for DS2. Thus, it is illustrated that all the statistical patterns produced promising results with minimum loss. The plot also shows that the training process converged well. It is also inferred that the plot

for loss is smooth, which reciprocates the continuous nature of the error between the probability distributions. It is also illustrated that, in all the methods, the highest accuracy and minimum loss are produced at in the 40th epoch. The performance of the proposed GAPCNN with all the distribution pattern images are compared with other networks such as CNN without GAP and other pretrained networks [40] such as AlexNet [11], GoogLeNet [11], EfficientNet [41], VGG-16 [13], MobileNet v2 [41] and the results are tabulated in Table 4. The experimental results shown in Table 4 demonstrate that the GAPCNN with all the statistical distribution patterns produced highest performance measures, i.e., ~98% PRE, ~98% REC, ~98% ACC, ~99% SP and ~98% FS for both the datasets. Table 4 also illustrates that, CNN without using GAP produced ~97% PRE, ~96% REC, ~98% ACC, ~97% SP and ~95% FS for both the datasets. It is also examined that, AlexNet produced ~93% PRE, ~94% REC, ~95% ACC, ~93% SP and ~94% FS for both the datasets. Moreover, it is also inferred that GoogLeNet produced ~92% PRE, ~93% REC, ~94% ACC, ~92% SP and ~90% FS for both the datasets. It is also observed that, EfficientNet produced ~89% PRE, ~98% REC, ~90% ACC, ~89% SP and ~86% FS for both the datasets. Similarly, VGG-16 produced ~85% PRE, ~83% REC, ~86% ACC, ~84% SP and ~82% FS for both the datasets. Likewise, MobileNet produced ~80% PRE, ~81% REC, ~82% ACC, ~80% SP and ~81% FS for both the datasets. Following the evaluation approach in Manit et al. [42], this work presents epoch-wise accuracy and loss plots for the GAPCNN model across different statistical distribution features, as shown in Figure 13, to analyze training behavior and model performance.

A comparative performance table structure, similar to Jayasree et al. [4, 37], has been adapted here to present the evaluation results across various models and statistical distributions; however, the results, datasets, and proposed GAPCNN architecture are original contributions of this study. The results shown in Table 4 reveal that all the statistical distribution patterns with GAPCNN provided better outcomes compared with the pre-trained networks, such as AlexNet, EfficientNet, GoogleNet, VGG-16, MobileNet, SVM, and KNN. This is because the pre-trained CNN needs a large input size, and they are not optimized for tasks. This tends to produce lower accuracy and performance. It is observed that performance improvements of ~more than 12% are achieved using the proposed GAPCNN. Thus, it is concluded that all the distribution patterns with GAPCNN produced the highest performance compared to CNN alone. In the proposed CNN design, the last fully connected layer is replaced by the GAP layer. All the performance results illustrate that the proposed GAPCNN model outperforms others. The GAP helps to reduce the overfitting of the model and improves the performance measures. The experimental results obtained using the proposed methods for the classification of infant cry signals are validated with the Paediatrician in the Government Hospital in Tiruchendur, Thoothukudi District, located in Tamil Nadu in India. A total of 5 paediatricians with diverse clinical experience in infant care participated in the validation of the proposed model. The paediatricians evaluated the model based on four primary criteria: accuracy, precision, recall and F score which assessed the models ability to correctly classify various types of infant cries consistency, which measured the models reliability in classifying similar cry signals across repeated instances; practical relevance, evaluating how well the model's predictions aligned with clinical expectations for cry type differentiation; and ease of use, which focused on how intuitive and user-friendly the model's output was for the paediatricians. The validation involved both quantitative metrics and qualitative feedback. Paediatricians were presented with a set of infant cry signals alongside the models predicted classifications, and they assessed whether the predictions corresponded with their clinical judgment. This valuable input will inform future refinements to the model, ensuring its clinical effectiveness and enhancing its deployment in real-world infant monitoring systems. The proposed MSDP-GAPCNN methodology can be integrated into real-time infant monitoring systems that detect different cry types, enabling caregivers or medical professionals to respond quickly to specific needs. The system could be embedded in smart baby monitors or wearable devices, where the model continuously analyzes the infant's cry signals, providing alerts or detailed feedback on the baby's condition. By incorporating the statistical patterns and GAPCNN classification, the system would offer more accurate identification of cry types, thereby improving infant care and safety. This balances performance and computational efficiency, which reduces the number of parameters compared to conventional CNNs. Hence, it is manageable on devices with moderate processing power, allowing for efficient real-time operation. Given the reduced model size and computational efficiency of GAPCNN, our methodology is suitable for deployment on edge devices, such as smartphones or low-power embedded systems. The model can run on these devices in real time, enabling immediate processing of cry signals without the need for constant connectivity to cloud servers, ensuring both privacy and fast responses.

