2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS), 8-9 March, Dhaka, Bangladesh

2024 International Conference on Advances in Computing, Communication, Electrical, and Smart Systems (iCACCESS) | 979-8-3503-5028-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICACCESS61735.2024.10499449

From Cries to Answers: A Comprehensive CNN+DNN Hybrid Model for Infant Cry Classification with Enhanced Data Augmentation and Feature Extraction

Khorshed Alam
Dept. of CSE
United International University
Dhaka, Bangladesh
mohdkhurshed120@gmail.com

Khondaker A. Mamun
Computer Science & Engineering
United International University
Dhaka, Bangladesh
mamun@cse.uiu.ac.bd

***Abstract*— Infant cry classification involves utilizing machine learning techniques to automatically recognize and categorize various types of infant cries. The goal of infant cry classification using deep learning is to develop a robust and accurate system capable of automatically recognizing and categorizing the different types of infant cries. This technology can potentially assist healthcare professionals or caregivers in identifying the underlying reasons for the crying, such as hunger, discomfort, pain, etc., aiding in prompt and appropriate responses to the infant's needs. This paper presents an innovative approach to infant cry classification using deep learning technique, data augmentation and feature extraction methods. Our primary model is based on popular infant cry classification dataset named Donateacry and Infants Cry Sounds corpus. We use data augmentation techniques like noise, stretch, shift, pitch, time masking and frequency masking to enhance generalizability of our model. Then we use SMOTE to balance all the classes of dataset. Furthermore, we use feature extraction techniques like Chroma STFT , Root Mean Square, Mel-frequency Cepstral Coefficients (MFCC), and Zero Crossing Rate to capture distinctive patterns in children's cry signals. To validate our model, we employed k-fold cross-validation, yielding a mean accuracy of 0.9384 and a mean loss of 0.0652.**

Keywords*— *Infant cry, MFCC, healthcare, Infant cry classification, Infant cry detection, Cry Classification

I. INTRODUCTION

Infants predominantly express themselves through crying, a sophisticated and nuanced form of communication that parents and guardians try to understand. These screams often represent a variety of demands or feelings, from hunger to uneasiness, and vary in strength, length, and pitch [1]. The motivation for the study likely arises from a need to improve pain assessment in pediatric care. Since children, especially those who are pre-verbal or have communication difficulties, cannot always express their pain, assessing pain levels accurately can be challenging for healthcare providers. It is, however, difficult for new parents, as well as untrained clinicians and adults, to understand infant cries. Understanding them correctly is important so we can help the baby in the right way. Consequently, the crucial task of discerning cries with specific meanings relies on the identification of cry-related audio characteristics [2,3].

Artificial Intelligence (AI) provides effective solutions in various real-world domains, including healthcare, agriculture, emotion analysis, and manufacturing [4-7]. It involves creating computer systems that can perform tasks requiring human-like intelligence. Because of this, there is a rising interest in using technology like advanced machine learning models to help accurately detect infant cries, which will then enable parents to provide timely and appropriate treatment. This method has the potential to improve babies' general comfort and well-being while also offering parents and medical professionals valuable support.

Machine learning and deep learning techniques have significantly advanced cry classification research in baby care and health. Despite these strides, a common constraint persisted in the field. Most of the benchmark dataset contains noise free audios, which is not always a suitable scenario in real life. In real life there can be present of background noises. Most of the previous studies did not consider working with noisy audio datasets. Data gathering is also one of the most difficult aspects of infant cry analysis.

As the recording of cry sounds is no longer a standard practice in medical records, the establishment of a cry sound database requires dedicated initiatives, typically through clinical research. Such endeavors incur substantial costs and often demand the collaboration of various hospital personnel over an extended period, spanning several years. In this study we collect the primary dataset from here [8,9]. We propose a CNN+DNN hybrid model which can classify hunger, pain, eructation, tiredness, discomfort, colic, and pathology from infant's cry data. The contributions of this study:

- In real life audio data may have background noises. In order to make model more generalized with real life scenario, we use data augmentation techniques like noise, stretch, shift, pitch, time masking and frequency masking which is unseen in previous studies.
- In order to extract the most value from the data, we use feature extraction techniques like Chroma STFT , Root Mean Square, Mel-frequency Cepstral Coefficients

979-8-3503-5028-9/24/$31.00 ©2024 IEEE

(MFCC), and Zero Crossing Rate to capture distinctive patterns in children's cry signals.

