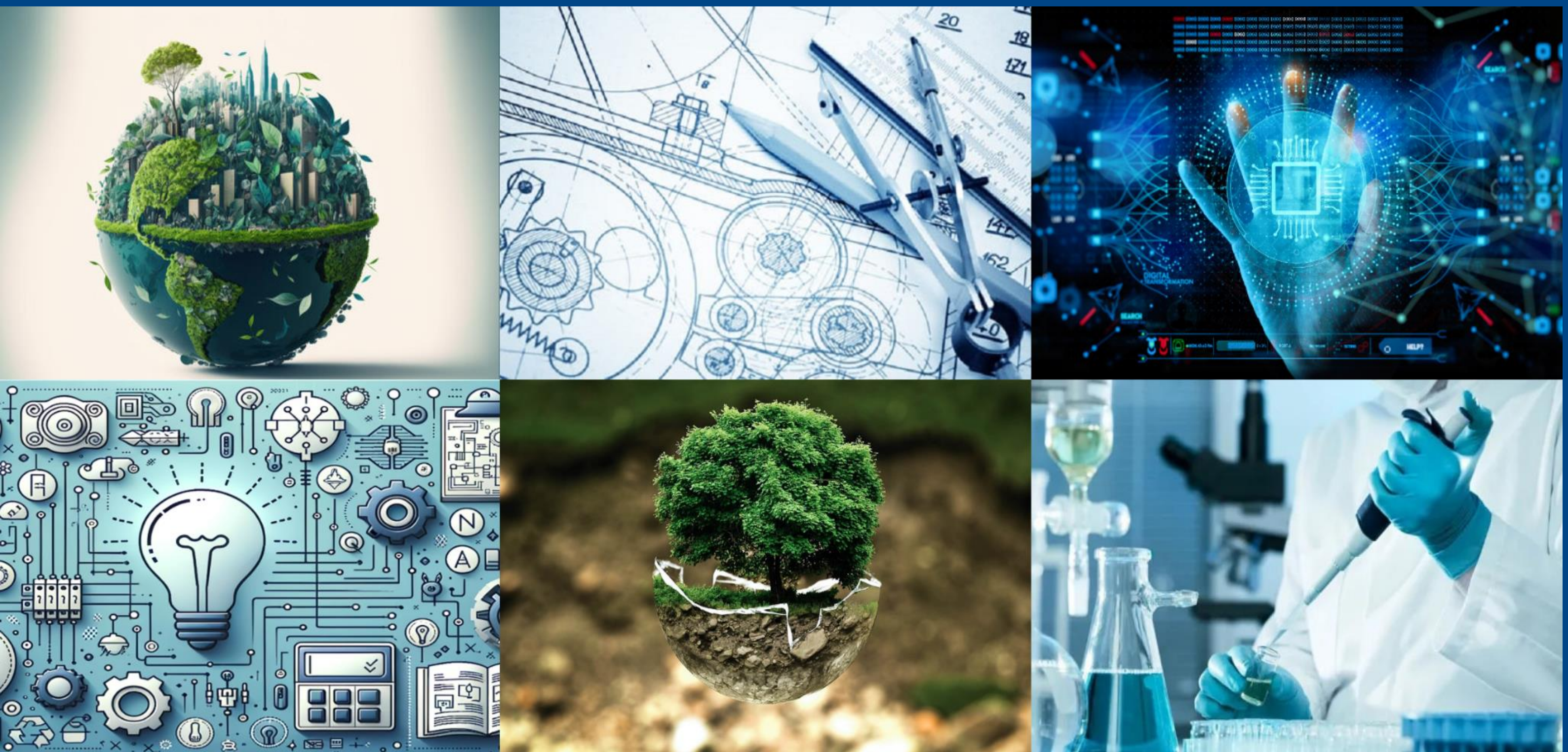
ISSN: 2582-7219

International Journal of Multidisciplinary Research in Science, Engineering and Technology

