International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Machine Learning Approach for Infant Emotion Classification through Audio Signals.

*Dhanyatha Y [a], Tejaswini N Reddy [a], Shaista Naseem Kausur [a], M Keerthana [a], Dr. Vani A [b]**

[a] *PG Student, B.M.S College of Engineering, Bengaluru, India.*

[b] *Assistant Professor, B.M.S College of Engineering, Bengaluru, India.*

ABSTRACT

Understanding newborn communication patterns and emotional states greatly depends on the ability to recognize infant vocalizations like screams and chuckles. This study provides an automated method that uses machine learning and audio signal analysis to discriminate between newborn cries and laughs. For training and assessment, a datasets of pre- processed audio recordings with weep and laugh parts was used. The results of the studies show how effectively and precisely the suggested method can distinguish between baby cries and laughter. With an average accuracy of 94.7% on the test dataset, the generated model performed admirably. Childcare, pediatrics, and human-computer interaction can all benefit from this study's contributions to the field of infant vocalization analysis. The findings of the study lay the groundwork for creating intelligent systems that can recognize and react to newborn emotions, ultimately promoting improved caregiver-baby relationships.

Keywords: infant vocalization, signal analysis, machine learning.

Introduction

Infants express their wants, emotions, and general well-being through a variety of vocalizations, such as cries and laughter. The first indication of life and a baby's means of interaction with the outer world is crying [1]. Until they are ready to speak, an infant can communicate with a parent or caregiver by crying and laughing [2], [3]. Crying happens in the most delicate region of the ear in humans and incorporates voice characteristics, facial expressions, and bodily movements. Contrary to appearances, the acoustic signals of newborn tears and giggles are complex [4], dynamic, and connected to variations in the infant's requirements and states of physical, emotional, and psychological development [5,6]. In order to understand the baby's mood and health status, parents, caregivers, and healthcare professionals must be able to distinguish between various vocalizations.

According to the Dustan theory, a baby's language consists of five words that are tied to five basic needs: "Neh" (I'm hungry), "Eh" (I need to poop), "Owh/Oah" (I'm tired), "Eair/Eargghh" (I have cramps), and "Heh" (I'm physically uncomfortable; I feel hot or damp) [7]. Babies' vocal tracts are more sensitive than adults' because they lack the ability to control them. Each type of cry has a pattern, classifying infant cries can be compared to pattern recognition or speech recognition.

Furthermore, Cry audio signal (CAS) was discovered to be diagnostically useful in neonates [8]. In the study published by [9], the CAS characteristics of newborns with various disorders, such as hypoxia, deafness, etc., were compared to newborns without these abnormalities. Even during a physical examination, a CAS may reveal an infant's health may be at risk from a pathology [10]. Collecting data, analyzing signals, extracting features, and categorizing data are all part of infant cry research [11]. In addition, it's important to pick the most relevant features and scale them down in order to build successful classification models. Baby vocalizations can be subjective and time-consuming to analyze manually the traditional way. To analyze baby voice and audio signals, a variety of deep learning and ensemble models can be used [12], [13], [14]. Systems that use machine learning are capable of learning on their own without explicit programming, and as they gain experience, they perform better.

Therefore, Machine learning approaches are required for precise classification or detection of infants' crying events [11]. In this study, the infant's cries and laughter are classified by linear, polynomial, gaussian, KNN and random forest classifier.

Literature Survey:

Smart nurseries are crucial in the age of smart cities for automating tasks like housekeeping and cooking. This study suggests a smart cradle system that automates cradle operations depending on baby sounds. The method makes use of the Support Vector Classifier (SVC) with a Radial Basis Function (RBF) kernel and 18 retrieved features of baby sounds. The most effective model, SVC using RBF kernel function, outperforms earlier literature systems with an average accuracy of over 96%.[15]

The likelihood that preterm and babies with extremely low birth weights would survive has increased thanks to developments in perinatal and neonatal medicine. The objective of this work is to identify distinguishing characteristics of preterm and full-term infant cries using automatic sound analysis and data mining. To categorize the types and meanings of infant crying, the method employs deep support vector machines, fractal descriptors, and iterative neighborhood component analysis. When compared to other methods, the fractal method and optimal classification outperform them in terms of diagnostic accuracy by addressing problems like classification mistakes and uncertainty.[16]

The K-Nearest Neighbor (KNN) algorithm and Mel Frequency Cepstral Coefficients (MFCC) techniques are used in this study to demonstrate a baby cry identification system. In residential situations, the system uses machine learning algorithms to identify infant cries. In order to categorize cries, the method extracts MFCC features from audio data and uses KNN. Using a publically available dataset, the method showed great accuracy in recognizing different cry kinds.[17]

