Eastern Anatolian Journal of Science
Volume IX, Issue I, 2023, 16-26

Eastern Anatolian Journal of Science

Classification of Baby Cries Using Machine Learning Algorithms

Adem EKİNCİ*[1], Enver KÜÇÜKKÜLAHLI[2]

[1] *Düzce University, Mathematics Master's Degree Program, Düzceı, Turkiye,*
adem3ekinci@gmail.com (ORCID: 0009-0003-4702-4835)

[2] *Düzce University, Faculty of Engineering, Department of Computer Engineering, Düzceı, Turkiye,*
enverkucukkulahli@duzce.edu.tr (ORCID: 0000-0002-0525-0477)

Abstract

People are constantly engaged in communication with each other, and they mostly do so through language. The most effective form of communication for a newborn baby until they acquire this skill is crying. Although baby cries are often perceived as bothersome by adult individuals, they can contain a wealth of information. In this study, the information contained in infant cry signals was interpreted using audio processing methods and classified using machine learning algorithms. Feature extraction was performed on the dataset using the Mel Frequency Cepstral Coefficients (MFCC) method, and performance metrics were measured after using k-NN, SVM, Random Forest, and MLP classification algorithms

In the donate-a-cry dataset used in the study, before the feature extraction, the data was divided into 10 equal parts to increase the number of data and it has been observed that this method increases the classification success. For instance the k-NN algorithm achieved a performance value of 85.78% before the method was applied then after the method was applied, the performance value increased by 13.11% and reach 98.88%.

Keywords: Baby cries, Machine learning, SVM, Random Forest, MLP, k-NN

Received: 14.06.2023
Accepted: 26.06.2023
Published: 30.06.2023
**Corresponding author:* Adem EKİNCİ, *Düzce University, Mathematics Master's Degree Program, Düzceı, Turkiye.*
E-mail: adem3ekinci@gmail.com
Cite this article as: A. Ekinci and E. Küçükkülahlı., Classification of Baby Cries Using Machine Learning Algorithms, Eastern Anatolian Journal of Science, Vol. 9, Issue 1, 16-26, 2023.

1. Introduction

Crying is an innate response observed in all babies and serves as an effective means of communication for them. They can express their hunger, discomfort, sleeplessness, or when something is amiss through crying. Therefore, a baby's cry is one of the most important signals that caregivers need to pay attention to. When the indicated situation is understood correctly and the issues are addressed, babies become happier, which positively impacts their healthy development (Lahti et al., 2019). In this regard, it is crucial to identify the reason quickly and accurately behind a baby's crying. Mothers, experienced caregivers, or healthcare professionals may be competent in this regard, but not always. A father who doesn't spend much time with the baby, an inexperienced caregiver, or a healthcare professional may struggle to understand the cause of crying, which can prolong the duration of crying. Prolonged crying in babies leads to an increase in cortisol levels, which can have negative effects on the brain and cognitive functions (Halpern & Coelho, 2016). Therefore, flawless and prompt methods should be preferred to quickly stop crying by meeting the baby's genuine needs. In this respect the main purpose of this study is to determine the cause of baby crying with higher accuracy using machine learning. Considering that baby crying is a sound signal, sound processing can provide more effective results in this context.

Not every sound we hear in our daily lives carries meaning to us. An unknown language, the sound of a bird, or a baby crying. However, when these sounds are recorded and analyzed using sound processing methods, it is observed that they carry more information than what we perceive. Sound processing fundamentally involves operations such as organizing sound signals, reducing noise, conducting spectral analysis, filtering, visualization, and feature extraction (Purwins et al., 2019). Feature extraction is the process of distinguishing the significant characteristics of data in a dataset. Features emerge from the differences in time, frequency, and other properties of sound data. Various methods can be employed for feature extraction, including Mel Frequency Cepstral Coefficients (MFCC), Linear Predictive Coding (LPC), and Short-Time Fourier Transform (STFT) (Ji et al., 2021; Rezaee et al., 2023a). After this process, the data containing sound becomes compatible with machine learning algorithms. Consequently, numerous operations such as speech recognition, speaker recognition, emotion detection, disease diagnosis, and music recognition can be performed.

Machine learning is one of the most significant branches of artificial intelligence. It involves creating and developing systems that acquire learning capabilities by analyzing data or experiences. Machine learning algorithms are trained on datasets, and the quality, diversity, size, and accuracy of the data directly impact the performance of the algorithms. Therefore, the dataset serves as the fundamental basis for machine learning (Zhou, 2021).

Insufficient data in a dataset can lead to overfitting or a decrease in the model's generalization ability. Data augmentation methods can be employed to prevent such situations. Data augmentation techniques involve manipulating existing data to generate new data (Zhou, 2021). Some well-known data augmentation methods include random cropping, random shifting, mirroring, rotation, time stretching, and others (Huang et al., 2019).