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.206

Volume 8, Issue 3, March 2025

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Infant Cry Classification through Audio Signal using Machine Learning

Mr. M. Pardha Saradhi[1], P. Yaswanth Kumar[2], P. Jayavardhan Reddy[3], Y. Lakshmi Brahmaiah[4], L. Ravindra Reddy[5]

Associate Professor, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P., India[1]

Undergraduate students, Department of ECE, Vasireddy Venkatadri Institute of Technology, Nambur, Guntur, A.P., India[2-5]

ABSTRACT: Infant crying is a fundamental mode of communication, conveying essential information about an infant's needs, distress, or discomfort. Accurately identifying the reason behind an infant's cry can help caregivers and healthcare professionals provide timely and appropriate care. This study focuses on the development of an intelligent and automated infant cry classification system using machine learning techniques to analyze audio signals. The primary objective is to classify infant cries into distinct categories such as hunger, pain, or discomfort by employing advanced machine learning algorithms, including Cat Boost, Decision Trees, Random Forest, XG Boost, and Multilayer Perceptron (MLP)To ensure robust and reliable classification, the project involves a structured pipeline comprising multiple stages. The first stage includes preprocessing the raw audio signals to remove noise and enhance signal quality, thereby improving the accuracy of subsequent analysis. This is followed by feature engineering, where key acoustic features such as Mel-frequency cepstral coefficients (MFCCs), spectral features, and pitch variations are extracted to create meaningful representations of the cry patterns. These features are then fed into different machine learning models, which are trained and fine-tuned using a diverse dataset of infant cries.The performance of the models is rigorously evaluated using standard metrics such as accuracy, precision, recall, and F1-score to determine their effectiveness in distinguishing between different types of cries. The insights gained from this study have significant real-world applications, particularly in the development of smart baby monitoring systems and healthcare devices. By integrating machine learning with audio signal processing, this research aims to contribute to advancements in infant care technology, providing an automated and intelligent tool to assist caregivers in understanding and responding to an infant's needs more effectively.

KEYWORDS: Infant Health Monitoring, Biomedical Signal Processing, Smart Baby Monitoring System, Real-Time Classification, Edge AI & IoT Integration

I. INTRODUCTION

Infant cries have been a subject of research for a long time due to their potential to provide crucial insights into a baby's physical and mental state. Since newborns cannot verbally communicate their needs, cry analysis serves as a valuable tool for detecting distress, discomfort, or underlying health conditions. Researchers have observed that cry patterns contain significant information that can help in understanding an infant's well-being. By studying these patterns, healthcare professionals can identify early signs of discomfort, illness, or developmental issues, leading to timely intervention and improved infant care.Several approaches have been explored for detecting various infant conditions, including facial recognition-based methods used for drowsiness detection. In such systems, facial features such as eye closure or yawning are analyzed to determine a person's level of fatigue.

However, when it comes to infant monitoring, cry signal analysis is more effective as it directly focuses on the baby's vocal expressions rather than external facial cues. This requires advanced signal processing and machine learning techniques to accurately classify and interpret different cry types.According to the World Health Organization (WHO), nearly 40% of infant deaths occur within the first 30–50 days of life, with 72% of these deaths happening within the first week. Studies suggest that up to two-thirds of these fatalities could be prevented if early warning signs were detected in time. This highlights the importance of developing automated systems that can analyze infant cries to identify potential health risks at an early stage. Such a system could provide caregivers and healthcare professionals with crucial insights into an infant's condition, helping reduce infant mortality rates and improving neonatal care.

One of the key challenges in developing a cry analysis system is filtering out background noise and accurately distinguishing between different cry types. Traditional non-cry detection systems (NCDS) may misinterpret signals if they contain unwanted sounds, leading to inaccurate results. Therefore, an effective cry classification system must incorporate noise reduction techniques, feature extraction, and machine learning algorithms to enhance accuracy. By analyzing the inspiratory and expiratory components of a baby's cry, researchers can extract meaningful information that helps in diagnosing potential health issues.

