



Archives available at journals.mriindia.com

**International Journal on Advanced Computer Engineering and
Communication Technology**

ISSN: 2278-5140

Volume 14 Issue 01, 2025

Infant Crying Detection and Classification Using Deep Learning

Jyoti Dudhal¹, Shraddha Shinde², Kabir Kharade³

^{1,2}Research student, Computer Science Department, Shivaji University, Kolhapur, Maharashtra

³Assistant Professor, Computer Science Department, Shivaji University, Kolhapur, Maharashtra.
Jyotidudhal68@gmail.com¹, shraddha2shinde@gmail.com², kgk_csd@unishivaji.ac.in³

Peer Review Information	Abstract
<i>Submission: 16 Jan 2025</i> <i>Revision: 13 Feb 2025</i> <i>Acceptance: 12 March 2025</i> Keywords <i>Infant Cry</i> <i>Audio Signals</i> <i>Spectrogram Images</i> <i>MAMBA State Space Model</i> <i>Audio Spectrogram</i> <i>Transformer model</i> <i>Audio Classification</i>	Infant cry are crucial for various reasons. Cry pattern can be indicative of underlying medical issue such as belly-pain, discomfort, burping, hunger, tired and many more. By recognizing and responding different cry patterns caregivers can also faster a stronger bond with their baby. This technology can be particularly beneficial for parent with disability insuring their baby's needs are met even they cannot response immediately. Infants communicate via their cries. This research aims for classify the cry signals with using MAMBA state space model and Audio Spectrogram Transformer model with the help of Mel-Spectrogram images. The primary objective of this to compare the effectiveness model such as MAMBA and AST model. The proposed system utilizes a dataset of labeled audio signals from infants. We use a Donate-a-cry corpus dataset. The image based features through the spectrogram which are extracted using pre-trained model.

INTRODUCTION

In infant crying detection and classification the main task is to identify the reason of that cry such as hunger, discomfort, pain, tired and etc. Infant can't tell us his needs and emotions. Sometimes their cries can indicate a medical issue and sometimes their basic needs. Each new parents and all the related people can't understand infant languages or their cries. To helps the new parent and detect the potential health problem and basic

needs early, we have studied infant crying detection and classification. The process is start with collecting the audio recordings in different types then this audios can converted into spectrogram images which are visual representation of the audios of the cry frequency over the time. By analyzing this images pre-trained model can identify the patterns with their cry and relate with their cause. A limitation of infant cry research is the availability of good

quality data. On the online platforms datasets are available but it contains classes with imbalance number of audios and also some contains many more background voice. So, we modify the original dataset and create a own new dataset with the help of original dataset And further used the modified new dataset.

LITERATURE REVIEW

This research is to categories infant cries and identify their needs by using Convolutional Neural Networks (CNNs). Six CNN architectures were tested using the Donate-a-Cry Corpus and Dunstan Baby Language datasets: VGG-16, VGG-19, LeNet-5, AlexNet, ResNet-50, and ResNet-152. The results showed that AlexNet had the highest accuracy (84.78% and 72.73% on the two datasets). This study contributes to the advancement of infant cry detection apps by assisting parents and medical professionals in understanding and meeting the needs of newborns [1]. In order to track brain health, especially in preterm infants, this study suggests a novel method for analyzing infant screams. To extract discriminate characteristics and categories weeping types, the technique uses deep support vector machines, fractal descriptors, and iterative neighborhood component analysis (iNCA). The suggested approach recognized five different types of weeping perceptions with a high diagnostic accuracy of 98.34%. This method shows promise for enhancing neurologic health monitoring in newborns while overcoming obstacles related to analyzing infant cries, such as uncertainty and classification errors [2]. The primary means of communication for newborns is crying, with different types of crying indicating different requirements. It can be difficult for careers and novice pediatricians to accurately interpret these cries. This work suggests a novel method for classifying baby cries using cutting-edge deep learning algorithms. The suggested model classifies cry types related to hunger, discomfort, pain, fatigue, and the desire to burp with an accuracy of 98.33%. Real-time cry signal capturing using IoT-enabled sensors improves the accuracy and responsiveness of the system. This study highlights the significance of accurate cry classification in understanding and meeting infant needs, with potential implications for improving infant care practices [3]. This study proposes a novel approach for infant cry classification using time-, frequency-, and time-frequency-domain feature representations. The extracted features, including Mel-frequency cepstral coefficients

