Deciphering Infant Cry using Advanced Machine Learning Techniques

Kalyani Phursule

23229268

Master of Science in Data Analytics

National College of Ireland

**Abstract**

This paper presents the development of a reliable baby cry classification system using machine learning and deep learning techniques to classify various types of infant cries. The motivation for this research stems from the critical importance of accurately identifying the reasons behind a baby's cries, which is vital for parents and healthcare professionals. We utilized multiple models, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, AdaBoost, XGBoost, Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), Convolutional Neural Network (CNN), and ensemble models, to achieve this goal. To address the issue of class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) was applied to enhance the dataset. The evaluation revealed that Random Forest and XGBoost models provided the highest accuracy, with Random Forest (using SMOTE) achieving an accuracy of 99.6% and an F1 score of 99.6%. Ensemble models also demonstrated improvements in performance, although their accuracy was slightly lower than standalone models trained with SMOTE. The findings emphasize the significance of handling data imbalance to improve model performance, especially for minority classes. The study successfully demonstrated the potential of machine learning in classifying infant cries with high accuracy, but challenges such as computational requirements and real-time processing remain. Future work will explore transformer-based models, the inclusion of pathological cry types, and optimization for real-time applications in low-resource environments. This research contributes to enhancing automated infant care, providing a foundation for further improvements in smart baby monitoring systems.

**Keywords:** **Baby cry classification, Machine learning, Ensemble learning, Deep learning, MFCC, Audio features extraction, SMOTE, Data Imbalance**

# Introduction

Languages provide an ease for human to communicate their sentiments with each other. However, it is difficult for infants to express their feelings or discomfort due to lack of linguistic skill development. High pitched sound of screaming or crying is the primary means for kids to signal any unease or needs. Although parents remain most attentive, still sometimes it is difficult for them to know the exact reason for baby’s cry. In order to improve infant well-being as well as parent’s responsiveness to address infant needs promptly it is significantly important to understand and interpret cry of the baby. Parents and care givers traditionally rely on individual experiences and instincts to guess or understand the reason of baby cry, this may be inaccurate and inconsistent. Employing advanced machine learning techniques can help to offer a reliable solution, which can help to accurately detect the cry reason (Pradhan et al., 2022).

The automated cry classification system can help parents and care givers to understand exact reason for baby’s unease, significantly reducing response time to handle baby’s need, improving infant care and detecting potential health risk at early stage. These models can also be integrated with baby care instruments such as baby monitors which may lead to an intelligent tool in assisting parents. This study aims toexplore implementation of advanced machine learning techniques to differentiate the infant cries into classes : ‘hungry’, ’burping’, ’discomfort’, ‘belly-pain’ and ‘tired’. The existing work in this field shows variety of techniques to classify audio signals, including different feature extraction methods such as MFCC, Chrom feature extraction and Spectrograms, and machine learning models such as Support Vector Machine (SVM), Random Forest KNN, etc (Aggarwal et al., 2023; Riadi et al., 2024). Although deep learning methods specially those involving hybrid models and ensemble learning models are underexplored (Senthil et al., 2024; Narayanan et al., 2024). Moreover, the class imbalance issue in infant cry datasets remains inadequately addressed, leading to models that perform poorly when exposed to underrepresented classes. This gap motivates the need for a more comprehensive approach that not only leverages cutting-edge deep learning techniques but also addresses data imbalance issues.

The research question aiming in the study is : How can machine learning and deep learning models be effectively used to classify infant cries in different categories like hunger, discomfort and pain, especially when addressing class imbalance using technique like SMOTE. The proposed solution involves building and comparing different models from conventional machine learning classifiers to advanced deep learning models, and implementing Synthetic Minority Over-Sampling Technique (SMOTE) to balance the minority classes. This study targets to contribute to the field by providing an in-depth evaluation of these models and their robustness in noisy, real-world environments, eventually helping in developing effective infant cry monitoring systems.