Extensive research has been conducted to understand the relationship between cry patterns and infant health conditions. Automated segmentation of cry signals into inspiratory and expiratory parts has shown promising results in improving classification accuracy. By developing a robust system that can automatically detect and analyze these cry patterns, healthcare professionals can gain deeper insights into a baby's health and provide timely medical attention when necessary.

II.LITERATURE SURVEY

The classification of infant cries based on audio signals has been a growing area of research due to its potential to provide crucial insights into an infant's health and well-being. Over the years, researchers have explored various techniques, including traditional signal processing methods and modern machine learning approaches, to develop efficient cry classification systems. This section reviews significant contributions in the field, covering different methodologies, datasets, and algorithms used for infant cry analysis.

Early studies on infant cry analysis focused primarily on signal processing techniques to extract relevant features from cry sounds. Researchers used spectral, temporal, and cepstral features, such as Mel-Frequency Cepstral Coefficients (MFCCs), linear predictive coding (LPC), and fundamental frequency, to analyze cry patterns. These features helped differentiate between normal and abnormal cries, enabling early detection of health conditions like respiratory disorders, neurological impairments, and asphyxia. However, conventional methods often required manual feature extraction, making them less efficient and highly dependent on domain expertise.

With advancements in artificial intelligence, researchers began leveraging machine learning techniques for cry classification. Traditional supervised learning algorithms like Support Vector Machines (SVM), Decision Trees, k-Nearest Neighbors (k-NN), and Random Forest were widely used to classify infant cries based on predefined features. These models showed improved performance in distinguishing between different types of cries, such as hunger, pain, and discomfort. However, the effectiveness of these approaches depended heavily on the quality of feature selection, requiring extensive preprocessing and fine-tuning.

In recent years, deep learning methods have gained prominence in infant cry classification, overcoming the limitations of traditional machine learning approaches. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been particularly effective in processing and analyzing audio signals. CNNs extract hierarchical features from spectrogram representations of cry sounds, while LSTMs capture temporal dependencies in cry sequences. Studies have demonstrated that deep learning models outperform conventional classifiers in terms of accuracy and robustness, making them a promising solution for automated infant monitoring systems.

III. PROPOSED METHODOLOGY

The developed method (figure 1) Can handle imbalanced data using weighted loss functions or advanced sampling techniques. Automates feature extraction and selection using techniques like Neural embedding and pretrained audio models. End-to-end pipelines where the algorithm learns feature relevance directly from raw data. Better for large-scale or real time applications. Ideal for systems requiring high accuracy and robustness

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

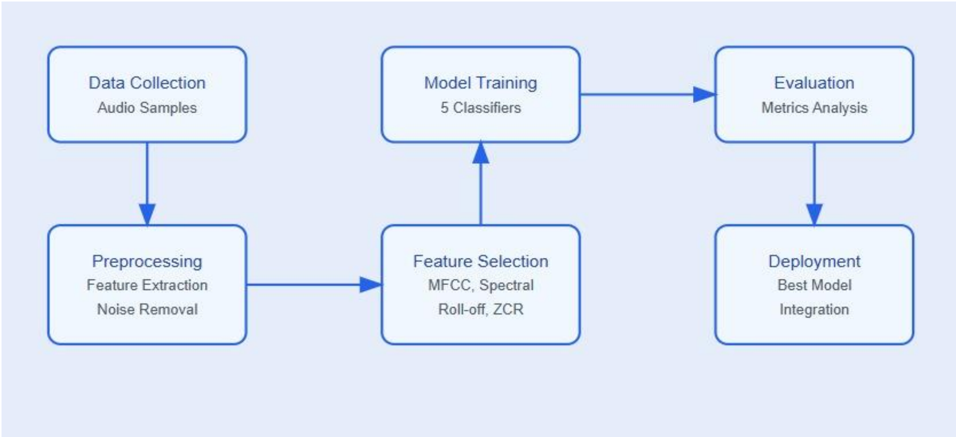


Figure 1: System Architecture

The Dataset

The first step in the proposed methodology involves acquiring a high-quality dataset of infant cries. This dataset should include a diverse range of cry sounds recorded in different environments and under various conditions to ensure robustness. The dataset is typically collected from:

- Publicly available infant cry datasets (e.g., Hungry Database, Pain Database).
- Hospital and neonatal intensive care unit (NICU) recordings for medical research purposes.