(MFCCs), zero-crossing rate, and root mean square, are fed into various machine learning classifiers. The optimal MFCC-based random forest classifier achieved an accuracy of 96.39%, outperforming the state-of-the-art scalogram-based shuffleNet classifier. This approach enables accurate interpretation of infant cries, facilitating effective care giving and potentially reducing severe problems caused by misinterpretation [4]. This research proposes a novel EHO-DCGR net for classifying cry signals from premature babies. The proposed approach preprocesses cry signals, extracts features using MFCC, PNCC, BFCC, and LPCC, and utilizes the Elephant Herding optimization algorithm for feature selection. The EHO-DCGR net achieves an impressive 98.45% classification accuracy, outperforming existing methods (MFCC-SVM, DFFNN, SVM-RBF, and SGDM) by 12.64%, 3.18%, 9.71%, and 3.50%, respectively. This study contributes to the development of accurate and reliable premature baby cry classification systems [5]. The emotional, behavioral, and relational development of infants depends on the ability to identify their needs through cry recognition. This study suggests a novel method for classifying five different sorts of demands or emotions that are represented through infant cries: burping, discomfort, hunger, tiredness, and stomachache. It does this by combining Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). The suggested CNN-RNN model outperforms existing techniques with an average classification accuracy of 94.97% on the Dunstan Baby Language dataset. 5-fold cross-validation guarantees reliable and accurate results, highlighting the method's potential for identifying and responding to infant cries [6]. This study investigates the use of machine learning algorithms and audio processing techniques to analyze baby cry signals. The Donate-a-Cry dataset was subjected to classification methods (k-NN, SVM, Random Forest, and MLP) and feature extraction using Mel Frequency Cepstral Coefficients (MFCC). By splitting the data into ten equal pieces, a data augmentation technique greatly increased classification success, raising the k-NN algorithm's performance from 85.78% to 98.88%. This study shows how machine learning and audio processing may be used to accurately classify baby cries [7]. Since weeping is a baby's main form of communication, parents and medical professionals must be able to determine its source. In order to categorize the causes of neonates' crying, this study suggests a novel method that

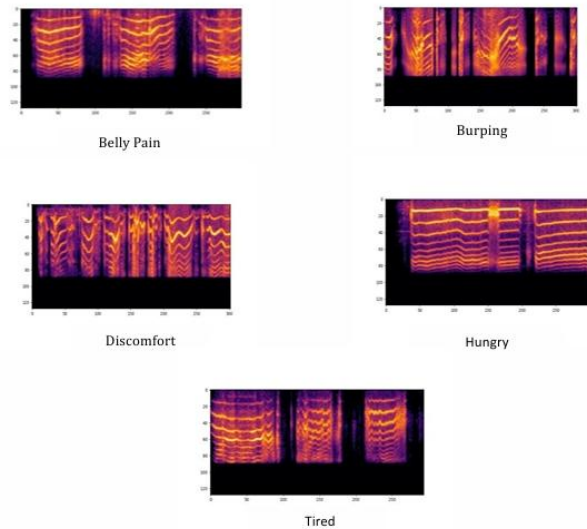
combines genetic algorithms and artificial neural networks with Linear Predictive Coding (LPC) and Mel-Frequency Cepstral Coefficients (MFCC). In comparison to earlier research, the results show how effective the suggested strategy is, attaining the highest categorization accuracy. This method demonstrates how audio signal analysis may be used to create efficient categorization techniques, which have promise uses in a number of fields, including neonatal care [8]. Accurate cry detection is essential for prompt parental action since infants use crying as a means of communicating their needs. Current baby monitoring devices frequently have trouble picking out cries over background noise. By creating and assessing a number of traditional and hybrid machine learning models for infant cry detection, this study overcomes this constraint. In order to attain exceptional performance, a unique stacked classifier called CNN-SCNet is presented, utilizing cutting-edge technologies. The suggested model outperforms all other generated models with an F1-score of 98.39%, precision of 98.72%, and recall of 98.05%. In order to precisely identify baby cries in noisy home settings, this classifier can be incorporated into infant monitoring systems [9]. A baby's primary form of communication is crying, which is used to express wants and emotional states. In order to fill a research gap, this study investigates how acoustic range affects machine learning classification of newborn cries. We used machine learning models (SVM, Random Forest, Naïve Bayes), MFCC feature extraction, cross-validation, and spectrogram approaches on the Donate-a-Cry-Corpus dataset. With 10 MFCC coefficients and a 6-second audio range, Random Forest produced the best accuracy (0.844) and F1 score (0.773), according to the results. This study establishes the groundwork for the creation of methods for identifying premature illnesses based on baby vocalizations, which will allow pediatric professionals to diagnose patients more quickly [10].