Neonatal babies communicate via crying, which calls for specialized care and the use of deep learning techniques for feature extraction and selection. The main goal of this research is to distinguish between pain, hunger, and tiredness in infant screams. Short-time Fourier transform and deep convolutional neural network are used to modify the spectrogram images, and features are then supplied to an SVM classifier. The experimental results are encouraging, and SVM-RBF, which has an accuracy of 88.89%, is the most accurate system for classifying infant cries based on kernels.[18]

With the help of feature vector techniques, this work preprocesses and extracts characteristics from Dunstan's baby cry data. Using SVM, MLP, and CNN classifiers, five different types of newborn cries are identified. The model developed by CNN performs the best, with a 92.1% average accuracy across all five classes. Increasing complexity leads to better results from the model.[19]

Methodology:

In this section, the datasets, feature extraction, Support Vector Machine, and data accuracy testing will all be briefly explained. The process for categorizing a baby's sound as a cry or a giggle is shown in Fig. 1.

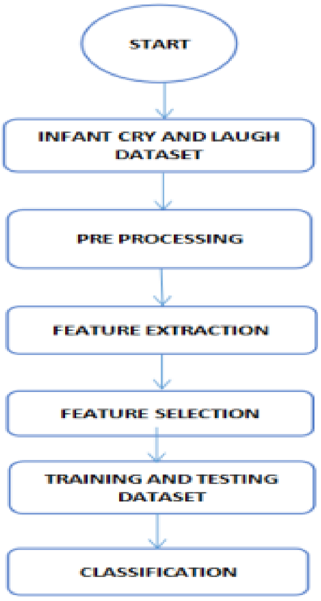

Fig. 1 - Flow chart

Datasets:

For this study, secondary data was employed for investigation. 50 recordings of the baby's cries and laughs were used. The dataset was retrieved from Kaggle.

Signal Pre -Processing:

The screening and normalization processes was performed. The frequency of all infant crying sound data is 44,100 Hz. The baby's cry has a different level of noise in the recording than it does in real life. In a few recordings, the original signal frequency is greater than the noise frequency. A few recordings, however, include noise frequencies that are higher than the original sound. The filtering process was carried out with the use of the Audacity software to eliminate or minimize the noise. Then, using a normalizing procedure, the amplitude interval was changed from -1 to 1 to equalize the maximum amplitude interval of each baby's crying sound signal so that the characteristics of the baby's cry sound extraction process would not be affected by the change in amplitude.

Feature Extraction:

The audio samples were used to extract spectral features (frequency-based features) which are acquired by applying the Fourier Transform to transform a time-based signal into a frequency-based signal.

Feature Selection:

Five features were taken from the audio signals: centroid, crest, decline, entropy, and flatness. The "center of mass" of the spectrum can be determined using the spectral centroid measurement. The peakness of the spectrum can be determined by the spectral crest. A greater spectral crest denotes more tonality, while a lower spectral crest denotes more noise. The rate-map representation's average spectral slope is described by the spectral decline, which places more focus on the low frequencies. A signal's spectral entropy (SE), which is based on the Shannon entropy, or information entropy, is a measure of its spectral power distribution. A measurement of how much a sound is noise-like as opposed to tone-like is called spectral flatness (or tonality coefficient).

Training and Testing dataset:

The dataset was split into a training dataset and a test dataset. A machine learning model is taught using training data, which is a very sizable dataset. Training data is used to educate prediction models that employ machine learning algorithms how to extract relevant features; the higher the quality of the training data, the better the performance of the model. A little over 60% of the entire data is made up of training data. The test dataset is a different subset of the original data that is unrelated to the training dataset and is used to check (test) whether the model was appropriately selected. A training and test data collection can be used with Principal Component Analysis (PCA). This method is helpful for reducing the dimensionality of data and is used to improve data visualization and speed up machine learning model training.

Classification:

As it only employs two variables, classification is said to be binary. Utilizing criteria including accuracy, precision, recall, specificity, and F1 score, the trained models' performance was assessed. The classification methods employed were SVM and confusion matrix. The classification techniques used to evaluate the model's accuracy include linear kernel, polynomial kernel, gaussian approach, KNN, and random forest [20]. The performance of a classification model at all classification thresholds is displayed using a ROC curve (receiver operating characteristic curve) together with a confusion matrix. The Higuchi fractal dimension value of each baby's sobbing sound signal provides the foundation for the KNN and SVM processes. In this study, the KNN technique experiment values were 1, 2, 3, 5, and 10. While the linear kernel, polynomial kernel, and RBF kernel were used as trials with the SVM algorithm with c = 1, 10, 20, 40, and 70. To determine the c value that is appropriate for the infant sobbing sound data, a value of c between 1 and 100 is selected. The misclassification penalty's amount (threshold) is specified in parameter c. Degree 1 is the linear kernel, and Degree 10 is the polynomial kernel. Gamma (γ)=1 and 10 are utilized as experiments in the RBF kernel, in contrast.