- We propose a custom CNN and DNN based hybrid model for classifying the infant's cry sounds.
- We use k-fold cross validation to validate our model and obtained a mean accuracy of 0.9299 and a mean loss of 0.0760 which ensures the robustness of the model.

II. Related Work

Since the 1960s, researchers have explored the link between an infant's cry characteristics and various health conditions or needs. Authors from [10] use k-nearest neighbors (KNN)-based model for infant cry classification. The highest accuracy achieves up to 85%. However, Dataset size is too small and having 85% on small dataset is not enough to make conclusion. Comparison with other state of art machine learning methods are missing.

Authors [11] conducted research on the classification of Infant's Cry. The authors used Support Vector Machine (SVM) for hierarchical classification. In terms of feature extraction, the authors convert the audio file into a spectrogram, emphasizing visual characteristics. To address data imbalance, they incorporate SMOTE (Synthetic Minority Over-sampling Technique). The study focuses on distinguishing between two classes: pain and no pain. Notably, the precision scores for the respective classes are 61.12% and 67.39%, while the recall scores are 26.19% and 90.14%.

In the [12] study, it focuses Self-Supervised Learning for Infant Cry Analysis. The study demonstrates that large-scale CNN based Semi-Supervised Learning (SSL) for infant cry analysis. The highest AUC score 75.6% achieve by their CNN-SSL model. While SSL shows promise, it may require careful tuning and validation to achieve optimal results. Comparison with other benchmark approach and dataset are missing. In the study referenced as [13], the authors employ Mel Frequency Cepstral Coefficients (MFCC) for extracting features, and they apply Support Vector Machine (SVM), Decision Tree (DT), and Naïve Bayes (NB) classifiers to categorize infant cries. Highest accuracy achieved by SVM classifier with 71%. However, Dataset size is too small which contains only 113 audio files and having 71% on small dataset is not enough to make conclusion.

Authors of [14], used various feature extraction techniques, including Linear Predictive Cepstral Coefficients (LPCC), Linear Predictive Coding (LPC), Bark Frequency Cepstral Coefficients (BFCC), and Mel Frequency Cepstral Coefficients (MFCC), are employed to capture distinctive characteristics from the data. Furthermore, they use Nearest Neighborhood (NN) and Artificial Neural Network (ANN) for classification. The paper mentions using practical data from hospitals, but the dataset's size and diversity may be limited. A larger and more diverse dataset could improve model generalization. A transfer learning based model is proposed by [15] using CNN+lstm, CNN-ANN, VGG16, CNN, MI+SVM, ANOVA+RF based architecture and MFCCs, Linear Frequency Cepstral coefficients (LFCC), and Bark Frequency Cepstral Coefficients (BFCC)-13 based feature extraction methods for infant cry classification. However, Model performance can be low in noisy data as they did not mention about any Data Augmentation process to add noise.

To the extent of the authors' comprehension, we studied a few more papers in the domain of speech recognition [16, 19] and saw usage of data augmentations and feature extractions for categorizing audio data into distinct categories [17,20]. After a proper literature review, the proposed approach acknowledges the importance of the presence of background noise, simulating scenarios where the beginning of the cry might not be captured or is less audible, and pitch changes in real-life audio data. To enhance the model's adaptability to these scenarios, data augmentation techniques including noise, stretch, shift, pitch, time masking, and frequency masking are used, setting this study apart from previous research. Furthermore, the study optimizes data utilization by using advanced feature extraction methods like Chroma STFT, ZCR, RMS and MFCC. These techniques enable the model to effectively capture unique patterns in children's cry signals, contributing to a more accurate and robust classification system.

