Child Cry Classification – A study of features and models

Prathamesh Kulkarni
prathameshsp17.extc@coep.ac.in

Sarthak Umarani
sarthakns17.extc@coep.ac.in

Vaishnavi Diwan
ravindradr17.extc@coep.ac.in

Vishakha Korde
subhodks17.extc@coep.ac.in

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

Abstract — **This paper presents a study on the classification of child cries on the basis of 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 human body is a source of various signals related to different functions such as cardiological and nervous systems and many more. The analysis of biological signals is important for medical diagnoses. Although a child has no explicit way of communication, it can inform through a cry about a need to be satisfied. Before crying, the baby will try to communicate with specific language which known as Dustan Baby Language (DBL) that has some meaning like ―I am hungry‖ as ―Neh‖, ―I am sleepy‖ as ―Owh‖ likewise others. The sound of crying baby contains a lot of information about emotional and physical condition.

Crying is the first oral communication of babies. Babies express their feelings by crying before they learn how to express their feelings through speech. Sometimes, however, it becomes hard to find out why an infant cries. This can certainly be frustrating for a caregiver or mother.

Infant mortality rate is inferred as child death rate before completion of first year. The datum released by World Health Organization (WHO) infers that the infant mortality rate is 4.1 million. The main reason behind this is due to health issues. Nearly 75% of infant deaths shall be avoided if 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.

II RELATED WORK

In [1] a classification model is developed based on 3 criteria: degree of overfitting, accuracy, conformability. Total 468 cry audios are used in the dataset. Different features were extracted for each cry (which included first 6 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 [16] total 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 was 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.[2]

In [10] 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 [11], 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 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.

In continuation with research work in the above [2],[16] and 12 used all the features extracted in above papers, and trained the models first with the individual features and then by combining them. Features were extracted on the frame basis with the frame 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.

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 ranged 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 the reduced number of features.

In [3], 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, study was also done on using Dynamic Time Warping (DTW) 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

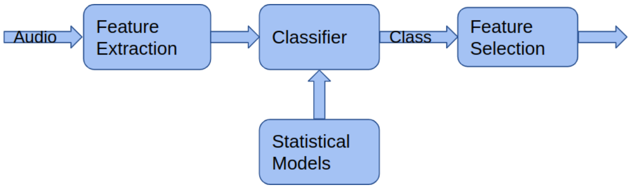
The general structure of the designed system**.** The procedure followed by us consists of the following two steps.

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, which consists of 457 child cry samples of different classes such as “hungry”, “belly pain”, “discomfort”, “burping”.

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 figure several audio signal features are extracted and by selecting the best features we classify them using classifying models which are explained later.

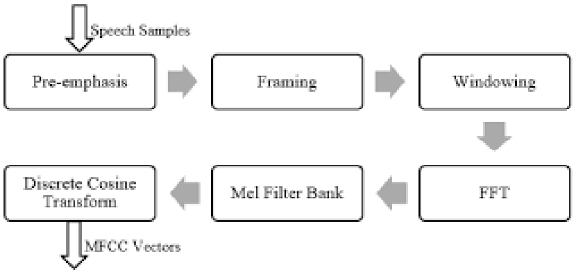


There are two types of features of a speech signal:

- The temporal features (time domain features), which are simple to extract and have easy physical interpretation, like: the energy of signal, zero crossing rate, maximum amplitude, minimum energy, etc.
- The spectral features (frequency based features), which are obtained by converting the time based signal into the frequency domain using the Fourier Transform, like: fundamental frequency, frequency components, spectral centroid, spectral flux, spectral density, spectral roll-off, MFCC, LPC etc. These features can be used to identify the notes, pitch, rhythm, and melody.

MFCC :

Any sound produced by humans is determined by the shape of vocal tract including tongue, teeth and other organs in human speech production system. If this shape of 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 vocal tract and MFCC accurately represents this envelope. Short Time Fourier Transform is used to calculate MFCCs for each frames. Following steps are used to calculate MFCCs:



By using following formula frequency is converted from Hertz to Mel scale:

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

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

SPECTRAL FLATNESS:

Spectral flatness is nothing but tonality coefficient, also known as Wiener entropy, which is used in digital signal processing to characterize an audio spectrum. It is typically measured in decibels (dBs). It also provides a way to quantify how tone-like a sound is, as opposed to being noise-like. This is measure of uniformity in the frequency distribution of the power spectrum. The spectral flatness is calculated by following formula:

$$\text{Flatness} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\frac{\sum_{n=0}^{N-1} x(n)}{N}} = \frac{\exp\left(\frac{1}{N}\sum_{n=0}^{N-1} \ln x(n)\right)}{\frac{1}{N}\sum_{n=0}^{N-1} x(n)}$$

LOUDNESS:

Loudness is an attribute of sound that determines the intensity of auditory sensation produced during proecess. The loudness as perceived by human ears is proportional to the logarithm of sound intensity. The perceived loudness depends on the nature of the sound, due to loading of the vocal tract system on 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.

