Child Cry Classification – An Analysis of Features and Models

Conference Paper · April 2021

DOI: 10.1109/I2CT51068.2021.9418129

5 authors, including:

Prathamesh Kulkarni
College of Engineering, Pune
2 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Sarthak Umarani
College of Engineering, Pune
1 PUBLICATION **4** CITATIONS
SEE PROFILE

Vaishnavi Diwan
College of Engineering, Pune
1 PUBLICATION **4** CITATIONS
SEE PROFILE

Vishakha Korde
College of Engineering, Pune
1 PUBLICATION **4** CITATIONS
SEE PROFILE

2021 6th International Conference for Convergence in Technology (I2CT)
Pune, India. Apr 02-04, 2021

Child Cry Classification – An Analysis of Features and Models

Prathamesh Kulkarni
College of Engineering Pune(COEP),
Pune, India
prathameshsp17.extc@coep.ac.in

Sarthak Umarani
College of Engineering Pune(COEP),
Pune, India
sarthakns17.extc@coep.ac.in

Vaishnavi Diwan
College of Engineering Pune(COEP),
Pune, India
ravindradr17.extc@coep.ac.in

Vishakha Korde
College of Engineering Pune(COEP),
Pune, India
subodhks17.extc@coep.ac.in

Priti P. Rege
College of Engineering Pune(COEP),
Pune, India
ppr.extc@coep.ac.in

2021 6th International Conference for Convergence in Technology (I2CT) | 978-1-7281-8876-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/I2CT51068.2021.9418129

***Abstract* —This paper presents a study on the classification of child cries based on various features extracted through speech and auditory processing. Certain spectral and descriptive features vary significantly in a child's cry intended for a specific purpose. Firstly, the model was trained using individual features. Later, the best features were selected and the model was again trained by combining these features. Logistic regression, SVM, KNN and Random Forest models were used for classification. A total of 457 samples were used for training/testing the models from the dataset Donate-a-cry corpus.**

Keywords — MFCC, GFCC, KNN, SVM, random forest, feature extraction, spectrogram

I. INTRODUCTION

The analysis of biological signals is important for medical diagnoses. A child does not know an explicit way of communication, it can indicate its needs through a cry. The information about the emotional and physical condition of a baby can be extracted from the sound.

The first oral communication of babies is through crying before they start expressing their feelings through speech. At times, it becomes difficult to know the reason why an infant cries. This leads to frustration for a caregiver or mother.

Infant mortality rate is inferred as child death rate before completion of the first year. The datum released by World Health Organization (WHO) infers that the infant mortality rate is million. The main reason behind this is due to health issues. Nearly 75% of infant deaths shall be avoided if the disease is predicted at an earlier stage. Therefore, to rectify this issue, designed a child cry classification system to classify infant cries is in need. From the cry signals of the baby it is possible to classify the need of the baby by extracting specific features from that sound signal.[2]

II. RELATED WORK

In [26], a classification model is developed based on three criteria: the degree of overfitting, accuracy, conformability. Total of 468 cry audios are used in the dataset. Different features were extracted for each cry (which included first six formants, intensity, jitter and shimmer values, Harmonic to Noise Ratio (HNR), degree and number of voice breaks, unvoiced pitch fraction and cry duration) Decision Tree, KNN, LDA, LR, SVM, ANN were used for observing the results of the classification models.

In [13] total of 1615 cry samples were analyzed. First, they were pre-processed (silence removal and filtering with 4th order LPF to remove the noise). Each frame was of 25ms with 20% overlap. Different audio features such as pitch, intensity, jitter were proposed and MFCCs were extracted to carry out the classification using PNN (Probabilistic Neural Network).

A radial basis function (RBF) network is implemented for cry classification of infants. Features are extracted from each cry included F0, F1, voiced-ness, energy, first latency, rising melody type, stridor and shift occurrences.[27]

In [7], the work is done to classify the birds in their species according to their sounds. Different features were extracted to train the model for the classification. All the features were extracted on the small frames of the audio. Some of them are described below.

MFCC (Mel Frequency Cepstral Coefficients) - The mel-scale filter banks are used to get the non-linear frequency response like human auditory system.