A baby's crying is essentially a sound signal. Like other sound signals, it has a specific frequency, tone, and intensity. Any operation used in sound processing can be applied to these sound signals. Feature extraction is one of these operations. Baby cry signals with extracted features can be processed using machine learning. Thus, a machine learning model trained with sufficient data and the right methods can accurately identify the reasons why babies cry. Although the notion of Gröbner basis firstly handled with the current name by Buchberger in his PhD thesis (Buchberger 1965), in 1927, Macaulay had been used this idea in his famous paper (Macaulay 1927).

2. *Related Work*

In a study conducted by Michelsson and colleagues, the analysis of newborn infants' crying sounds using spectrographic analysis was discussed, highlighting the differences between healthy and sick infants (Michelsson & Michelsson, 1999).

Sharma and colleagues achieved an accuracy of 81.27% using Gaussian mixture model clustering for the Donate-a-cry dataset (Sharma et al., 2019).

Kulkarni and colleagues classified infant cries using various features such as MFCC, Spectral Flatness, GTCC, STACF, Spectral Roll-off, etc., along with LR, SVM, k-NN, and RF algorithms, obtaining an accuracy of 84% with MFCC and LR combination (Kulkarni et al., 2021).

Dewi and colleagues compared the performance of MFCC and LFCC for infant cry samples using the K-NN algorithm, finding that LFCC was more effective (Dewi et al., 2019).

Sutanto and colleagues discussed an Internet of Things (IoT)-based baby incubator monitoring system to identify infant cries and monitor their health conditions, utilizing the donate-a-cry dataset (Sutanto et al., 2021).

Messaoud and colleagues achieved a 71.4% accuracy by segmenting 1615 sound samples from 13 infants and using the MFCC method with PNN (Messaoud & Tadj, 2010).

Izmirli employed a spectral flatness-based audio segmentation method in their study (Izmirli, 2000).

Bashiri and colleagues achieved a 99.9% accuracy using the MFCC and ANN algorithms with the Baby-Chilanto dataset (Bashiri & Hosseinkhani, 2020). Other studies using the same dataset include Hariharan et al., who achieved 99% accuracy with STFT-GRNN (Hariharan et al., 2012), Hariharan et al., who obtained 99.49% accuracy with Wavelet Packets-PNN (Hariharan et al., 2011), and Sahak et al., who achieved 95.07% accuracy with BPSO-ANN (Sahak et al., 2010).

In a study by Razaee and colleagues, using the donate-a-cry dataset and the deep-SVM algorithm with a segmentation method, an accuracy of 98.4% was obtained (Rezaee et al., 2023b).

Apart from the Baby-Chilanto and donate-a-cry datasets, some studies in the literature have utilized the Dunstan Baby Language dataset for baby sound classification (Bănică et al., 2016; Franti et al., 2018; Maghfira et al., 2020).

3. Material

3.1. Dataset

Donate-a-cry is created through parents recording their children's crying sounds and relevant information using a mobile application installed on IOS and Android smartphones. The contributors who uploaded the audio recording also labeled the crying according to the given instructions and indicated the cause of crying they suspected. It was initially published on the GitHub platform by Gaber Veres in 2015 (Veres, 2015/2023). The dataset was later updated, applying preliminary audio processing steps to each recording. The sound files consist of baby cries ranging from 1 to 22 months. The files have a frequency of 1800 and a bitrate value of 128 kbps. The updated version of the dataset was used in this study. The distribution of data into classes is provided in Table 1. The database is published under the ODbL.

Reason	Records
Belly pain	16
Burping	8
Discomfort	27
Hungry	382
Tired	24
Total	457

Table 1 The distribution of classes in the Donate-a-cry dataset.

3.2. Feature Extraction

Feature extraction is essentially creating numerical representations that capture the important characteristics of the sound. In this study, Mel-Frequency Cepstral Coefficients (MFCC) were used for feature extraction.

MFCC (Mel-Frequency Cepstral Coefficients) is a representation of the short-term power spectrum of a sound signal on the Mel scale, which represents the perceptual changes in human hearing of sound frequencies (Figure 2). The MFCC method divides the sound signal into short time frames and performs a power spectrum analysis (Figure 1) for each frame. In this analysis, the energy of the sound in different frequency bands is calculated and weighted in a manner similar to the frequency sensitivity of the human ear. Subsequently, logarithmic processing and Discrete Cosine Transform (DCT) are applied to transform the frequency scale, resulting in the extraction of MFCC coefficients. These coefficients represent different features of the sound signal and can be used in various applications such as speech recognition, speaker identification, and music classification.