Each cry sample is labelled according to its underlying cause (hunger, pain, discomfort, illness) based on expert medical opinions or caregiver observations. The dataset is split into training, validation, and test sets to ensure unbiased evaluation.

Feature Extraction

Feature extraction is a crucial step in infant cry classification, as it transforms raw audio signals into meaningful numerical representations that machine learning models can interpret. Extracted features help in distinguishing different types of cries, such as those caused by hunger, pain, discomfort, or illness. These features are generally categorized into time-domain, frequency-domain, and cepstral features, each providing unique insights into the cry signal.

Time-domain features analyze variations in the amplitude of the cry signal over time. The Zero-Crossing Rate (ZCR) measures how frequently the waveform crosses the zero amplitude axis, with higher values indicating more abrupt changes, which may be linked to distress cries. Energy and Root Mean Square (RMS) Energy capture the intensity of the sound, helping differentiate between normal and high-intensity cries. Short-Term Energy further provides insight into loudness variations over time.

Frequency-domain features examine how energy is distributed across different frequency components. The Spectral Centroid represents the center of mass of the spectrum and indicates the dominant frequency component in the cry signal. Spectral Bandwidth measures the range of frequencies present, helping to distinguish normal from abnormal cries. Spectral Roll-off identifies the frequency below which a significant portion of the total spectral energy is contained, and Pitch and Formants help analyze tonal variations, which are crucial for identifying different emotional states in infant cries.

To enhance model performance, feature selection techniques such as Principal Component Analysis (PCA) and Recursive Feature Elimination (RFE) are used to reduce redundancy and retain the most relevant features. Mutual Information Analysis ensures that only informative features are included in the classification process. Once extracted, these features are structured into feature vectors and fed into machine learning models, such as Decision Trees, Random Forest, XG Boost, Multilayer Perceptron (MLP), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks, for classification.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803293

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Mel recurrence cepstral coefficients (MFCC)

	filename	label	mfcc_1	mfcc_2	mfcc_3	mfcc_4	mfcc_5
0	4-180337-A.ogg	noise	-423.012360	112.444450	25.158617	39.886257	20.326664
1	5-188495-A.ogg	noise	-495.349670	168.771286	87.187874	34.451836	12.474512
2	2-116400-A.ogg	noise	-387.639404	112.270485	12.134430	-11.460219	-15.379561
3	3-216284-A.ogg	noise	-635.321472	-2.435241	-6.372802	1.830521	-0.474074
4	5-156999-D.ogg	noise	-291.775665	166.504593	79.650230	30.511665	20.203709

Figure 2: Mel recurrence cepstral coefficients

MFCC is one in all of the principal famous component extraction strategies utilized in programmed discourse or speaker acknowledgment frameworks utilizing the Mel scale which depends on the human ear scale. it's upheld the nonlinear human impression of the recurrence of sounds. These coefficients address sound upheld discernment. they're gotten from the Mel recurrence cestrum The otherworldly data can after that be changed over to MFCC by going the signs through band pass channels where higher frequencies are misleadingly supported, so applying a backwards Fast Fourier Transform (FFT) consequently. It consolidates the advantages of the cestrum examination with a perceptual recurrence scale upheld basic bands. Thus, the upper frequencies are turning out to be more conspicuous. Since the Mel recurrence cestrum can address an audience's reaction framework obviously, hence MFCC (figure 2) is normally viewed as the least complex accessible estimation of the human ear.