This paper is organized as follows: Section 1 provides an overview of the topic, outlining the research problem, motivation and research question. In Section 2 a comprehensive review of related previous work is provided, which discusses audio classification and highlighting challenges and techniques used in infant cry analysis. Followed by Section 3 where research methodology is discussed including data collection, pre-processing, feature extraction, model training. Section 4, showcases the research findings providing evaluation metrics to perform comparative analysis of different models. Section 5 discusses the results, providing insights into model strengths, weaknesses, and the implications of using SMOTE. In Section 6 the conclusion and future work is discussed, which includes summary of the key findings, research outcome and outline for future work recommendations.

# Related Work

The related work on baby cry classification and analysis has explored a wide range of methods, including machine learning, deep learning, hybrid approaches, and IoT-based systems, each aiming to improve the identification and understanding of infant cries. This section presents a critical analysis of these works, highlighting their objectives, contributions, limitations, and their relevance to the current research on baby cry analysis.

**2.1 Machine Learning Approaches**

Several studies utilized traditional machine learning methods to classify baby cries. Aggarwal et al. (2023) compared Random Forest, SVM, and Decision Tree models using spectral features and found that SVM performed best in cry classification. Their study, however, was limited by a small dataset, affecting the generalizability of results. Riadi et al. (2024) also explored multiple machine learning models, including SVM and Random Forest, finding that Random Forest achieved the highest accuracy. Similarly, Mahmoud et al. (2020) proposed a semi-supervised K-Nearest Neighbor (SSKNN) approach to expand the training set using unlabeled data, achieving higher accuracy for hunger-related cries. While the detailed feature extraction in these studies was a strength, they were all constrained by dataset limitations, impacting their broader applicability. This similarity highlights a key challenge in the field: the availability of sufficiently diverse datasets for training and validating models. The focus of these studies on spectral features informs the current research by underlining the need to explore richer datasets to improve robustness in classification.

2.2 Deep Learning Techniques

Recent advancements in deep learning have been pivotal in addressing some of the limitations of traditional machine learning methods for baby cry analysis. Narayanan et al. (2024) utilized spectrograms with an LSTM-based model to enhance classification accuracy, achieving high sensitivity and specificity. The robustness of LSTM models in noisy environments was a notable improvement compared to the models used by Aggarwal et al. (2023). Similarly, Gülmez et al. (2024) employed CNNs for automatic cry classification, incorporating data augmentation to boost accuracy. Özseven (2024) provided a review of deep learning methods, noting the increasing adoption of these techniques compared to traditional approaches. Despite their promising results, deep learning studies often faced challenges with computational demands, highlighting the trade-off between model complexity and real-world feasibility. The effectiveness of LSTM and CNN in these studies suggests that deep learning can significantly improve classification performance, particularly when combined with techniques like data augmentation. However, the computational burden calls for exploring lighter architectures that maintain high accuracy, a focus area for the current research.

2.3 Hybrid and Transfer Learning Approaches

Hybrid deep learning approaches have also been investigated for baby cry detection. Senthil et al. (2024) combined CNN and LSTM techniques to classify emotional states of infants, achieving high accuracy. Reddy et al. (2024) also adopted a hybrid approach with CNN and LSTM, achieving a 99% accuracy rate in distinguishing baby cries from other noises. Similarly, Sharma & Malhotra (2020) introduced an Intelligent Infant Cry Classifier (IICC) that utilized CNN and Decision Trees, focusing on identifying reasons behind infant cries. While hybrid models often outperform single architectures by leveraging their complementary strengths, they can be computationally intensive, which was a common limitation observed in these studies. Transfer learning has also been explored as a means to address dataset limitations. Anjali et al. (2024) used VGG16 for transfer learning, achieving the highest accuracy among their compared models. Zhang et al. (2024) proposed a BCRNet model that combines transfer learning and feature fusion to mitigate overfitting issues. The strength of these methods was their ability to leverage pretrained networks to improve results despite limited data availability. However, the models' sensitivity to real-world noise conditions remains a challenge. The exploration of hybrid and transfer learning techniques in these works informs the current research by suggesting potential avenues for improving model accuracy without the need for extensive training data, albeit with attention to real-world performance considerations.