Human Factor Cepstral Coefficients (HFCC) - In HFCC, fc of filters in the filter bank is on the mel scale, but bandwidth for it is different, which computed by the formula, ERB(equivalent rectangular bandwidth) = 6.23*fc^2 + 93.39*fc +28.52)

In [8], all the audios were preprocessed by removing low frequency noise by High Pass Filter and interfering noise by cepstral subtraction. Also, silence was removed by comparing the average energy of a segment with a threshold value)

Wavelet Packet Decomposition (WPD) was used to derive the feature from the audio, and K Nearest Neighbors Classifier was used. Along with this, use of 2D cepstral coefficients was also discussed.

Above research work in [7], [8] continued in [9] which used all the features extracted in above papers, and trained models first with the individual features and then by combining them. Features were extracted on the frame basis with the duration of 10ms and overlap of 50%.

In addition to the above features, PLP (Perceptual linear Prediction) was also used as one of the feature sets.

The potential set of features was selected out of all these features to see the results after combining the features using two methods, namely SVD (Singular Value Decomposition) and QRcp. With individual feature sets, RA (Recognition Accuracy) ranged from 80.23% to 86.74%. When the features were combined, RA increased significantly and ranges from 83.71% to 90.76%. The performance of HFCC features was found out to be the maximum, even greater than the most widely used MFCC feature set. Also, along with perceptual features like PLP, MFCC, HFCC; time and frequency-based features played an important role in increasing RA. Maximum RA was achieved by taking a reduced number of features.

In [25], gamma tone frequency cepstral coefficients (GFCC) were explored to classify the bird species. MFCC and GFCC were extracted from the frames of the bird sounds and ANN and SVM models were used for the classification.

Along with this, a study on using Dynamic Time Warping (DTW) was also carried out to get the difference between the spectrograms which could be a useful feature.

LPC coefficients, cepstral coefficients derived from LPC, LPC reflection coefficients, MFCCs, linear mel-filter bank channel and log mel-filter bank channel could also be the useful features.

PCA (Principal Component Analysis), mean computation, Vector Quantization using LBG could be used as the data reduction techniques.

III. PROPOSED WORK

The general structure of the designed system is given in Figure 1. The procedure followed by us consists of the following two steps.

First, we implemented signal processing to extract distinct acoustic features characteristic to the specific cry. Secondly, we used a classifier to assign each cry to one of labelled reasons for crying babies as given in the dataset. Then we trained and tested real cries recorded for this purpose in Android and iOS phones.

In order to build and train a classifier it was imperative to obtain an appropriate dataset. We used Donate-A-Cry corpus dataset,[29] which consists of 457 child cry samples of different classes such as “hungry”, “belly pain”, “discomfort”, “burping”.

A. *Feature Extraction:*

The primary goal of this step is the conversion of audio signals into a set of numeric values which represent the signal in a very unique and compact way conveying only relevant information. As shown in the figure1 several audio signal features are extracted and by selecting the best features we classify them using classifying models which are explained later.[4]

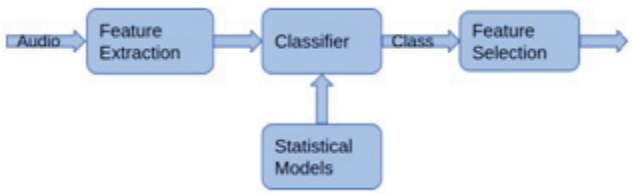

Fig. 1. Methodology

- The temporal features (time domain features)like the energy of signal, zero crossing rate, maximum amplitude, minimum energy, etc.
- The spectral features (frequency domain features)like MFCC, LPC, fundamental frequency, frequency components, spectral flux, spectral density, spectral centroid, spectral roll-off, etc.