SPECTRAL CENTROID:

In digital signal processing, 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 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.

$$SC = \frac{\sum_{k=0}^{N/2} k|X(k)|^2}{\sum_{k=0}^{N/2} |X(k)|^2} \qquad (3)$$

where $X(k)$ is the DFT of the frame, N is the size of the DFT and k varies from 0, 1, …$N - 1$.

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.

$$\mathrm{SF}_i = \sum_{k=0}^{N/2} |||X_{(i+1)} - X_i|||$$

BANDWIDTH:

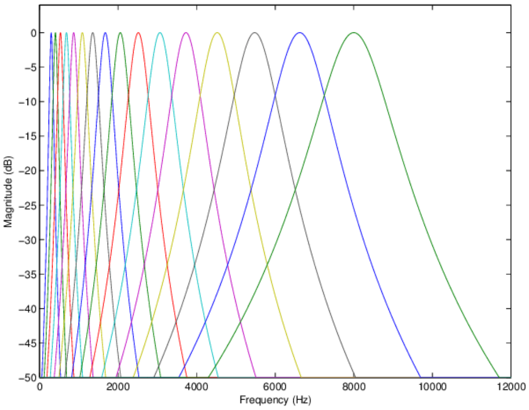
The width of the band of frequency of the frame around the middle point of the spectrum is termed as Bandwidth.

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

GAMMA-TONE FREQUENCY CEPSTRAL COEFFICIENTS (GFCC):

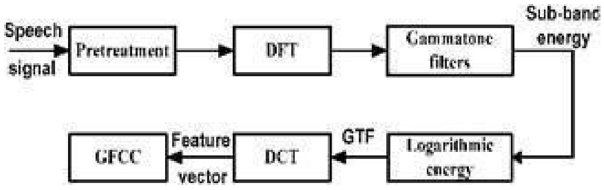
Availability of a set of features becomes indispensable for studying the characteristics efficiently in case of non-speech signals. Since a long time now, Mel Frequency Cepstral Coefficients(MFCCs) have been considered as the touchstone for parameterizing these signals. On the basis of MFCC computation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

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.



The block diagram for obtaining GFCCs is similar to that of obtaining 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.

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 a consistent analysis.

$$\hat{\rho}_k = \frac{\sum_{t=k+1}^{T}(r_t - \bar{r})(r_{t-k} - \bar{r})}{\sum_{t=1}^{T}(r_t - \bar{r})^2}$$

ZERO CROSSING RATE:

The zero-crossing rate or ZCR is the rate of change of sign of a signal In simpler words it is the rate at which the signal changes from positive to negative (or zero) or from negative to positive (or zero). This feature is instrumentally used in both speech recognition and information retrieval, hence it is a major feature to classify percussive sounds. ZCR can be used to determine the smoothness of a signal and can also determine whether a given audio signal is voiced or unvoiced.

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

LINEAR PREDICTIVE COEFFICIENTS:

To represent the spectral envelope of the speech signal and compress the signal for transmission, Linear Predictive Coding is widely used. This is a great method to achieve the very accurate estimates of speech and encode it with the low bit rate. LPC is the most widely used method in speech coding and speech synthesis.

LPC works on the human speech production model in which there is an impulse train generator for voiced speech and this passes through the vocal tract, which change its shape to utter different utterances and 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 finite number of coefficients, there is a difference between predicted and actual sample which is denoted by an additional coefficient which is error ―e‖.

Thus, N order LPC calculates this N coefficients along with an additional error term „e‟ which is the difference between the predicted and actual value of the sample.

LPC coefficients are evaluated over one frame, and as number of LPC coefficients are much less than the actual number of samples in the frame, it achieves good compression rate for the transmission of speech signal. Using the LPC coefficients and the error term, the original signal can be constructed again.

As we saw earlier, that LPC coefficients represent the filter coefficients of the all-pole filter in the human speech production mechanism, these coefficients can be effectively used as the features which model the shape of the vocal tract and hence the shape of the spectral envelope as well and can be useful in the audio classification task.