2.4 IoT and Real-World Implementations

Several studies have focused on practical implementations of baby cry detection systems, integrating IoT components to enhance applicability. Younis et al. (2024) employed Vision Transformers and CNNs, combined with IoT-enabled sensors, to facilitate real-time cry signal acquisition. The integration of IoT improved system responsiveness, although computational resource requirements posed a significant limitation for deployment in low-resource settings. Similarly, Kolandaisamy et al. (2024) developed a baby cry detection system using Raspberry Pi and wireless sensor networks, which was practically implemented in real-world monitoring scenarios. Visvesvaran et al. (2024) developed an IoT-based smart baby monitoring system with features such as automatic cradle swinging and email notifications to parents, emphasizing practical real-world deployment. While these studies have demonstrated the feasibility of integrating cry analysis with IoT technologies, they have also revealed the complexities involved in hardware integration and scalability. The insights from these implementations highlight the need for lightweight and scalable solutions that can be effectively used in residential environments, which aligns with the objectives of the current research project on practical cry detection.

2.5 Cry Detection in Challenging Environments

Several studies specifically addressed cry detection in challenging environments, focusing on background noise and real-world applicability. Nimbarte et al. (2023) and Mala & Darandale (2024) both highlighted issues related to handling background noise, a critical factor for practical deployment. Models often struggle with varying noise levels, which affects their accuracy and reliability. The frequent occurrence of false alarms, as noted by Narayanan et al. (2024), also limits the practicality of these models. Jamal & Al-Azani (2023) introduced a hybrid feature approach combining prosodic and spectral acoustic features, which significantly improved classification accuracy for pain-related cries, but faced scalability issues. Khandelwal et al. (2024) proposed a low-complexity Convolutional Recurrent Neural Network (CRNN) model that aimed to address computational efficiency while detecting baby cries in domestic environments, though with some trade-offs in accuracy. The use of augmented datasets, as demonstrated by Bella & Sanjaya (2024), shows potential in enhancing model robustness. This informs the current research by underscoring the importance of developing models capable of distinguishing cry types even under challenging acoustic conditions.

2.6 Pathological and Specialized Cry Detection

Some studies focused on detecting pathological cries or specific infant needs. Kumari & Mahto (2024) provided a narrative review of pathological cry detection techniques, emphasizing the scarcity of pathological datasets as a major challenge. You et al. (2023) utilized LSTM networks to classify multiple types of infant cries, expanding the scope to include awake, diaper change, hunger, sleepy, and discomfort cries. The inclusion of multiple cry types makes this approach more practical for real-world applications. The focus on pathological detection and multiple cry types informs the current research by suggesting the importance of expanding cry categories beyond general needs, which can enhance the practical utility of baby cry detection systems.

**Summary of Literature Review**

The reviewed literature reveals that machine learning and deep learning models have been extensively used for infant cry classification, with approaches ranging from classical methods like SVM and Random Forest to more advanced deep learning models like LSTM and Vision Transformers. Machine learning models are simpler and less computationally demanding, but they struggle with complex cry signals and background noise. On the other hand, deep learning models demonstrate higher accuracy but require large datasets and significant computational resources.

Hybrid approaches, which combine traditional machine learning and deep learning techniques, show promising results by leveraging the strengths of both methods. Techniques such as data augmentation, transfer learning, and hybrid models enhance accuracy but often come with increased computational demands and challenges in handling real-world noise. Notable gaps include the need for more robust models that can maintain high accuracy without excessive resource requirements and better generalizability to diverse and noisy environments.

This research aims to bridge these gaps by exploring more efficient hybrid models and noise-robust techniques to improve the practical applicability of infant cry classification systems in real-world settings.