B. *MFCC :*

Any sound produced by humans is determined by the shape of the vocal tract including tongue, teeth and other organs in human speech production system. If this shape of the vocal tract can be determined correctly, any sound produced can be accurately represented by that shape. The envelope of the time power spectrum of the speech signal represents a vocal tract and MFCC accurately represents this envelope. Short Time Fourier Transform is used to calculate MFCCs for each frames[5]. Following steps are used to calculate MFCCs:

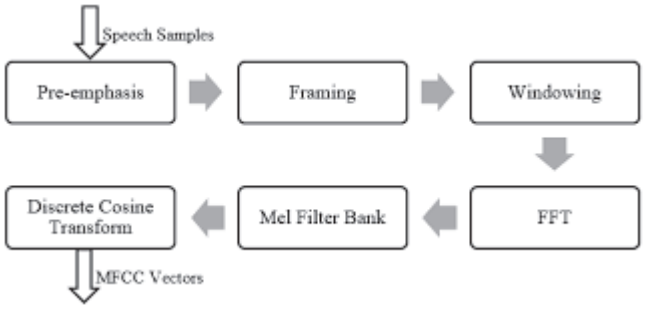

Fig. 2. Block Diagram for MFCC

The frequency is converted from Hertz to Mel scale using equation 1

$$Mel(f) = 2595\left(1 + \frac{f}{700}\right) \dots\dots\dots\dots\dots (1)$$

The final step in the calculation is Direct Cosine Transform (DCT)after results of the previous process is converted again to the time domain and then These results form a row of an acoustic vector which is MFCC [1].

C. *Spectral Flatness:*

Spectral flatness, also known as Wiener entropy is the tonality coefficient, which is used in characterizing an audio spectrum.[14] It represents a way to quantify how close the tone is like a sound, as opposed to being noise-like. This is a measure of uniformity in the frequency distribution of the power spectrum. The spectral flatness is calculated by following formula:

$$Flatness = \frac{\sqrt[N]{\prod_{n-0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} = exp \quad \text{..... (2)}$$

D. Loudness:

Loudness of sound that determines the intensity of auditory sensation produced during a process. The loudness as perceived by human ears varies as logarithm of sound intensity. The perceived loudness depends on the nature of the sound, due to the loading of the vocal tract system on the source during the production of speech. It is also affected by the behavioral characteristics of the speaker, such as emotional or mental state of the speaker, which in turn useful for classification of child cry.

E. Spectral Centroid:

In digital signal processing, a spectral centroid is used as a measure for signal characterization. As evident from the terminology, an indication of the center of mass of the signal is provided by it. It is robustly connected with the impression of the brightness of a sound.

This basically means higher the SC, more intense/bright the sound and a lower SC corresponds to a less intense audio signal.

$$SC = \frac{\sum_{k=0}^{\frac{N}{2}} k|X(k)|^2}{\sum_{k=0}^{\frac{N}{2}} |X(k)|^2} \quad \text{..... (3)}$$

F. Spectral Flux:

In case the power spectrum of a signal is changing, Spectral Flux determines the rate at which it changes. By comparing the power spectrum of the current frame to that of the previous frame, flux can be computed.

$$SF_i = \sum_{k=0}^{N/2} |X_{i+1} - X_i| \quad \text{..... (4)}$$

G. BANDWIDTH:

The width of the band of frequency of the frame around the middle point of the spectrum is termed as Bandwidth.

$$\sqrt{\frac{\sum_{k=0}^{N/2} (k-SC)^2 |X(k)|}{\sum_{k=0}^{N/2} |X(k)|^2}} \quad (5)$$

H. Gamma-Tone Frequency Cepstral Coefficients(Gfcc):

Availability of a set of features becomes indispensable for studying the characteristics efficiently in case of non-speech signals. Since a long time now, Mel Frequency Cepstral Coefficients(MFCCs) have been considered as the touchstone for parameterizing these signals. On the basis of MFCC computation scheme, the GFCCs have been introduced which utilize Gammatone filters with commensurate rectangular bandwidth bands.[12]

An impulse response that is obtained by the multiplication of a gamma distribution and sinusoidal tone is used to describe the Gammatone filter, which is a linear filter.

