Journal of Electronics, Electromedical Engineering, and Medical Informatics
Multidisciplinary: Rapid Review: Open Access Journal **Vol. 6, No. 4, October 2024, pp: 355-369; eISSN: 2656-8632**

RESEARCH ARTICLE **OPEN ACCESS**

Manuscript received May 29, 2024; revised August 8, 2024; accepted August 15, 2024; date of publication October 20, 2024
Digital Object Identifier (**DOI**): https://doi.org/10.35882/jeeemi.v6i4.465

How to cite: Ridha Fahmi Junaidi, Mohammad Reza Faisal, Andi Farmadi, Rudy Herteno1, Dodon Turianto Nugrahadi, Luu Duc Ngo and Bahriddin Abapihi , "Baby Cry Sound Detection: A Comparison of Mel Spectrogram Image on Convolutional Neural Network Models", Journal of Electronics, Electromedical Engineering, and Medical Informatics, vol. 6, no. 4 pp: 355-369, October 2024.

Baby Cry Sound Detection: A Comparison of Mel Spectrogram Image on Convolutional Neural Network Models

Ridha Fahmi Junaidi[1], Mohammad Reza Faisal[1], Andi Farmadi[1], Rudy Herteno[1], Dodon Turianto Nugrahadi[1], Luu Duc Ngo[2] and Bahriddin Abapihi[3]

[1]Computer Science Department, Lambung Mangkurat University, Banjarbaru, South Kalimantan, Indonesia
[2]Faculty of Information Technology, Bac Lieu University, Bac Lieu, Vietnam
[3]The Department of Statistics, Faculty of Mathematics and Natural Sciences, Halu Oleo University, Kendari, Indonesia

Corresponding author: Mohammad Reza Faisal (e-mail: reza.faisal@ulm.ac.id).

This paragraph of the first footnote will contain support information, including sponsor and financial support acknowledgment. For example, "This work was supported in part by the U.S. Department of Commerce under Grant BS123456."

ABSTRACT Baby cries contain patterns that indicate their needs, such as pain, hunger, discomfort, colic, or fatigue. This study explores the use of Convolutional Neural Network (CNN) architectures for classifying baby cries using Mel Spectrogram images. The primary objective of this research is to compare the effectiveness of various CNN architectures such as VGG-16, VGG-19, LeNet-5, AlexNet, ResNet-50, and ResNet-152 in detecting baby needs based on their cries. The datasets used include the Donate-a-Cry Corpus and Dunstan Baby Language. The results show that AlexNet achieved the best performance with an accuracy of 84.78% on the Donate-a-Cry Corpus dataset and 72.73% on the Dunstan Baby Language dataset. Other models like ResNet-50 and LeNet-5 also demonstrated good performance although their computational efficiency varied, while VGG-16 and VGG-19 exhibited lower performance. This research provides significant contributions to the understanding and application of CNN models for baby cry classification. Practical implications include the development of baby cry detection applications that can assist parents and healthcare provide.

INDEX TERMS baby cry sound detection, Convolutional Neural Network, Mel Spectrogram, audio classification

I. INTRODUCTION

Meeting a baby's needs is a top priority in childcare, where understanding and responding to their cries is key. However, distinguishing the reasons behind a baby's cry is often challenging, especially for new parents [1]. Baby cries contain patterns that indicate their needs, such as pain, hunger, discomfort, colic, or fatigue [2]. Accurately classifying these cries, however, remains difficult, particularly for parents who may struggle to discern the specific reason for their baby's distress. This difficulty underscores the need for reliable methods to automatically identify the meaning behind infant cries.

Baby crying is one of the most basic and primary forms of communication for a baby. Given that babies are not yet able to use verbal language to express their wants and needs, crying becomes a vital tool to convey messages to adults around them. Crying can reflect various conditions and needs of the baby, ranging from hunger, pain, discomfort, to the need for attention and emotional warmth.

Sometimes babies cry also to seek attention, because they begin to understand feelings, sounds, smiles, and also eye contact. Crying is also a form of communication from the baby to the people around them. Understanding the meaning of a baby's cry has a significant impact not only on the well-being of the baby, but also on the well-being of parents and caregivers. The inability to interpret a baby's cry accurately can cause stress and confusion for parents, which in turn can affect the quality of care provided to the baby.

Infant development is a complex and dynamic process, involving various physical, cognitive, emotional, and social aspects. Infant crying is essential for healthy emotional and cognitive development. Warm and loving interactions

between infants and caregivers help to form a strong foundation for a sense of security and self-confidence in infants. A safe and stimulating environment allows infants to explore the world around them and develop motor and cognitive skills.

In addition, community support and access to quality health services also contribute positively to infant development. A holistic approach to caring for and supporting infant development. By providing comprehensive attention to the physical, emotional, and cognitive needs of infants, we can ensure that they grow into healthy, happy, and competitive individuals. Therefore, it is necessary to continue to develop effective strategies and interventions to support the holistic development of infants [3]. One way to achieve this is by recognizing baby cries with the assistance of Machine Learning [14].

In recent years, advances in voice processing technology and machine learning have opened up significant opportunities to address this challenge. Machine learning (ML) has become a crucial milestone in technological development, enabling computers to learn from data automatically [4]. Convolutional Neural Networks (CNNs), a dominant approach in ML, have demonstrated remarkable success in various fields, including image processing, facial recognition, and, importantly, sound classification [5], [6], [7]. In sound processing, CNNs have proven highly effective in tasks such as music classification, speech recognition, and audio signal detection [8], [9], [10]. Their application to baby cry analysis represents a significant breakthrough in our ability to understand and respond to babies' needs based on their vocal expressions [9].

Through learning from data, CNNs can extract relevant features from baby cries and identify patterns that are difficult to recognize manually [11]. Besides using CNNs, other machine learning methods such as Support Vector Machine, Random Forest, and Naïve Bayes can also be used to identify baby cries [14] . However, compared to CNNs, these methods have shown results that are still below those of CNNs in other studies [36],[42]. Thus, using CNNs to classify baby cries represents a significant breakthrough in efforts to understand and respond to babies' needs based on their vocal expressions [9]. The difference between this study and previous research lies in the use of various CNN architectures to classify baby cries using Mel Spectrogram images, whereas previous studies have not specifically compared the performance of different CNN architectures for this task.

Previous research has shown that CNNs are effective in detecting baby cries in audio data, although they tend to overfit on small datasets [12]. Selecting the appropriate model architecture can reduce the risk of overfitting and improve classification performance [13]. However, no study has systematically compared the performance of various CNN architectures in classifying baby cries using Mel Spectrogram images, a visual representation of audio signals well-suited for CNN analysis.

Previous studies have compared the performance of various CNN architectures in sound processing but have not specifically examined baby cry classification [12]. Although there is research using CNNs to classify the meaning of baby cries, the use of Mel Spectrogram images has not been directly covered. Related studies also focus on other fields such as detecting COVID-19 based on cough sounds or retinal image classification, which are not directly relevant to classifying baby cries using Mel Spectrogram images [14], [15], [16].

Previous research [14] explains that from the audio of baby cries is considered unstructured and requires a feature extraction procedure to produce structured data suitable for machine learning algorithms. In [14] Mel Frequency Cepstral Coefficients (MFCC) serves as the basis for feature extraction techniques, with various coefficient values used to process baby cry audio covering a range of 1 to 7 seconds.

More specific research is needed to explore the effectiveness of different CNN architectures in classifying baby cries using Mel Spectrogram images. The aim of this research is to provide a deeper understanding of the capabilities of CNN architectures in detecting and responding to babies' needs based on their cries. This study will test the performance of several CNN architectures, including VGG-16, VGG-19, LeNet-5, AlexNet, ResNet-50, and ResNet-152, in the task of classifying baby cries. The novelty of this research is expected to provide better insights into how to optimally use CNNs to support the health and well-being of babies, as well as reduce parents' stress and anxiety.

This study contributes to research provides significant contributions to the understanding and application of CNN models for baby cry classification. Practical insights for developing more accurate and reliable baby cry detection applications, potentially improving childcare practices and supporting parents. Not only that, this study also contributes to providing understanding to parents, caregivers and people closest to the baby to know the meaning of the baby's cry. So that it can help in determining effective actions to relieve crying or meet the baby's needs. This study also contributes to the field of education that is relevant to the baby's cry. So that this study can be a basis or material for further research.[17].

Based on the strengths of CNNs in sound classification tasks, we hypothesize that the detection of a baby's crying sound using these architectures and Mel Spectrogram images will yield superior results in classifying baby cries compared to other machine learning methods and traditional feature extraction techniques. The findings of this research have the potential to significantly impact the development of assistive technologies for parents and caregivers, ultimately contributing to the well-being of infants.

II. MATERIAL AND METHODS

Dunstan Baby Language is a theory proposed by Priscilla Dunstan that claims that babies have five basic types of cries that can be identified by different sounds, namely bellypain, burping, discomfort, hungry and tired.

While the Donate-a-Cry Dataset is a collection of baby cry recordings collected from various sources with the aim

of training a machine learning model to classify types of baby cries. This dataset is often used to validate the Dunstan Baby Language theory or to develop a more accurate classification model.