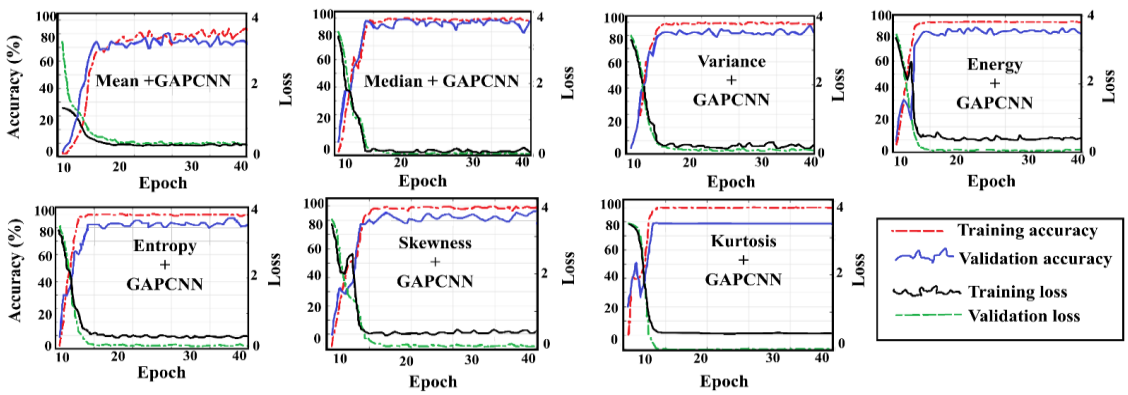

Figure 13. Training accuracy, testing accuracy, training loss and testing loss curves for GAPCNN with all distribution patterns

