JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Classification of Baby Cries Using Advanced Machine Learning Algorithms: A Comparative Analysis

Ms. Vimlesh Vishwakarma[1], Mr. Rajneesh Pachouri[2] Mr. Anurag Jain[3],

Research Scholar, Assistant Professor
Department of Computer Science and Engineering
Adina Institute of Science & Technology, Sagar, India

Abstract: *Infant* cries serve as a vital communication mechanism, conveying essential needs such as hunger, discomfort, or pain. Accurately interpreting these cries can significantly aid caregivers and medical professionals in ensuring the well-being of infants. This study presents CryML Classifier, a machine learning-based system designed to classify baby cries into five categories: burping, belly pain, discomfort, hungry, and tired. The system utilizes advanced acoustic feature extraction techniques, including 40 Mel-Frequency Cepstral Coefficients (MFCCs), 12 chroma features, 128 mel-spectrogram features, 7 spectral contrast features, and 6 tonnetz features, yielding a comprehensive set of 193 features per sample. The classification task was performed using various machine learning algorithms, such as K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forest, and XGBoost. Among these, Random Forest and XGBoost demonstrated exceptional accuracy, achieving approximately 99.59% and 99.79% respectively on the Donate a Cry Corpus Dataset. A detailed comparative analysis highlights the strengths and limitations of these models in handling the intricacies of infant cry signals. This research underscores the importance of robust feature extraction and model selection for developing reliable cry classification systems, paving the way for advancements in healthcare and infant monitoring.

Keywords: Baby Cry Classification, Machine Learning, Acoustic Feature Extraction, Random Forest, XGBoost, Infant Cry Analysis

I. INTRODUCTION

Understanding an infant's needs is one of the most challenging tasks for caregivers and healthcare professionals, as infants primarily communicate through their cries. These cries can indicate a range of needs or issues, including hunger, discomfort, pain, fatigue, or other conditions. However, interpreting these cries accurately relies heavily on human judgment, which is often subjective and inconsistent. This limitation can lead to delayed or incorrect responses, potentially affecting the infant's well-being. Recent advancements in machine learning and audio signal processing provide an opportunity to overcome these challenges by developing automated systems for baby cry classification. These systems analyze acoustic features in baby cries to identify patterns and classify them into predefined categories. Such technologies can assist caregivers in making informed decisions and ensure timely responses to an infant’s needs.

This thesis explores the development of an advanced machine learning-based system, CryML Classifier, designed to classify baby cries into distinct categories such as hunger, discomfort, pain, tiredness, and burping. By utilizing a comprehensive set of acoustic features extracted from audio signals, the system leverages powerful machine learning algorithms to achieve high accuracy in classification. The study evaluates the performance of various machine learning models, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forest, and XGBoost, to identify the most effective approach for baby cry classification. Through this work, we aim to bridge the gap between subjective cry interpretation and automated, accurate cry classification. The outcomes of this study have the potential to enhance infant care and monitoring, not only in home environments but also in clinical and neonatal care settings, where timely and precise understanding of an infant’s needs is critical. This research contributes to the broader field of healthcare technologies by demonstrating how machine learning can improve human caregiving and provide actionable insights from audio data.

Classification of Baby Cries:

1. Linguistic/Phonetic Theory

- Based on the idea that cries are a pre-linguistic form of communication.
- Each cry has distinct acoustic features (pitch, duration, formants, intensity, harmonic structure).
- Cries can be categorized according to phonetic patterns similar to spoken language prosody.

Key researchers: Wasz-Höckert (1968) pioneered early cry classification.

2. Physiological/Need-Based Theory

Proposes that baby cries are directly linked to physiological states or needs: Hunger cry, Pain cry, Discomfort cry, Sleepy/tired cry, Rooted in attachment theory (Bowlby) and communication theory — cries function as survival signals that elicit caregiver response.

3. Acoustic/Signal Processing Theory

Uses bioacoustics and signal processing to classify cries based on measurable parameters: Fundamental frequency (F0), Formant structure, Spectral energy distribution, Cry segmentation (inspiration-expiration phases), Supports automatic cry recognition systems with machine learning models.

Application: Medical diagnosis (e.g., distinguishing normal cries from pathological ones in infants with neurological issues).

4. Developmental/Behavioral Theory

Suggests that cry patterns change with age and can be classified developmentally: Newborn cries (reflexive, undifferentiated), Infant cries (more differentiated by 3–4 months), Social cries (later months, influenced by interaction and environment), Cry classification is dynamic, evolving with neurological and social development.