III. Methodology

This is the research methodology we use to classify infant cries in order to assess children's suffering. Our approach is broken down into three sections:

- Data Preprocessing: In this phase, we collect infant cry classification speech data from the sources referenced in papers [8,9]. Subsequently, we construct a data frame that includes audio paths and their corresponding classes, such as belly pain, cry, discomfort, hungry, laugh, noise, silence, and burping. Additionally, prior to feature extraction, we use data augmentation algorithms to enhance the diversity and richness of the dataset. This augmentation step aims to improve the model's performance and generalization by introducing variations in the training data.
- Feature Extraction: In this segment, we maximize the utility of our dataset by using feature extraction techniques like ZCR, MFCC, Chroma STFT and RMS to capture distinctive patterns in children's cry signals before jumping into model development.
- CNN+DNN Model Development : In this segment, we formulate a hybrid model for Infant Cry classification by combining CNN and DNN components, depicted in Figure 1. Getting high accuracy on a specific dataset does not ensure robustness. In order to prove robustness of the model, we use k-fold cross validation.

A. Data Preprocessing

We use donate-a-cry-corpus-features-dataset [8] and Infants cry sounds dataset [9] for this work. After merging these dataset we get the infant cry sound types belly_pain', cry, discomfort, Hungry, laugh, noise, silence, burping. In Fig. 2, we can see the data distribution of each class after merging these dataset. After collecting the data, to increase the amount and variety of training available for our model, some data augmentations are done.

Post data collection, a variety of data augmentation techniques are implemented to augment both the quantity and variety of the training data available for model training. Instead of memorizing specific examples, we augment training data with more randomness and non-linearity to enable our model to learn the complex features needed to identify the context of an infant's cry from speech data. Instead

of evaluating weather learning's efficacy, this enables us to determine the machine's intelligence by analyzing its performance on unknown data. Performing effectively with known data is often linked to verifying weather learning perfections. However, we aim to create an intelligent model that can demonstrate its effectiveness using real-world data that is unseen. Real-world audio scenarios may include noise and other elements. We use data augmentation techniques like noise, stretch, shift, pitch, time masking and frequency masking to enhance generalizability of our model. Noise Injection adds random noise to the signal. Infants may cry in various environments with different background noises shown in Fig 3. Adding noise helps the model become more robust to such variations. Stretch alters the time duration of the cry.

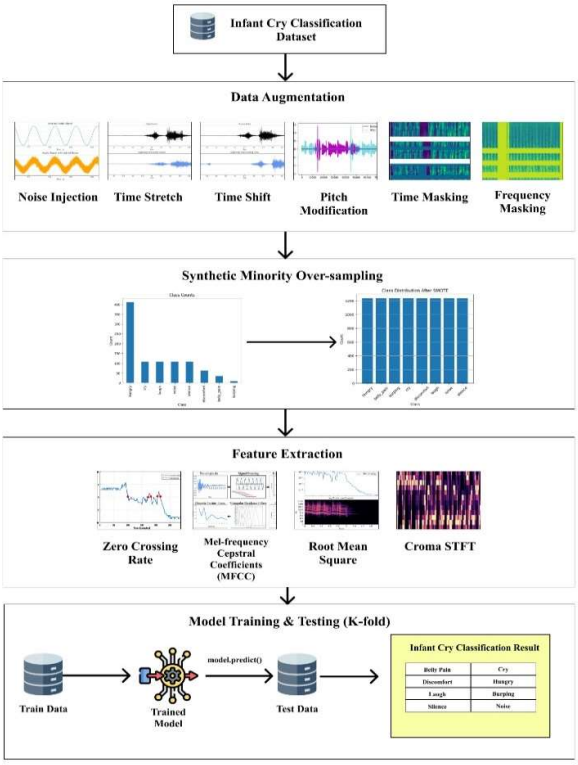

Fig. 1 Proposed Methodology

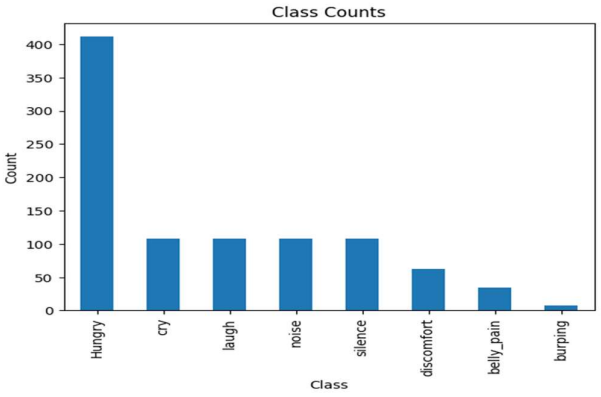

Fig. 2 Data Distribution of each Class in Merged Dataset

Infant cries may vary in length. Stretching or compressing the duration helps the model learn to classify cries of different lengths shown in Fig 4. Shifts the signal in time. This could represent changes in the intensity or onset of the cry shown in Fig 5. It helps the model learn to recognize cries irrespective of their starting point. Pitch modification modifies the pitch of the cry. Infants' cries can differ in pitch, and this augmentation simulates those variations to make the model invariant to pitch changes shown in Fig 6.

