Infant Cry Classification: Data Processing, Training, and Evaluation Report

Ganesh Islavath

February 25, 2025

1 Introduction

Infant cry classification is crucial for early detection of discomfort or medical conditions in newborns. This report details the data processing pipeline, feature extraction, model training, and evaluation of different classification models, including XGBoost, Wav2Vec2, and multiple ensemble learning methods.

2 Data Processing

2.1 Preprocessing

We applied noise reduction, silence trimming, resampling (16kHz), and segmentation (5-15s) to ensure uniform input.

2.2 Feature Extraction

Two feature extraction methods were used:

- **YAMNet** - Extracted 1024-dimensional embeddings from Google's YAMNet model.
- **Wav2Vec2** - Extracted contextualized speech representations from Facebook's Wav2Vec2 model.

3 Model Training

3.1 XGBoost

Trained on YAMNet features with 150 trees, a learning rate of 0.05, and early stopping at 10 rounds.

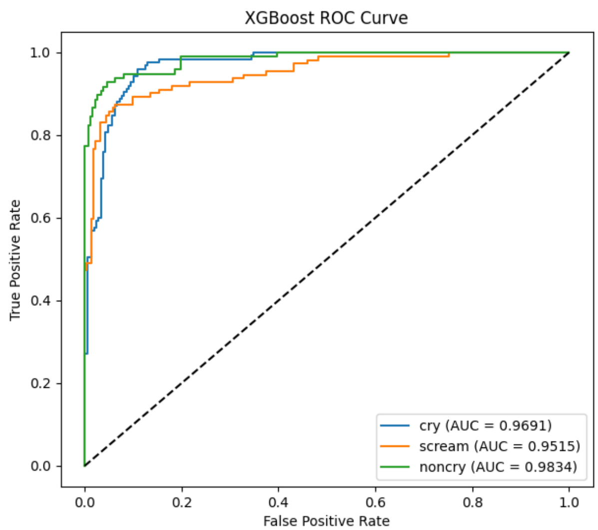

Figure 1: XGBoost training loss curve.

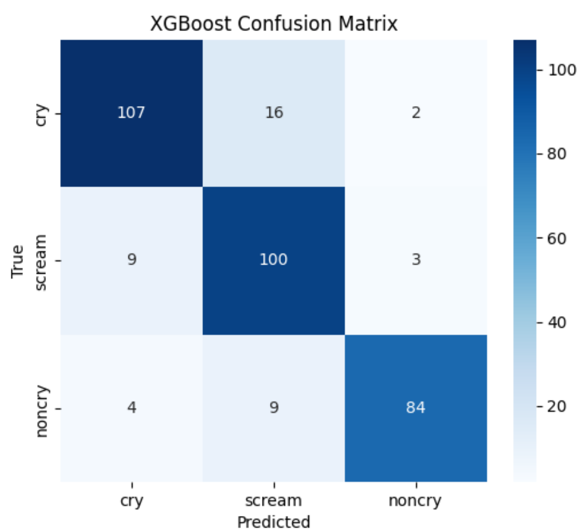

Figure 2: XGBoost training confusion Matrix.

3.2 Wav2Vec2 Fine-Tuning

Wav2Vec2 was fine-tuned using a batch size of 8 for 8 epochs with AdamW optimizer.

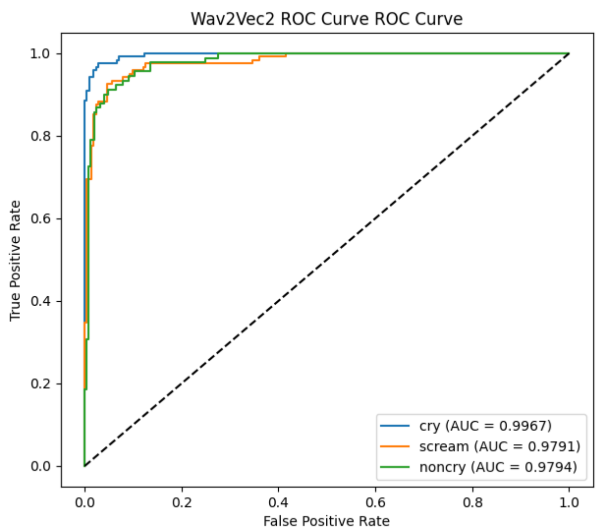

Figure 3: Wav2Vec2 training loss curve.