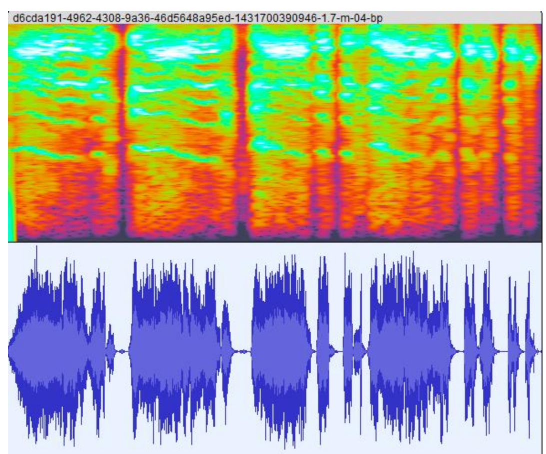

Figure 1 A sound signal and its spectrogram from the Belly pain class.

MFCC consists of the following steps (Tiwari, n.d.):

- First, the sound signal is converted from analog to digital, and the sampling frequency is determined.
- Then, the sound signal is passed through a pre-emphasis filter. This filter enhances the energy in high frequencies and improves speech recognition performance.

$$y[n] = x[n] - \alpha\, x[n-1]$$

where $y[n]$ is the filtered signal, $x[n]$ is the input signal, and α is the pre-emphasis coefficient.

- Next, the sound signal is divided into short time intervals called frames. Typically, a frame length of 25 ms and a frame shift of 10 ms are used. A Hamming or Hanning window is applied to the frames to prevent abrupt transitions between frames.

$$w[n] = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$$

where, $w[n]$ is the window function, n is the frame index, and N is the frame length.

- Then, the spectrum is calculated for each frame. The spectrum represents the frequency components of the sound signal. Fast Fourier Transform (FFT) is used to calculate the spectrum.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{\frac{-j2\pi kn}{N}}$$

where, $X[k]$ is the spectrum, $x[n]$ is the time signal, N is the frame length, and k is the frequency index.

- The spectrum is then weighted by a Mel-filter bank. The Mel-filter bank consists of a series of triangular filters that mimic the frequency scale perceived by the human ear.

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k-f(m-1)}{f(m)-f(m-1)} & f(m-1) \le k < f(m) \\ \frac{f(m+1)-k}{f(m+1)-f(m)} & f(m) \le k < f(m+1) \\ 0 & k \ge f(m+1) \end{cases}$$

where, $H_m(k)$ is the value of the $m-th$ filter at the $k-th$ frequency, and $f(m)$ is the center frequency of the $m-th$ filter.

- Finally, the logarithm of the spectrum passed through each filter is taken, and the inverse Fourier transform is applied to obtain the MFCC.

$$E = \sum_{n=0}^{N-1} x^2[n]$$

where, E is the energy and $x[n]$ is the time signal.

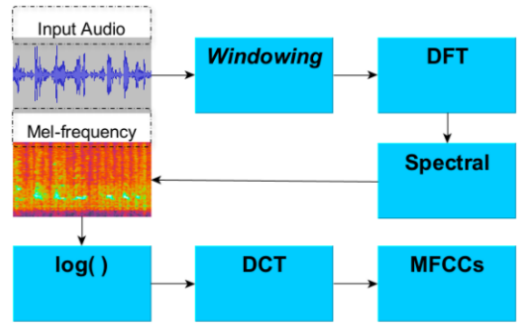

Figure 2 MFCC process diagram.

3.3. Machine Learning

Machine learning, one of the most used subfields of artificial intelligence, has found its place in many studies related to the classification of baby cries. In this study, four different classification algorithms of machine learning were modeled, and their classification performances were examined.

3.3.1. Random Forest

Random forest algorithm is a powerful and flexible machine learning technique used in supervised learning. This algorithm can solve classification or regression problems by combining multiple decision trees. It is based on the concept of ensemble learning, which involves aggregating multiple classifiers together to improve model performance and solve complex problems.

During training, the algorithm constructs multiple decision trees and works by determining the class that is most frequent (for classification) or the average prediction (for regression) among the individual tree's classes. The decision trees are constructed using a random subset of features and data with replacement (bootstrapping). This introduces randomness and diversity among the trees, reducing the correlation and variance of predictions (Probst et al., 2019).

3.3.2. Support Vector Machine (SVM)

SVM is a supervised learning algorithm used for classification problems. The SVM algorithm draws a hyperplane to separate two or more classes. This hyperplane maximizes the margin, which is the gap between the two classes. The margin represents the distance between the closest points of the two classes.

The SVM algorithm can project data points into higher dimensions to expand the margin. It achieves this by using functions called kernels. Kernel functions measure how similar or distant data points are from each other. Different kernel models, such as Linear kernel, Polynomial kernel (Poly), Radial Basis Function (RBF) kernel, Sigmoid kernel, are available.