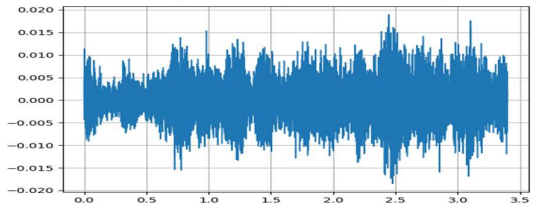

Fig. 3 Noise Injection

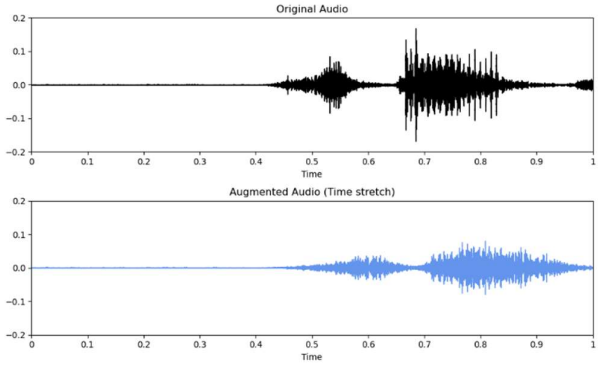

Fig. 4 Time Stretch

Time Masking involves concealing a segment of the signal within the temporal domain. This could simulate situations where certain parts of the cry are not captured or are less audible. It helps the model focus on the relevant segments for classification. Frequency Masking entails obscuring a segment of the signal within the frequency domain, as illustrated in Figure 7. Similar to time masking, this alteration helps the model learn to focus on relevant frequency components for classification while being invariant to certain frequency ranges.

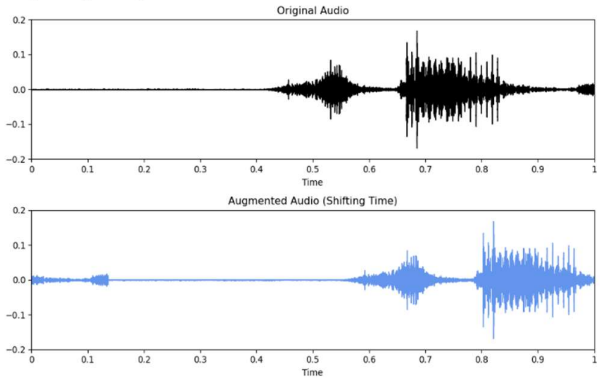

Fig. 5 Time Shift

Figure 2 illustrates a highly imbalanced distribution of data. To address this issue, we use the Synthetic Minority Over-sampling Technique (SMOTE), as depicted in Figure 8. SMOTE generates synthetic samples by considering feature similarity, thereby introducing new information to the dataset. This technique proves effective in mitigating the

impact of imbalanced data, aiding in preventing overfitting to some extent. The resulting data distribution, as demonstrated in Figure 8, reflects a more balanced representation of each class post the application of SMOTE.