IV. ALGORITHM USED

1. Decision Tree (DT)

Decision Trees are widely used for classification tasks due to their simplicity and interpretability. They work by recursively splitting the dataset based on feature values, forming a tree-like structure where each node represents a decision rule. In infant cry classification, Decision Trees analyze extracted features such as MFCCs, spectral characteristics, and pitch variations to classify cry types. However, they are prone to overfitting, which can be mitigated using ensemble methods like Random Forest.

2. Random Forest (RF)

Random Forest is an ensemble learning technique that improves upon Decision Trees by training multiple trees on different subsets of the dataset and averaging their predictions. This approach enhances model robustness and generalization. In the context of infant cry classification, Random Forest efficiently handles complex feature relationships, providing high accuracy and stability in distinguishing different cry types.

3. XG Boost (Extreme Gradient Boosting)

XG Boost is an optimized gradient boosting algorithm that sequentially builds multiple decision trees to minimize classification errors. It employs regularization techniques to prevent overfitting, making it highly effective for structured data classification tasks. XG Boost is particularly useful in handling high-dimensional features extracted from infant cry signals, improving classification accuracy through feature importance weighting.

4. Cat Boost

Cat Boost is another gradient boosting algorithm specifically designed to handle categorical data efficiently. It reduces the need for extensive preprocessing and is highly effective in handling imbalanced datasets, a common challenge in infant cry classification. By leveraging categorical feature encoding and boosting techniques, Cat Boost enhances model accuracy and reduces computational complexity.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5. Multilayer Perceptron (MLP)

MLP is a type of artificial neural network that consists of multiple layers of neurons, including input, hidden, and output layers. It processes audio features by learning complex non-linear relationships between features and cry categories. MLPs are particularly useful when combined with feature extraction techniques like MFCCs and spectral analysis, providing a deep learning-based approach to classification.

V. RESULT AND DISCUSSION

The results of the infant cry classification system are evaluated based on various machine learning models, analyzing their performance in distinguishing different types of cries, such as hunger, pain, discomfort, and illness. The evaluation includes accuracy, precision, recall, F1-score, and computational efficiency. Additionally, a comparative analysis of different models is presented to determine the most effective approach for real-time implementation**.** The result will observed in web application of its flask deployment (figure 3) the predicted results as shown in figure 4.

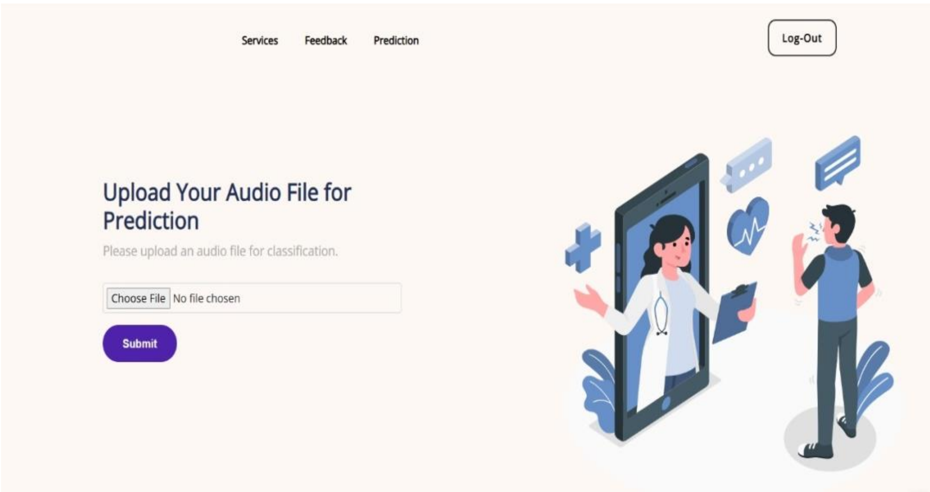


Figure 3: The developed model, Where we will input the audio signal then displayed appropriate Result with required Probability.