# Research Method & Specification

This section provides an in-depth explanation of the research procedure, including the rationale behind chosen methods, data collection, techniques used, and evaluation methodology. The aim is to ensure clarity and transparency in the research process, demonstrating adherence to a scientific approach as highlighted in related works.

**1. Research Process**

This study follows a **quantitative research approach** designed to classify infant cries into categories such as hunger, discomfort, and pain. The research process is divided into multiple phases: data collection, data pre-processing, feature extraction, model selection, model training, and evaluation. Each phase has been implemented to address the research gaps identified in Section 2, specifically focusing on improving noise robustness and classification accuracy.

The research relied heavily on experimental simulations for model training and testing, which involved combining traditional machine learning with advanced deep learning models, inspired by related works (Aggarwal et al., 2023; Senthil et al., 2024; Narayanan et al., 2024).

**2. Evaluation Methodology**

Evaluation metrics and methods were chosen to address the shortcomings found in existing works. Cross-validation, particularly **5-fold cross-validation**, was used to validate the consistency and robustness of machine learning models. This was inspired by similar practices in related studies, which emphasize reducing the variability of model performance due to dataset size (Riadi et al., 2024).

To assess the effectiveness of deep learning models, metrics such as **accuracy, precision, recall and F1-score** were used. The choice of these metrics was based on their ability to provide comprehensive insights into model performance, particularly in imbalanced data settings (Bella & Sanjaya, 2024; Younis et al., 2024).

**3. Data Collection**

The data for this study was obtained from the **Kaggle dataset "Infant Cry Audio Corpus"** (https://www.kaggle.com/datasets/warcoder/infant-cry-audio-corpus). This dataset includes audio recordings of infant cries, classified into different categories. The class distribution was highly imbalanced: Counter({3: 310, 2: 20, 4: 17, 0: 12, 1: 6}). **Data augmentation** was used to expand the dataset, simulating real-life conditions and enhancing model training. **SMOTE (Synthetic Minority Over-sampling Technique)** was applied to the dataset to generate additional samples for the underrepresented classes (Kumari & Mahto, 2024; Zhang et al., 2024).

**4. Techniques and Equipment**

The following techniques and tools were used to implement this research:

* **Techniques**:
  + **Noise Reduction**: Used spectral gating to reduce background noise in audio recordings.
  + **Feature Extraction**: **MFCC** (Mel Frequency Cepstral Coefficients) and **Spectrograms** were extracted to capture the audio characteristics (Aggarwal et al., 2023; Nimbarte et al., 2023).
  + **SMOTE**: Applied to address class imbalance by generating synthetic samples.
* **Tools and Equipment**:
  + **Jupyter Notebook**: The primary environment for executing code, data visualization, and experimentation.
  + **Librosa**: For audio analysis, noise reduction, and feature extraction.
  + **Scikit-learn**: For implementing machine learning models, applying cross-validation, and generating evaluation metrics.
  + **TensorFlow/Keras**: For designing, training, and validating deep learning models.
  + **NumPy and Pandas**: For data manipulation and storage.
  + **Matplotlib and Seaborn**: For visualization of data and model performance.

The rationale for using these tools is that they are well-established in the field of machine learning and provide comprehensive functionality for audio analysis and model evaluation.

**5. Data Analysis**

**Data analysis** involved both statistical and machine learning techniques. The **raw audio data** was processed to extract meaningful features using MFCC and spectrograms. These features were then standardized and normalized to ensure uniformity across the dataset. Statistical analysis included calculating **mean, variance, and standard deviation**, which helped in understanding the variability in the dataset.

Machine learning models were trained using the extracted features. **Cross-validation** was employed to evaluate model consistency, while **confusion matrices** were used to examine misclassification patterns (Senthil et al., 2024). The deep learning models were trained using **categorical cross-entropy** as the loss function, and early stopping was implemented to prevent overfitting by halting training when the validation loss did not improve.