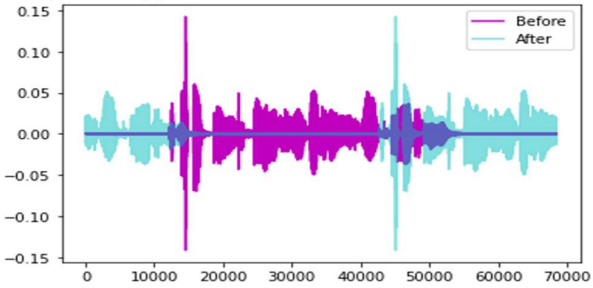

Fig. 6 Pitch Modification

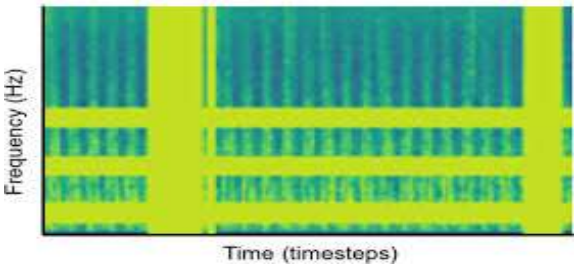

Fig. 7 Time & frequency Masking

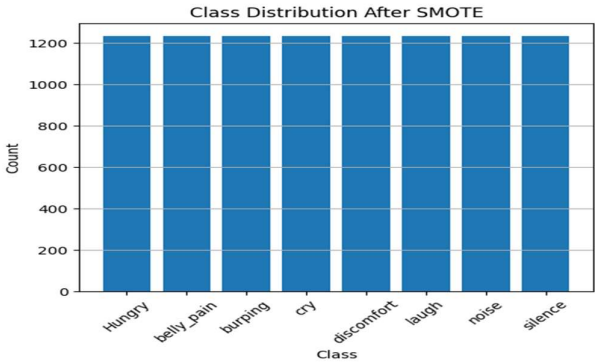

Fig. 8 Data distribution of each class after using SMOTE

B. *Feature Extraction*

To maximize the utilization of the available data, our feature extraction process incorporates Mel-frequency Cepstral Coefficient (MFCC) [18], Zero Crossing Rate (ZCR), Chroma STFT, and Root Mean Square. ZCR quantifies the rate at which the audio signal undergoes sign changes, essentially counting the occurrences of signal crossings through zero within a given frame. This set of features aims to capture diverse aspects of the audio signal, enhancing the model's ability to discern key patterns and characteristics. Reason to choose this feature extraction method is the infants' cries can exhibit varied tonal changes. ZCR captures abrupt changes or transitions in the cry signal, which can be indicative of different cry types. Chroma features capture the distribution of energy in different pitch classes.

Chroma STFT is designed to specifically compute chroma features derived from the short-time Fourier transform of the audio signal. The reason is it characterizes the harmonic content of the cry. Changes in pitch and tonal qualities in infant cries can provide cues about the underlying cause or nature of the cry, aiding in classification. MFCCs serve as a representation of the short-term power spectrum of a sound. These coefficients are obtained through the process of taking the Fourier transform of the logarithm of the power spectrum of the signal. Reason to choose this feature extraction method is the MFCCs capture both spectral and temporal characteristics of the cry signal. These coefficients are widely used in speech and audio processing due to their ability to represent the unique aspects of different sounds, which can be essential for distinguishing various cry patterns. RMS measures the average power of the signal. It calculates the square root of the average of squared values of the signal within a frame. The rationale of choose this technique, it quantifies the overall amplitude or energy of the cry signal. Variations in RMS values might indicate changes in the intensity or loudness of the cry, which could be indicative of different emotional states or needs of the infant.

C. *CNN+DNN Model Development*

We introduce a novel hybrid model, incorporating both CNN and DNN layers, designed for the classification of infant cry sounds. This unique model architecture synergizes CNN layers with DNN layers, as outlined in Table I, to effectively discern and categorize infant cries from speech data.. Let's break it down and understand its rationale:

TABLE I. MODEL ARCHITECTURE

Layer (type)	Output Shape	Param
conv1d_166 (Conv1D)	(None, 160, 16)	64
max_pooling1d_162 (MaxPooling1D)	(None, 80, 16)	0
conv1d_167 (Conv1D)	(None, 78, 32)	1568
max_pooling1d_163 (MaxPooling1D)	(None, 19, 32)	0
conv1d_168 (Conv1D)	(None, 17, 64)	6208
max_pooling1d_164 (MaxPooling1D)	(None, 2, 64)	0
flatten_50 (Flatten)	(None, 128)	0
dense_201 (Dense)	(None, 16)	2064
dropout_151 (Dropout)	(None, 16)	0
dense_202 (Dense)	(None, 32)	544
dropout_152 (Dropout)	(None, 32)	0
dense_203 (Dense)	(None, 64)	2112
dropout_153 (Dropout)	(None, 64)	0
dense_204 (Dense)	(None, 8)	520
Total Params		13080
Trainable Params		13080
Non-trainable Params		0