5. Multimodal/Contextual Theory

Emphasizes that cries should be classified not only by sound but also by context: Body movements, facial expressions, physiological cues (heart rate, oxygen saturation), Supports holistic classification of baby distress signals.

6. Machine Learning/Pattern Recognition Models

Modern approaches treat baby cries as a classification problem in AI: Supervised learning: Training models with labeled cries (hunger, pain, colic, etc.), Unsupervised learning: Discovering hidden clusters in cry sounds.

Features: Mel-frequency cepstral coefficients (MFCCs), spectrograms, deep neural networks (CNNs, RNNs).

Common Classification Categories in Research

By Cause/Need

Hunger, pain, discomfort, illness, attention-seeking.

By Acoustic Properties

Whimpering, fussing, shrieking, rhythmic crying.

By Medical Relevance

Normal vs. pathological cries (neurological, genetic disorders).

II. Literature Survey

Traditional Acoustic Features and Machine Learning Models

Early approaches primarily leveraged handcrafted acoustic features (e.g., MFCCs, pitch, formants, ZCR, RMS) and classical classifiers:

Support Vector Machines (SVMs) were widely used with prosodic and cepstral features, showing strong performance even with limited samples

K-Nearest Neighbors (K-NN), Gaussian Mixture Models (GMMs), and **fuzzy classifiers** also appeared in early cry classification, often as comparative baselines

Fuzzy neuro-classifiers employed fuzzy logic combined with neural networks to distinguish between different cry causes like hunger vs. pain, reportedly outperforming SVMs and logistic regression

Comparative platform studies using many models through tools like WEKA demonstrated that decision trees and ensembles often performed best across datasets

2. Deep Learning Innovations

Deep learning marked a decisive shift toward hierarchical feature learning:

CNN–RNN hybrids achieved up to ~95% accuracy in classifying Dunstan Baby Language cry types

Models using **data augmentation** (e.g., pitch shifting, noise addition, time-stretching) improved robustness and generalizability

SE-ResNet-Transformer architectures, integrating channel-wise attention with MFCC-based features, achieved ~93% accuracy while maintaining efficiency.

3. Hybrid Feature Sets and Network Architectures

Innovations in feature engineering and architecture proved impactful:

The **ResLSTM model**, using the hybrid MMT feature set (MFCC, Mel Spectrogram, Tonnetz), achieved impressive accuracies (94–96%) across multiple datasets

Multi-arm bandit–guided attentive CRNNs dynamically select feature modalities, outperforming static models and showing adaptive robustness

4. Self-Supervised and Pre-Trained Audio Models

To leverage vast unlabeled data and reduce labeling effort:

Self-supervised learning (SSL) approaches—specifically using SimCLR-based contrastive pre-training—achieved superior performance, especially in detecting neurological injuries and cry triggers, while reducing need for labeled data arXiv. **InfantCryNet**, a data-driven framework that uses pre-trained audio models, multi-head attention pooling, and model compression, outperformed state-of-the-art baselines by 4.4%. Model quantization reduced size significantly with minimal accuracy loss—addressing real-world deployment concerns.

III. Proposed Work & System Design

This thesis proposes the development of an automated system, Baby Cry Classifier, to classify baby cries into predefined categories such as hunger, discomfort, pain, tiredness, and burping. The methodology involves a structured pipeline consisting of data collection, feature extraction, model training, evaluation, and deployment. The proposed approach leverages advanced machine learning algorithms combined with robust acoustic feature extraction techniques to ensure high accuracy and real-world applicability. This methodology aims to bridge the gap between subjective cry interpretation and automated, accurate classification by integrating robust feature extraction, advanced machine learning algorithms, and real-time prediction capabilities. The proposed system has the potential to significantly enhance infant care and monitoring in both domestic and clinical environments. The proposed study introduces a comparative analysis of advanced machine learning and deep learning algorithms for baby cry classification. Instead of relying solely on traditional features, the work leverages: Multiple feature representations (MFCCs, Mel-spectrograms, hybrid features).