Table 4. Performance comparison of proposed GAPCNN with other methods

Distribution Patterns and Network Models	DS1 (Donat-a-Cry-Corpus Dataset)					DS2 (Neo-Cry Dataset)				
	PRE	**REC**	**ACC**	**SP**	**FS**	**PRE**	**RE**	**ACC**	**SP**	**FS**
Mean + GAPCNN	**98.15**	**98.76**	**99.75**	**97.20**	**97.85**	**98.35**	**97.40**	**98.17**	**98.34**	**98.11**
Mean + CNN	96.40	97.45	97.10	96.45	95.75	96.56	95.54	97.40	96.11	96.90
Mean + AlexNet	93.56	91.34	95.34	92.30	93.90	94.89	94.22	95.76	93.20	93.11
Mean +GoogleNet	90.90	90.29	92.10	91.99	91.97	92.00	92.23	92.45	91.99	91.90
Mean+EfficientNet	88.32	87.89	89.20	88.20	87.17	87.20	86.44	89.43	87.40	86.11
Mean+VGG-16	85.90	84.28	86.34	84.70	85.70	87.98	84.60	87.70	85.79	85.70
Mean+MobileNet v2	85.60	84.29	85.20	84.90	84.40	85.20	84.00	86.43	86.09	85.67
Mean+SVM	83.20	82.67	83.10	83.00	82.15	82.90	81.99	82.20	81.21	82.03
Mean+KNN	81.79	81.00	81.07	80.67	80.45	82.90	81.67	80.34	81.27	80.13
Median + GAPCNN	**98.35**	**98.75**	**99.75**	**98.20**	**97.22**	**98.85**	**98.90**	**99.20**	**98.02**	**98.20**
Median + CNN	93.13	92.21	93.78	92.55	91.20	94.56	92.11	96.49	93.67	94.11
Median + AlexNet	90.89	90.20	91.45	90.45	90.34	90.12	90.32	92.90	91.09	91.18
Median +GoogleNet	88.90	87.90	89.70	88.70	86.90	90.00	89.95	90.90	89.97	87.90
Median+EfficientNet	86.78	85.88	86.78	85.30	84.19	88.89	87.98	88.90	87.89	88.60
Median+VGG-16	84.69	85.66	85.34	84.44	83.20	86.12	85.44	87.30	84.67	83.33
Median+MobileNet v2	83.98	83.89	84.30	83.89	82.10	85.99	85.23	86.20	85.40	85.02
Median+SVM	82.67	81.98	82.10	82.20	81.45	82.90	81.99	81.98	82.00	81.70
Median+KNN	81.19	81.01	80.90	80.11	81.29	81.90	80.23	80.21	80.11	80.55
Variance+ GAPCNN	**98.00**	**97.66**	**98.65**	**97.03**	**98.08**	**98.33**	**98.13**	**98.76**	**97.32**	**98.32**
Variance+ CNN	94.98	94.37	95.99	94.22	93.10	95.91	95.11	96.90	95.01	93.34
Variance+ AlexNet	94.01	94.01	94.12	92.01	90.21	94.59	94.09	95.00	94.20	93.29
Variance +GoogleNet	92.10	92.99	93.01	92.10	91.00	91.99	91.20	92.19	91.03	89.12
Variance+EfficientNet	91.10	90.22	92.45	91.25	91.00	90.21	89.32	90.46	89.34	89.23
Variance+VGG-16	85.21	84.22	86.22	83.20	81.00	85.89	84.89	86.99	83.45	84.20
Variance+MobileNet v2	80.19	81.20	82.10.	80.20	80.02	81.09	81.07	82.99	81.89	80.03
Variance+SVM	81.98	80.80	81.07	81.00	80.32	82.12	80.12	81.22	80.97	80.20
Variance+KNN	81.07	80.09	81.08	81.79	80.09	81.99	80.98	81.01	80.11	80.00
Energy+ GAPCNN	**99.55**	**99.45**	**99.85**	**99.07**	**98.59**	**99.34**	**98.31**	**99.22**	**98.43**	**98.40**
Energy+ CNN	97.01	96.20	97.90	96.29	97.11	94.99	93.40	95.08	92.99	91.90