The SVM algorithm can be mathematically expressed as follows:

$$\min_{w,b} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} \xi_i$$

where w and b are the parameters of the hyperplane. The parameter C adjusts the balance between expanding the margin and reducing classification errors. ξ_i is a loss function that determines how much a data point enters the margin region or deviates from the correct class. This formula aims to optimize both the margin and the classification errors of the hyperplane (Pisner & Schnyer, 2020).

3.3.3. Multilayer Perceptron (MLP)

MLP (Multilayer Perceptron) is an important model in the context of the approximation theory within neural networks. It consists of an input layer, one or more hidden layers, and an output layer. Units (neurons) in each layer are interconnected with connections from the input layer to the output layer. The basic structure of an MLP is to process input data through the hidden layers using weights and activation functions to produce an output. In each hidden layer, the inputs are multiplied by weights, subjected to an activation function, and passed on to subsequent layers. This process is repeated as it passes through the hidden layers, ultimately resulting in an output in the output layer. The MLP model is used to capture complex and non-linear relationships in input data. This is done by employing various activation functions such as sigmoid, ReLU, tanh. The MLP adjusts its weights through a learning process that fits the data, and it can then make predictions on new data. The MLP has a wide range of applications and can achieve successful results in various problems such as pattern recognition, classification, regression, and time series analysis. Additionally, the presence of multiple hidden layers in an MLP allows for the learning of more complex relationships and features (Taud & Mas, 2018).

3.3.4. k-Nearest Neighbors (k-NN)

The k-NN (k-Nearest Neighbors) algorithm is a supervised learning method that predicts the class or value of a given instance based on its closest neighbors. The steps of this algorithm are as follows:

First, the k parameter is selected. This parameter represents the number of nearest neighbors to consider for a given instance. For example, if k=3, the closest 3 neighbors will be considered.
The distance between the given instance and other instances is calculated. Various distance measures can be used for this purpose, such as Euclidean distance, Manhattan distance, Minkowski distance, and so on.
Next, the distances are sorted in ascending order, and the k instances with the smallest distances are identified. These instances are referred to as the nearest k neighbors.
If the problem is a classification problem, the classes of the nearest k neighbors are examined, and the most commonly occurring class is predicted.
If the problem is a regression problem, the average of the target variable values of the nearest k neighbors is taken, and this value is presented as the prediction.
The k-NN algorithm is a simple and easy-to-implement method. However, it also has some disadvantages. For example, choosing the right value for the k parameter is important, as having a value that is too large or too small can reduce the accuracy. Additionally, computing distances with the entire dataset for each new instance can be computationally expensive in terms of time and memory (Kramer, 2013).

3.4. Performance Evaluation Scores

Confusion Matrix: It is a matrix used to evaluate the performance of a classification model and shows the number of correct and incorrect predictions made by the model. The confusion matrix consists of four main categories: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). These categories indicate in which cases the classification model made correct or incorrect predictions.

Accuracy: Represents the ratio of correctly classified examples to the total number of examples. In other words, it shows the percentage of correctly classified examples among all the examples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: It is a value that indicates how many of the examples predicted as positive by a classification model are truly positive. In simpler terms, precision represents the ratio of correct positive predictions to the total positive predictions made by the model.

$$Precision = \frac{TP}{TP + FP}$$

Recall (Sensitivity): It is a metric that shows how much of the true positive examples a classification model correctly identifies. Recall is also referred to as "sensitivity".

$$Recall = \frac{TP}{TP + FN}$$

F1 Score: It is the harmonic mean of precision and recall metrics and helps evaluate the performance of the model from a broader perspective.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

4. Method

As a data augmentation method, each data in the dataset has been divided into 10 parts to create a second dataset. This method has been used for music genre classification in (Velardo, 2020). For each part in this resulting dataset, feature extraction is applied using MFCC (Mel-frequency cepstral coefficients). To apply MFCC, the audio signals of each part are divided into windows with a sampling rate of 22050 Hz and a 50% overlap. Windowing helps us visualize the time and frequency components of the signal better and reduces side effects in Fourier transformation. Then, a Mel-frequency filter bank and discrete cosine transformation are applied to each window. As a result, 13 MFCC coefficients are calculated for each part.