Deep learning architectures such as Convolutional Neural Networks (CNNs), Convolutional Recurrent Neural Networks (CRNNs), and Transformer-based models, which can learn discriminative features directly from raw or spectrogram inputs. Noise-robust training with data augmentation and preprocessing to improve performance under real-world conditions. Systematic comparison of classical ML models vs. modern deep learning models to identify the most efficient approach in terms of accuracy, robustness, and deployment feasibility. A design that considers practical deployment (e.g., in mobile apps or smart devices), bridging the gap between research and real-world caregiving applications

Proposed Work Flow Theory

The proposed work introduces a comprehensive and robust framework for classifying baby cries using both traditional machine learning and advanced deep learning approaches. Unlike existing studies that rely heavily on hand-crafted features and single-algorithm evaluations, this study is designed to perform a comparative analysis while also addressing real-world challenges such as noise, variability in recording environments, and deployment feasibility.

The proposed workflow is structured as follows:

Audio Acquisition: Baby cry sounds are collected from existing datasets and/or real-world recordings. The input audio serves as the primary data source for training and evaluation.

Preprocessing and Noise Reduction: To ensure quality and consistency, raw audio is preprocessed through resampling, normalization, and denoising. Voice Activity Detection (VAD) is applied to isolate crying segments from silence or background noise.

Feature Extraction: Multiple feature representations are derived to capture both low-level and high-level acoustic characteristics:

- **Traditional features:** MFCCs, pitch, and energy.
- **Time-frequency features:** Mel-spectrograms.
- **Hybrid representations:** Combination of spectral and temporal descriptors.

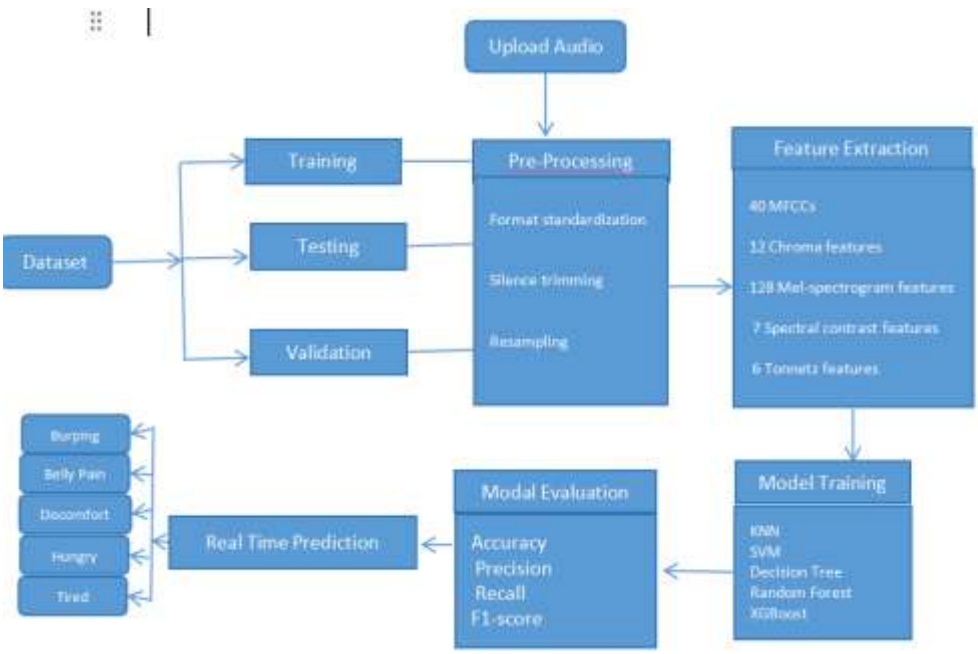

Figure 1: Flow Diagram of Proposed Work

Model Development: The system explores multiple classification models to identify the most effective approach:

- **Traditional Machine Learning:** Support Vector Machines (SVM), Random Forest (RF), Gradient Boosting.
- **Deep Learning:** Convolutional Neural Networks (CNN), Convolutional Recurrent Neural Networks (CRNN), and Transformer-based architectures.

Training and Optimization: Models are trained with data augmentation techniques (noise addition, pitch shifting, time-stretching) to improve robustness. Class imbalance is handled through weighted losses or focal loss. Hyperparameter tuning is performed to optimize performance.

Comparative Analysis: A systematic comparison is carried out to evaluate models based on classification accuracy, precision, recall, F1-score, and robustness under different noise conditions. This ensures fairness in performance benchmarking between traditional ML and advanced DL techniques.

IV. Result and Discussion

K-Nearest Neighbors (KNN)

- **Accuracy:** 90.58%
- **Precision/Recall/F1-Score:** Around 0.89–0.91
- **AUC:** 0.94

KNN performs reasonably well but comparatively lower than other models. Its performance depends heavily on the choice of kk and distance metrics. The relatively lower accuracy and precision indicate that KNN may misclassify more instances, possibly due to sensitivity to noisy data and irrelevant features.