There are different methods to calculate the LPC coefficients, some of them are having time complexities while some of them are recursive and reduce the time complexity to much extent Levinson-Durbin algorithm, Burg's method are widely used algorithms for computing LPCs.

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. Skewness of the spectral shape is represented by it. It is given by the formula

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

If 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

MULTICLASS LOGISTIC REGRESSION

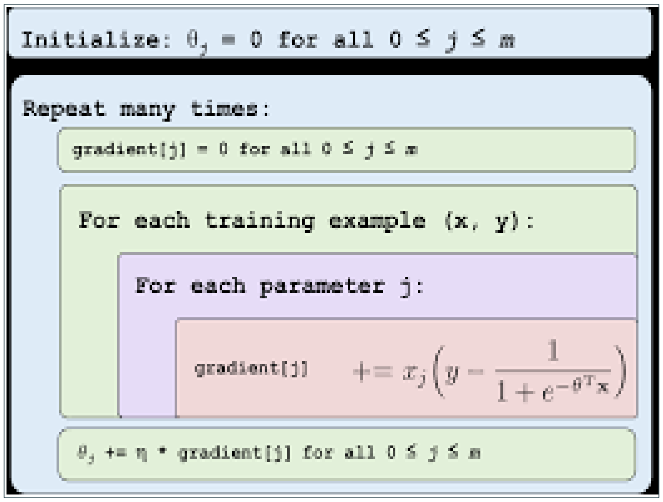
Multiclass Classification is required when the output categorical variable belong to more than two classes. Multiclass Logistic Regression is one such algorithm. In One vs All algorithm we need N classifiers whereas in One vs One algorithm we will require O(N2) classifiers.

If we substitute the data-point value in the LHS part of the boundary equation (where RHS is 0) then we get the value giving idea of the perpendicular distance of that point from the boundary. More is the distance from the boundary, more is its probability to belong to any one of the 2 classes. This is given by Z = WX + B, hΘ(x) = sigmoid (Z)

Thus, if a point is lying on the boundary, then it will evaluate

$$S(x) = \frac{1}{1 + e^{-x}}$$

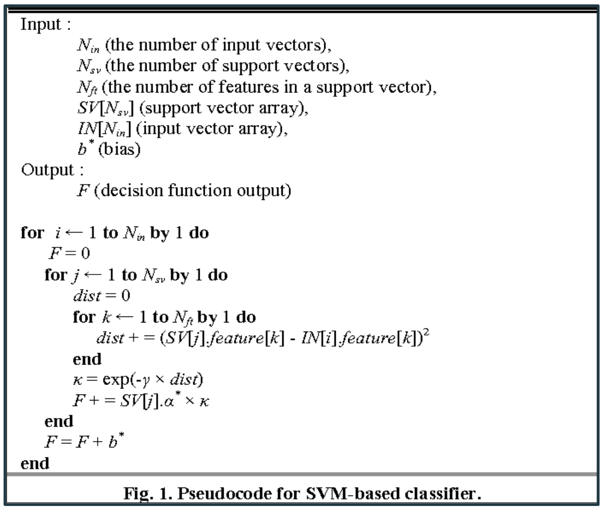
to 0 and after applying sigmoid function on it, probability of it to belong to the class 1 will turn out to be 0.5. Sigmoid function‘s output ranges from 0 to 1 as it gives the probabilistic measure.



```
Initialize: $\theta_j = 0$ for all $0 \le j \le m$
Repeat many times:
    gradient[j] = 0 for all $0 \le j \le m$
    For each training example (x, y):
        For each parameter j:
            gradient[j]   $\mathrel{+}= x_j\left(y - \frac{1}{1+e^{-\theta^T \mathbf{x}}}\right)$
    $\theta_j$ += $\eta$ * gradient[j] for all $0 \le j \le m$
```

SUPPORT VECTOR MACHINE

In a Support Vector Machine we construct a hyperplane in a higher order dimensional space. Optimization objective 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.



