IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Deep Learning-Based Infant Cry Analysis for Need and Distress Identification

[1]Asst. Prof. Arunadevi S. Khaple, Department of Computer Engineering, ZCOER , Pune
[2]Om C. Shinde, Department of Computer Engineering, ZCOER , Pune
[3]Anish B. Talole, Department of Computer Engineering, ZCOER , Pune
[4]Sairaj M. Shilimkar, Department of Computer Engineering, ZCOER , Pune
[5] Shreejit B. Sawant, Department of Computer Engineering, ZCOER , Pune

Abstract - Abstract - Infant care is a critical responsibility, and accurately identifying the reasons behind an infant's cry is essential for timely and effective caregiving. Traditional approaches primarily depend on parental observation and experience, which may lead to inconsistent interpretations, delayed responses, and increased caregiver stress. To address these challenges, this paper presents an enhanced real-time infant cry classification system that utilizes deep learning techniques to classify infant cries into categories such as hunger, discomfort, pain, and tiredness. The proposed system combines YAMNet-based feature extraction with an optimized Convolutional Neural Network (CNN) architecture to improve classification accuracy and robustness in diverse acoustic environments [1][3][5].. Real-time audio recordings captured through the mobile application are securely processed by the backend and analyzed by the deep learning model to generate prediction probabilities for each cry category. In addition to classification, the system incorporates an intelligent care recommendation engine that provides caregivers with personalized precautionary suggestions based on the predicted infant need [2][6].

Furthermore, prediction results and associated metadata are stored within the database, enabling historical trend analysis and comprehensive reporting. These features allow caregivers to monitor infant behavioral patterns over time and support informed healthcare decisions. By combining advanced audio processing, transfer learning, cloud infrastructure, and mobile accessibility, the proposed system enhances caregiving efficiency, reduces uncertainty, and provides actionable insights into infant well-being [5][7][8].The experimental evaluation demonstrates improved classification performance, real-time responsiveness, and scalability compared to conventional infant cry analysis approaches. Future enhancements include the integration of transformer-based audio models, explainable artificial intelligence techniques, larger and more diverse datasets, and IoT-enabled smart infant monitoring systems to further improve the effectiveness of infant care solutions [2][6][10].

Key Words: Infant Cry Classification, Deep Learning, YAMNet, Convolutional Neural Network (CNN), Real-Time Audio Analysis, React Native, Java Spring Boot, Amazon S3, PostgreSQL, Healthcare AI.

1.INTRODUCTION

The need for effective infant care has increased significantly as parents and caregivers seek reliable ways to understand and respond to an infant's needs. Since infants cannot communicate verbally, crying serves as their primary means of expressing hunger, discomfort, pain, tiredness, and other physical conditions. However, interpreting the exact reason behind a cry often depends on parental experience and observation, which can result in misinterpretation and delayed responses. This challenge creates a demand for intelligent systems capable of automatically analyzing infant cries and providing accurate insights in real time.

Recent advancements in artificial intelligence and deep learning have shown promising results in audio classification tasks, including infant cry analysis. Traditional machine learning approaches relied on handcrafted acoustic features and classifiers such as Support Vector Machines (SVMs) and Random Forests [4][7]. Although these methods achieved moderate success, their performance was often limited in noisy environments and required extensive feature engineering. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have

demonstrated superior capability in automatically learning meaningful patterns from audio signals, leading to improved classification accuracy [1][3].

To address these challenges, this paper presents an enhanced real-time infant cry classification system that combines YAMNet-based feature extraction with an optimized CNN architecture [5][9]. YAMNet provides robust audio embeddings from infant cry recordings, while the CNN model classifies the cries into categories such as hunger, discomfort, pain, and tiredness. The proposed approach incorporates improved preprocessing and data augmentation techniques to enhance model robustness and generalization performance.

The system integrates a React Native mobile application, a Java Spring Boot backend, Amazon S3 cloud storage, and a PostgreSQL database to provide a complete end-to-end solution. Parents can record infant cries in real time and receive classification results along with intelligent care recommendations. In addition, prediction histories are stored and analyzed to generate reports that help caregivers monitor infant behavioral patterns over time [2][6].

By combining deep learning, cloud technology, and mobile accessibility, the proposed system aims to reduce caregiver uncertainty and improve responsiveness in infant care. Furthermore, the scalable architecture supports future enhancements such as transformer-based audio models, explainable AI techniques, and IoT-enabled smart monitoring devices, making it a practical and future-ready healthcare solution [6][10].

2. RELATED WORK

Infant cry analysis has been widely studied as an effective approach for understanding infant needs and supporting early healthcare interventions. Initial research primarily focused on traditional audio signal processing techniques, where acoustic features such as Mel-Frequency Cepstral Coefficients (MFCCs), pitch, frequency, and energy patterns were extracted and classified using machine learning algorithms including Support Vector Machines (SVMs) and Random Forests [4][7]. While these approaches demonstrated reasonable classification performance, they required extensive feature engineering and often struggled in noisy real-world environments.