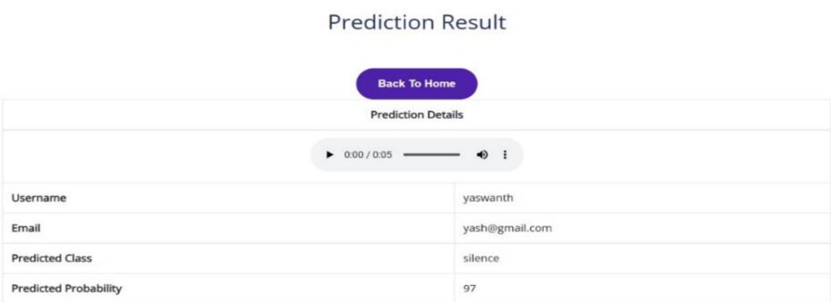


Fig. 4 Silence Predication Class

A confusion matrix is a powerful tool to evaluate the performance of a classification model. It shows how well the model is at predicting different categories by comparing the actual (true) labels with the predicted labels. The rows represent the actual or true labels. The columns represent the predicted labels. Each cell in the matrix represents the number of data points that were actually in a certain category (row) and predicted to be in another category (column).

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803293

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The Random Forest model demonstrated varying degrees of success in classifying infant audio. For the "cry" category, 14 instances were accurately identified, while one instance was mistakenly labeled as "noise." The model performed exceptionally well in classifying "laugh" sounds, correctly identifying all 24 instances. Similarly, the "noise" category saw 28 accurate classifications, with only two instances being mislabeled as "cry." Finally, the model correctly classified all 18 instances of "silence."

Notably, there were no misclassifications between "laugh," "noise," and "silence," indicating a strong ability to distinguish between these categories. However, the model showed some confusion between "cry" and "noise," suggesting a potential area for improvement in future iterations.

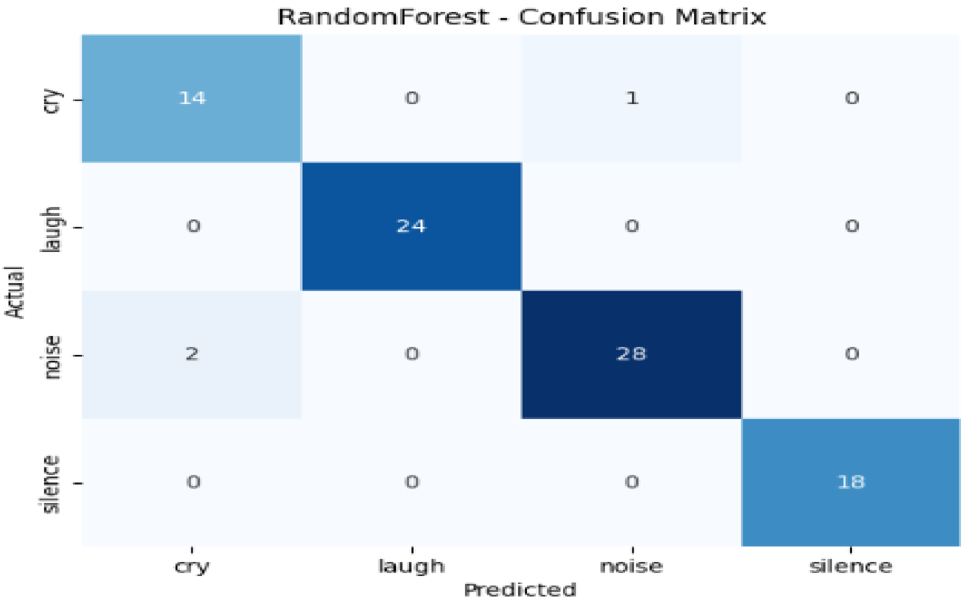


Figure 5: Normalized Confusion Matrix

The confusion matrix (Figure 5) provides an overview of the model’s performance. The system correctly detected 87% signal. The overall accuracy of the model is 87`%, showing strong performance in real-world conditions as shown in figure 6