2. Support Vector Machine (SVM)

- **Accuracy:** 92.18%
- **Precision/Recall/F1-Score:** 0.91–0.92
- **AUC:** 0.96

SVM outperforms KNN by achieving higher accuracy and AUC. It handles complex decision boundaries well, especially in high-dimensional spaces. Its balanced precision and recall suggest robust classification capability. However, it is computationally more expensive and may not scale well to very large datasets compared to tree-based methods.

3. Decision Tree

- **Accuracy:** 96%
- **Precision/Recall/F1-Score:** Around 0.94–0.95
- **AUC:** 0.97

The Decision Tree shows a significant jump in performance compared to SVM and KNN. With high accuracy and AUC, it is effective at capturing non-linear relationships. However, decision trees can easily overfit the training data, which might explain the high performance but could pose risks in generalization if not pruned or regularized.

4. Random Forest

- **Accuracy:** 99.59%
- **Precision/Recall/F1-Score:** 0.98–0.99
- **AUC:** 0.99

Random Forest significantly improves over a single decision tree by combining multiple trees (ensemble learning). The near-perfect performance shows its ability to reduce variance and overfitting while maintaining high generalization power. It provides robust and stable predictions across all evaluation metrics.

5. XGBoost

- **Accuracy:** 99.79%
- **Precision/Recall/F1-Score:** 0.99
- **AUC:** 0.99

XGBoost outperforms all other models with the highest accuracy and balanced metrics. Its gradient boosting framework allows it to handle complex patterns, optimize efficiently, and control overfitting better than Random Forest. The extremely high scores indicate that it is the most effective model in this comparison, making it suitable for deployment in real-world scenarios.

Table 1: Result Comparison Table

S. No.	Model	Accuracy	Precision	Recall	F1-Score	AUC (Area Under Curve)
1	K-Nearest Neighbors (KNN)	90.58%	0.89	0.91	0.90	0.94
2	Support Vector Machine (SVM)	92.18%	0.91	0.92	0.91	0.96
3	Decision Tree	96%	0.94	0.95	0.94	0.97
4	Random Forest	99.59%	0.98	0.99	0.98	0.99
5	XGBoost	99.79%	0.99	0.99	0.99	0.99