METHODOLOGY

Dataset

Download the dataset in from of audio of infant crying from the online platforms such as Kaggle. We used the dataset of Donate-a-cry-corpus dataset. In this dataset the audios of the different types like hungry, tired, burping, belly-pain, discomfort are present. To increase the size of dataset we added the duplicate audios of original dataset. Now the each of the different type contain 100 audios.

Data Preprocessing



Fig(1) Spectrogram images of different types of crying audios

The preprocessing is done by converting the audio into the spectrogram images which is a visual representation of the audio signal frequency over the time. By analyzing these images the pertained model can extract the features and identify the patterns and structures. With the help of spectrogram images we can visualize the complex audio data also. By using the spectrogram images we can easily understand and analyze the cause of cry. For converting the audios into the spectrogram we use the “librosa. feature. mels pectrogram” function which takes the audio signals can convert it into the visual representation of the audios frequency over the time. Also it can uses the special filters to differ the audio signals which are changes over the time. Then our new dataset of spectrogram images are created for the further use. This dataset contains the different classes according to the type of cry.

Split Data

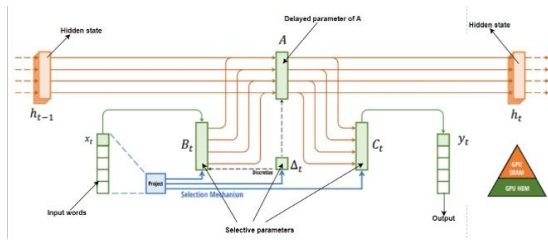
Our dataset contains the 500 images which having the 5 types contains the 100 images per type. We split this dataset in three parts to improve the model generalization and obtain accurate prediction. The train set contain 70% (350 images), test set contain 20% (100 images) and test set contain only 10% (50 images). This is the common ratio for the splitting while the train set helps to train model and learn cry patterns and optimize its parameters. Validation set helps to evaluate and fine tune model. Test set helps to find the final model performance on unseen data.

MODEL USED

A MAMBA State Space Model

A MAMBA model is a special deep learning model which is used for the classifying the spectrogram images. For extracting the low-level features, the MAMBA model uses a sequential structure, starting with convolutional layer (Conv2D) that has 32 filters and ReLU activation function. For mapping the downsample features a max-pooling layer (MaxPooling2D) is used. For avoiding the the overfitting, the model employs a dropout layer with 0.2 rate, and dense layer 128 units with ReLU and activation and flatten layer. At last the output layer generates a probability distribution across the classes. This model is handle the spectrogram images as input which are before preprocessed and normalized before passing to the model. It uses the Adam optimizer. With all this things the MAMBA model is reliable and effective for the classifying the audio in different classes.

Architecture of MAMBA Model:



Fig(2) Architecture of MAMBA Model

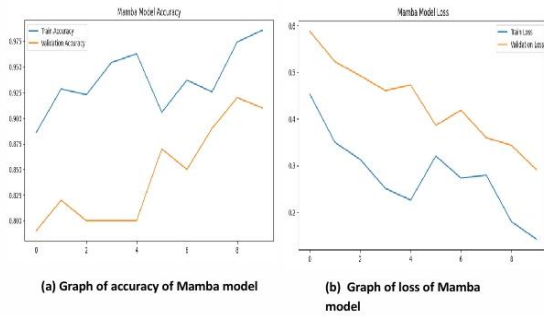


Fig (3) Graphs of loss and accuracy of MAMBA model

Audio Spectrogram Transformer

The Audio Spectrogram Transformer model is a deep learning based convolutional neural network (CNN) designed to classify infant cry audio signal. The model takes the audio then converted in to the spectrogram images and output will be the classified labels such as, hungry, tired, belly-pain, and etc. This model consist of a Conv2D layer, MaxPooling2D layer, Flatten layer, Dense layer and

output layer and the softmax activation. For the compilation of the model the Adam optimizer, categorical cross-entropy, loss function and the accuracy metric are used. The AST model is trained by using the dataset of the spectrogram images of infant cry audio signals with batch size 32, 10 epochs. This model achieves an accuracy of the training accuracy is 0.9181 and the validation accuracy of 0.8400 .