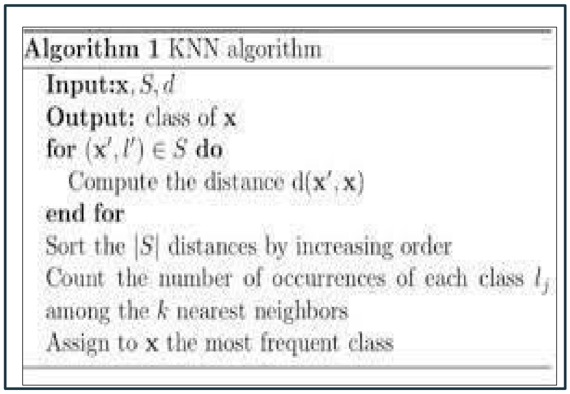
```
Input :
        $N_{in}$ (the number of input vectors),
        $N_{sv}$ (the number of support vectors),
        $N_{ft}$ (the number of features in a support vector),
        $SV[N_{sv}]$ (support vector array),
        $IN[N_{in}]$ (input vector array),
        $b^*$ (bias)
Output :
        $F$ (decision function output)

for $i \leftarrow 1$ to $N_{in}$ by 1 do
    $F = 0$
    for $j \leftarrow 1$ to $N_{sv}$ by 1 do
        $dist = 0$
        for $k \leftarrow 1$ to $N_{ft}$ by 1 do
            $dist \mathrel{+}= (SV[j].feature[k]$ - $IN[i].feature[k])^2$
        end
        $\kappa = \exp(-\gamma \times dist)$
        $F \mathrel{+}= SV[j].\alpha^* \times \kappa$
    end
    $F = F + b^*$
end
```

Fig. 1. Pseudocode for SVM-based classifier.

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 x_1 and x_2. Our goal is to classify the pair (x_1, x_2) considering x_1 and x_2 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 $z=x_1^2 + x_2^2$. In this way we can convert the data to become linearly separable.

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. Distance can be calculated in continuous variables by using Euclidean distance. We can use other metrics such as overlap metric also known as Hamming distance and l-n norms among others.



```
Algorithm 1 KNN algorithm
    Input: $\mathbf{x}, S, d$
    Output: class of $\mathbf{x}$
    for $(\mathbf{x}', l') \in S$ do
        Compute the distance $d(\mathbf{x}', \mathbf{x})$
    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 $\mathbf{x}$ the most frequent class
```

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 below:

For each test data vector −

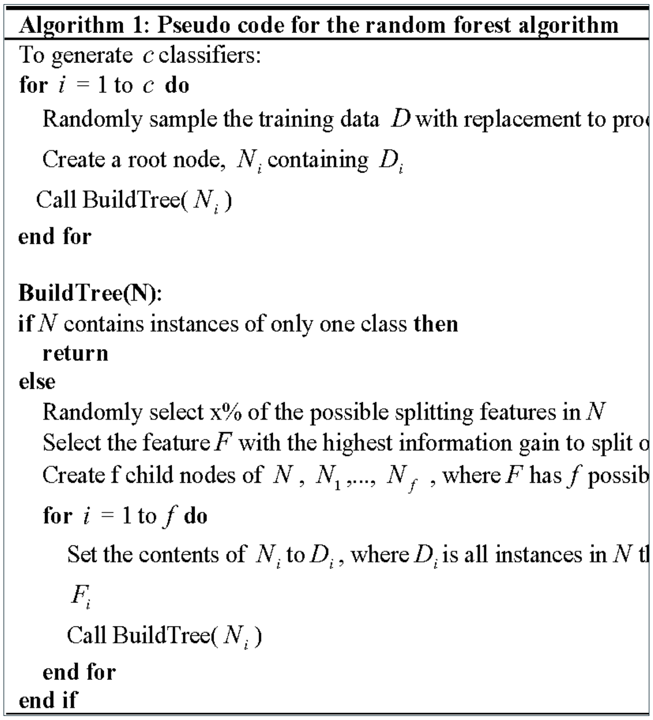
1. Calculate the distance between the test vector and each training data sample. Distance can be estimated using Euclidean, Manhattan or Hamming distance. The most popular metric used is Euclidean distance.
2. Sort the distance values in ascending order.
3. For the top K values, check the labels.
4.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

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. 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 algorithms 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 1: Pseudo code for the random forest algorithm

```
To generate $c$ classifiers:
for $i$ = 1 to $c$ do
    Randomly sample the training data $D$ with replacement to pro
    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 select x% of the possible splitting features in $N$
    Select the feature $F$ with the highest information gain to split o
    Create f child nodes of $N$ , $N_1$,..., $N_f$ , where $F$ has $f$ possib
    for $i$ = 1 to $f$ do
        Set the contents of $N_i$ to $D_i$ , where $D_i$ is all instances in $N$ t
        $F_i$
        Call BuildTree( $N_i$ )
    end for
end if
```

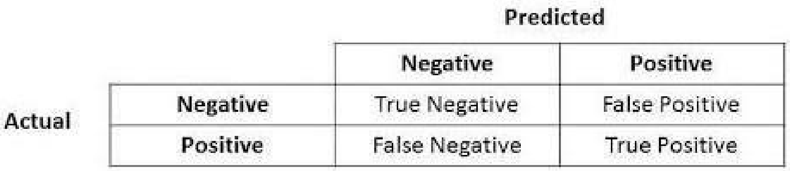
V. MODEL PERFORMANCE PARAMETERS

PRECISION: Precision gives the probability of true positive results among all positive results.