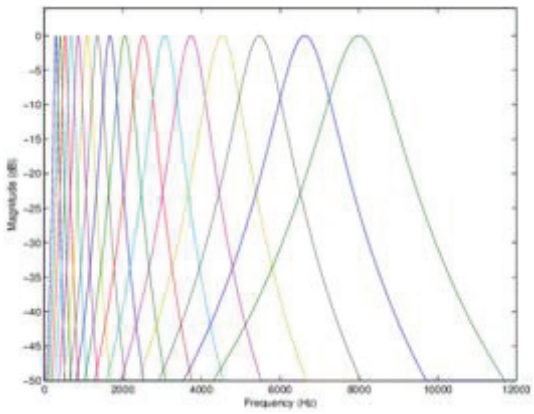

Fig. 3. Gammatone Filter

The block diagram for obtaining GFCCs is similar to that of getting MFCCs, the major difference being the usage of Gammatone filters instead of the Mel Frequency Filter Bank in case of MFCC.

Steps to calculate GFCC:

With a similar computational cost, the GFCC are more effective than MFCC in representing the spectral characteristics of non- speech audio signals, especially at low frequencies.[11]

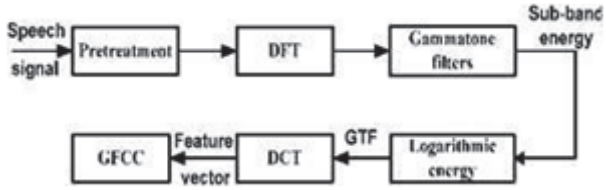

Fig. 4. Block Diagram for GFCC

I. STACF:

Autocorrelation is the mathematical representation used to assess the degree of similarity between a signal and a shifted version of itself over successive time intervals. We perform autocorrelation on each frame and hence it is called Short Term Autocorrelation Function. STACF can be used to determine whether a given audio signal is periodic or aperiodic as it finds repeating events and hence is an indicator of pitch. It is also used in pattern recognition. We often normalize this measure for consistent analysis.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T} (r_t - \overline{r})(r_{t-k} - \overline{r})}{\sum_{t=k+1}^{T} (r_t - \overline{r})^2} \quad \text{.....(6)}$$

J. Zero Crossing Rate:

The zero-crossing rate or ZCR is the rate at which the signal magnitude changes from positive to negative (or zero) or from negative to positive (or zero).ZCR can be used to determine the smoothness of a signal and can also determine whether a given audio signal is voiced or unvoiced.

$$ZCR = \frac{1}{2N} \sum_{n=1}^{N} |sgn(x[n]) - sgn(x[n-1])| \text{... (7)}$$

K. Linear Predictive Coefficients:

LPC represents the spectral shape of the signal and predicts the signal with very few coefficients. It works on the human speech production model in which, an impulse train passes through the vocal tract, which changes its shape. It is represented by the all-pole synthesis filter, whose filter coefficients are represented by LPC coefficients.

As it is the all-pole filter, the next sample can be predicted by the weighted sum of previous samples and as we are taking a finite number of coefficients, there is a difference between predicted and actual sample which is denoted by an additional coefficient which is error e.

LPC coefficients can be effectively used as the features which model the shape of the vocal tract and hence the shape of the spectral envelope. Levinson-Durbin algorithm, Burg's method are widely used to compute the LPCs.

L. SPECTRAL ROLL-OFF:

Frequency below which the particular percentage of the total energy in the spectrum, e.g., 95%, lies is called as Spectral Roll-Off [31]. The skewness of the spectral shape is represented by it. It is given by the formula

$$SRF = max\left(M \sum_{k=0}^{M} |X(k)|^2 < 0.95 \sum_{k=0}^{N/2} |X(k)|^2\right). \quad (8)$$

If the roll-off factor is 100%, we get the maximum frequency and if it is 0%, then we get the minimum frequency. This was evaluated over each frame's spectrum.

IV. MODELS USED FOR CLASSIFICATION

A. Multiclass Logistic Regression

Multiclass classification is required when the categorical output variable belongs to more than two classes. In One vs All algorithm we need N classifiers whereas in One vs One algorithm we will require O(N2) classifiers.[24]

Multiclass Logistic Regression finds the equation of boundary which gives the perpendicular distance. This is given by,

Z = WX + B

To get the probabilistic measure, sigmoid function is used.

$h_\Theta(x)$ = sigmoid(Z)

$$Sigmoid(x) = \frac{1}{1+e^{-x}} \ldots\ldots\ldots\ldots \quad \ldots.(9)$$

Sigmoid function's output ranges from 0 to 1 as it gives the probabilistic measure.