Classification Report:

	precision	recall	f1-score	support
0	0.57	0.80	0.67	15
1	0.96	0.96	0.96	24
2	0.92	0.73	0.81	30
3	1.00	1.00	1.00	18
accuracy			0.86	87
macro avg	0.86	0.87	0.86	87
weighted avg	0.89	0.86	0.87	87

Fig 6. Overall Performance

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The figure 7 represents a comparison of five machine learning models—Decision Tree, Random Forest, XGBoost, CatBoost, and MLP—used in an infant cry project, evaluating their performance across four key metrics: accuracy, precision, recall, and F1-score. Notably, Random Forest, XGBoost, and CatBoost models exhibit near-perfect performance across all metrics, indicating highly effective classification capabilities for infant audio. In contrast, the Decision Tree model shows slightly lower, though still strong, performance, while the MLP model, despite high accuracy, demonstrates a noticeable dip in precision, recall, and F1-score. Overall, the ensemble models (Random Forest, XGBoost, and CatBoost) prove to be the most robust and reliable for this application, suggesting they are excellent choices for accurately distinguishing between different infant sounds like cries, laughter, noise, and silence.

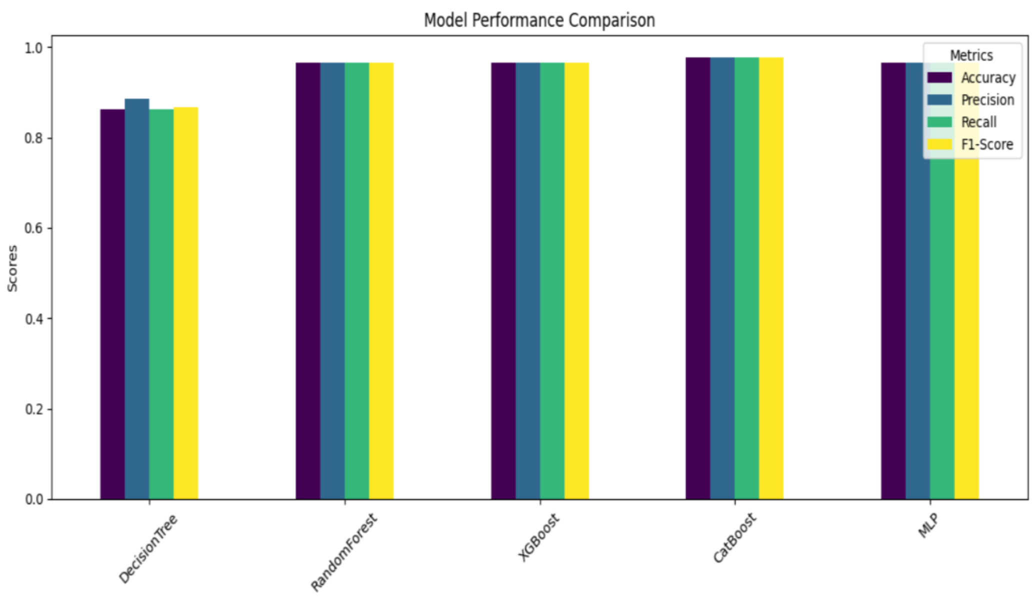


Fig 7. Model Performance Comparson

VI. CONCLUSION

The study on infant cry classification using machine learning demonstrates the potential of various algorithms in accurately distinguishing different types of cries, such as hunger, pain, discomfort, and illness. Traditional machine learning models like Decision Trees, Random Forest, XG Boost, and Cat Boost provide good interpretability and efficiency, making them suitable for low-power embedded systems. However, their performance is limited when dealing with complex audio features and sequential data. Deep learning models, including Multilayer Perceptron (MLP), Convolutional Neural Networks (CNNs), and Long Short-Term Memory (LSTM) networks, show superior performance by automatically learning both spatial and temporal patterns in infant cry signals.