The emergence of deep learning has significantly improved infant cry classification by enabling automatic feature learning from audio signals. Convolutional Neural Networks (CNNs) have been extensively used to analyze cry spectrograms and identify patterns associated with different infant conditions such as hunger, discomfort, pain, and tiredness [1][3]. These models have achieved higher accuracy compared to traditional methods by learning complex temporal and spectral characteristics directly from the data. However, their performance often depends on the availability of large labeled datasets and significant computational resources.

Recent studies have explored advanced architectures including recurrent neural networks, attention-based models, and transformer-based approaches to further enhance classification performance. Adaptive CNN-RNN frameworks and SE-ResNet-Transformer models have shown improved capability in capturing both temporal and contextual information from infant cries [2][6]. Similarly, deep convolutional recurrent neural networks with attention mechanisms have demonstrated strong performance in complex audio classification tasks, highlighting the potential of hybrid architectures for infant cry recognition [10].

Transfer learning and self-supervised learning have also emerged as promising solutions for addressing limited data availability. Models such as YAMNet and self-supervised audio learning frameworks leverage knowledge acquired from large-scale audio datasets to generate robust feature embeddings [5][9]. These approaches reduce the dependency on extensive labeled infant cry datasets while improving generalization and noise robustness. As a result, they have become increasingly popular in modern infant cry classification systems.

Despite these advancements, many existing studies primarily focus on classification accuracy and do not provide complete end-to-end solutions for caregivers. Features such as real-time mobile accessibility, cloud-based data management, historical report generation, and intelligent care recommendations are often missing [1][8]. To address these limitations, the proposed system integrates YAMNet-based feature extraction, an optimized CNN architecture, cloud infrastructure, and a mobile application to deliver accurate cry classification along with actionable caregiving insights. This combination enhances both technical performance and practical usability, making the system suitable for real-world infant care applications.

3. METHODOLOGY

3.1 Data Collection and Preprocessing

The dataset used in this study consists of infant cry recordings categorized into four major classes: hunger, discomfort, tiredness, and pain [1][4]. Since raw audio recordings often contain background noise, silence, and recording inconsistencies, several preprocessing techniques were applied to improve data quality and model performance [2]. The preprocessing pipeline included silence removal, noise reduction, audio normalization, and resampling to maintain a consistent audio format. Additionally, data augmentation techniques such as pitch shifting and time stretching were employed to increase dataset diversity and improve the model's ability to generalize across different recording conditions [1][3]. These preprocessing steps

enhanced feature quality and reduced overfitting during training [5].

3.2 Feature Extraction

To extract meaningful acoustic representations from infant cry recordings, YAMNet, a pre-trained deep neural network trained on Google's AudioSet dataset, was utilized [5][9]. YAMNet generates high-dimensional embeddings that capture important temporal and spectral characteristics of audio signals. Compared to traditional handcrafted features, these embeddings provide richer and more robust representations, particularly when working with limited labeled data. The extracted embeddings are then used as input for the downstream classification model [6][10].

3.3 Model Design

The proposed architecture combines YAMNet embeddings with an optimized Convolutional Neural Network (CNN) for multi-class infant cry classification [1][3]. The CNN consists of convolutional layers with ReLU activation functions, max-pooling layers for feature reduction, dropout layers for regularization, and fully connected layers followed by a softmax classifier. This architecture enables the model to learn discriminative patterns associated with different infant cry categories while maintaining computational efficiency.

• Training: Optimizer – Adam, Loss Function – Categorical Cross-Entropy

• Validation: Early Stopping applied to prevent overfitting

• Output: Probability distribution across hunger, discomfort, tiredness, and pain classes

3.4 Backend and Cloud Integration

The system follows a cloud-enabled architecture for real-time processing and secure data management. When an infant cry is recorded through the mobile application, the audio file is transmitted to the Java Spring Boot backend, which securely stores the recording in Amazon S3 cloud storage. The backend generates a secure access link and forwards it to the model API for inference. After classification, the prediction probabilities are returned to the backend and subsequently delivered to the mobile application. All prediction results and metadata are stored in a PostgreSQL database to facilitate historical reporting and future analysis.

3.5 Intelligent Care Recommendation Engine

To improve usability and practical relevance, an intelligent recommendation engine has been integrated into the system. Based on the predicted cry category, the engine generates appropriate caregiving suggestions for parents and caregivers. For example, if the model predicts hunger, the system recommends feeding the infant, while discomfort-related predictions encourage caregivers to check diapers, clothing, or environmental conditions. This functionality transforms classification outputs into actionable insights that support faster and more informed decision-making.