**6. Procedure**

The research procedure can be summarized in the following steps:

1. **Data Collection**: Audio recordings were sourced from the Kaggle dataset, which provided labeled data representing different types of baby cries. Data augmentation was applied to increase dataset diversity.
2. **Data Pre-Processing**: Noise reduction techniques, such as spectral gating, were applied to reduce unwanted background noise. The data was then normalized, and silent sections were removed.
3. **Feature Extraction**: MFCC and spectrograms were extracted as features from the pre-processed audio data using **Librosa**.
4. **Model Training**: Machine learning and deep learning models were trained using both the original and SMOTE-augmented datasets. **5-fold cross-validation** was implemented for traditional machine learning models to ensure robustness, while dropout regularization and batch normalization were used in deep learning models to reduce overfitting.
5. **Evaluation**: Model performance was evaluated using metrics such as **accuracy, precision, recall and F1-score**. Real-time testing was conducted with added background noise to simulate real-world conditions and assess model robustness.

The entire process ensured that a systematic and well-evaluated approach was employed to address the research question of effectively classifying infant cries using computational techniques. The evaluation metrics, model training approaches, and real-world testing phases were designed to provide a comprehensive understanding of model performance under various scenarios, thus contributing significantly to the field of infant cry analysis.

# 4. Design Specification

The design of the baby cry classification system involves the integration of various machine learning and deep learning models, supported by the data pre-processing, feature extraction, and augmentation techniques discussed previously (Aggarwal et al., 2023; Narayanan et al., 2024; Riadi et al., 2024). The underlying framework includes both traditional machine learning and advanced deep learning models, aimed at creating a robust classification system that can be effectively deployed for real-time applications.

**4.1 System Architecture**

The overall architecture of the system comprises several stages:

1. **Data Pre-Processing**: Audio data is first pre-processed using spectral gating for noise reduction, trimming silent sections, and normalizing the audio to ensure consistency across samples (Younis et al., 2024; Nimbarte et al., 2023). This ensures that the data fed into the models is clean and standardized.
2. **Feature Extraction**: Mel Frequency Cepstral Coefficients (MFCC) and spectrograms are extracted from the audio data (Aggarwal et al., 2023; Riadi et al., 2024). MFCC is particularly suitable for capturing characteristics of infant cries, while spectrograms are used to provide visual frequency distribution data, which deep learning models like Convolutional Neural Networks (CNNs) can effectively learn from.
3. **Modeling and Training**: The system leverages a combination of the following models (Senthil et al., 2024; Zhang et al., 2024; Bella & Sanjaya, 2024):
   * **Support Vector Machine (SVM)** for linear classification.
   * **K-Nearest Neighbors (KNN)** for proximity-based classification.
   * **Random Forest, AdaBoost, and XGBoost** for ensemble-based classification, which helps reduce bias and improve model robustness.
   * **Deep Learning Models** such as CNN, Long Short-Term Memory (LSTM), and Recurrent Neural Network (RNN) are employed for feature learning from spectrograms and time-series patterns in the audio data.
   * **Hybrid and Ensemble Models**: CNN-XGBoost, CNN-LSTM, and an ensemble of all models (CNN, Random Forest, XGBoost, SVM, KNN) are also included to explore potential performance improvements through model combination.
4. **Data Augmentation and Balancing**: Given the significant class imbalance present in the dataset, Synthetic Minority Over-sampling Technique (SMOTE) is used to generate synthetic samples for underrepresented classes (Kumari & Mahto, 2024; Zhang et al., 2024). This helps in training the models on balanced data and reducing the bias towards majority classes.

**4.2 Requirements**

**4.2.1 Functional Requirements**

* **Accurate Classification**: The system must classify infant cries into various categories such as hunger, discomfort, and pain with high accuracy.
* **Noise Robustness**: The system must be capable of functioning effectively even in noisy environments, such as households with typical background noise.
* **Real-Time Processing**: The system should be capable of performing real-time classification to assist caregivers and monitoring systems.