Algorithm:

1.Initialize: $\theta_2 = 0$ for all $0 \leq j \leq m$.

2.Repeat many times:

Gradient(j) =0 for all $0 \leq j \leq m$

For each training example (x,y):

For each parameter j:

Gradient (j) += $xj(y - \frac{-1}{1+e^{-\theta^T x}})$

$\theta_2 + * gradient(j)$, for all $0 \leq j \leq m$.

B. Support VECTOR MACHINE

In a Support Vector Machine we construct a hyperplane in a higher order dimensional space. The objective of optimization is to maximize the perpendicular distance (margin) between the positive and negative sample space. The plane which maximizes this perpendicular distance defines the boundary between the 2 classes. The 2 samples which result in this maximum margin are called the support vectors and hence the name. This same model can be extended to N classes where we need to define N – 1 boundaries.[21]

To gain a better understanding of this algorithm, let us consider an example having only two classes. We use the data to extract, say, two features x1 and x2. Our goal is to classify the pair (x1, x2) considering x1 and x2 as the axes.

In 2D space, we can plot the data and separate it into two classes by a straight line. There are many lines to choose from which could separate the classes. We must choose the line providing maximum margin between the classes. In case of non-linear data, we cannot separate using a straight line. Hence we can add one more dimension. For example, we can use z=x2+y 2. In this way we can convert the data to become linearly separable.

Algorithm:

Input:
Nin (Number of input vector)
Nsv (Number of support vector)
Nft (Number of features in a support vector)
SV [Nsv] (support vector array).
IN [Nin](Input vector array)
b * (bias)
Output:
F (decision function output)
For i←1 to Nin by 1 do
F=0
For j←1 to Nsv by 1 do
dist =0
For k ←1 toNft by 1 do
dist + = $\left(SV[j].feature[k] - \in [i].feature[k]\right)^2$
end
k=exp(-y × dist)
F+= SV[j] .∝* × k
End
F= F + b *
End

C. K Nearest Neighbours Classifier

We can say that the training data are vectors in a multidimensional feature space and it is labelled with the correct class. In order to classify, first we assign a constant value to k. A test vector is classified by observing k training vectors nearest to that point and then it is assigned the class which is most frequent among these k samples.[22] The distance can be calculated in continuous variables by using Euclidean dis-

tance. We can use other metrics such as overlap metric also known as Hamming distance and l-n norms using Euclidean distance. We can use other metrics such as overlap metric also known as Hamming distance and l-n norms, among others.

KNN algorithm uses feature similarity (proximity) to predict the class of test data based on its proximity to the testing data vectors. This algorithm and its implementation follow the steps as mentioned:

For each test data vector −

1. Calculate the distance between the test vector and each training data sample. The distance can be estimated using Euclidean, Manhattan or Hamming distance. The most popular metric used is the Euclidean distance.
2. Sort the distance values in ascending order.
3. For the top K values, check the labels.
4. Assign the most frequently occurring label to the test vector.

Algorithm:

Input: x,S d

Output : class of X

For $(x', l') \in$ S do

Compute distance d(x', x)

End for

Sort the |S| distances by increasing order

Count the number of occurrences of each class l_j

Among the k nearest neighbors

Assign to x the most frequent class.

D. Random Forest Classifier

Random forest classifier involves multiple decision trees, which collectively predict the label of a sample. Each individual tree in the random forest predicts one of the available classes and the most frequently occurring class is assigned to that sample.[23] The individual decision trees must be random with low correlation. This protects against and prevents the likelihood of misclassification. Except for a few trees, a most decision trees give an accurate prediction and the collective error is low.

This algorithm proceeds with the following steps −

1. Select random samples from the database.
2. Construct a decision tree for each sample. Obtain the prediction from each decision tree.
3. Count the frequency of results for each class.
4. Select the most frequent result as the final prediction.