Energy+ AlexNet	90.99	90.00	92.30	91.45	90.14	93.01	92.30	93.09	91.98	90.86
Energy +GoogleNet	89.89	88.89	90.10	90.00	90.23	89.11	89.90	90.47	87.78	90.32
Energy+EfficientNet	88.78	87.90	89.98	87.89	86.67	86.20	87.79	88.90	86.30	83.56
Energy+ VGG-16	85.56	87.34	86.67	83.56	84.23	85.23	76.60	87.34	84.30	82.30
Energye+MobileNet v2	82.30	81.30	83.80	85.90	82.45	80.23	75.20	81.90	80.12	82.90
Energy+SVM	82.07	81.90	81.00	81.90	81.06	81.95	80.76	80.32	81.29	80.07
Energy+KNN	81.98	81.77	80.65	81.09	80.97	80.98	80.43	80.01	80.00	80.21
Entropy+ GAPCNN	**99.65**	**98.54**	**99.96**	**99.34**	**98.21**	**99.09**	**98.87**	**99.11**	**98.20**	**98.32**
Entropy+ CNN	98.23	97.65	97.56	95.60	96.23	97.54	95.67	98.09	97.50	95.45
Entropy+ AlexNet	95.90	93.56	96.78	93.56	94.56	96.90	95.56	97.44	96.88	93.78
Entropy+GoogleNet	94.56	92.45	95.67	92.00	93.20	94.50	93.00	94.40	95.12	90.34
Entropy+EfficientNet	89.89	90.00	90.50	89.56	90.09	90.43	90.26	90.23	88.09	89.56
Entropy+ VGG-16	87.54	87.09	89.90	87.89	86.89	88.54	84.65	89.89	86.51	82.45
Entropy+MobileNet v2	86.89	85.81	84.56	80.76	85.56	82.90	83.00	84.36	82.67	80.45
Entropy+ SVM	84.45	84.09	84.20	83.90	82.22	83.90	82.98	81.76	82.90	83.09
Entropy+ KNN	82.89	82.19	81.98	81.67	82.09	81.99	80.23	81.80	81.03	81.65
Kurtosis+ GAPCNN	**99.95**	**98.85**	**99.97**	**98.93**	**99.52**	**98.20**	**98.28**	**99.98**	**99.11**	**99.09**
Kurtosis+ CNN	97.09	96.08	97.65	95.66	96.09	92.67	90.04	93.20	91.23	92.33
Kurtosis+ AlexNet v2	93.56	92.33	94.78	93.40	94.11	90.08	88.90	91.33	90.22	90.00
Kurtosis+GoogleNet	90.10	91.90	92.94	92.11	91.00	86.88	85.67	90.00	86.34	84.56
Kurtosis+EfficientNet	85.78	85.89	90.11	89.09	84.56	87.98	87.90	89.90	83.55	83.45
Kurtosis+ VGG-16	86.57	85.89	87.34	87.77	82.20	87.08	86.89	88.45	82.60	85.67
Kurtosis+MobileNet v2	82.80	82.49	84.69	86.90	82.07	84.09	84.34	85.12	84.70	83.33
Kurtosis+SVM	81.90	80.76	81.65	80.67	80.09	82.90	82.23	82.90	81.60	81.80
Kurtosis+KNN	80.95	80.77	80.01	80.00	80.32	81.89	80.98	80.45	81.90	80.98
Skewness+ GAPCNN	**98.55**	**98.46**	**99.22**	**99.10**	**99.09**	**99.70**	**99.75**	**99.85**	**98.90**	**98.89**
Skewness+ CNN	95.89	94.20	97.90	95.67	97.33	97.22	96.33	97.90	95.44	95.22
Skewness + AlexNet	93.56	91.90	95.40	93.56	92.01	93.70	92.45	95.43	92.40	91.09
Skewness EfficientNet	86.90	83.89	87.78	84.51	83.40	84.23	83.12	88.76	83.04	85.45
Skewness + VGG-16	85.09	85.23	85.54	83.37	82.20	83.00	81.34	84.56	82.11	80.44
Skewness +MobileNet v2	80.00	81.98	82.56	80.12	81.88	81.90	80.31	82.00	80.10	81.99
Skewness+SVM	81.90	80.87	80.55	80.12	80.21	82.90	82.07	81.80	81.08	81.01
Skewness+KNN	80.95	80.88	80.54	80.00	80.21	81.99	81.80	81.75	81.00	80.95