**4.2.2 Non-Functional Requirements**

* **Scalability**: The system must be scalable to accommodate large datasets and handle real-time data input for practical applications.
* **Resource Efficiency**: The models should be optimized for resource efficiency, particularly in terms of computational power, to enable deployment on edge devices or embedded systems.
* **Model Interpretability**: The system should provide insights into the decision-making process of the model, especially in situations where the classification might have significant implications for infant care.

**4.3 Model Functionality**

The proposed hybrid models, such as CNN-LSTM and CNN-XGBoost, combine the strength of both CNNs for feature extraction from spectrograms and LSTMs for temporal sequence learning (Zhang et al., 2024; Bella & Sanjaya, 2024). The CNN is responsible for identifying relevant features in the audio frequency distribution, while the LSTM or XGBoost adds predictive power by learning temporal dependencies or combining multiple weak learners for better generalization.

The **Ensemble Model** brings together the strengths of different algorithms to provide a robust and generalized classifier (Senthil et al., 2024; Reddy et al., 2024). By combining multiple models such as CNN, Random Forest, SVM, and KNN, the ensemble aims to minimize the individual weaknesses of each model, thereby improving overall accuracy and stability.

The overall goal of the design is to create a reliable, robust, and accurate classification system that can be utilized for monitoring infants and aiding caregivers in understanding their needs in real-time, especially in environments that may have various background noises and distractions.

**5. Implementation**

The implementation of the proposed baby cry classification system involved the final stage of integrating the data pre-processing, feature extraction, model training, and evaluation into a functional pipeline. This section outlines the key outputs, models developed, and tools used during the implementation.

**5.1 Outputs Produced**

* **Transformed Data**: The raw audio data underwent pre-processing, which included noise reduction, normalization, and silent trimming, resulting in clean audio segments suitable for analysis. The processed data was transformed into feature representations, specifically Mel Frequency Cepstral Coefficients (MFCC) and spectrograms, which were then used for training the models.
* **Feature Extraction Outputs**: MFCC features and spectrogram images were extracted from each audio recording. These features were subsequently used for both machine learning and deep learning models. Spectrograms provided a visual frequency representation, ideal for deep learning models such as CNNs.
* **Models Developed**: Various machine learning and deep learning models were implemented, including Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Random Forest, AdaBoost, XGBoost, Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and Convolutional Neural Network (CNN). Additionally, hybrid models like CNN-XGBoost and CNN-LSTM, as well as an ensemble model combining CNN, Random Forest, XGBoost, SVM, and KNN, were developed to improve classification accuracy.
* **Class Balancing Outputs**: To address the class imbalance issue, SMOTE was applied to the training data, producing balanced datasets that enhanced the performance of the models. This resulted in improved accuracy and robustness in detecting minority class cries, such as 'burping' and 'belly\_pain'.
* **Evaluation Metrics**: The outputs included detailed performance metrics for each model, such as accuracy, precision, recall, F1-score, and Receiver Operating Characteristic (ROC) curves. Confusion matrices were also generated to provide insights into the types of errors made by the models.

**5.2 Tools and Languages Used**

* **Programming Language**: The implementation was done using **Python**, which provided a versatile environment for data processing, model training, and evaluation.
* **Development Environment**: **Jupyter Notebook** was used for interactive coding, allowing for efficient experimentation, visualization, and debugging of the code.
* **Libraries and Frameworks**:
  + **Librosa**: Used for audio analysis, including noise reduction and feature extraction (MFCC and spectrograms).
  + **Scikit-learn**: Utilized for implementing machine learning models (SVM, Random Forest, KNN, etc.), applying cross-validation, generating evaluation metrics, and using SMOTE for data balancing.
  + **TensorFlow/Keras**: Used for designing and training deep learning models, including CNNs, LSTMs, and hybrid models.
  + **NumPy and Pandas**: Employed for data manipulation, storage, and processing.
  + **Matplotlib and Seaborn**: Utilized for data visualization during Exploratory Data Analysis (EDA) and for plotting model performance metrics such as accuracy curves and confusion matrices.