REFERENCES

[1] J. Saraswathy, M. Hariharan, Sazali Yaacob and Wan Khairunizam, "Automatic Classification of Infant Cry: A Review", School of Mechatronic Engineering, niversiti Malaysia Perlis (UniMAP),02600 Perlis, Malaysia,2011. Kevin Kuo, "Feature extraction and recognition of infant cries". 2018.

[2] Jose Orozco Garcia, Carlos A. Reyes Garcia," Mel Frequency Cepstrum Coefficients Extraction from Infant Cry for Classification of Normal and Pathological Cry with Feed-forward Neural Networks", lnstituto Nacional de Astrofisica Optica y Electronica (INAOE) Luis Enrique Err 23 Page 11-25 © MANTECH PUBLICATIONS 2019. All Rights Reserved Journal of Artificial Intelligence, Machine Learning and Soft Computing Volume 4 Issue 1 I,Tonantzintla, Puebla, Mexico.

[3] Jose Orozco, Carlos A. ReyesGarcia," Implementation and Analysis of Training Algorithms for the Classification of Infant Cry with Feed-forward Neural Networks", Instituto Nacional de Astrofisica Optica y Electronica (INAOE) Luis Enrique Erro # 1,Tonantzintla, Puebla, Mexico,2018.

© 2025 IJMRSET | Volume 8, Issue 3, March 2025| DOI: 10.15680/IJMRSET.2025.0803293

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 8.206| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[4] Pritam Pal, Ananth N. Iyer and Robert E. Yantorno, "Emotion Detection From Infant Facial Expressions And Cries" Electrical & Computer Engineering Department, Temple University 12th & Norris Streets, Philadelphia, 2019.

[5] Ioana-Alina Bănică, Horia Cucu, Andi Buzo, Dragoş Burileanu, and Corneliu Burileanu," Baby Cry Recognition in Real-World Conditions", Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Romania, 2019.

[6] Andi Buzo, Dragos Burileanu, Cornelu Burileanu," Automatic Methods for infants cry classification", Speech and Dialogue Research Laboratory, University Politehnica of Bucharest, Bucharest, Romania 2019.

[7] Chuan-Yu Chang, Jia-Jing Li,"Application of Deep Learning for Recognizing Infant cries ",Chuan-Yu Chang, Jia-Jing Li National Yunlin University of Science & Technology, Taiwan,2018.

[8] J. Saraswathi, N.Thiyagar, M.Hariharan," Infant Cry Classification: Time Frequency Analysis", International conference on control system, computing and engineering, Malaysia, 2020.

[9] Shubham Asthana, Naman Varma, Vinay Kumar Mittal, "Peliminary Analysis of causes of Infant Cry ",Electronics and Communication Engineering 1,2The LNM Institute of Information Technology 1,2Rupa Ki Nangal, Post-Sumel, Via Jamdoli, Jaipur 302 031, Rajasthan, India 3 Indian Institute of Information Technology 3Sri City, Satyavedu Mandal, Chittoor District 517 588, Andhra Pradesh, India,2018.

[10] [10]. Mrs. Sameena Bano, Dr. K.M. RaviKumar, "Decoding Baby Talk: A Novel Approach for Normal Infant Cry Signal Classification", 2015 International Conference on Soft-Computing and Network Security (ICSNS 2015), Feb. 25 – 27, Coimbatore, INDIA,2019.

[11] Suryakanth V.Gangashetty, Pruthvi Raj Myakala, Rajashree Nalumachu, Shivam Sharma, VK Mithal," Acoustic Analysis of Infant Cry Signal Toward Automatic Detection of the cause of Crying", Indian Institute of Information Technology Chittoor, Sri City, A.P., India1,2,3 International Institute of Information Technology, Hyderabad, Telangana, India4 CEO, Ritwik Software Technologies Pvt. Ltd., Hyderabad, India5,2020.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com