Algorithm:

To generate C classifiers:

For i= 1 to c do:

Randomly sample the training Data D with replacement to produce D_i

Create a root node, N_i containing D_i

Call BuildTree (N_i)

End for

BuildTree(N):

If N contains instances of only one class then

Return

Else

Randomly selected x% of the possible splitting features in N

Select the feature F with the highest information gain to split on

Create f child nodes of $N, N_i, \ldots, N_f$, where F has f possible values $(F_1, \ldots, F_f)$

For i=1 to f do

Set the contents of N_i to D_i ,where D_i is all instances in N

Match F_i

Call BuildTree(N_i)

End for

End if

End

V. MODEL PERFORMANCE PARAMETERS

We have used the following parameters to evaluate the performance of our model:

PRECISION: Precision gives the probability of true positive results among all positive results. RECALL: Recall gives the probability of true positive results among all expected positive results. F1 SCORE: A metric that combines precision and recall as the harmonic mean of precision and recall is known as the traditional F-measure or balanced F-score.ACCURACY: Accuracy is also used as a statistical measure of how well a classification test correctly identifies a class. The accuracy is the probability of correct results (true positives and true negatives) among all examined samples.

TABLE I.

		Predicted	
		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

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

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

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

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

VI. RESULTS

A. Results of Individual Features and Best Model

TABLE II. RESULTS FOR INDIVIDUAL FEATURES

Features	Accuracy	F1 Score	Recall	Precision	Best Model
MFCC	**84%**	**76%**	**84%**	**70%**	**RF**
Spectral Flatness	**82%**	**76%**	**80%**	**71%**	**RF**
GTCC	**84%**	**80%**	**84%**	**76%**	**RF**
STACF	**84%**	**76%**	**84%**	**70%**	**KNN**
ZCR	**83%**	**76%**	**83%**	**70%**	**KNN**
LPC	**83%**	**77%**	**83%**	**72%**	**KNN**
Spectral Roll-off	**82%**	**75%**	**82%**	**70%**	**RF**
Sp. Centroid	**80%**	**75%**	**80%**	**70%**	**KNN**
Sp. Flux	**79%**	**78%**	**79%**	**77%**	**RF**
Bandwidth	**84%**	**74%**	**81%**	**70%**	**KNN**

We first extracted individual features of the audio signals and judged the performance of each model on the basis of each individual feature. All the performance parameters were calculated. K nearest neighbors (KNN) and Random Forest provided promising results in classification. The accuracy all the features was around 80- 85%. Taking an aggregate of other performance parameters, we shortlisted 3 main features, namely MFCC, GFCC and Zero crossing rate (ZCR) and then trained our models on these combined features. In this case, the most effective classification was provided by SVM as evident from the results.

TABLE III. RESULTS FOR COMBINED FEATURES

MODEL	ACCURACY	F1-SCORE	RECALL	PRECISION
RF	**84%**	**76%**	**84%**	**70%**
KNN	**82%**	**77%**	**82%**	**76%**
SVM	**71%**	**72%**	**71%**	**75%**
LR	**42%**	**53%**	**42%**	**74%**

VII. CONCLUSION

Firstly, based on the results of individual features in classifying the cries, GTCC outperforms MFCC in most of the training models. This highlights the primal importance of GTCC in recognizing the emotions signified by an audio signal. In case of individual feature performance, Random Forest and K- Nearest Neighbors algorithms give the best results. MFCC, GTCC and ZCR proved to be the most efficient features for most accurate classification of the lot.

VIII. FUTURE WORK

Some of the machine learning models which were not studied in depth in the field of classifying child cries are seen to provide promising results as compared to conventional approaches. Furthermore, we would try to study the suitability of Deep Learning models for efficiently classifying the infants' cries. Concluding, the findings of our research show that certain ML models exist that perform well in classifying cries of infants. Such models could further assist in developing a screening instrument on the basis of auditory characteristics of cries. The development of such an instrument would help in detecting any pathological development earlier. These instruments would also be helpful for professions dealing with babies like babysitters, nurses, pediatricians and even parents of that very child.

REFERENCES

[1] The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods S. P. Dewi, AnggunmekaLuhurPrasasti, BudhiIrawan Published 2019 Computer Science 2019 IEEE International Conference on Signals and Systems (ICSigSys)

[2] L. Liu, W. Li, X. Wu and B. X. Zhou, "Infant cry language analysis and recognition: an experimental approach," in IEEE/CAA Journal of AutomaticaSinica, vol. 6, no. 3, pp. 778-788, May 2019, doi: 10.1109/JAS.2019.1911435.