5. CONCLUSION

Infant cries convey information about the infant's feelings. This article discusses different types of statistical distribution patterns and the GAPCNN for the classification of infant cry signals. The audio cry signals are converted into mean, median, variance, energy, entropy, skewness, and kurtosis pattern distribution images using MODWT. Different types of distribution pattern images are fed into the GAPCNN for further classification. The performance of the proposed methods is compared with other CNN models, and the experimental results reveal that the statistical base methods produce promising results compared to other methods.

REFERENCES

[1] Gu, G., Shen, X., Xu, P. (2018). A set of DSP system to detect baby crying. In 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, pp. 411-415. https://doi.org/10.1109/IMCEC.2018.8469246.

[2] Wu, K., Zhang, C., Wu, X., Wu, D., Niu, X. (2019). Research on acoustic feature extraction of crying for early screening of children with autism. In 2019 34rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), Jinzhou, China, pp. 290-295. https://doi.org/10.1109/YAC.2019.8787725

[3] Dewi, S.P., Prasasti, A.L., Irawan, B. (2019). The study of baby crying analysis using MFCC and LFCC in different classification methods. In 2019 IEEE International Conference on Signals and Systems (ICSigSys), Bandung, Indonesia, pp. 18-23. https://doi.org/10.1109/ICSIGSYS.2019.8811070

[4] Renisha, G., Jayasree, T. (2019). Cascaded feedforward neural networks for speaker identification using perceptual wavelet based cepstral coefficients. Journal of Intelligent & Fuzzy Systems, 37(1): 1141-1153. https://doi.org/10.3233/JIFS-182599

[5] Rajesh, S., Nalini, N.J. (2021). Combined evidence of MFCC and CRP features using machine learning algorithms for singer identification. International Journal of Pattern Recognition and Artificial Intelligence, 35(1): 2158001. https://doi.org/10.1142/S0218001421580015

[6] Chang, C.Y., Chang, C.W., Kathiravan, S., Lin, C., Chen, S.T. (2017). DAG-SVM based infant cry classification system using sequential forward floating feature selection. Multidimensional Systems and Signal Processing, 28: 961-976. https://doi.org/10.1007/s11045-016-0404-5

[7] Lim, W.J., Muthusamy, H., Vijean, V., Yazid, H., Nadarajaw, T., Yaacob, S. (2018). Dual-tree complex wavelet packet transform and feature selection techniques for infant cry classification. Journal of Telecommunication, Electronic and Computer Engineering, 10(1-16): 75-79.

[8] Khalilzad, Z., Kheddache, Y., Tadj, C. (2022). An entropy-based architecture for detection of sepsis in newborn cry diagnostic systems. Entropy, 24(9): 1194. https://doi.org/10.3390/e24091194

[9] Franti, E., Ispas, I., Dascalu, M. (2018). Testing the universal baby language hypothesis-automatic infant speech recognition with CNNs. In 2018 41st International Conference on Telecommunications and Signal Processing (TSP), Athens, Greece, pp. 1-4. https://doi.org/10.1109/TSP.2018.8441412

[10] Ricossa, D., Baccaglini, E., Di Nardo, E., Parodi, E., Scopigno, R. (2019). On the automatic audio analysis and classification of cry for infant pain assessment. International Journal of Speech Technology, 22: 259-269. https://doi.org/10.1007/s10772-019-09601-0

[11] Ozseven, T. (2023). Infant cry classification by using different deep neural network models and hand-crafted features. Biomedical Signal Processing and Control, 83: 104648. https://doi.org/10.1016/j.bspc.2023.104648

[12] Gujral, A., Feng, K., Mandhyan, G., Snehil, N., Chaspari, T. (2019). Leveraging transfer learning techniques for classifying infant vocalizations. In 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA, pp. 1-4. https://doi.org/10.1109/BHI.2019.8834666

[13] Rezaeian, A., Rezaeian, M., Khatami, S.F., Khorashadizadeh, F., Moghaddam, F.P. (2022). Prediction of mortality of premature neonates using neural network and logistic regression. Journal of Ambient Intelligence and Humanized Computing, 13(3): 1269-1277. https://doi.org/10.1007/s12652-020-02562-2

[14] Ting, H.N., Choo, Y.M., Kamar, A.A. (2022). Classification of asphyxia infant cry using hybrid speech features and deep learning models. Expert Systems with Applications, 208: 118064. https://doi.org/10.1016/j.eswa.2022.118064

[15] Lahmiri, S., Tadj, C., Gargour, C., Bekiros, S. (2022). Deep learning systems for automatic diagnosis of infant cry signals. Chaos, Solitons & Fractals, 154: 111700. https://doi.org/10.1016/j.chaos.2021.111700