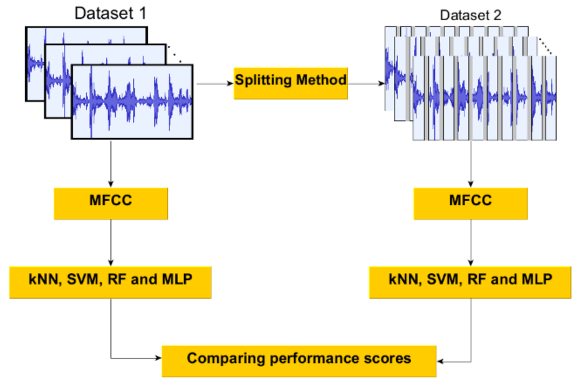

Figure 3 Flowchart of the model proposed

Using these generated MFCC coefficients, performance evaluation scores are calculated for the four metrics of the confusion matrix in algorithms.

Some important parameters used in the algorithms are provided in the table below.

	P1	P2	P3	P4
k-NN	k = 5	Distance = Minkowski	-	-
SVM	Kernel = RBF	c = 12	-	-
RF	Trees = 100	Tree Depth = 32	Criterion = Entropy	-
MLP	Alpha = 0.01	Max Iterations = 600	Hidden Layer Size = 12	Algorithm = LBFGS

To ensure more reliable and unbiased detection of each metric, a 10-fold cross-validation method is used. This method creates different training and test sets, utilizing the entire dataset and preventing data wastage. Also Each algorithm was run 30 times with the same parameters for result verification.

5. Findings

The scores identified in the study are provided in the tables 2 and 3 below. Table 2 has been examined for four algorithms. As seen in Table 2, using the data augmentation method, a 98.88% accuracy was achieved with the k-NN algorithm. The impact of the data augmentation method on the results is more clearly seen in Table 3. The performance values have increased by 15.69% for SVM, 13.11% for k-NN, 11.91% for Random Forest, and 11.42% for MLP. Additionally, the confusion matrices for dataset1 and dataset2 are provided in Figures 4, 5, 6 and 7 for four algorithms.

		Accuracy Score	Recall (Sensitivity) Score	Precision Score	F1 Score
k-NN	Max	98,88%	98,88%	98,89%	98,89%
	Min	98,88%	98,88%	98,89%	98,89%
	Mean	**98,88%**	**98,88%**	**98,89%**	**98,89%**
SVM	Max	98,60%	98,60%	98,60%	98,60%
	Min	98,60%	98,60%	98,60%	98,60%
	Mean	**98,60%**	**98,60%**	**98,60%**	**98,60%**
Random Forest	Max	98,58%	98,88%	98,56%	98,54%
	Min	98,18%	98,18%	98,14%	98,13%
	Mean	**98,33%**	**98,33%**	**98,30%**	**98,29%**
MLP	Max	98,42%	98,42%	98,45%	98,44%
	Min	97,86%	97,86%	97,88%	97,86%
	Mean	**98,23%**	**98,23%**	**98,24%**	**98,23%**

Table 2 Performance scores obtained after running the algorithms 30 times.

	Accuracy Score for 10 Segments	One Piece Accuracy Score	Difference
k-NN	98,88%	85,78%	**13,11%**
SVM	98,18%	82,49%	**15,69%**
Random Forest	98,38%	86,47%	**11,91%**
MLP	98,60%	87,18%	**11,42%**