In this study, the methodology is comparison and divided into four main phases: data collection, data pre-processing, classification, and evaluation, as shown in FIGURE 1

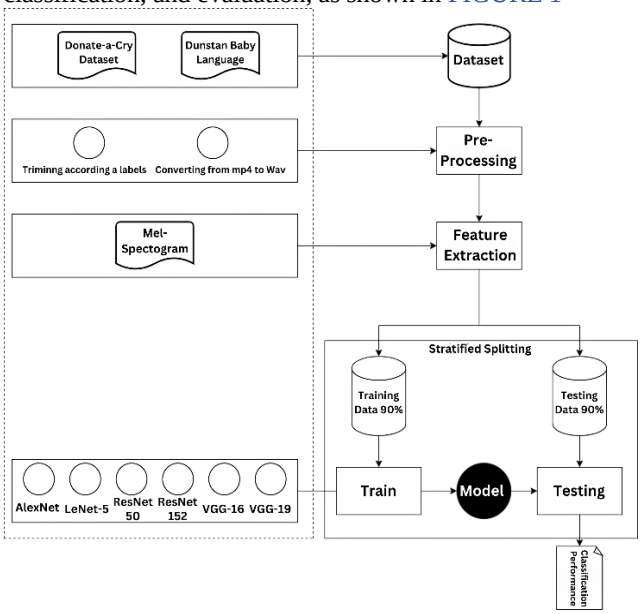


FIGURE 1 Research Flow

This study uses two datasets, namely the Donate-a-Cry dataset and the Dunstan Baby Language dataset. This dataset contains recordings of baby cries with 5 types of cries, namely Bellypain, Burping, Discomfort, Hungry, and Tired. In the data pre-processing phase, the audio data is converted from MP4 to Wav and then the grouping process is carried out based on its label.

In the feature extraction stage, the audio data is processed to obtain the Mel Spectrogram image to be used as data in the train and test model process to obtain evaluation results from the model used. Mel Spectrogram Feature Extraction converts audio signals into visual representations that capture frequency information over time.

In this study, we propose the use of Mel spectrogram, a transformation that details the frequency composition of a signal over time. This is because it can produce an image representation of an audio signal, Mel spectrogram is the input to our machine learning model. This allows us to use well-researched image classification techniques[18].

A. DATASET

In this study, two datasets were used: the Donate-a-cry-corpus Dataset and the Dunstan Baby Language Dataset. The "Donate-a-cry-corpus" dataset is one of the data sources in this research, obtained from the official GitHub repository (https://github.com/gveres/donateacry-corpus). Several studies that have used this dataset include [19], [20], [21]. This dataset has five class labels: Bellypain, Burping, Discomfort, Hungry, and Tired. The label distribution of this dataset can be seen in TABLE 1. The total recordings in this dataset are 457 audio recordings, each lasting 7 seconds and formatted as WAV files.

TABLE 1
Donate a cry corpus Dataset Distribution.

Label	Quantity
belly-pain	16
Burping	8
Discomfort	27
Hungry	382
Tired	24
Total	457

The "Dunstan Baby Language" dataset originates from the research and observations conducted by Priscilla Dunstan, a child development expert. Previous studies that have used this dataset include [22], [23], [24]. This dataset also has five class labels: Bellypain, Burping, Discomfort, Hungry, and Tired. The details of this dataset can be seen in TABLE 2. The total data in this dataset comprises 156 recordings, with durations ranging from 1 to 26 seconds. The variation in recording lengths is due to the manual data collection process by the researcher, which involved extracting segments from Priscilla Dunstan's videos that illustrate examples of baby cries for each label.

TABLE 2
Dunstan Baby Language Dataset Distribution

Label	Quantity
belly-pain	20
Burping	26
Discomfort	27
Hungry	45
Tired	38
Total	156

B. PREPROCESSING

Preprocessing is the initial step taken to prepare raw data into a more structured and clean form before entering the analysis or modeling stage in machine learning and data mining processes [25]. The main objective of preprocessing is to enhance the quality of the data and ensure that it is ready for use in predictive models, thereby making the analysis or modeling results more accurate and reliable. An illustration of the Waveplot of the data can be seen in FIGURE 2

For the Donate-a-cry-corpus dataset, preprocessing involved converting audio files into Mel-Spectrogram images. The image dimensions of the Mel-Spectrograms were adjusted to match the default size required by the CNN architectures to be used. After transforming the audio into Mel-Spectrogram images, the dataset was split into training and testing data using a stratified technique to ensure proper distribution of data across each class. The data split ratio was 90% for training and 10% for testing.

For the Dunstan Baby Language dataset, preprocessing began by collecting videos from Priscilla Dunstan that explain and demonstrate various types of baby cries along with examples. These videos were then segmented according to their respective labels, resulting in varying video lengths from 1 second to 26 seconds. After segmentation, the video data was converted into WAV format, followed by transformation into Mel-Spectrogram images. The image dimensions of the Mel-Spectrograms were adjusted to match the default size

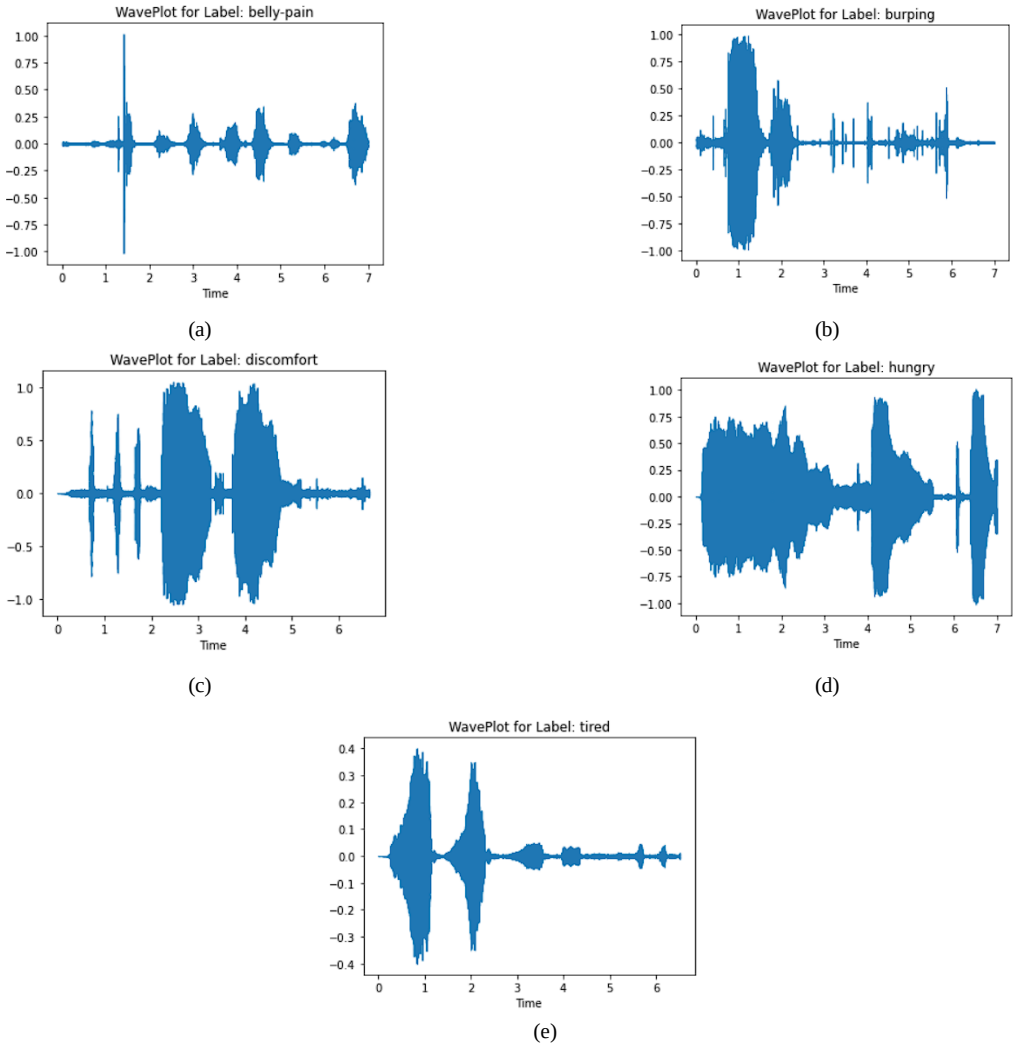


FIGURE 2 The sound wave for baby a) crying burping, b) cry bell-pain, c)cry discomfort, d)cry hungry, and e) cry tired

required by the CNN architectures to be used. This dataset was also split into training and testing data using a stratified technique to ensure proper distribution of data across each class. The data split ratio was 90% for training and 10% for testing.