[16] Díaz-Pacheco, A., Reyes-García, C.A., Chicatto-Gasperín, V. (2021). Granule-based fuzzy rules to assist in the infant-crying pattern recognition problem. Sādhanā, 46(4): 199. https://doi.org/10.1007/s12046-021-01736-8

[17] Novamizanti, L., Prasasti, A.L., Utama, B.S. (2020). Study of linear discriminant analysis to identify baby cry based on DWT and MFCC. IOP Conference Series: Materials Science and Engineering, 982(1): 012009. https://doi.org/10.1088/1757-899X/982/1/012009

[18] Patnaik, S. (2023). Speech emotion recognition by using complex MFCC and deep sequential model. Multimedia Tools and Applications, 82(8): 11897-11922. https://doi.org/10.1007/s11042-022-13725-y

[19] Anders, F., Hlawitschka, M., Fuchs, M. (2020). Automatic classification of infant vocalization sequences with convolutional neural networks. Speech Communication, 119: 36-45. https://doi.org/10.1016/j.specom.2020.03.003

[20] Ji, C., Xiao, X., Basodi, S., Pan, Y. (2019). Deep learning for asphyxiated infant cry classification based on acoustic features and weighted prosodic features. In 2019 Int Conf on Internet of Things (iThings) and IEEE Green Comp and Comm (GreenCom) and IEEE Cyber, Physical and Social Comp (CPSCom) and IEEE Smart Data (SmartData): Atlanta, GA, USA, pp. 1233-1240. https://doi.org/10.1109/iThings/GreenCom/CPSCom/SmartData.2019.00206

[21] Cohen, R., Ruinskiy, D., Zickfeld, J., IJzerman, H., Lavner, Y. (2020). Baby cry detection: Deep learning and classical approaches. In Development and Analysis

of Deep Learning Architectures, pp. 171-196. https://doi.org/10.1007/978-3-030-31764-5_7

[22] Mala, B.M., Darandale, S.S. (2024). Effective infant cry signal analysis and reasoning using IARO based leaky Bi-LSTM model. Computer Speech & Language, 86: 101621. https://doi.org/10.1016/j.csl.2024.101621

[23] Hariharan, M., Sindhu, R., Vijean, V., Yazid, H., Nadarajaw, T., Yaacob, S., Polat, K. (2018). Improved binary dragonfly optimization algorithm and wavelet packet based non-linear features for infant cry classification. Computer Methods and Programs in Biomedicine, 155: 39-51. https://doi.org/10.1016/j.cmpb.2017.11.021

[24] Kheddache, Y., Tadj, C. (2019). Identification of diseases in newborns using advanced acoustic features of cry signals. Biomedical Signal Processing and Control, 50: 35-44. https://doi.org/10.1016/j.bspc.2019.01.010

[25] Jeyaraman, S., Muthusamy, H., Khairunizam, W., Jeyaraman, S., Nadarajaw, T., Yaacob, S., Nisha, S. (2018). A review: Survey on automatic infant cry analysis and classification. Health and Technology, 8: 391-404. https://doi.org/10.1007/s12553-018-0243-5

[26] Gurumoorthy, S., Muppalaneni, N.B., Kumari, G.S. (2020). EEG signal denoising using haar transform and maximal overlap discrete wavelet transform (MODWT) for the finding of epilepsy. In Epilepsy-Update on Classification, Etiologies, Instrumental Diagnosis and Treatment. https://doi.org/10.5772/intechopen.93180

[27] Dey, S.K., Uddin, K.M.M., Howlader, A., Rahman, M. M., Babu, H.M.H., Biswas, N., Siddiqi, U.R., Mazumder, B. (2025). Analyzing infant cry to detect birth asphyxia using a hybrid CNN and feature extraction approach. Neuroscience Informatics, 5(2): 100193. https://doi.org/10.1016/j.neuri.2025.100193

[28] Islam, M.A., Olm, G. (2024). Deep learning techniques to detect rail indications from ultrasonic data for automated rail monitoring and maintenance. Ultrasonics, 140: 107314. https://doi.org/10.1016/j.ultras.2024.107314