Table 3 The Impact of Data Augmentation Method on Performance Metrics

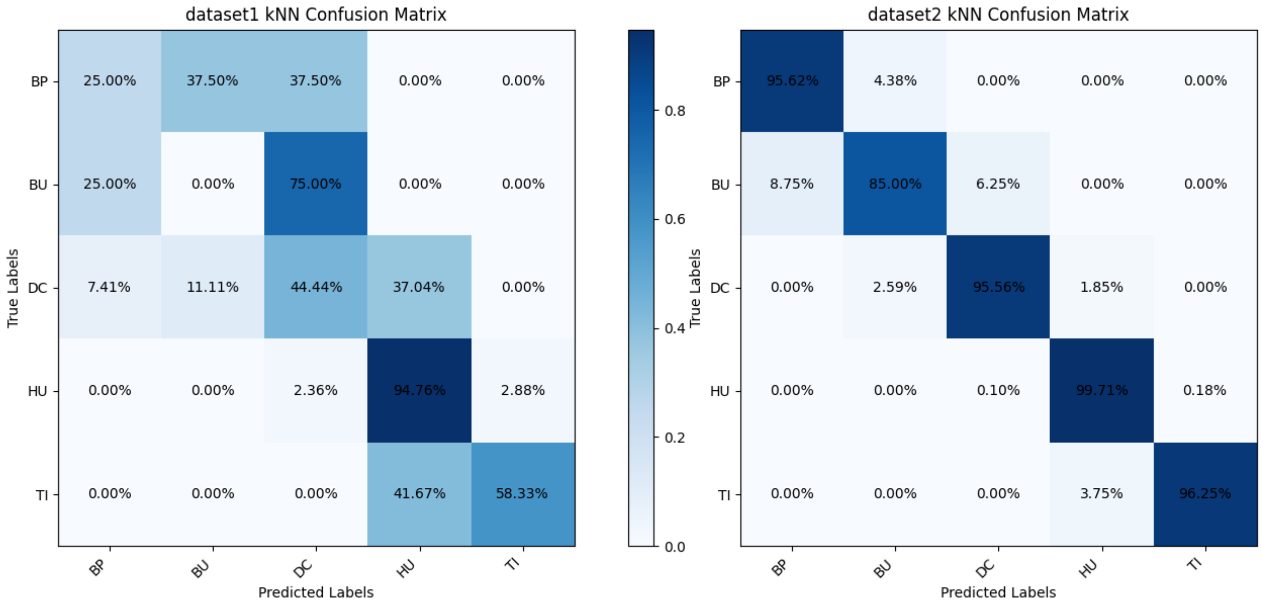

Figure 4: Confusion matrix of dataset1 and dataset2 with k-NN algorithm

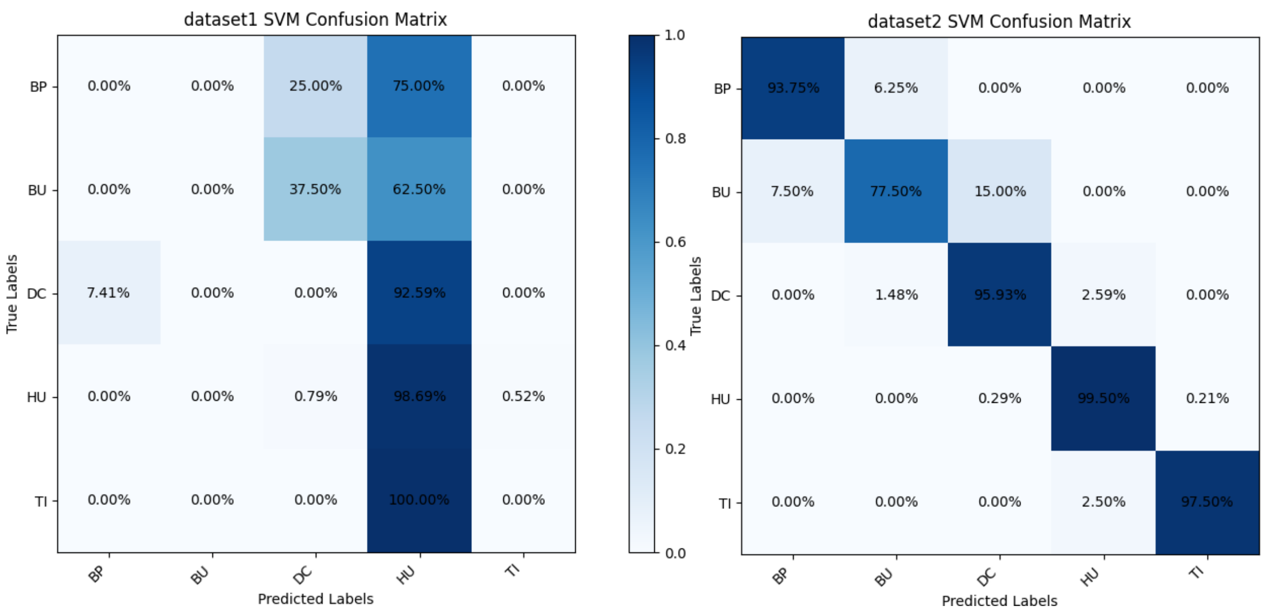

Figure 5:SVM Confusion matrix of dataset1 and dataset2 with SVM algorithm

Figure 6: Confusion matrix of dataset1 and dataset2 with RF algorithm

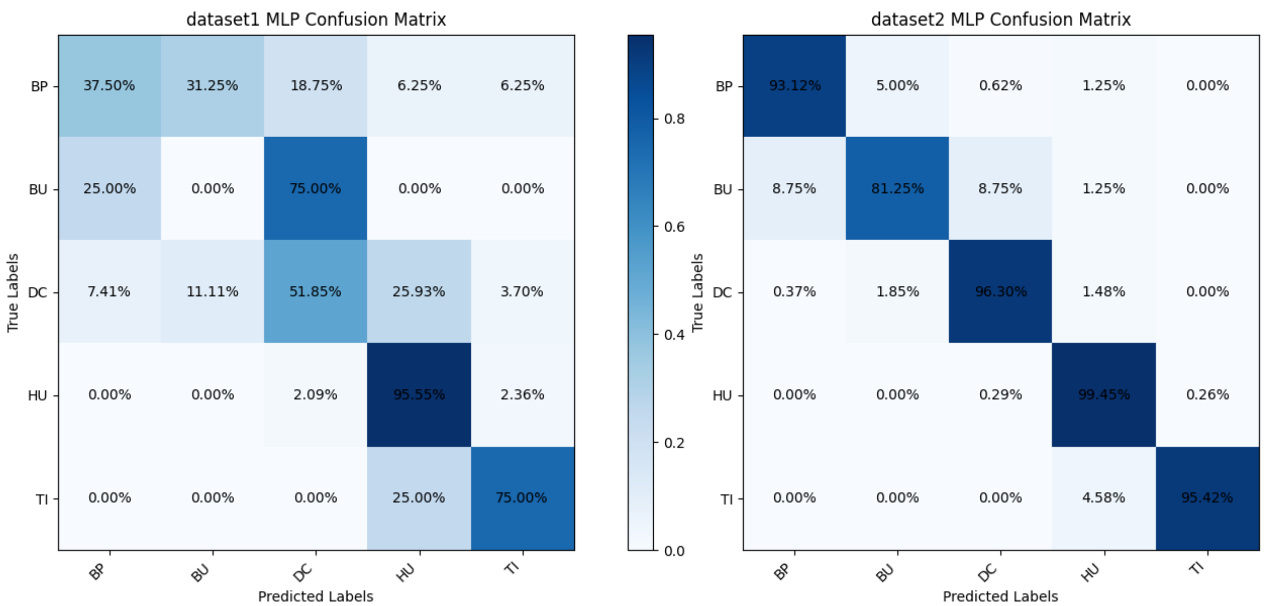

Figure 7: Confusion matrix of dataset1 and dataset2 with MLP algorithm

5. Discussion and Conclusions

In our study, the method of augmenting the size of the database through dividing the data into equal segments was employed for classification of baby cries which used in (Velardo, 2020) for music genre classification. The results indicate that this method has had a positive impact on identifying the causes of baby cries.

The Donate-a-cry dataset used in our study has also been employed in previous research (Sharma et al., 2019; Kulkarni et al., 2020; Sutanto et al., 2021; Rezaee et al., 2023b). Among these studies, the work by Rezaee et al. (2023b) achieved the highest performance with a deepSVM algorithm, reaching an accuracy of 98.34%. In this study, by employing data augmentation techniques and the k-NN algorithm, a performance of 98.88% was attained, surpassing the previous works that utilized the donate-a-cry dataset, thereby achieving the highest success rate.

Based on the performance scores, it has been considered as a strong assumption that the sound of baby's crying for a specific reason exhibits characteristic features even in small segments.

References

BĂNICĂ, I.-A., CUCU, H., BUZO, A., BURILEANU, D., & BURILEANU, C. (2016). Automatic methods for infant cry classification. 2016 International Conference on Communications (COMM), 51-54. https://doi.org/10.1109/ICComm.2016.7528261