C. MEL-SPECTROGRAM

A Mel Spectrogram is a visual representation of the frequency spectrum of an audio signal over time, designed to mimic human hearing. In the context of analyzing sounds, including infant cries, a Mel Spectrogram provides invaluable information about the frequency and temporal characteristics of the signal. Each type of infant cry has different frequency and temporal characteristics. A Mel Spectrogram is able to capture these subtle differences, and can be used to distinguish between types of cries (e.g., hunger, pain, or discomfort). By analyzing patterns in a Mel Spectrogram, machine learning algorithms can be trained to classify infant cries with high accuracy [26].

According to the image, a spectrogram is a visual representation of the frequency spectrum of a signal. Spectrograms can be formed using the Fourier Transform. A spectrogram is defined as the magnitude of the square of the STFT, giving the sound power for a given frequency and time in the third dimension. After that, each part will be adjusted to the vertical line in the image, which is a comparison of magnitude with frequency in a given time. After that, the spectra will be plotted side by side to form an image[5]. Mel spectrogram has better sound signal information compared to others. In addition, the use of Mel spectrogram can avoid cost computation due to complex calculations [27].

The spectrogram was chosen because of its compatibility with the human auditory system in perceiving logarithmic frequencies. In addition, the use of mel-spectrograms generally provides better performance than other audio representations when used as input to audio problems involving deep learning[28].

A Spectrogram is a two-dimensional image that displays changes in sound intensity across various frequencies over time, obtained through the Fast Fourier Transform (FFT) on multiple signal windows. The FFT is an algorithm that computes the Discrete Fourier Transform (DFT) of a

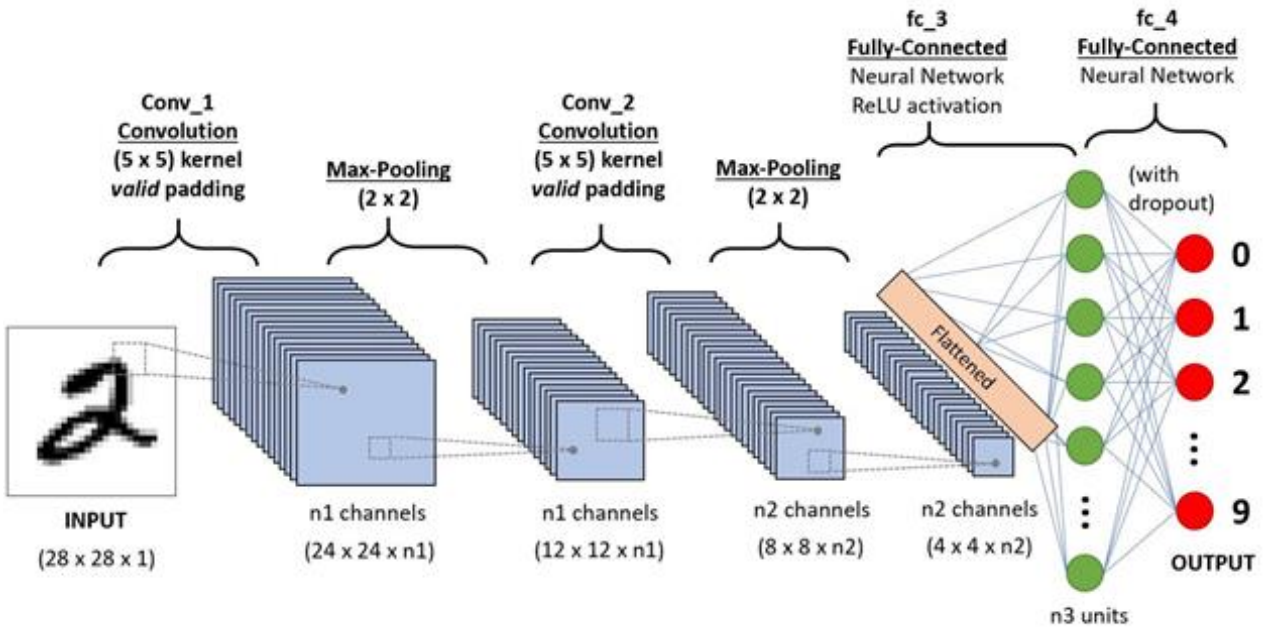


FIGURE 3. Default CNN Architecture

sequence, transforming a signal from its original time domain into a frequency domain representation. [29], [30]. The DFT formula are given in equation (1)

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-j2\pi kn/N} \quad (1)$$

where $x[n]$ is the input signal, $X[k]$ is the DFT result, N is the number of points in the DFT, j is the imaginary unit, and k is the index of the output frequency component. The result is a collection of frequency spectra interconnected and represented with color or brightness scales, where higher sound intensity is displayed brighter or with stronger colors. Mel-Spectrogram is a type of spectrogram where the frequencies are adjusted to the Mel frequency scale, which is more aligned with human auditory sensitivity, being more responsive to changes in low frequencies compared to high frequencies [31]. The conversion from frequency f (in Hz) to the Mel scale m is given by the formula (2):

$$m = 2595 \times log_{10}(1 + \frac{f}{700}) \quad (2)$$

Some studies that have used Mel-Spectrogram for classification include research by [32], [33], [34], [35].

D. CONVOLUTIONAL NEURAL NETWORK

Convolutional Neural Network (CNN) is a type of artificial neural network architecture that has been very successful in processing visual data. CNNs are designed to recognize patterns in image data in a way that is similar to how the human brain processes visual information. In the context of infant crying research, CNNs can be used to extract important features from the Mel Spectrogram of infant cries, which can then be used for classification or detection.

Convolutional Neural Network (CNN) can automatically learn to extract the most relevant features from the Mel Spectrogram, without the need for manual feature extraction. NNs can be trained to classify the type of infant cry into categories, such as hunger, pain, or discomfort.

Convolutional Neural Network (CNN) is a deep learning architecture highly effective for image data analysis [14]. CNN consists of multiple hidden layers involving convolutional, pooling, and fully-connected layers, which progressively extract features from input images as seen in FIGURE 3[37].

The Convolutional Layer functions to extract features from input images using filters or kernels [38]. The convolution operation can be represented on Equation (3):

$$I_i^l = \left(\sum_j I_j^{l-1} \otimes w_{ij}^l + b_i^l\right) \quad (3)$$

where I_i^l is the output with *mxn* size, $\otimes$ indicates convolution operator, w_{ij}^l represents convolution kernels and b_i^l is bias value [39]. Each filter scans the entire image by sliding the filter window across the image, producing feature maps that identify basic patterns such as edges, lines, and textures [40]. The main function of the convolutional layers is to detect various visual elements in the image that can be used for further classification.

Each filter scans the entire image by sliding the filter window across the image, producing feature maps that identify basic patterns such as edges, lines, and textures [41]. The main function of the convolutional layers is to detect various visual elements in the image that can be used for further classification.

The Pooling Layer is responsible for reducing the dimensionality of the feature maps from the convolutional

layer while retaining important information. Commonly used techniques include max pooling and average pooling [42]. Max pooling takes the maximum value from each region defined by the pooling filter, while average pooling takes the average value from that region [43]. The formulas for max pooling and average pooling are given in Equations (4) and (5):

$$P_{Max} = max\ (x_1, x_2, \dots, x_n) \quad (4)$$

$$P_{avg} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (5)$$

P_{Max} is the maximum value from the pooling region $x_1, x_2, \dots, x_n$. And P_{avg} is the average of the values x_i in the pooling region with n being the number of value. Pooling helps reduce the number of parameters and computations required, as well as handling small translations in the image, making the model more efficient and robust to variations in object positions within the image [44].

The Fully Connected Layer functions to combine all the features extracted by the previous layers and produces the final output used for classification [45]. Each neuron in this layer is connected to all neurons in the previous layer, allowing combinations of various features to form the final representation of the image. The output of a fully connected neuron is computed on formula (6):

$$y_{jk}(\mathrm{x}) = f\left(\sum_{i=1}^{n_H} \omega_{jk} x_i + w_{j0}\right) \quad (6)$$

$y_{jk}(x)$ is the output of the neuron, ω_{jk} is the weight corresponding to the input x_i, and n_H is the number of inputs to the neuron. The function f is a non-linear activation function applied to the linear combination. The bias w_{j0} is added to this linear combination [46]. The fully connected layer is often followed by an activation function, such as softmax, which generates a probability distribution for each possible class, allowing the network to make accurate predictions or classifications [47].

This study encompassed a comparative analysis of six widely-used CNN architectures to determine their effectiveness in baby cry classification. LeNet-5, a foundational CNN architecture known for its simplicity and efficiency, served as a baseline model. AlexNet, a deeper CNN that introduced key concepts like ReLU activation and dropout for improved performance, was also included in the comparison. VGG-16 and VGG-19, architectures recognized for their use of small convolutional filters and increased depth for enhanced feature extraction, were evaluated. Finally, ResNet-50 and ResNet-152, deep residual networks designed to address the vanishing gradient problem and enable the training of very deep models, were investigated to explore the impact of significant network depth on classification accuracy [12], [27], [28], [29]. These architecture have different input sizes, characteristics, and complexities, each with its own advantages and specific applications in image classification and pattern recognition, as shown more comprehensively in TABLE 3.