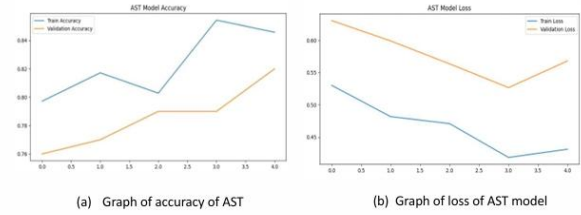


Fig (4) Graphs of accuracy and loss of AST model

Training and Testing Accuracy of Both Model

sr. no	Model Name	Training Accuracy	Testing Accuracy
1	MAMBA model	0.9892	0.9100
2	AST model	0.9181	0.8400

Prediction:

```
audio from: /content/drive/MyDrive/audios/hungry_78.wav
1/1 ————— 0s 45ms/step
Predicted Class: hungry
```

Fig (5) Prediction for hungry audio

```
audio from: /content/drive/MyDrive/audios/discomfort_24.wav
1/1 ————— 0s 34ms/step
Predicted Class: discomfort
```

Fig (6) Prediction for discomfort audio

CONCLUSION

This research “Infant Crying Detection and classification”, successfully examined the application using two deep learning model separately i.e. MAMBA and AST to detect and classify the infant cry sound. Our result shows that both models can effectively extract features from spectrogram which is created by converting the infant cry audios and achieve high accuracy to classify the different types of infant cries.

References

Junaidi, R. F., Faisal, M. R., Farmadi, A., Herteno, R., Nugrahadi, D. T., Ngo, L. D., & Abapihi, B. (2024). Baby Cry Sound Detection: A comparison of MEL spectrogram image on convolutional neural network models. *Journal of Electronics*

Electromedical Engineering and Medical Informatics, 6(4), 355–369. <https://doi.org/10.35882/jeeemi.v6i4.465>

Khosravi, M., &Asaei, A. (2024). Can You Understand Why I Am Crying? A Decision-making System for Infant Cries Based on Improved Deep Support Vector Machine. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 23(1). <https://doi.org/10.1145/3579032>

Younis, S. A., Sobhy, D., & Tawfik, N. S. (2024). Evaluating Convolutional Neural Networks and Vision Transformers for Baby Cry Sound Analysis. *Future Internet*, 16(7), 242. <https://doi.org/10.3390/fi16070242>

Hammoud, M., Getahun, M. N., Baldycheva, A., & Somov, A. (2024). Machine learning-based infant crying interpretation. *Frontiers in Artificial Intelligence*, 7. <https://doi.org/10.3389/frai.2024.1337356>

Vaishnavi, V., Braveen, M., Muthukumaran, N., & Poonkodi, P. (2024). Premature Infant Cry Classification via Elephant Herding Optimized Convolutional Gated Recurrent Neural Network. *Circuits, Systems, and Signal Processing*, 43, 6592–6619. <https://doi.org/10.1007/s00034-024-02764-5>

Maghfira, T., &Basaruddin, T. (2020). *Infant Cry Classification Using CNN–RNN*. Retrieved from https://www.researchgate.net/publication/342069767_Infant_cry_classification_using_CNN_-_RNN

Ekinci, A., &Küçükkülahlı, E. (2023). Classification of Baby Cries Using Machine Learning Algorithms. *Eastern Anatolian Journal of Science*, 9(1), 16–26. <https://dergipark.org.tr/en/pub/eajs/issue/78712/1314296>

A. Bashiri and R. Hosseinkhani, (2020). Infant Crying Classification by Using Genetic Algorithm and Artificial Neural Network. Retrieved from https://www.researchgate.net/publication/347830876_Infant_Crying_Classification_by_Using_Genetic_Algorithm_and_Artificial_Neural_Network

Jahangir, R. (2023). A CNN net-based deep learning framework for infant cry detection in a household setting. *Engineering Reports*, e12786. <https://doi.org/10.1002/eng2.12786>

Riadi, P. A., Faisal, M. R., Kartini, D., Nugroho, R. A., Nugrahadi, D. T., &Magfira, D. B. (2024). A Comparative Study of Machine Learning Methods for Baby Cry Detection Using MFCC Features. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 6(1). <https://doi.org/10.35882/jeeemi.v6i1.350>

Hegde, P. (2024, February 24). Mamba architecture: A Leap Forward in SequenceModeling. *Medium*. <https://medium.com/@puneetthegde22/mamba-architecture-a-leap-forward-in-sequence-modeling-370dfcbfe44a>