The final implementation brought together various models into a cohesive system that could accurately classify infant cries based on different categories, such as hunger, discomfort, and pain. The combination of traditional machine learning and advanced deep learning models, along with hybrid and ensemble approaches, ensured that the classification system was both accurate and robust, capable of handling real-world noise and variability effectively.

**6. Evaluation**

The evaluation of the baby cry classification system focused on a comprehensive analysis of the models' performance, examining accuracy, F1-score, precision, recall, and ROC curves for all implemented models. The findings are presented with a focus on their implications for both academic and practical use, supported by statistical analysis and visual representations (Aggarwal et al., 2023; Narayanan et al., 2024; Younis et al., 2024).

**6.1 Key Evaluation Metrics and Results**

The following key metrics were used to evaluate the models:

* **Accuracy**: Represents the ratio of correctly predicted instances over the total number of instances. It provides an overall performance view but can be misleading in the case of imbalanced data.
* **Precision**: Indicates the ratio of true positive predictions to the sum of true positive and false positive predictions. It reflects how well the model identifies relevant results.
* **Recall**: Represents the ratio of true positive predictions to the sum of true positive and false negative predictions, providing insight into the model's ability to capture all relevant instances.
* **F1-Score**: Combines precision and recall to provide a balanced evaluation metric, particularly useful for imbalanced datasets.

**6.2 Model Comparison and Analysis**

The performance results of each model are summarized in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Accuracy** | **F1 Score** | **Precision** | **Recall** |
| SVM (5-fold CV) | 0.767 | 0.749 | 0.732 | 0.767 |
| KNN (5-fold CV) | 0.838 | 0.775 | 0.720 | 0.838 |
| Random Forest (5-fold CV) | 0.849 | 0.780 | 0.721 | 0.849 |
| AdaBoost (5-fold CV) | 0.759 | 0.736 | 0.739 | 0.759 |
| XGBoost (5-fold CV) | 0.844 | 0.777 | 0.721 | 0.844 |
| LSTM | 0.790 | 0.741 | 0.699 | 0.790 |
| RNN | 0.826 | 0.757 | 0.699 | 0.826 |
| CNN | 0.793 | 0.741 | 0.718 | 0.793 |
| CNN-XGBoost | 0.793 | 0.741 | 0.718 | 0.793 |
| CNN-LSTM | 0.783 | 0.687 | 0.612 | 0.783 |
| Ensemble Model (CNN, RF, XGB, SVM, KNN) | 0.728 | 0.670 | 0.620 | 0.728 |
| SVM (SMOTE) | 0.977 | 0.976 | 0.978 | 0.977 |
| KNN (SMOTE) | 0.844 | 0.803 | 0.870 | 0.844 |
| Random Forest (SMOTE) | 0.996 | 0.996 | 0.996 | 0.996 |
| AdaBoost (SMOTE) | 0.294 | 0.208 | 0.259 | 0.294 |
| XGBoost (SMOTE) | 0.989 | 0.989 | 0.989 | 0.989 |
| LSTM (SMOTE) | 0.754 | 0.731 | 0.710 | 0.754 |
| RNN (SMOTE) | 0.717 | 0.702 | 0.688 | 0.717 |
| CNN (SMOTE) | 0.685 | 0.658 | 0.640 | 0.685 |
| CNN-XGBoost (SMOTE) | 0.983 | 0.983 | 0.984 | 0.983 |
| CNN-LSTM (SMOTE) | 0.370 | 0.433 | 0.571 | 0.370 |
| Ensemble Model (CNN, RF, XGB, SVM, KNN, SMOTE) | 0.707 | 0.684 | 0.679 | 0.707 |