E. CONFUSSION MATRIX

The confusion matrix is a crucial tool in evaluating the performance of classification algorithms, especially when comparing their predictions with the actual values tested [48], [49]. In TABLE 4, there are four types of entries in the confusion matrix: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). In the confusion matrix table, columns reflect the actual values while rows reflect the predictions made by the algorithm [50], [51], [52]. From the values of TP, FP, FN, and TN, we can calculate several important performance metrics.

TABLE 3
Detail Architecture Convolutional Neural Network

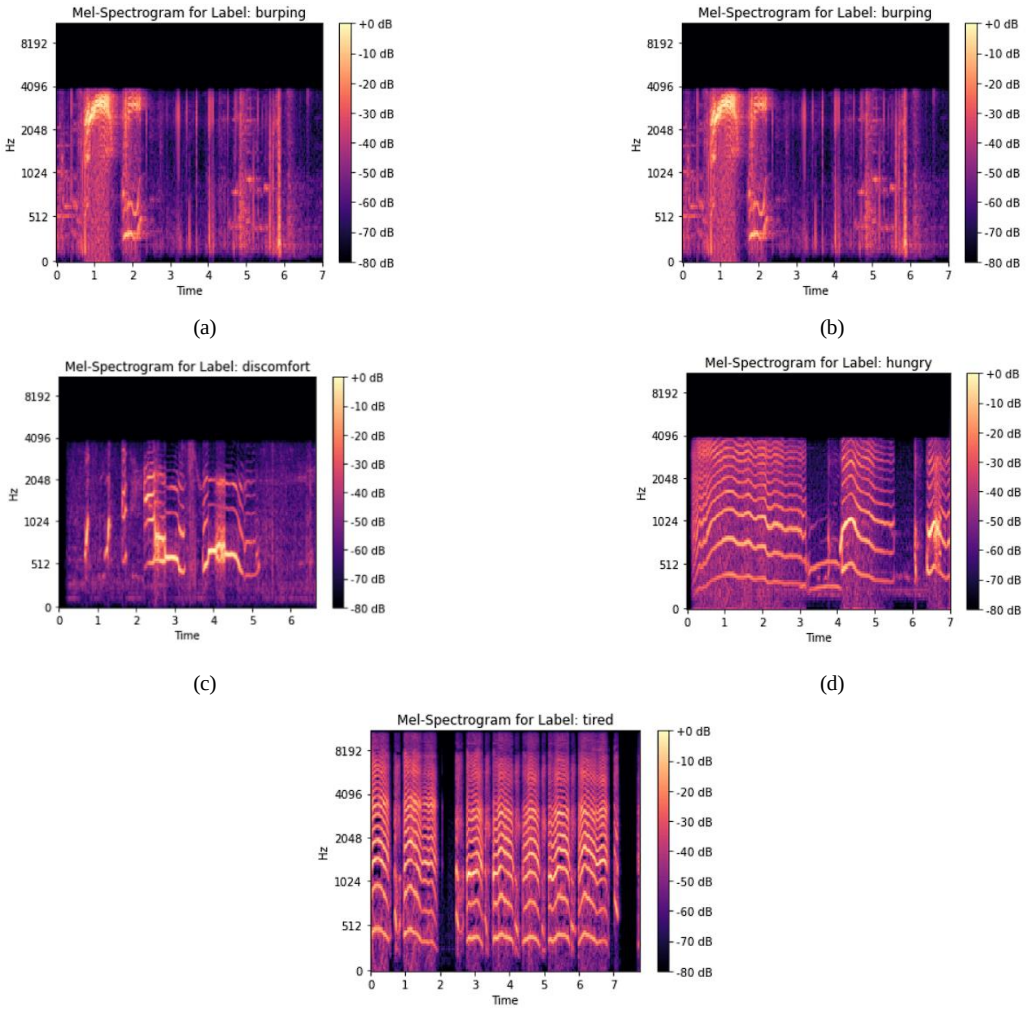
Architecture	Input Size	Convolutional Layers	Pooling Layers	Fully Connected Layers	Residual Blocks	Total Layers
LeNet-5	32x32	3	2	2	-	7
AlexNet	227x227	5	3	3	-	8
VGG-16	224x224	13	5	3	-	16
VGG-19	224x224	16	5	3	-	19
ResNet-50	224x224	49	1	1	Yes	50
ResNet-152	224x224	151	1	1	Yes	152

TABLE 4
Confussion Matrix

	Prediction	
Actual	True	False
True	TP	FP
False	FN	TN

In this study, we focus on 4 performance metrics:
Accuracy measures the extent to which an algorithm can correctly predict all cases, both positive and negative. It is calculated using the formula (7) :

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (7)$$



(a) (b) (c) (d) (e)

FIGURE 4 The Mel Spectrogram of Donate-a-Cry-Dataset for baby a) crying bell-pain, b) crying burping, c) cry discomfort, d) cry hungry, and e) cry tired

Recall also known as Sensitivity or True Positive Rate, measures the extent to which an algorithm can identify all true positive cases. It is calculated using the formula (8):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (8)$$

Precision also known as Positive Predictive Value, provides information about how precise or accurate the algorithm is in classifying an instance as positive. It is calculated using the formula (9):

$$Precission = \frac{TP}{TP+FP} \quad (9)$$

The F1 Score is a measure that combines Precision and Recall. It provides a single value that reflects the balance between these two metrics. It is calculated using the formula (10).

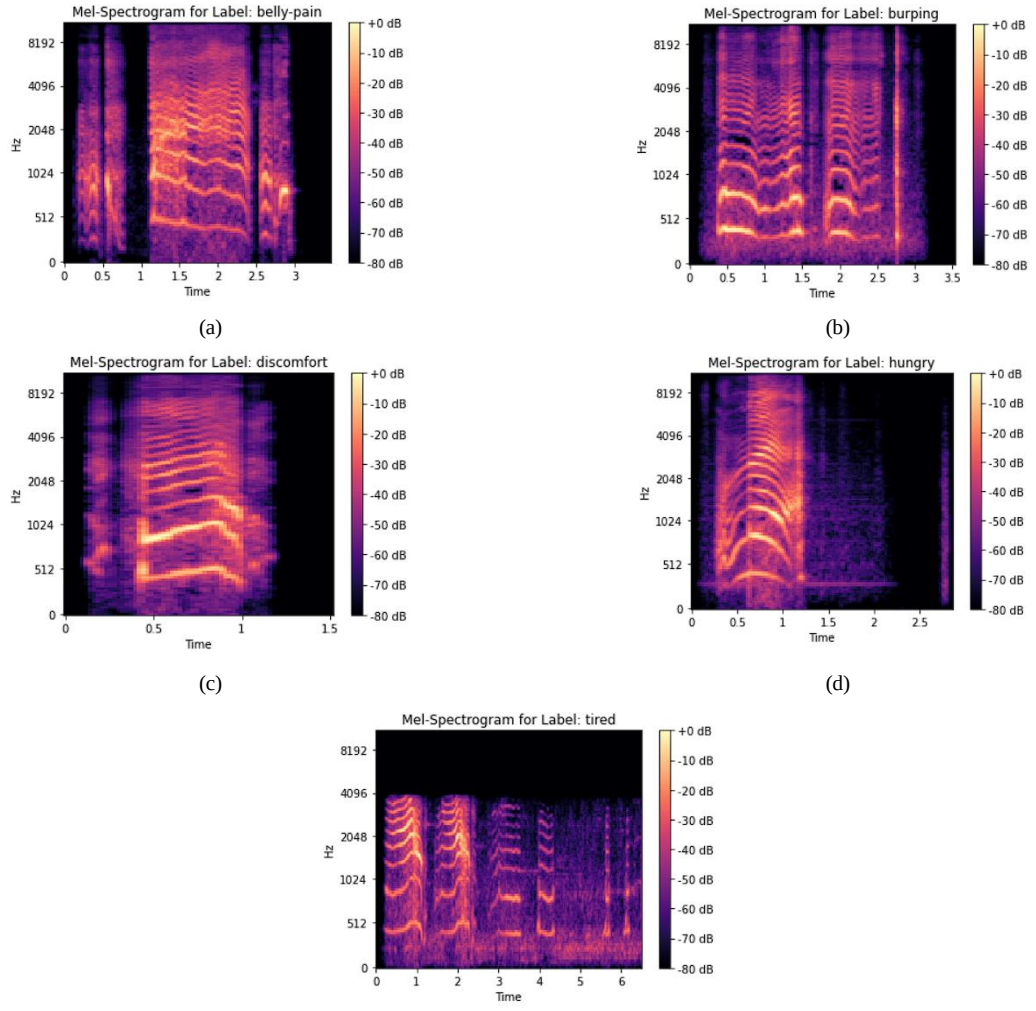
$$F1\ Score = \frac{2 \times Precission \times Recall}{Precission+Recall} \quad (10)$$

III. RESULT

In the Corpus-Donate-a-Cry dataset, data originally in .wav audio format is processed using the Librosa library from the Python programming language to obtain Mel-Spectrogram images. An example of the processed Corpus-Donate-a-Cry dataset can be seen on FIGURE 4

For the Dunstan Baby Language dataset, which initially comes in the form of videos, trimming is performed first to group them according to their labels. The data is then converted into .wav format for further processing to obtain Mel-Spectrogram images using the same method as applied to the Donate-a-Cry dataset. An example of a Mel-Spectrogram from the Dunstan Baby Language dataset can be seen on FIGURE 5

After the preprocessing process, the data is then split into train and test sets with a ratio of 90% for training and 10% for testing. The splitting applies the stratify method to ensure



(a) (b) (c) (d) (e)

FIGURE 5 The Mel Spectrogram of Dunstan Baby Language Dataset for baby a) crying bell-pain, b) crying burping, c) cry discomfort, d) cry hungry, and e) cry tired

that the class proportions in the training and testing datasets are the same as the class proportions in the overall dataset. Then, the classification process is performed using the proposed CNN architecture. The parameters of all experimental processes can be viewed in the provided documentation TABLE 5.