3.6 Model Evaluation

The proposed model was evaluated using standard performance metrics including accuracy, precision, recall, and F1-score. These metrics provide a comprehensive assessment of classification performance across all cry categories. In addition, a confusion matrix was generated to visualize prediction outcomes and identify classes that are frequently misclassified. The evaluation process also considered inference speed and real-time responsiveness to ensure the model's suitability for deployment in mobile healthcare applications.

Figure 1 shows the confusion matrix of the proposed model, illustrating the classification performance for each infant cry category. The analysis provides valuable insights into model strengths and areas for future improvement.

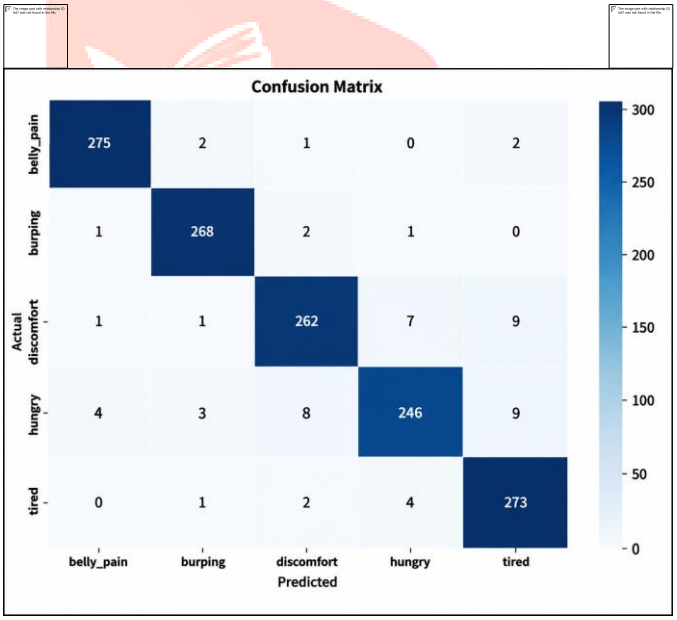

Figure 1.

4. IMPLEMENTATION

The enhanced real-time infant cry classification system integrates mobile, cloud, backend, and deep learning components into a unified architecture, as illustrated in Figure 2. The mobile application, developed using React Native, serves as the primary interface for parents and caregivers. It enables users to record infant cries in real time and securely transmit audio data to the backend server. The application displays predicted cry categories along with confidence scores, provides intelligent caregiving recommendations, and allows users to access

historical reports for monitoring infant behavioral patterns over time.

The backend infrastructure is implemented using Java Spring Boot and acts as the central communication layer between the mobile application, cloud storage, database, and deep learning service. When an audio recording is received, the backend securely uploads the file to Amazon S3 and generates a secure access link. This link is then forwarded to the model API for inference. After receiving the classification results, the backend processes the prediction data and returns it to the mobile application. Simultaneously, prediction results, timestamps, and related metadata are stored in a PostgreSQL database for future analysis and report generation.The deep learning service consists of a preprocessing module, a YAMNet-based feature extraction layer, and an optimized CNN classifier. During inference, the uploaded audio undergoes preprocessing to remove noise and normalize audio quality. YAMNet then extracts meaningful acoustic embeddings, which are passed to the CNN model for classification. The model generates probability scores for each cry category, including hunger, discomfort, tiredness, and pain. The category with the highest probability is selected as the final prediction and is used by the recommendation engine to provide appropriate caregiving guidance.

In addition to cry classification, the system incorporates an intelligent care recommendation engine that maps predicted categories to relevant precautionary actions. This feature enhances practical usability by helping caregivers respond quickly and appropriately to infant needs. The historical reporting module further strengthens the system by enabling trend analysis of infant behavior, allowing parents to identify recurring patterns and share meaningful information with healthcare professionals when necessary.

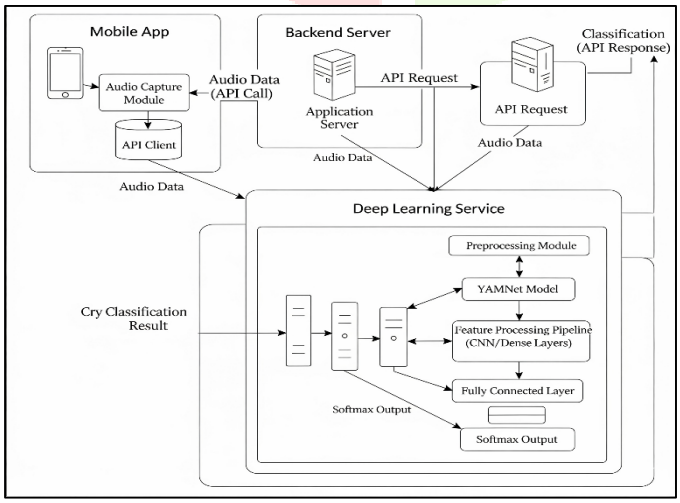

Figure 2.