CNN layers are effective for learning hierarchical representations and detecting patterns in images, audio, and sequential data. In the context of infant cry classification, these layers are expected to learn important spectral and temporal features from the cry recordings, aiding in classification. The DNN layers following the CNN layers perform higher-level feature learning and classification based on the features extracted by the CNN layers. These layers capture more abstract representations derived from the convolutional layers, enhancing the model's ability to discern complex features. Dropout layers play a pivotal role in averting overfitting by randomly deactivating neurons during training, thereby enhancing the model's generalization capabilities. In the output layer, SoftMax activation is applied, generating probabilities for each class. This configuration

enables the model to make predictions across multiple classes, offering a nuanced understanding of the input data's likelihood of belonging to each respective class.

Conv1D layers in Table I with increasing filters and decreasing kernel sizes, followed by MaxPooling1D layers perform feature extraction by applying convolution operations on the input data. The filters detect patterns at different levels of abstraction. The pivotal function of MaxPooling layers lies in reducing the spatial dimensions of the output. By retaining the most pertinent information, these layers contribute to a reduction in computational complexity, optimizing the overall efficiency of the model. CNNs are effective in learning hierarchical representations from input data, capturing local and global patterns, which can be beneficial for analyzing spectrogram-like representations of audio data. The Flatten layer plays a crucial role in transforming the output from the convolutional layers into a 1D array, preparing it for input into subsequent Dense layers. These Dense and Dropout layers collectively process the flattened features extracted by the Convolutional Neural Network (CNN) layers, engaging in higher-level feature learning and classification.

The final Dense layer, equipped with SoftMax activation, produces output probabilities corresponding to each class, offering insights into the probability of the input being associated with a specific class. To optimize the model, Adam optimizer is employed with binary cross-entropy loss and accuracy metric. A learning rate of 0.0001 is set to regulate the optimization process. For model assessment, k-fold cross-validation with k=5, epochs=30, and a batch size of 64 is conducted. This method involves random shuffling of the dataset and its division into k groups. The training-validation process is iteratively performed, Furnishing an all-encompassing assessment of the model's performance on diverse subsets of the data.

IV. Result Analysis

Ensuring the robustness of a model extends beyond achieving high accuracy on a particular dataset. Validation of robustness necessitates the implementation of k-fold cross-validation as a means to substantiate the model's resilience across diverse data subsets. Here in Table II, the performance Analysis of our model shown. Our model achieves an average accuracy of 0.9384 and average loss 0.0652. By utilizing k-fold cross-validation, we obtain a more reliable estimation of the model's performance across various data subsets. The model is performing consistently well across various subsets, enhancing the confidence in its robustness and ability to generalize to unseen data. It's a strong indicator of a well-performing and stable model. The assessment of the proposed model's effectiveness incorporates the following performance metrics: Accuracy, Recall, Precision, and F1-Score.

TABLE II. Evaluation of the efficacy of our model's performance

Class	Precision	Recall	F1-Score	Support
belly_pain	0.86	0.65	0.75	1233
cry	0.93	0.99	0.96	1233
discomfort	0.99	1.00	0.99	1233
Hungry	0.91	0.99	0.95	1233
laugh	0.87	0.93	0.90	1233
noise	0.99	1.00	0.99	1233
silence	0.96	0.96	0.96	1233
burping	1.00	1.00	1.00	1233
ACCURACY			0.94	9864
MACRO AVG	0.94	0.94	0.94	9864
WEIGHTED AVG	0.94	0.94	0.94	9864

TABLE III. PREDICTED VS ACTUAL Labels of Our Model (First 10 row)

Sl.	PREDICTED LABELS	ACTUAL LABELS
0	Hungry	Hungry
1	noise	noise
2	belly_pain	belly_pain
3	belly_pain	belly_pain
4	laugh	laugh
5	discomfort	discomfort
6	discomfort	discomfort
7	Hungry	Hungry
8	Hungry	Hungry
9	discomfort	discomfort

In Table III, we can see the actual vs predicted output of our model on unseen data. This table consists 9864 rows, only first 10 rows is being shown. In Fig. 9, Accuracy \& Loss Across 5 Fold is being shown. In Fig 10, we can see the confusion matrix of our model's performance.