TABLE 5
Default Parameter Convolutional neural network

Parameter	Value
Batch size	8
Epochs	50
Learning Rate	0.0001
Optimizer	Adam
Activation	Relu
Dropout Rate	0.5

The classification process is conducted utilizing the proposed Convolutional Neural Network (CNN) architecture. Following this, the results are carefully evaluated to measure key performance metrics, including Accuracy, F1 Score, Precision, and Recall. These metrics provide a comprehensive understanding of the model's performance. The detailed classification outcomes are illustrated in TABLE 7, where each metric is clearly presented. Additionally, TABLE 6 offers a comparative analysis of the time required by each method, highlighting the efficiency of the CNN architecture in contrast to other approaches.

TABLE 6
Comparison Training Time (s) process of CNN Architecture on Datasets

Model	Training Time (Second)	
	Donate-a-Cry	Dunstan Baby Language
AlexNet	1180.35	542.28
LeNet-5 (RGB)	**12.04**	**7.04**
LeNet-5 (GrayScale)	13.2	11.38
ResNet50	5883.62	2658.37
ResNet50V2	5217.32	2668.15
ResNet152	12863.46	5690.04
ResNet152V2	14110.45	5334.54
VGG16	3670	3544.22
VGG19	9991.18	4568.42

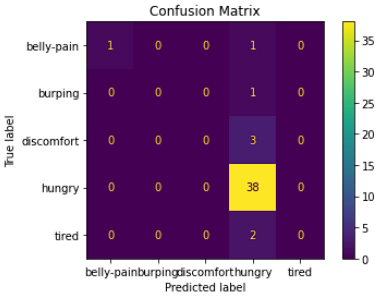
TABLE 7
Result Accuracy, F1 Score, Precission and Recall of CNN Architecture on Dataset

Dataset	Method	Accuracy	F1 Score	Precission	Recall
Donate-a-Cry dataset	AlexNet	**0.848**	**0.785**	**0.741**	**0.848**
	LeNet-5 (RGB)	0.826	0.747	0.682	0.826
	LeNet-5 (GrayScale)	0.826	0.747	0.682	0.826
	ResNet50	0.826	0.747	0.682	0.826
	ResNet50V2	0.826	0.747	0.682	0.826
	ResNet152	0.826	0.756	0.698	0.826
	ResNet152V2	0.804	0.737	0.679	0.804
	VGG16	0.826	0.747	0.682	0.826
	VGG19	0.826	0.747	0.682	0.826
Dunstan Baby Languange dataset	AlexNet	0.727	0.711	0.799	0.727
	LeNet-5 (RGB)	0.636	0.534	0.461	0.636
	LeNet-5 (GrayScale)	0.430	0.422	0.452	0.430
	ResNet50	0.591	0.586	0.648	0.591
	ResNet50V2	0.500	0.467	0.440	0.500
	ResNet152	0.364	0.258	0.273	0.364
	ResNet152V2	**0.773**	**0.773**	**0.800**	**0.773**
	VGG16	0.455	0.331	0.284	0.455
	VGG19	0.273	0.117	0.074	0.273

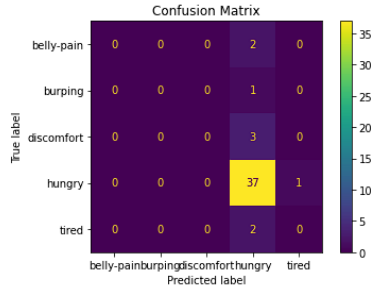
It can be observed from TABLE 7 that the highest accuracy on the Donate-a-Cry dataset was achieved when using the AlexNet method, which is 0.848, while the lowest accuracy was 0.804 when using the ResNet152V2 method. Other architectures obtained the same accuracy of 0.826. AlexNet also outperformed other methods in terms of F1 Score, Precision, and Recall. Regarding training time, LeNet-5 (RGB) was the fastest with a time of 12.04 (s), and the slowest was ResNet152V2 with a time of 14110.45 (s).

On the other hand, for the Dunstan Baby Language dataset, the highest accuracy performance was achieved by ResNet152V2 with an accuracy of 0.773, while the lowest was obtained by the VGG-19 method with an accuracy of 0.273. Looking at the F1-Score, Precision, and Recall values, ResNet152V2 still outperformed the others, although the numbers are almost similar to AlexNet. In terms of time, LeNet-5(RGB) remained the fastest with a time of 7.04 (s), while the slowest was ResNet152 with a time of 5334.52 (s).

In addition to the metrics of Accuracy, F1 Score, Precision, and Recall, we also evaluated the performance of the classifiers using confusion matrices. FIGURE 6 and FIGURE 7 show the confusion matrices for the CNN Architecture with Best and lack accuracy on Donate-a-Cry dataset and the Dunstan Baby Language dataset, respectively. The confusion matrices reveal significant performance variations across the different CNN architectures and datasets. On the Donate-a-Cry dataset, both AlexNet and ResNet-152V2 can FIGURE 6 demonstrated strong performance for the "hungry" class, accurately classifying the majority of samples. However, both models exhibited some misclassifications, particularly confusing "tired" cries for "hungry" and, in the case of ResNet-152V2, misclassifying all "belly-pain" instances. For the Dunstan Baby Language dataset FIGURE 7, ResNet-152V2 performed relatively well, correctly identifying most samples from the "belly-pain," "burping," and "tired" classes. In contrast, VGG-19 exhibited poor performance, misclassifying all samples as "tired".



(a)



(b)

FIGURE 6. The Confusion Matrix for Donate-a-Cry Dataset classification with CNN Architecture (a) AlexNet (b) ResNet-152V2

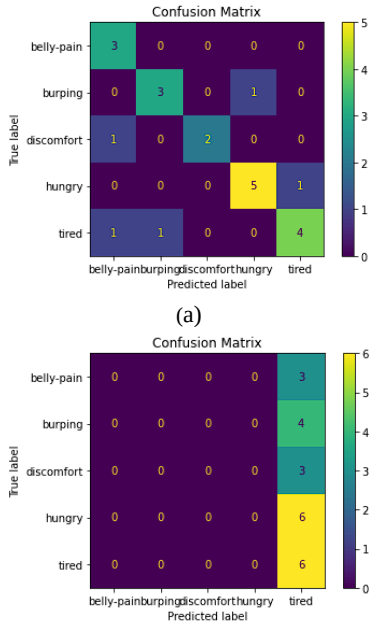


FIGURE 7. The Confusion Matrix for Dunstan Baby Language Dataset classification with CNN Architecture (a) ResNet-152V2 (b) VGG-19.

IV. DISCUSSION

From the research results presented above, the best performance values are obtained from the models built using various CNN architectures for classifying baby cries. These results can be seen in TABLE 6 for Accuracy, F1 Score, Precision, and Recall, and TABLE 7 for the comparison of training times. From TABLE 6, it can be observed that AlexNet performs the best on the Donate-a-Cry dataset with the highest accuracy of 0.848, F1 Score of 0.785, Precision of 0.741, and Recall of 0.848. Although its performance on the Dunstan Baby Language dataset is surpassed by ResNet152V2, AlexNet still demonstrates a good balance between performance and training time. ResNet152V2 shows the highest accuracy of 0.773, F1 Score of 0.773, Precision of 0.800, and Recall of 0.773 on the Dunstan Baby Language dataset, but its long training time indicates high computational costs.

FIGURE 8 dan FIGURE 9 show the comparison between the time and accuracy of CNN architecture models on the Donate-a-Cry dataset. These results are consistent with previous research [12] which indicated that AlexNet performs the best when classifying using Mel-Spectrogram images. In TABLE 7, LeNet-5 is noteworthy because compared to other model architectures, both RGB and GrayScale LeNet-5 obtained very fast times with accuracy results that can still compete with other models. The training time achievement of LeNet aligns with previous research [15], [33], which considered LeNet-5 to have fast training times. One of the factors influencing training speed is the input size of the model architecture, and LeNet-5 has a relatively small input size compared to others, which is 32x32.