Figure 2 illustrates the complete workflow of the proposed system. Audio data flows from the mobile application to the backend server, then to the deep learning service for preprocessing, feature extraction, and classification. The prediction results are subsequently returned to the backend and displayed within the mobile application along with caregiving recommendations and historical reports. This architecture ensures real-time performance, secure data management, scalability, and improved support for caregivers in everyday infant care scenarios.The modular design of the system enables easy integration of future enhancements without affecting existing functionalities.

5. CONCLUSION

This paper presents an enhanced real-time infant cry classification system that combines deep learning, cloud computing, and mobile technology to support parents and caregivers in understanding infant needs more effectively. By utilizing YAMNet-based feature extraction and an optimized CNN architecture, the proposed system accurately classifies infant cries into categories such as hunger, discomfort, tiredness, and pain. The integration of a React Native mobile application, Java Spring Boot backend, Amazon S3 cloud storage, and PostgreSQL database provides a scalable and efficient platform for real-time cry analysis.

In addition to classification, the system incorporates an intelligent care recommendation engine that assists caregivers by providing appropriate suggestions based on the predicted cry category. The storage of prediction history and report generation features further enhance the system by enabling long-term monitoring of infant behavioral patterns. Experimental results indicate that the proposed approach delivers reliable classification performance and real-time responsiveness, making it suitable for practical healthcare applications.

Future work may focus on expanding the dataset, improving classification accuracy through transformer-based architectures, incorporating explainable AI techniques, and integrating IoT-enabled smart monitoring devices. These enhancements have the potential to further improve the effectiveness and accessibility of intelligent infant care systems.

ACKNOWLEDGEMENT

We would like to express our sincere gratitude to Prof. Arunadevi S. Khaple for her valuable guidance, continuous encouragement, and insightful suggestions throughout the development of this research work. Her expertise and support greatly contributed to the successful completion of the project.We are thankful to Zeal College of Engineering and Research, Pune, for providing the necessary infrastructure, technical resources, and academic environment required to carry out this research effectively. The facilities and opportunities offered by the institution played a significant role in the implementation of the proposed system.

We also extend our heartfelt appreciation to all faculty members, colleagues, and friends for their cooperation, constructive feedback, and support during various stages of the project. Their encouragement and assistance helped us overcome numerous challenges encountered throughout the research process.Finally, we acknowledge the contributions of researchers and authors whose published work provided valuable knowledge, inspiration, and guidance for this study.

REFERENCES

[1] Khorshed Alam, Khondaker A. Mamun, "From Cries to Answers: A Comprehensive CNN+DNN Hybrid Model for Infant Cry Classification with Enhanced Data Augmentation and Feature Extraction," IEEE, 2024.

[2] Geofrey Owino, Timothy Kamanu, John Ndiritu, Conlet Biketi Kikechi,"Adaptive Infant Cry Classification Using Multi-Armed Bandit Modality Selection in an Attentive Convolutional Recurrent Neural Network Model," Springer, 2025.

[3] Dayanand Alagundi, Satish Chikkamath, Anchal Patil, Nirmala S R, Varadesh Haridas, Suneeta Budihal,"Infant Cry Classification Using CNN-MFCC Fusion," IEEE, 2024.

[4] K. Teeravajanadet, N. Siwilai, K. Thanaselanggul, N. Ponsiricharoenphan, S. Tungjitkusolmun, P. Phasukkit,"An Infant Cry Recognition based on Convolutional Neural Network Method", *IEEE, 2019.*

[5] Arsenii Gorin, Cem Subakan, Sajjad Abdoli, Junhao Wang, Samantha Latremouille, Charles Onu,"Self-Supervised Learning for Infant Cry Analysis",IEEE, 2023.

[6] Feng Li , Chenxi Cui and Yashi Hu Classification of Infant Crying Sounds Using SE-ResNet-Transformer. Springer, 2024.

[7] Vincent PM, Srinivasan K, Chang CY Deep learning assisted neonatal cry classification via support vector machine models. Frontiers, 2021.

[8] Ji C, Mudiyanselage TB, Gao Y, Lui S, Li HA review of infant cry analysis and classification. J Audio Speech Music Process. IEEE, 2021.

[9] A. Gorin, C. Subakan, S. Abdoli, J. Wang, S. Latremouille and C. Onu, "Self-Supervised Learning for Infant Cry Analysis", IEEE 2023.

[10] C. W. Huang and S. S. Narayanan, "Deep convolutional recurrent neural network with attention mechanism for robust speech emotion recognition", IEEE 2017.