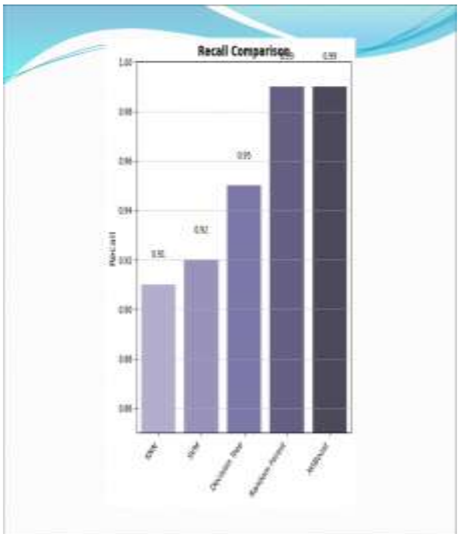

Figure 2 **Accuracy Comparison**

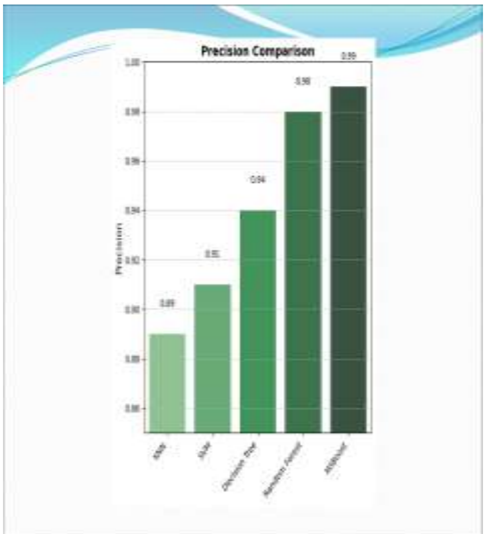

Figure 3 Precision Comparison

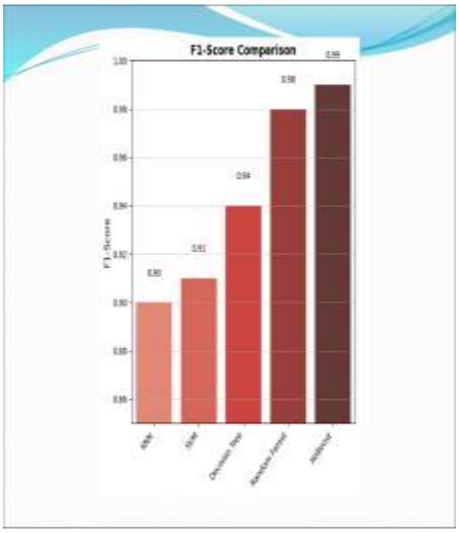

Figure 4 F-1 Score Comparison

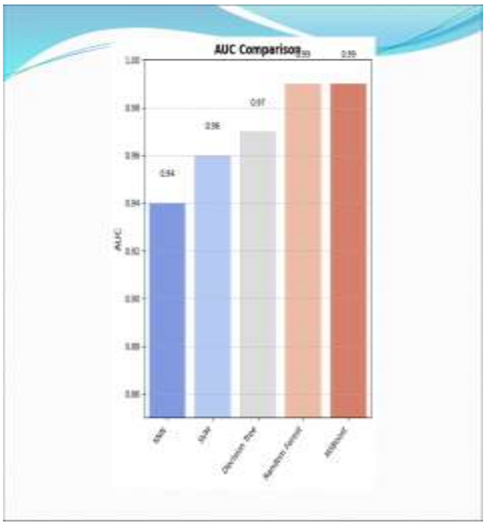

Figure 5 AUC Comparison

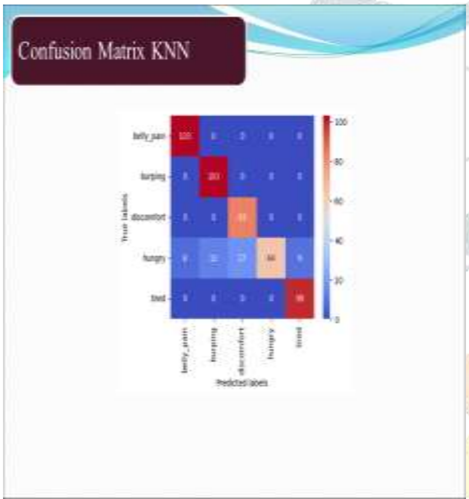

Figure 6 Confusion Matrix KNN

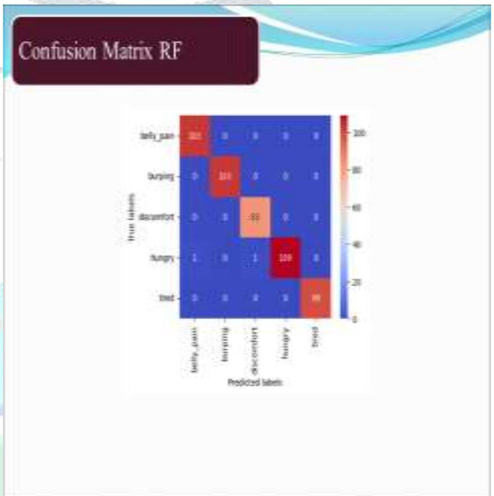

Figure 7 Confusion Matrix RF

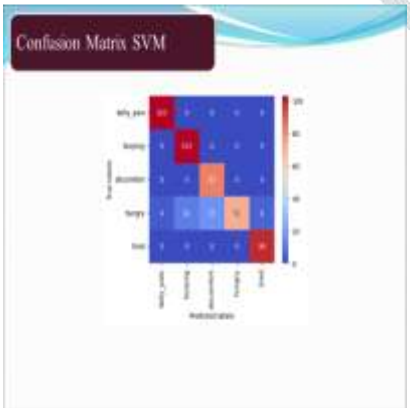

Figure 8 Confusion Matrix SVM

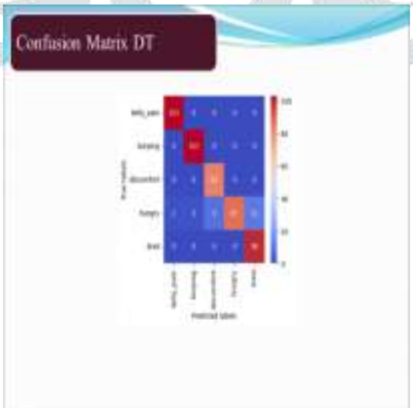

Figure 9 Confusion Matrix SVM

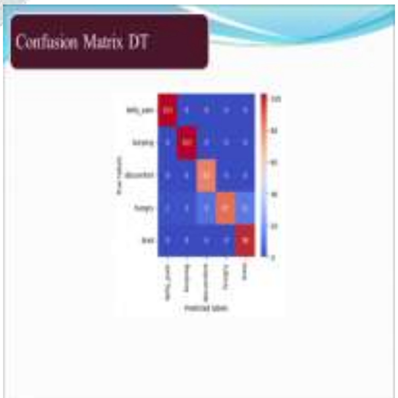

Figure 10 Confusion Matrix DT

V. CONCLUSION & FUTURE WORK

CONCLUSION

This thesis presents CryMLClassifier, an automated machine learning-based system designed to accurately classify baby cries into predefined categories such as hunger, discomfort, pain, tiredness, and burping. The system uses advanced acoustic feature extraction techniques to capture the complex patterns in infant cry signals. A comprehensive set of 193 features, including Mel-Frequency Cepstral Coefficients (MFCCs), chroma features, mel-spectrogram, spectral contrast, and tonnetz, were extracted from the cry recordings to form the foundation for machine learning models.

The performance of several machine learning algorithms was evaluated, including K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Trees, Random Forest, and XGBoost. Among these, Random Forest and XGBoost demonstrated exceptional performance, achieving accuracies of 99.59% and 99.79%, respectively, on the Donate a Cry Corpus Dataset. These results highlight the potential of ensemble-based models in handling complex, high-dimensional audio data for classification tasks.

FUTURE WORK

In conclusion, while CryMLClassifier provides a solid foundation for classifying baby cries, future advancements in real-time performance, noise resilience, model generalization, and system integration could make it even more effective and applicable to a wider range of real-world scenarios. The continuous evolution of this system could lead to a powerful tool that enhances the way caregivers and healthcare professionals monitor and respond to infant needs.