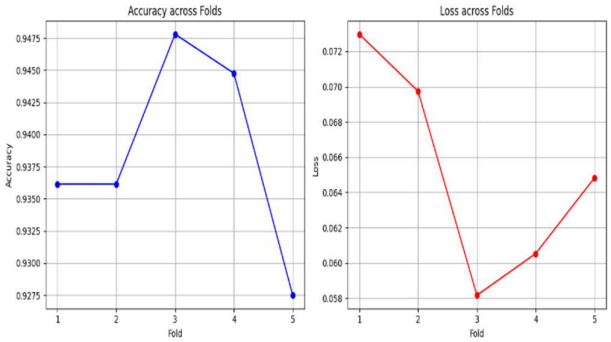

Fig. 9 Accuracy & Loss Across Fold

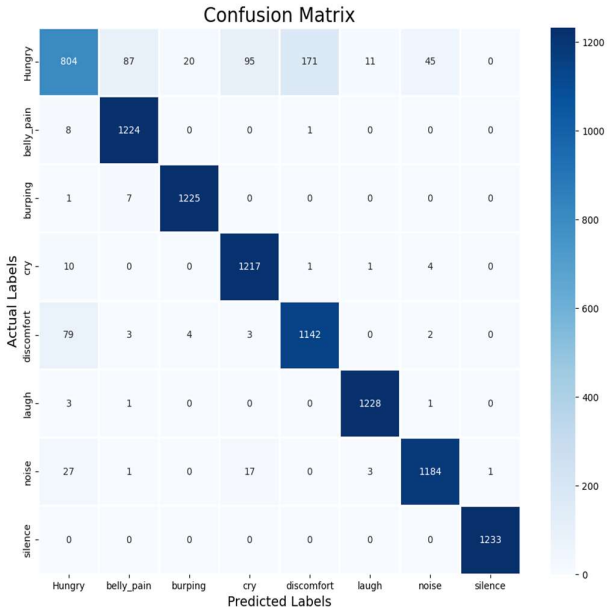

Fig. 10 Confusion Matrix of Proposed model

V. Conclusion

In conclusion, this study demonstrates an innovative and robust approach to infant cry classification using CNN+DNN hybrid model, data augmentation, and feature extraction methods. Leveraging popular datasets like Donateacry and

Infants Cry Sounds corpus, we applied diverse augmentation strategies and balanced class distributions using SMOTE to enhance model generalizability. The integration of feature extraction techniques, including Chroma STFT,MFCC, ZCR and RMS facilitated the capturing of distinctive patterns in cry signals. Our comprehensive validation, including k-fold cross validation, underscores the efficacy of our model in automatically recognizing and categorizing different types of infant cries with 93.84% accuracy and loss 0.0652. This technology holds promise in assisting healthcare professionals and caregivers by swiftly identifying underlying reasons for crying, potentially enabling more prompt and accurate responses to infants' needs, whether related to hunger, discomfort, pain, or other factors.

REFERENCES

[1] J. Saraswathy, M. Hariharan, S. Yaacob and W. Khairunizam, "Automatic classification of infant cry: A review," 2012 International Conference on Biomedical Engineering (ICoBE), Penang, Malaysia, 2012, pp. 543-548, doi: 10.1109/ICoBE.2012.6179077.

[2] J. A. Green, P. G. Whitney and M. Potegalb, “Screaming, Yelling, Whining and Crying: Categorical and intensity differences in Vocal Expressions of Anger and Sadness in Children’s Tantrums,” Emotion. vol.5, no. 11, pp.1124–1133 Oct. 2011.

[3] H. Karp, The Happiest Baby on the Block; Fully Revised and Updated Second Edition: The New Way to Calm Crying, New York City, NY, USA, Bantam, 2015.

[4] S. Al-Azani and E. -S. M. El-Alfy, "Enhanced Video Analytics for Sentiment Analysis Based on Fusing Textual, Auditory and Visual Information," in IEEE Access, vol. 8, pp. 136843-136857, 2020, doi: 10.1109/ACCESS.2020.3011977.