[3] Subramanian, Hariharan, P. Rao and D. Roy. "AUDIO SIGNAL CLASSIFICATION." (2004).

[4] Vibhute, Anup. (2014). Feature Extraction Techniques in Speech Processing A Survey.International Journal of Computer Applications. 107. 1-8. 10.5120/18744-9997.

[5] Jasleen and Dawood Dilber. "Feature Selection and Extraction of Audio Signal." (2016).

[6] http://www.ijirset.com/upload/2016/march/64_Feature.pdf

[7] Bang, Arti&Rege, Priti. (2018). Automatic Recognition of Bird Species Using Human Factor Cepstral Coefficients.10.1007/978-981-10-5544-7_35.

[8] Bang, Arti V. and P. Rege. "Classification of Bird Species based on Bioacoustics." (2013).

[9] Bang, Arti&Rege, Priti. (2017). Evaluation of various feature sets and feature selection towards automatic recognition of bird species.International Journal of Computer Applications in Technology. 56. 172. 10.1504/IJCAT.2017.088197.

[10] Bang, Arti&Rege, Priti. (2017). Recognition of Bird Species from their Sounds using Data Reduction Techniques.111-116. 10.1145/3154979.3155002.

[11] Valero, Xavier &Alías, Francesc. (2012). Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification. Multimedia, IEEE Transactions on. 14. 1684-1689. 10.1109/TMM.2012.2199972.

[12] Evaluating Gammatone Frequency Cepstral Coefficients with Neural Networks for Emotion Recognition from Speech,

[13] Messaoud, Ali &Tadj, Chakib. (2010). A Cry-Based Babies Identification System. 6134. 192-199. 10.1007/978-3-642- 13681-8_23.

[14] Izmirli, Özgür. (2000). Using a Spectral Flatness Based Feature for Audio Segmentation and Retrieval.

[15] Přibil, Jiří&Přibilová, Anna. (2009). Spectral Flatness Analysis for Emotional Speech Synthesis and Transformation. Lecture Notes in Computer Science. 5641. 106-115. 10.1007/978-3-642-03320-9_11
[16] https://link.springer.com/chapter/10.1007/978-3-642- 03320- 9_11
[17] Auditory-model based robust feature selection for speech recognition .The Journal of the Acoustical Society of America 127, EL73 (2010), Christos Koniaris, Marcin Kuropatwinski, W. Bastiaan Kleijn
[18] https://towardsdatascience.com/feature-selection-using- random-forest-26d7b747597f
[19] https://www.researchgate.net/post/How_to_do_support_vector_machine_based_feature_variable_selection.
[20] Dewi, Sita&Prasasti, Anggunmeka&Irawan, Budhi. (2019). The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods.10.1109/ICSIGSYS.2019.8811070.'
[21] https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm
[22] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_knn_algorithm_finding_nearest_neighbors.htm
[23] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_random_forest.htm
[24] https://towardsdatascience.com/multi-class-classification- one-vs-all-one-vs-one-94daed32a87b
[25] RashmikaPatole and PritiRege. Acoustic Classification of Bird Species
[26] Fuhr, Tanja, H. Reetzand C. Wegener. "Comparison of Supervised-learning Models for Infant Cry Classification / Vergleich von KlassifikationsmodellenzurSäuglingsschreianalyse." International Journal of Health Professions 2 (2015): 15 – 4.
[27] Cano O, Sergio & Escobedo Beceiro, Daniel &Ekkel, Taco. (2004). A Radial Basis Function Network Oriented for Infant Cry Classification.Progress in Pattern Recognition, Image Analysis and Applications.Vol.3287 of Lecture Notes in Computer Science. 3287. 374-380. 10.1007/978-3-540- 30463-0_46.
[28] https://musicinformationretrieval.com/spectral_features.html#:~:text=Spectral%20rolloff%20is%20the%20frequency,spectral_rolloff%20%3D%20librosa
[29] https://github.com/gveres/donateacry-corpus