These results indicate that model performance heavily depends on the characteristics of the dataset used. AlexNet shows superior performance on the Donate-a-Cry dataset, while ResNet152V2 excels on the Dunstan Baby Language dataset. This difference indicates that the proper model selection should consider the specific characteristics of the dataset used. Additionally, this research highlights the importance of considering training time when selecting a model, especially when computational resources are limited. LeNet-5, with its fast training time and competitive performance, becomes an efficient choice in this context.

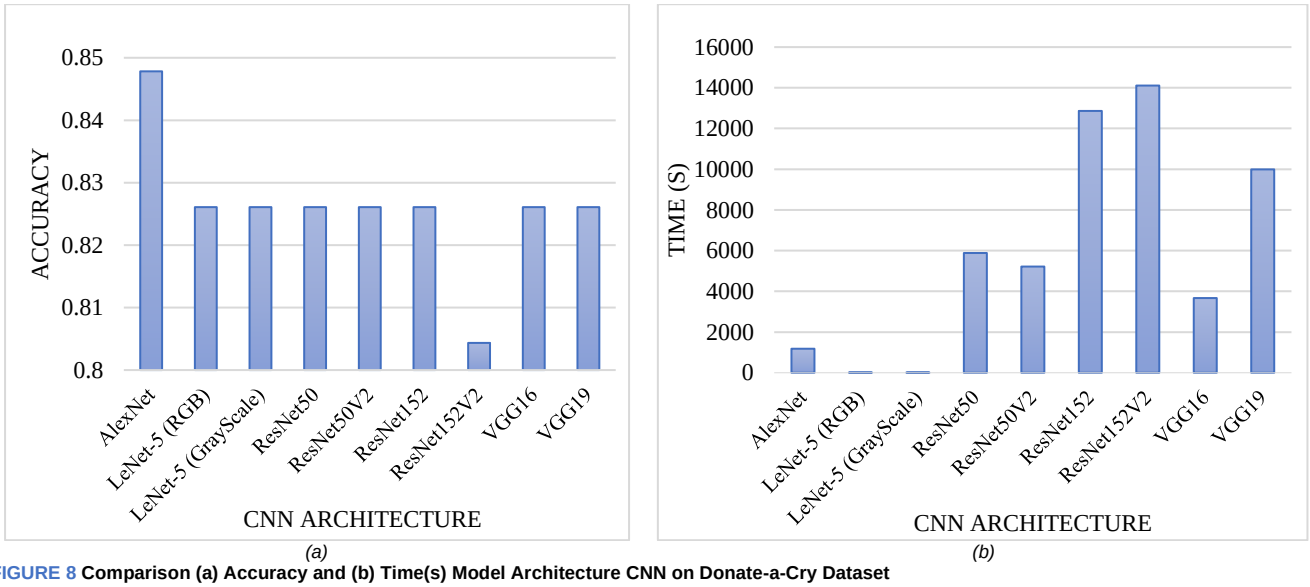
A closer examination of the confusion matrices provides valuable insight into model performance beyond overall metrics. For example, while AlexNet achieves high accuracy on the Donate-a-Cry dataset, the confusion matrix shown in FIGURE 6 reveals that it misclassifies some "tired" cries as "hungry." This confusion is also evident in ResNet-152V2's performance on the same dataset. These misclassifications could be attributed to the significant class imbalance in the Donate-a-Cry dataset, where "hungry" cries make up a large proportion.

Furthermore, the confusion matrices highlight the varying suitability of different architectures for the datasets. ResNet-152V2 performs well on the Dunstan Baby Language dataset, as shown in FIGURE 7, achieving good accuracy for several classes. Conversely, VGG-19 demonstrates poor performance on this dataset, misclassifying all samples as "tired." This discrepancy emphasizes the need to consider dataset-specific characteristics when selecting a model.

Various studies have utilized the Donate-a-Cry and Baby Chillanto datasets for baby cry classification using different methods and feature extraction techniques. Research [33] utilizing DWT and FFT with KNN and SVM classification resulted in the lowest accuracies, namely 53% and 37% respectively. Research [31] using Pitch and MFCC features with KNN classification achieved an accuracy of 76.16%. Research [22] with the Baby Chillanto and Donate-a-Cry datasets using MFCC features with additional features achieved almost the same accuracy for SVM and KNN, around 78%. Research [16] using MFCC features and CNN achieved the highest accuracy of 84.52% for the 'Hungry' label classification. Lastly, research [15] using a combination of MFCC, GFCC, and ZCR features with several classification methods showed that Random Forest (RF) provided the best results with an accuracy of 84%, while other methods like KNN, SVM, and Linear Regression (LR) yielded lower accuracies of 82%, 71%, and 41% respectively. The comparison between previous studies and the results of this research can be seen in TABLE 8.

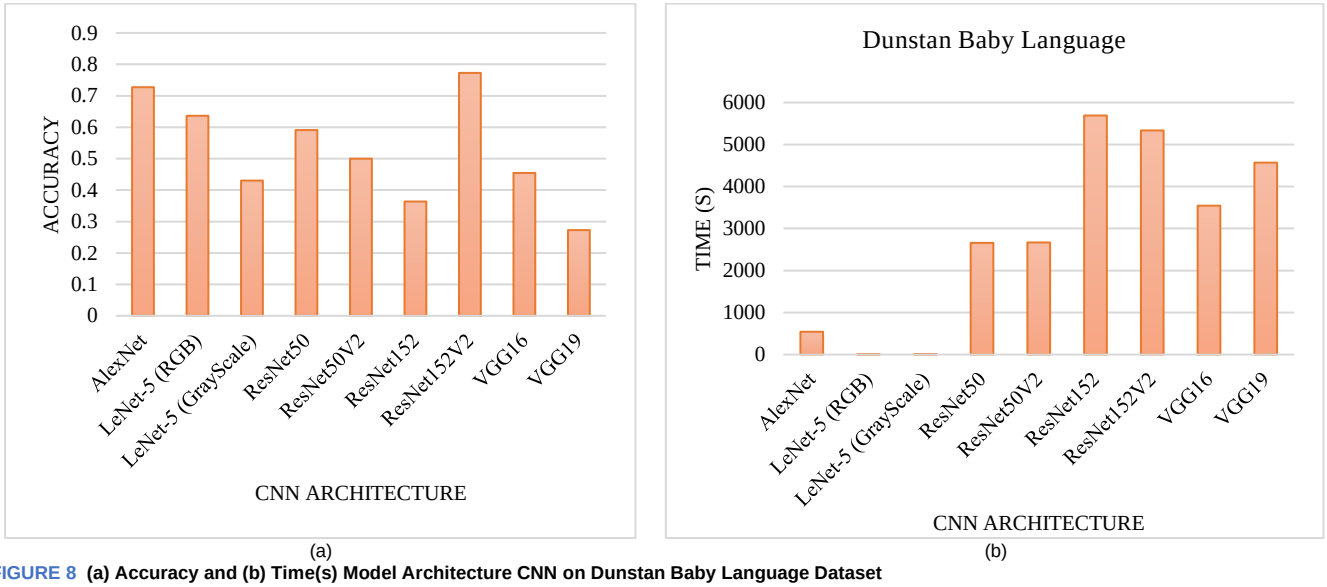
The proposed methods in this research can outperform or achieve comparable results to previous studies. This research introduces new aspects, such as using Mel-Spectrogram images for classification using CNN, which are still relatively unexplored in baby cry classification.

It's important to note the limitations of this research. The achieved results are still not optimal, as seen from the accuracy



(a) (b)

FIGURE 8 Comparison (a) Accuracy and (b) Time(s) Model Architecture CNN on Donate-a-Cry Dataset



(a) (b)

FIGURE 8 (a) Accuracy and (b) Time(s) Model Architecture CNN on Dunstan Baby Language Dataset

which is still less than 85% and F1 Score which has not reached 0.8. Some factors that may contribute to these suboptimal results include data imbalance, particularly in the Donate-a-Cry dataset, where the data for the 'hungry' label accounts for 83% of the total data. Additional methods are required to handle imbalanced data. Additionally, for the Dunstan Baby Language dataset, more comprehensive data processing is needed to make the data cleaner. Besides data improvements, further research can explore the use of parameters used for CNN models.

The findings of this study have implications, for the development of applications that detect baby cries. Despite some performance limitations this research contributes to using Mel Spectrogram images and Convolutional Neural Networks (CNN) for classifying baby cries. The potential advantages, such as identification of baby health issues and better care quality underscore its significance in childcare.

Nevertheless there are weaknesses and constraints that need attention. An unbalanced dataset, where certain cry types are not well represented can result in biased models that struggle to generalize in real world situations. Background noise may disrupt the detection and classification of cries. The proposed model may also face challenges in handling variations in baby cries caused by differences among babies like their age and health status. Moreover the computational demands of CNN models could limit their feasibility on low power devices. Achieving real time processing and response is complex due, to the intricacies of CNNs. Ethical and privacy issues related to data privacy and parental consent must also be taken into account.