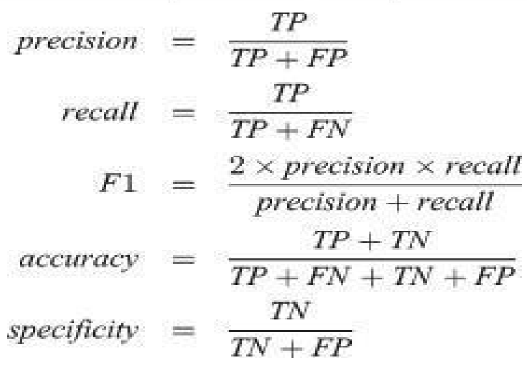
RECALL: Recall gives the probability of true positive results among all expected positive results.

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.



		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive



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

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

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

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

$$specificity = \frac{TN}{TN + FP}$$

VI. RESULTS

RESULTS OF INDIVIDUAL FEATURES AND BEST MODEL

FEATURES	ACCURACY	F1- SCORE	RECALL	PRECISION	BEST MODEL
MFCC	84%	76%	84%	70%	RF
Sp. 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
Sp. 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 from 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.

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 weren’t 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.

IX. REFERENCES

1. Duration in infants' communication by cries J.Rosenhouse
2. The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification MethodsSitaPurnamaDewi.
3. Infant Cry Language Analysis and Recognition: An Experimental Approach Lichuan Liu, Senior Member, IEEE, Wei Li, Senior Member, IEEE, Xianwen Wu, Member, IEEE and Benjamin X. Zhou
4. Speech/Music Classification using MFCC and KNN R. Thiruvengatanadhan Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Tamil Nadu, India.
5. AUDIO SIGNAL CLASSIFICATION Hariharan Subramanian (Roll No: 04307909) Supervisors: Prof. Preeti Rao and Dr. Sumantra. D. Roy
6. An Automatic Approach to Extract Features from the Infant‘s Cry Signals,V.Vaishnavi1 , P. Suveetha Dhanaselvam2
7. Feature Extraction Techniques in Speech Processing: A Survey
8. Feature Selection and Extraction of Audio SignalJasleen 1 , Dawood Dilber 2 P.G. Student, Department of Electronics and Communication Engineering, Amity University, Noida, U.P, India1
9. http://www.ijirset.com/upload/2016/march/64_Feature.pdf
10. BangA.V., Rege P.P. (2018) Automatic Recognition of Bird Species Using Human Factor Cepstral Coefficients.
11. Bang, Arti V. and P. Rege. —Classification of Bird Species based on Bioacoustics.‖ (2013).
12. Bang, Arti & Rege, Priti. (2017). Evaluation of various feature sets and feature selection towards automaticrecognition of bird species
13. Arti V. Bang and Priti P. Rege. 2017. Recognitionof Bird Species from their Sounds using Data Reduction Techniques.
14. X. Valero and F. Alias, "Gammatone Cepstral Coefficients: Biologically Inspired Features for Non-Speech Audio Classification"
15. arXiv:1806.09010 [cs.SD]
16. A Cry-Based Babies Identification System Ali Messaoud and Chakib Tadj
17. Using a Spectral Flatness Based Feature for Audio Segmentation and Retrieval
18. https://link.springer.com/chapter/10.1007/978-3-642-03320-9_11
19. https://link.springer.com/chapter/10.1007/978-3-642-03320-9_11
20. Auditory-model based robust feature selection forspeech recognition .The Journal of the Acoustical Society of America 127, EL73 (2010)
21. https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f
22. https://www.researchgate.net/post/How_to_do_support_vector_machine_based_feature_variable_selection
23. https://www.researchgate.net/publication/335388902_The_Study_of_Baby_Crying_Analysis_Using_MFCC_and_LFCC_in_Different_Classification_Methods