[5] J. Villalba-Diez, D. Schmidt, R. Gevers, J. Ordieres-Meré, M. Buchwitz, and W. Wellbrock, “Deep Learning for Industrial Computer Vision Quality Control in the Printing Industry 4.0,” Sensors, vol. 19, no. 18. MDPI AG, p. 3987, Sep. 15, 2019. doi: 10.3390/s19183987.

[6] A. Esteva et al., “Deep learning-enabled medical computer vision,” npj Digital Medicine, vol. 4, no. 1. Springer Science and Business Media LLC, Jan. 08, 2021. doi: 10.1038/s41746-020-00376-2.

[7] A. Bauer et al., “Combining computer vision and deep learning to enable ultra-scale aerial phenotyping and precision agriculture: A case study of lettuce production,” Horticulture Research, vol. 6, no. 1. Oxford University Press (OUP), Jun. 01, 2019. doi: 10.1038/s41438-019-0151-5.

[8] B. Valani, “Donate-a-cry-corpus-features-dataset,” Kaggle.

[9] R. Jahangir, “Infant cry sounds,” Kaggle.

[10] A. Jamal and S. Al-Azani, "A Machine-Learning Approach for Children’s Pain Assessments Using Prosodic and Spectral Acoustic Features," 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 2023, pp. 1-6, doi: 10.1109/ICECCME57830.2023.10252478.

[11] M. H. C. Sudul et al., "Automatic Classification of Infant’s Cry Using Data Balancing and Hierarchical Classification Techniques," 2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP), Ohrid, North Macedonia, 2023, pp. 1-5, doi: 10.1109/IWSSIP58668.2023.10180257.

[12] A. Gorin, C. Subakan, S. Abdoli, J. Wang, S. Latremouille and C. Onu, "Self-Supervised Learning for Infant Cry Analysis," 2023 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW), Rhodes Island, Greece, 2023, pp. 1-5, doi: 10.1109/ICASSPW59220.2023.10193421.

[13] N. Meephiw and P. Leesutthipornchai, "MFCC Feature Selection for Infant Cry Classification," 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, 2022, pp. 123-127, doi: 10.1109/ICSEC56337.2022.10049328.

[14] Z. Firas, A. A. Nashaat and G. Ahmad, "Optimizing Infant Cry Recognition: A Fusion of LPC and MFCC Features in Deep Learning Models," 2023 Seventh International Conference on Advances in Biomedical Engineering (ICABME), Beirut, Lebanon, 2023, pp. 01-06, doi: 10.1109/ICABME59496.2023.10293083.

[15] G. Anjali, S. Sanjeev, A. Mounika, G. Suhas, G. P. Reddy and Y. Kshiraja, "Infant Cry Classification using Transfer Learning," TENCON 2022 - 2022 IEEE Region 10 Conference (TENCON), Hong Kong, Hong Kong, 2022, pp. 1-7, doi: 10.1109/TENCON55691.2022.9977793.

[16] K. Alam, M. H. Bhuiyan and M. F. Monir, "Bangla Speaker Accent Variation Classification from Audio Using Deep Neural Networks: A Distinct Approach," TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), Chiang Mai, Thailand, 2023, pp. 134-139, doi: 10.1109/TENCON58879.2023.10322411.

[17] K. Alam, N. Nigar, H. Erler and A. Banerjee, "Speech Emotion Recognition from Audio Files Using Feedforward Neural Network," 2023 International Conference on Electrical, Computer and Communication Engineering (ECCE), Chittagong, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ECCE57851.2023.10101492.

[18] Z. K. Abdul and A. K. Al-Talabani, "Mel Frequency Cepstral Coefficient and its Applications: A Review," in IEEE Access, vol. 10, pp. 122136-122158, 2022, doi: 10.1109/ACCESS.2022.3223444.

[19] D. HABA, Data Augmentation with Python: Enhance Accuracy in Deep Learning with Practical Data Augmentation... for Image, Text, Audio & Tabular Data. S.l.: PACKT PUBLISHING LIMITED, 2023.

[20] M. Muthumari, C. A. Bhuvaneswari, J. E. N. S. Kumar Babu, and S. P. Raju, “Data Augmentation Model for Audio Signal Extraction,” 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC). IEEE, Aug. 17, 2022. doi: 10.1109/icesc54411.2022.9885539.