TABLE 8
Comparison with Previous Research On Baby cry Classification

Research	Dataset	Feature Extraction	Labels	Classification Method	Result
					Accuracy
[33]	Donate a cry corpus	DWT.FFT	Belly-pain. Burping. Discomfort. Hungry. Tired.	KNN	53%
				SVM	37%
[31]	Donate a cry corpus	Pitch. MFCC	Awake. Belly torment. Burping. Discomfort. Hug. Hungry. Sleepy and Tired	KNN	76.16%
[22]	Baby Chillanto. Donate-a-Cry	MFCC + Extra	Hunger.Pain	SVM (RBF kernel)	78.08%
				K-Nearest Neighbors	78.03%
[16]	Donate a Cry	MFCC	Hungry	CNN	84.52%
[15]	Donate a Cry	MFCC. GFCC and Zero crossing rate (ZCR)	Belly-pain. Burping. Discomfort. Hungry. Tired.	RF	84%
				KNN	82%
				SVM	71%
				LR	41%
Our Research	Donate a Cry	Mel-Spectrogram	Belly-pain. Burping. Discomfort. Hungry. Tired	AlexNet	**84,80%**
				LeNet-5 (RGB)	82,60%
				LeNet-5 (GrayScale)	82,60%
				ResNet50	82,60%
				ResNet50V2	82,60%
				ResNet152	82,60%
				ResNet152V2	80,40%
				VGG16	82,60%
				VGG19	82,60%
	Dunstan Baby Language			AlexNet	72,70%
				LeNet-5 (RGB)	63,60%
				LeNet-5 (GrayScale)	43,00%
				ResNet50	59,10%
				ResNet50V2	50,00%
				ResNet152	36,40%
				ResNet152V2	77,30%
				VGG16	45,50%
				VGG19	27,30%

By recognizing these weaknesses and constraints future research can concentrate on overcoming these obstacles enhancing the reliability and precision of baby cry detection systems and ultimately improving childcare quality.

V. CONCLUSION

The research findings indicate that AlexNet performs best on the Donate-a-Cry dataset, achieving the highest accuracy of 0.848, F1 Score of 0.785, Precision of 0.741, and Recall of 0.848. Meanwhile, ResNet152V2 shows the best performance on the Dunstan Baby Language dataset, with an accuracy of 0.773, F1 Score of 0.773, Precision of 0.800, and Recall of 0.773. Despite its longer training time, ResNet152V2 incurs high computational costs. LeNet-5, with fast training time and competitive performance, proves to be an efficient choice, especially with limited computational resources.

The study identifies several weaknesses, particularly in terms of suboptimal model performance, with accuracy below

85% and F1 Score not reaching 0.8, indicating room for improvement. Data imbalance and the need for more comprehensive data processing are identified as major contributors to suboptimal performance.

Future research could focus on addressing data imbalance and using more optimal parameters for CNN models, with efforts expected to significantly enhance model performance. This research has significant implications for the development of baby cry detection applications, aiding parents in detecting infant health issues faster and improving the quality of baby care. It lays a strong foundation for the development of applications that can provide significant benefits to parents, caregivers, and child health services.

ACKNOWLEDGMENT

We sincerely thank all the individuals from the Computer Science Department, Lambung Mangkurat University, and Faculty of Information Technology, Bac Lieu University, Vietnam, who have contributed to successfully completing this collaboration research. Their valuable inputs and suggestions have significantly improved the quality of our work. We are also indebted to our project team members for their collaboration and hard work, which have been vital to the research's accomplishments.

REFERENCES

[1] Y. G. Tefera and A. A. Ayele, "Newborns and Under-5 mortality in Ethiopia: the necessity to revitalize Partnership in Post-COVID-19 era to meet the SDG targets," *J Prim Care Community Health*, vol. 12, p. 2150132721996889, 2021.

[2] K. Rezaee, H. G. Zadeh, L. Qi, H. Rabiee, and M. R. Khosravi, "Can you understand why i am crying? a decision-making system for classifying infants' cry languages based on deepsvm model," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 1, pp. 1–17, 2024.

[3] A. Pramod, H. S. Naicker, and A. K. Tyagi, "Machine learning and deep learning: Open issues and future research directions for the next 10 years," *Computational analysis and deep learning for medical care: Principles, methods, and applications*, pp. 463–490, 2021.

[4] S. Tripathy and R. Singh, "Convolutional neural network: an overview and application in image classification," in *Proceedings of Third International Conference on Sustainable Computing: SUSCOM 2021*, 2022, pp. 145–153.

[5] R. E. Saragih, Q. H. To, and others, "A survey of face recognition based on convolutional neural network," *Indonesian Journal of Information Systems*, vol. 4, no. 2, 2022.

[6] D. Issa, M. F. Demirci, and A. Yazici, "Speech emotion recognition with deep convolutional neural networks," *Biomed Signal Process Control*, vol. 59, p. 101894, 2020.

[7] M. Ashraf *et al.*, "A hybrid cnn and rnn variant model for music classification," *Applied Sciences*, vol. 13, no. 3, p. 1476, 2023.

[8] M. K. Gourisaria, R. Agrawal, M. Sahni, and P. K. Singh, "Comparative analysis of audio classification with MFCC and STFT features using machine learning techniques," *Discover Internet of Things*, vol. 4, no. 1, p. 1, 2024.

[9] G. Owino, A. Waititu, A. Wanjoya, and J. Okwiri, "Autonomous Surveillance of Infants' Needs Using CNN Model for Audio Cry Classification," *Journal of Data Analysis and Information Processing*, vol. 10, no. 4, pp. 198–219, 2022.

[10] T. N. Maghfira, T. Basaruddin, and A. Krisnadhi, "Infant cry classification using CNN–RNN," in *Journal of Physics: Conference Series*, 2020, p. 12019.

[11] W. Bian, J. Wang, B. Zhuang, J. Yang, S. Wang, and J. Xiao, "Audio-based music classification with DenseNet and data augmentation," in *PRICAI 2019: Trends in Artificial Intelligence: 16th Pacific Rim International Conference on Artificial Intelligence, Cuvu, Yanuca Island, Fiji, August 26-30, 2019, Proceedings, Part III 16*, 2019, pp. 56–65.

[12] M. F. Nafiz, D. Kartini, M. R. Faisal, F. Indriani, and T. Hamonangan, "Automated Detection of COVID-19 Cough Sound using Mel-Spectrogram Images and Convolutional Neural Network," *J. Ilm. Tek. Elektro Komput. dan Inform*, vol. 9, no. 3, pp. 535–548, 2023.

[13] Y. Yohannes and R. Wijaya, "Klasifikasi Makna Tangisan Bayi Menggunakan CNN Berdasarkan Kombinasi Fitur MFCC dan DWT," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 8, no. 2, pp. 599–610, 2021.

[14] P. A. Riadi, M. R. Faisal, D. Kartini, R. A. Nugroho, D. T. Nugrahadi, and D. B. Magfira, "A Comparative Study of Machine Learning Methods for Baby Cry Detection Using MFCC Features," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 1, pp. 73–83, 2024.

[15] P. Kulkarni, S. Umarani, V. Diwan, V. Korde, and P. P. Rege, "Child cry classification-an analysis of features and models," in *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, pp. 1–7.

[16] E. Sutanto, F. Fahmi, W. Shalannanda, and A. Aridarma, "Cry Recognition for Infant Incubator Monitoring System Based on Internet of Things using Machine Learning.," *International Journal of Intelligent Engineering & Systems*, vol. 14, no. 1, 2021.

[17] C. A. Bratan *et al.*, "Dunstan Baby Language Classification with CNN," in *2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*, 2021, pp. 167–171.

[18] X. Qiao, S. Jiao, H. Li, G. Liu, X. Gao, and Z. Li, "Infant cry classification using an efficient graph structure and attention-based model," *Kuwait Journal of Science*, vol. 51, no. 3, p. 100221, 2024.

[19] D. Ćirić, Z. Perić, J. Nikolić, and N. Vučić, "Audio signal mapping into spectrogram-based images for deep learning applications," in *2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH)*, 2021, pp. 1–6.

[20] R. Yunida *et al.*, "LSTM and Bi-LSTM Models For Identifying Natural Disasters Reports From Social Media," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 5, no. 4, pp. 241–249, 2023.

[21] D. Joshi, J. Pareek, and P. Ambatkar, "Comparative study of Mfcc and Mel spectrogram for raga classification using CNN," *Indian J Sci Technol*, vol. 16, no. 11, pp. 816–822, 2023.

[22] H. de S. Moura, "Automatic Recognition of Baby Cry," 2022.

[23] A. Ustubioglu, B. Ustubioglu, and G. Ulutas, "Mel spectrogram-based audio forgery detection using CNN," *Signal Image Video Process*, vol. 17, no. 5, pp. 2211–2219, 2023.

[24] B. Ustubioglu, G. Tahaoglu, and G. Ulutas, "Detection of audio copy-move-forgery with novel feature matching on Mel spectrogram," *Expert Syst Appl*, vol. 213, p. 118963, 2023.

[25] T. Adhikari, "Designing a Convolutional Neural Network for Image Recognition: A Comparative Study of Different Architectures and Training Techniques," *Available at SSRN 4366645*, 2023.