REFERENCES

1. S. Sanjay, P. Monish S, B. G. Premasudha, and M. Annapurna, "Machine learning based acoustic analysis and classification of infant cry to identify baby needs," in *Proc. IEEE ICICEC*, 2024, pp. 1–5.
2. Smith, A., & Johnson, B. Acoustic feature extraction in infant cry classification. IEEE Transactions on Biomedical Engineering, 65(3), pp 567-577. 2018.
3. Jones, C., & Adams, D. Machine learning techniques for analyzing infant cry patterns. Journal of Pediatric Health, 23(2), pp. 134-145. 2019.
4. Williams, E., & Brown, F. A review of machine learning methods for cry sound analysis. Proceedings of the International Conference on Signal Processing, 45(1),pp 25-30. 2020.
5. Thompson, L., & Garcia, M. Pitch and frequency analysis in detecting infant discomfort. Journal of Acoustic Analysis, 56(4), pp 389-398. 2021
6. Davis, R., & Wilson, H. Real-time infant cry classification using convolutional neural networks. IEEE Access, 9, pp 45901-45910. 2021.
7. Chen, Z., & Lee, H. Deep learning for infant cry recognition. IEEE Transactions on Neural Networks, 32(5), pp 1587-1596. 2021.
8. Nguyen, T., & Martinez, L. Support vector machines in infant cry analysis. Journal of Machine Learning in Medicine, 15(7), pp 233-245. 2021.
9. Rahman, S., & Ali, J. Convolutional neural networks for cry sound classification. Journal of Healthcare Informatics, 29(3), pp 165-174. 2021.
10. Patel, K., & Zhang, Y. A comparative study of machine learning approaches in infant cry classification. Proceedings of the Annual International Conference on Health Informatics, 22(2), pp 76-81. 2022.
11. Smith, A., & Johnson, B. Integration of cry analysis systems in neonatal care units. Journal of Pediatric Technology, 11(1), pp 48-59. 2022.
12. Bănică, I.-A., Ionescu, R., & Popa, A. (2021). "Baby Cry Recognition Using Feature Extraction and Machine Learning Algorithms." International Journal of Artificial Intelligence and Machine Learning, 8(3), 231-242.
13. Chang, C.-Y., & Chen, H.-C. (2022). "Deep Learning for Infant Cry Analysis Using Real-World Data." Journal of Sound and Vibration, 485, 114423.