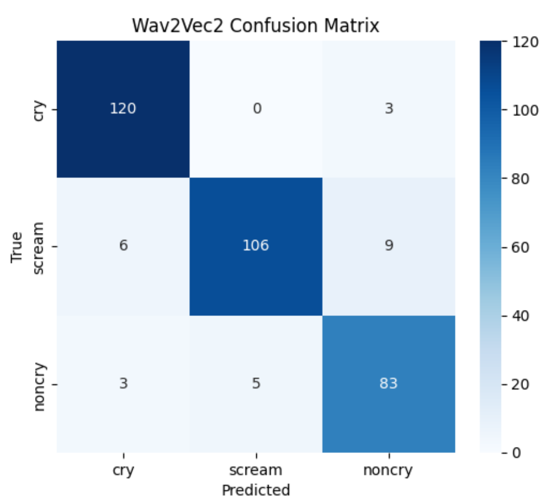

Figure 4: Wav2vec2 Confussion Matrix.

3.3 Ensemble Models

Four ensemble methods were tested:

- **Platt Scaling** - Logistic regression-based calibration.
- **Stacking** - Combines Wav2Vec2 and XGBoost using a meta-classifier.
- **Majority Voting** - Chooses the most frequently predicted class.
- **Weighted Ensemble** - A weighted combination of Wav2Vec2 and XGBoost (70% and 30% respectively).

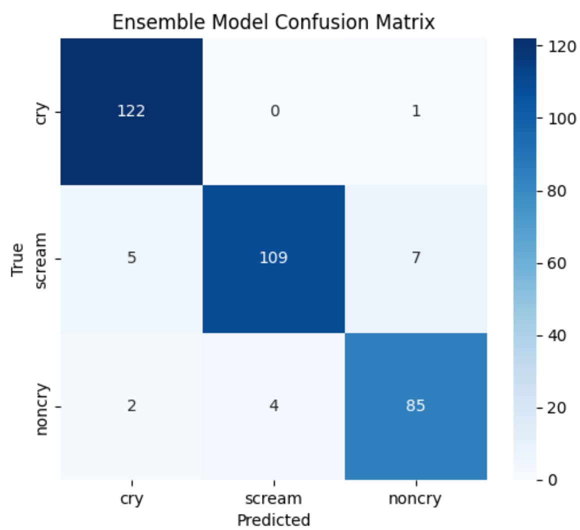

Figure 5: Weighted Ensemble Confusion Matrix.

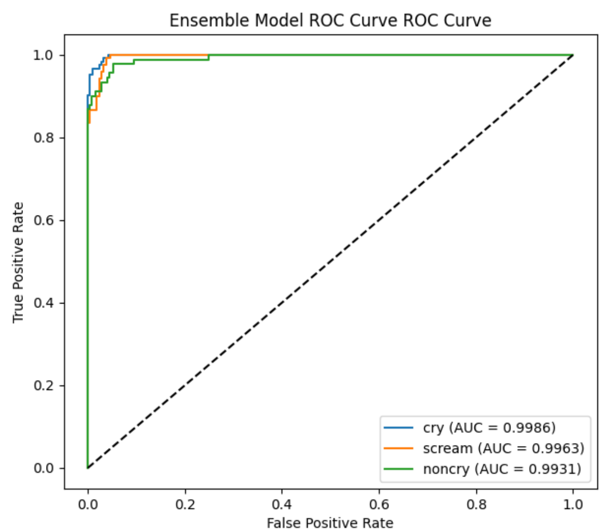

Figure 6: Weighted Ensemble ROC.

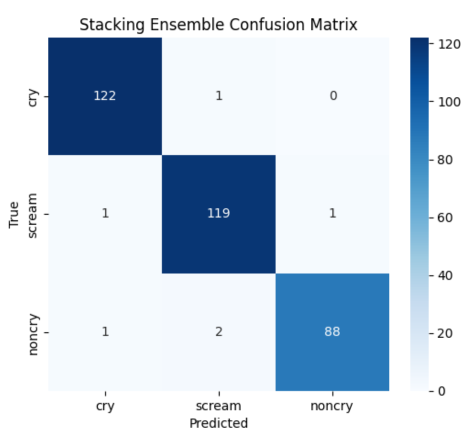

Figure 7: Stacking Ensemble Confusion Matrix.

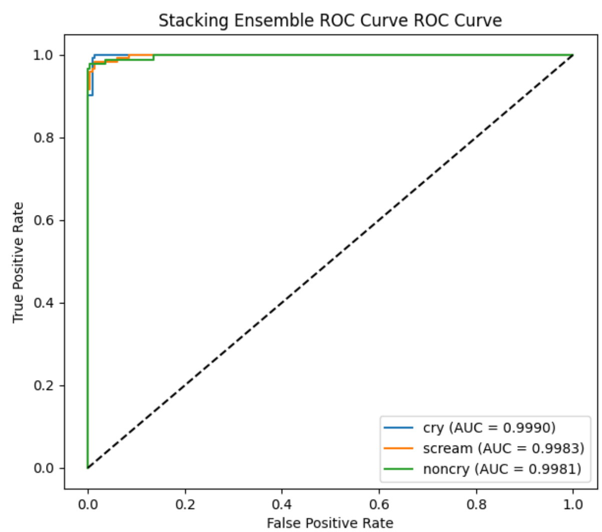

Figure 8: Stacking Ensemble ROC.