**6.3 Analysis of Results**

* **Handling Class Imbalance**: Models trained on SMOTE-augmented datasets, such as **Random Forest (SMOTE)** and **XGBoost (SMOTE)**, showed significantly higher accuracy and F1-scores compared to models trained on the original imbalanced dataset. This highlights the importance of addressing class imbalance to improve model performance, especially in detecting minority cry types like 'burping' and 'belly\_pain' (Kumari & Mahto, 2024; Zhang et al., 2024).
* **Deep Learning vs Machine Learning Models**: Deep learning models like **LSTM** and **CNN** demonstrated moderate performance, with accuracies between 75-80%. However, hybrid models like **CNN-XGBoost** achieved better results, indicating that combining feature extraction capabilities with ensemble learning methods can yield improvements in classification accuracy (Senthil et al., 2024; Zhang et al., 2024).
* **Model Robustness**: The **Ensemble Model** combining various classifiers, including **CNN, Random Forest, XGBoost, SVM, and KNN**, provided more robust results but did not outperform individual models trained with SMOTE. This suggests that while ensemble models reduce overfitting and generalize better, addressing data imbalance directly yields the most benefit (Reddy et al., 2024; Bella & Sanjaya, 2024).

**6.4 Visual Representation of Results**

To better understand the model performance, several visual aids were generated:

1. **Confusion Matrices**: Confusion matrices were created for each model to visually assess classification performance. These matrices helped identify which classes were most frequently misclassified and provided insight into areas where the models could be improved.
2. **ROC Curves**: The ROC curves for the top-performing models were plotted to show the trade-off between the true positive rate and false positive rate at various threshold settings.
3. **Performance Comparison**: A bar chart was created to compare the accuracy of all models, clearly showing the improvement in models trained on SMOTE-augmented datasets.

**6.5 Implications of Findings**

* **Academic Perspective**: The results demonstrate that hybrid models and ensemble techniques can effectively improve the classification of baby cries when combined with data augmentation methods like SMOTE. This contributes to the body of knowledge by showing that noise robustness and model generalizability can be significantly enhanced through careful dataset balancing (Aggarwal et al., 2023; Narayanan et al., 2024).
* **Practical Perspective**: From a practical standpoint, the implementation of SMOTE significantly improved the detection of minority classes, making the system more reliable for real-world applications (Jindal et al., 2024; Mala & Darandale, 2024). This means that caregivers using such systems can expect more consistent performance across different cry types, even those that are less common.

The evaluation demonstrates that addressing dataset imbalance, leveraging hybrid models, and employing deep learning techniques can lead to a robust and reliable baby cry classification system. The insights provided can guide future research and practical implementations in the field of infant care and monitoring systems.

**7. Conclusion and Future Work**

This study aimed to develop an effective baby cry classification system capable of distinguishing between different types of infant cries, using various machine learning and deep learning models. The research question addressed was: How can the classification of infant cries be improved to provide accurate and reliable information for caregivers?

The key objectives were to explore different models, tackle the challenges of class imbalance, and enhance classification accuracy through effective feature extraction techniques and oversampling methods. The experimental work successfully demonstrated that models trained on SMOTE-augmented datasets, particularly Random Forest and XGBoost, performed well, achieving high accuracy, precision, recall, and F1 scores.

Key findings of this study indicate that handling data imbalance using SMOTE significantly improves model performance, and hybrid approaches involving CNN and XGBoost yield superior results compared to standalone models. However, the study also highlighted the limitations associated with computational costs and challenges in processing real-time cry signals, particularly for deep learning models such as LSTM.

**Future Work**: There are several avenues for future research that could further advance the baby cry classification domain. One potential direction is to explore the use of transformer-based architectures, which have shown considerable success in sequence modeling tasks and could be well-suited for cry classification. Additionally, a more extensive dataset, possibly including pathological cry types, could be used to increase the robustness and applicability of the models in real-world scenarios (Kumari & Mahto, 2024). Finally, developing an edge-compatible version of the classification system, allowing for real-time processing on low-power devices, could help bridge the gap between research and practical, commercial applications.