[29] Alaie, H.F., Abou-Abbas, L., Tadj, C. (2016). Cry-based infant pathology classification using GMMs. Speech communication, 77, 28-52. https://doi.org/10.1016/j.specom.2015.12.001

[30] Zhang, Y., Huang, J., Xie, F., Huang, Q., Jiao, H., Cheng, W. (2024). Identification of plant microRNAs using convolutional neural network. Frontiers in Plant Science, 15: 1330854. https://doi.org/10.3389/fpls.2024.1330854

[31] Matikolaie, F.S., Tadj, C. (2024). Machine learning-based cry diagnostic system for identifying septic newborns. Journal of Voice, 38(4): 963-e1. https://doi.org/10.1016/j.jvoice.2021.12.021

[32] Sharma, K., Gupta, C., Gupta, S. (2019). Infant weeping calls decoder using statistical feature extraction and gaussian mixture models. In 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, pp. 1-6. https://doi.org/10.1109/ICCCNT45670.2019.8944527

[33] Vaishnavi, V., Suveetha Dhanaselvam, P. (2022). Neonatal cry signal prediction and classification via dense convolution neural network. Journal of Intelligent & Fuzzy Systems, 42(6): 6103-6116. https://doi.org/10.3233/JIFS-212473

[34] Sailor, H.B., Patil, H.A. (2018, September). Auditory filterbank learning using ConvRBM for infant cry classification. In INTERSPEECH, Hyderabad, India, pp. 706-710. https://doi.org/10.21437/Interspeech.2018-1536

[35] Moharir, M., Sachin, M.U., Nagaraj, R., Samiksha, M., Rao, S. (2017). Identification of asphyxia in newborns using GPU for deep learning. In 2017 2nd International Conference for Convergence in Technology (I2CT), Mumbai, India, pp. 236-239. https://doi.org/10.1109/I2CT.2017.8226127

[36] Jayasree, T., Shia, S.E. (2021). Combined signal processing based techniques and feed forward neural networks for pathological voice detection and classification. Sound Vib, 55: 141-161. https://doi.org/10.32604/sv.2021.011734

[37] Jayasree, T., Devaraj, D., Sukanesh, R. (2010). Power quality disturbance classification using Hilbert transform and RBF networks. Neurocomputing, 73(7-9): 1451-1456. https://doi.org/10.1016/j.neucom.2009.11.008

[38] Zamzmi, G., Kasturi, R., Goldgof, D., Zhi, R., Ashmeade, T., Sun, Y. (2017). A review of automated pain assessment in infants: Features, classification tasks, and databases. IEEE Reviews in Biomedical Engineering, 11: 77-96. https://doi.org/10.1109/RBME.2017.2777907

[39] Zamzmi, G., Pai, C.Y., Goldgof, D., Kasturi, R., Ashmeade, T., Sun, Y. (2019). A comprehensive and context-sensitive neonatal pain assessment using computer vision. IEEE Transactions on Affective Computing, 13(1): 28-45. https://doi.org/10.1109/TAFFC.2019.2926710

[40] Sabitha, R., Poonkodi, P., Kavitha, M.S., Karthik, S. (2023). Premature infant cry classification via deep convolutional recurrent neural network based on multi-class features. Circuits, Systems, and Signal Processing, 42(12): 7529-7548. https://doi.org/10.1007/s00034-023-02457-5

[41] Matikolaie, F.S., Tadj, C. (2020). On the use of long-term features in a newborn cry diagnostic system. Biomedical Signal Processing and Control, 59: 101889. https://doi.org/10.1016/j.bspc.2020.101889

[42] Manit, J., Preuße, L., Schweikard, A., Ernst, F. (2020). Human forehead recognition: A novel biometric modality based on near-infrared laser backscattering feature image using deep transfer learning. IET Biometrics, 9(1): 31-37. https://doi.org/10.1049/iet-bmt.2019.0015