BASHIRI, A., & HOSSEINKHANI, R. (2020). Infant Crying Classification by Using Genetic Algorithm and Artificial Neural Network. Acta Medica Iranica, 531-539. https://doi.org/10.18502/acta.v58i10.4916

DEWI, S. P., PRASASTI, A. L., & IRAWAN, B. (2019). The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods. 2019 IEEE International Conference on Signals and Systems (ICSigSys), 18-23. https://doi.org/10.1109/ICSIGSYS.2019.8811070

FRANTI, E., ISPAS, I., & DASCALU, M. (2018). Testing the Universal Baby Language Hypothesis—Automatic Infant Speech Recognition with CNNs. 2018 41st International Conference on Telecommunications and Signal Processing (TSP), 1-4. https://doi.org/10.1109/TSP.2018.8441412

HALPERN, R., & COELHO, R. (2016). Excessive crying in infants. Jornal De Pediatria, 92(3 Suppl 1), S40-45. https://doi.org/10.1016/j.jped.2016.01.004

HARIHARAN, M., SINDHU, R., & YAACOB, S. (2012). Normal and hypoacoustic infant cry signal classification using time-frequency analysis and general regression neural network. Computer Methods and Programs in Biomedicine, 108(2), 559-569. https://doi.org/10.1016/j.cmpb.2011.07.010

HARIHARAN, M., YAACOB, S., & AWANG, S. A. (2011). Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. Expert Systems with Applications, 38(12), 15377-15382. https://doi.org/10.1016/j.eswa.2011.06.025

HUANG, L., PAN, W., ZHANG, Y., QIAN, L., GAO, N., & WU, Y. (2019). Data Augmentation for Deep Learning-based Radio Modulation Classification (arXiv:1912.03026; Versiyon 1). arXiv. http://arxiv.org/abs/1912.03026

IZMIRLI, Ö. (2000, Ocak 1). Using a Spectral Flatness Based Feature for Audio Segmentation and Retrieval.