[26] M. M. Taye, "Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions," *Computation*, vol. 11, no. 3, p. 52, 2023.

[27] Z. Guo, C. Yang, D. Wang, and H. Liu, "A novel deep learning model integrating CNN and GRU to predict particulate matter concentrations," *Process Safety and Environmental Protection*, vol. 173, pp. 604–613, 2023.

[28] S. L. Tan, G. Selvachandran, W. Ding, R. Paramesran, and K. Kotecha, "Cervical cancer classification from pap smear images using deep convolutional neural network models," *Interdiscip Sci*, vol. 16, no. 1, pp. 16–38, 2024.

[29] P. Kubi, A. Islam, M. A. H. Bin Zaher, and S. H. Ripon, "A Deep Learning-Based Technique to Determine Various Stages of Alzheimer's Disease from 3D Brain MRI Images," in *International Conference on Information Integration and Web Intelligence*, 2023, pp. 162–175.

[30] N. H. Arif, M. R. Faisal, A. Farmadi, D. Nugrahadi, F. Abadi, and U. A. Ahmad, "An Approach to ECG-based Gender Recognition Using Random Forest Algorithm," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 2, pp. 107–115, 2024.

[31] P. Rani *et al.*, "Baby Cry Classification Using Machine Learning," *Int. J. Innov. Sci. Res. Technol*, vol. 7, 2022.

[32] M. Mahmud *et al.*, "Implementation of C5. 0 Algorithm using Chi-Square Feature Selection for Early Detection of Hepatitis C Disease," *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, vol. 6, no. 2, pp. 116–124, 2024.
[33] D. Widhyanti and D. Juniati, "Classification of baby cry sound using higuchi's fractal dimension with K-nearest neighbor and support vector machine," in *Journal of Physics: Conference Series*, 2021, p. 12014.
[34] Sindi, Hatem, et al. "Random fully connected layered 1D CNN for solving the Z-bus loss allocation problem." Measurement 171 (2021): 108794.
[35] R. Lubis, C. Wardani, P. Nabila Daulay, A. Ayu Sahara, Y. Ritonga, and R. Wahyuni, "Laporan Mini Riset Perkembangan Peserta Didik Arti Tangis Bayi," *J. Pendidik. Inov.*, vol. 6, no. 3, pp. 633–643, 2024, [Online
[36] Aslan, Muhammet Fatih. "Comparative Analysis of CNN Models and Bayesian Optimization-Based Machine Learning Algorithms in Leaf Type Classification." Balkan Journal of Electrical and Computer Engineering, vol. 11, 2023, pp. 13–24, https://doi.org/10.17694/bajece.1174242.
[37] B. Zhang, J. Leitner, and S. Thornton, "Audio Recognition using Mel Spectrograms and Convolution Neural Networks," *Noiselab Univ. Calif.*, vol. 3, no. 4, pp. 1–5, 2019, [Online]. Available: http://noiselab.ucsd.edu/ECE228_2019/Reports/Report38.pdf.
[38] M. Farid, A. Rahman, and H. Wicaksono, "Analisis Pengaruh Kombinasi Fitur Spektral terhadap Tingkat Akurasi Speech Emotion Recognition," *J. Sist. Inf. dan Teknol.*, vol. 5, no. 2, pp. 120–129, 2023, doi: 10.37034/jsisfotek.v5i1.234.
[39] D. Lionel, R. Adipranata, and E. Setyati, "Klasifikasi Genre Musik Menggunakan Metode Deep Learning Convolutional Neural Network dan Mel- Spektrogram," *J. Infra Petra*, vol. 7, no. 1, pp. 51–55, 2019, [Online]. Available: http://publication.petra.ac.id/index.php/teknik-informatika/article/view/8044
[40] M. Putra, B. Suprapto, and S. Dwijayanti, "Pengenalan Dialek Di Sumatera Selatan Menggunakan Algoritma Deep Neural Network," in *Applicable Innovation of ...*, 2021, pp. 27–28. [Online].
[41] M. Diarsyah and D. Setiawan, "Implementasi CNN-LSTM untuk Music Captioning," *J. Media Inform.*, vol. 23, no. 1, pp. 21–33, 2024.
[42] Jiang, Ziyu, et al. "Comparisons of Convolutional Neural Network and Other Machine Learning Methods in Landslide Susceptibility Assessment: A Case Study in Pingwu." Remote Sensing, vol. 15, 2023, www.mdpi.com/2072-4292/15/3/798, https://doi.org/10.3390/rs15030798.

BIBLIOGRAPHY

RIDHA FAHMI JUNAIDI was born in Banjarmasin, South Kalimantan. Since 2018, he has pursued his academic endeavours as a student of the Computer Science Department at Universitas Lambung Mangkurat. His current area of research lies within the realm of voice data. Additionally, his final project entailed conducting research centring around audio classification within recordings sourced from medical settings. This research endeavour aimed to facilitate the recognition of infant emotions, thereby aiding in the medical field.

MOHAMMAD REZA FAISAL was born in Banjarmasin. After graduating from high school, he pursued his undergraduate studies in the Informatics department at Pasundan University in 1995 and later majored in Physics at Bandung Institute of Technology in 1997. After completing his bachelor's program, he gained experience as a training trainer in information technology and software development. Since 2008, he has been a lecturer in computer science at Universitas Lambung Mangkurat while also pursuing his master's program in Informatics at Bandung Institute of Technology in 2010. In 2015, he furthered his education by pursuing a doctoral degree in Bioinformatics at Kanazawa University, Japan. To this day, he continues his work as a lecturer in Computer Science at Universitas Lambung Mangkurat. His research interests encompass Data Science, Software Engineering, and Bioinformatics.

ANDI FARMADI serves as a faculty member within the Department of Computer Science at Lambung Mangkurat University. His primary area of research revolves around the field of Data Science. Prior to assuming his role as a lecturer, he successfully obtained his bachelor's degree in Physics from Hasanuddin University in 1999, followed by the completion of his master's degree at Bandung Institute of Technology in 2007. It was in 2008 that he commenced his tenure as a lecturer within the Department of Computer Science at Lambung Mangkurat University. The primary focus of his research endeavors lies within the realm of Data Science.

DODON TURIANTO NUGRAHADI is a lecturer in Department of Computer Science, Lambung Mangkurat University. His research interest is centred on Data Science and Computer Networking. He completed his bachelor's degree in Informatics Engineering in the UK. Petra, Surabaya in 2004. After that, he pursued a master's degree in Information Engineering at Gajah Mada University, Yogyakarta, in 2009. His current area of research revolves around Network, Data Science, Internet of Things (IoT), and network Quality of service (QoS)

RUDY HERTENO was born in Banjarmasin, South Kalimantan. After completing high school, he pursued his undergraduate studies in the Computer Science Department at Lambung Mangkurat University and graduated in 2011. Following his undergraduate program, he gained several years of experience as a software developer, particularly developing software for local governments. In 2017, he obtained his master's degree in Informatics from STMIK Amikom University. Currently, he serves as a lecturer in the Mathematics and Natural Sciences faculty at Lambung Mangkurat University. His research interests encompass software engineering, software defect prediction, and deep learning.

LUU DUC NGO was born and raised in Bac Lieu Province, Vietnam. In 1997, he graduated from Can Tho University, Vietnam, with an engineering degree in Informatics. In 2007, he received a master's degree in Computer Science from the University of Information Technology (UIT), National University of Ho Chi Minh City (VNU-HCM). From 2013 to 2016, he studied in the PhD program in the Bioinformatics Laboratory at Kanazawa University (KU), Japan. He graduated from KU and received a PhD Degree in Computer Science. He was a lecturer at Bac Lieu Continuous Education Center, Vietnam, from 1997 to 2007. Then, he moved to Bac Lieu University (BLU), Vietnam, where he worked as a lecturer and dean of the IT faculty. To this day, he is working at BLU as a lecturer and chairman of the University Council. His research interests encompass machine learning, deep learning, data mining, text mining, natural languate processing, and bioinformatics for various areas such as education, biology, medicine, agriculture, aquaculture, and climate.

BAHRIDDIN ABAPIHI was born in Buton Island, Southeast Sulawesi Province, Indonesia. In 1991, he studied in Faculty of Education at Halu Oleo University in Kendari, Indonesia, but one year later he got scholarship from EIUDP-CIDA (Eastern Indonesia University Development Project-Canadian International Development Agency) to switch his study to IPB University in Bogor, Indonesia. He then received his bachelor and master degrees in Statistics in 1997 and 2001, respectively. In 2015, he enrolled in the PhD program in the Bioinformatics Laboratory of Kanazawa University, Japan, and then received a PhD degree in Electrical Engineering and Computer Science in 2018. He worked at Pakuan University in Bogor, Indonesia, from 2001 to 2003 as a lecturer. He moved to Halu Oleo University in 2003 and work as a lecturer until this day. He was the head of Department of Statistics in 2011 to 2015. His research interests cover machine learning, data mining, bioinformatics, statistics and data sciences for various fields of application such as agriculture, health, economics, and education.