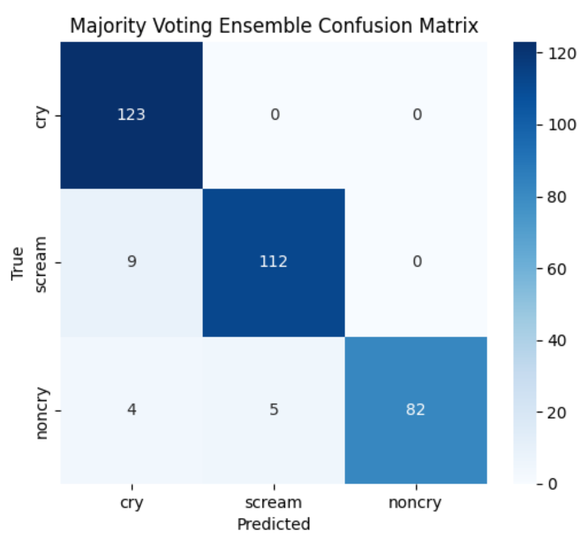

Figure 9: Majority Voting Ensemble Confusion Matrix.

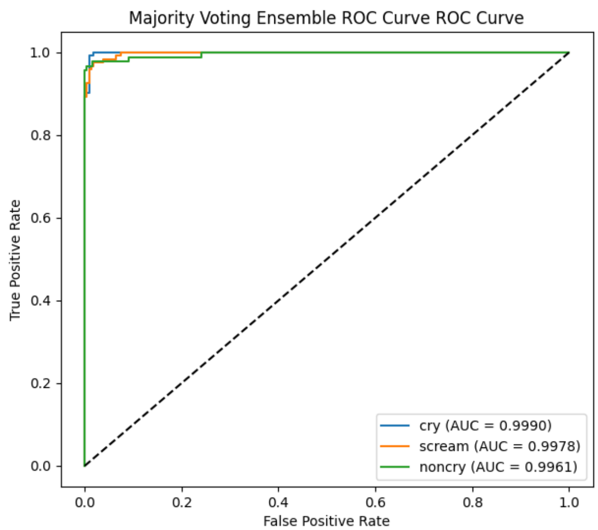

Figure 10: Majority Voting Ensemble ROC.

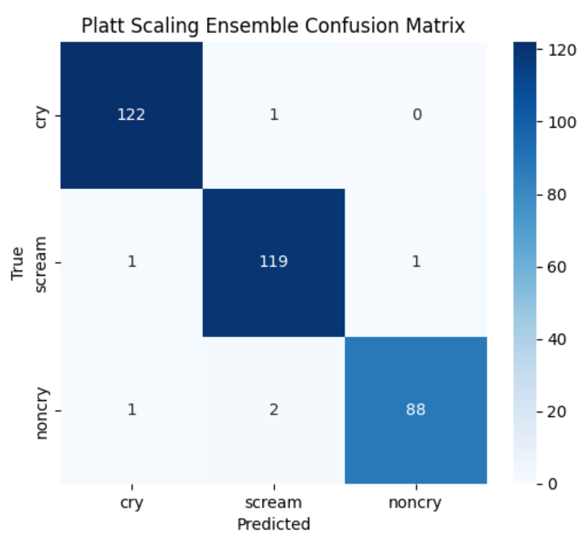

Figure 11: Platt Scaling Ensemble Confusion Matrix.

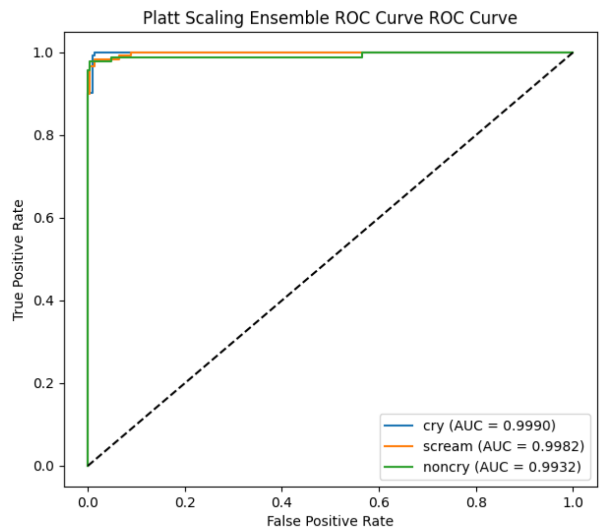

Figure 12: Platt Scaling Ensemble ROC.

4 Conclusion

This report presents a comprehensive approach to infant cry classification, integrating multiple models and inference methods. Our results demonstrate the feasibility of deep learning and ensemble techniques in real-world applications. Future work will focus on implementing real-time ensemble predictions and improving the adaptability of the models.