JI, C., MUDIYANSELAGE, T. B., GAO, Y., & PAN, Y. (2021). A review of infant cry analysis and classification. EURASIP Journal on Audio, Speech, and Music Processing, 2021(1), 8. https://doi.org/10.1186/s13636-021-00197-5

KRAMER, O. (2013). K-Nearest Neighbors. Içinde O. Kramer (Ed.), Dimensionality Reduction with Unsupervised Nearest Neighbors (ss. 13-23). Springer. https://doi.org/10.1007/978-3-642-38652-7_2

KULKARNI, P., UMARANI, S., DIWAN, V., KORDE, V., & REGE, P. P. (2021). Child Cry Classification—An Analysis of Features and Models. 2021 6th International Conference for Convergence in Technology (I2CT), 1-7. https://doi.org/10.1109/I2CT51068.2021.9418129

LAHTI, K., VÄNSKÄ, M., QOUTA, S. R., DIAB, S. Y., PERKO, K., & PUNAMÄKI, R. (2019). Maternal experience of their infants' crying in the context of war trauma: Determinants and consequences. Infant Mental Health Journal, imhj.21768. https://doi.org/10.1002/imhj.21768

MAGHFIRA, T. N., BASARUDDIN, T., & KRISNADHI, A. (2020). Infant cry classification using CNN – RNN. Journal of Physics: Conference Series, 1528(1), 012019. https://doi.org/10.1088/1742-6596/1528/1/012019

MESSAOUD, A., & TADJ, C. (2010). A Cry-Based Babies Identification System. 6134, 192-199. https://doi.org/10.1007/978-3-642-13681-8_23

MICHELSSON, K., & MICHELSSON, O. (1999). Phonation in the newborn, infant cry. International Journal of Pediatric Otorhinolaryngology, 49 Suppl 1, S297-301. https://doi.org/10.1016/s0165-5876(99)00180-9

PISNER, D. A., & SCHNYER, D. M. (2020). Chapter 6—Support vector machine. Içinde A. Mechelli & S. Vieira (Ed.), Machine Learning (ss. 101-121). Academic Press. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

PROBST, P., WRIGHT, M. N., & BOULESTEIX, A.-L. (2019). Hyperparameters and tuning strategies for random forest. WIREs Data Mining and Knowledge Discovery, 9(3), e1301. https://doi.org/10.1002/widm.1301

PURWINS, H., LI, B., VIRTANEN, T., SCHLÜTER, J., CHANG, S.-Y., & SAINATH, T. (2019). Deep Learning for Audio Signal Processing. IEEE Journal of Selected Topics in Signal Processing, 13(2), 206-219. https://doi.org/10.1109/JSTSP.2019.2908700

REZAEE, K., GHAYOUMI ZADEH, H., QI, L., RABIEE, H., & KHOSRAVI, M. R. (2023a). 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. https://doi.org/10.1145/3579032

REZAEE, K., GHAYOUMI ZADEH, H., QI, L., RABIEE, H., & KHOSRAVI, M. R. (2023b). 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. https://doi.org/10.1145/3579032

SAHAK, R., LEE, Y. K., MANSOR, W., YASSIN, A. I. M., & ZABIDI, A. (2010). Optimized Support Vector Machine for classifying infant cries with asphyxia using Orthogonal Least Square. 2010 International Conference on Computer Applications and Industrial Electronics, 692-696. https://doi.org/10.1109/ICCAIE.2010.5735023

SHARMA, K., GUPTA, C., & GUPTA, S. (2019). Infant Weeping Calls Decoder using Statistical Feature Extraction and Gaussian Mixture Models. 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 1-6. https://doi.org/10.1109/ICCCNT45670.2019.8944527

SUTANTO, E., FAHMI, F., SHALANNANDA, W., & ARIDARMA, A. (2021). Cry Recognition for Infant Incubator Monitoring System Based on Internet of Things using Machine Learning. International Journal of Intelligent Engineering and Systems, 14, 444-452. https://doi.org/10.22266/ijies2021.0228.41

TAUD, H., & MAS, J. F. (2018). Multilayer Perceptron (MLP). Içinde Geomatic Approaches for Modeling Land Change Scenarios (ss. 451-455). Springer, Cham. https://doi.org/10.1007/978-3-319-60801-3_27.

VERES, G. (2023). Donateacry-corpus. https://github.com/gveres/donateacry-corpus (Original work published 2015)

ZHOU, Z.-H. (2021). Machine Learning. Springer. https://doi.org/10.1007/978-981-15-1967-3

VELARDO, V. (2020). Deep Learning for Audio with Python, https://github.com/musikalkemist/DeepLearningForAudioWithPython/blob/master/12-%20Music%20genre%20classification:%20Preparing%20the%20dataset/code/extract_data.py