Confusion Matrix: An evaluation of a classification algorithm's performance is done using a table called a confusion matrix. A confusion matrix depicts and summarizes a classification algorithm's performance. The matrix's predictions is compared with the original class from the input data, to perform the study.

1) *Accuracy:* We shall use the Confusion Matrix to evaluate an algorithm's accuracy. The Confusion Matrix is a tool for assessing the accuracy of the classification outcomes produced by the classification procedure. The ratio of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), respectively. The accuracy of the system was calculated by the following equation:

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

2) *Precision:* Precision is the degree to which measurements of the same thing agree with one another. The following equation can be used to determine the precision.

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

3) *Recall:* Recall assesses a classification model's ability to correctly identify each pertinent instance within a dataset. It is the proportion of incidences of true positives (TP) to instances of both true positives and false negatives (FN).

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

4) *Specificity:* The ratio of true negatives to all other bad outcomes is known as specificity. The ideal specificity is 1.0, whereas the unfavorable value is 0.0.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN + FP}}$$

5) *F1 Score:* F1 score is a weighted average of precision and recall. As we know in precision and in recall there is false positive and false negative so it also consider both of them. F1 score is usually more useful than accuracy, especially for an uneven class distribution

$$\text{F1 Score} = \frac{\text{2TP}}{\text{2TP + FP + FN}}$$

Support Vector Machines (SVM):

Support Vector Machines (SVM) are supervised learning models with corresponding learning algorithms that examine data used for regression analysis and classification. Although it has been suggested that the key reason to use an SVM instead is that the problem may not be linearly separable, the SVM approach is frequently claimed to produce better results than alternative classifiers. Then, a Radial Basis Function (RBF)-based SVM with a non-linear kernel would be appropriate. SVMs, for instance, have reportedly been found to perform better in text categorization, despite the fact that training takes a long time. The SVM classifier is a kernel-based technique that is specifically created for binary classification and divides the data into two or more classes. It is not advised to use it when there are a lot of training instances. A kernel function is a mapping process that is applied to the training set to enhance its similarity to a set of data that may be linearly separated, which makes the data set more dimensional. The linear, RBF, quadratic, multi-layer perceptron, and polynomial kernels are a few of the frequently utilized kernel functions. It is not advised to use SVM when there are a lot of training examples because it divides the data into two or more classes.

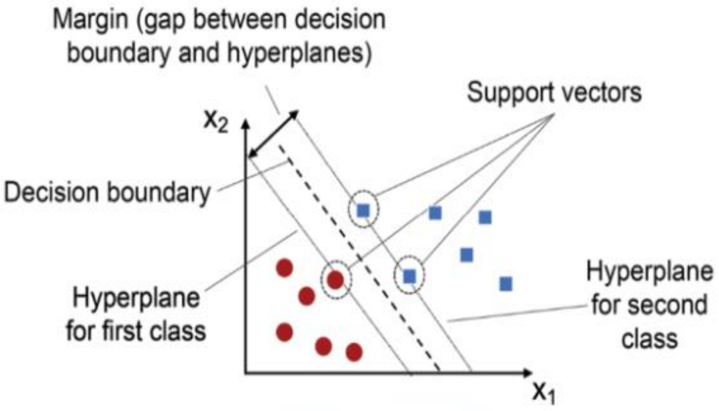
Margin (gap between decision boundary and hyperplanes)
Support vectors
x_2
Decision boundary
Hyperplane for first class
Hyperplane for second class
x_1

Fig. 2. A simple illustration of the Support Vector Machine (SVM) algorithm in 2-dimensions.

K-Nearest Neighbor (KNN): The KNN algorithm is an instance-based learning technique used in pattern recognition to categorize objects based on the nearest training instances in the feature space. An object is classed by a majority vote of its neighbors, or, more specifically, the object is assigned to the class that is most prevalent among its k-nearest neighbors, where k is a positive integer. Euclidean distance metrics are used to implement the KNN algorithm and find the nearest neighbor. CNN's main disadvantage is that it becomes significantly slower as the volume of data increases making it an impractical choice in environments where predictions need to be made rapidly. As the value of K decreases to 1, the predictions become less stable. Inversely, as the value of K is increased, the predictions become more stable due to majority voting/averaging, and thus, more likely to make more accurate predictions (up to a certain point). Eventually, an increasing number of errors is witnessed. It is at this point that one recognizes that the appropriate value of K has been exceeded. Problems involving classification, regression, and search can all be solved using the KNN algorithm. It is helpful in finding solutions to issues whose resolutions depend on recognizing related objects.

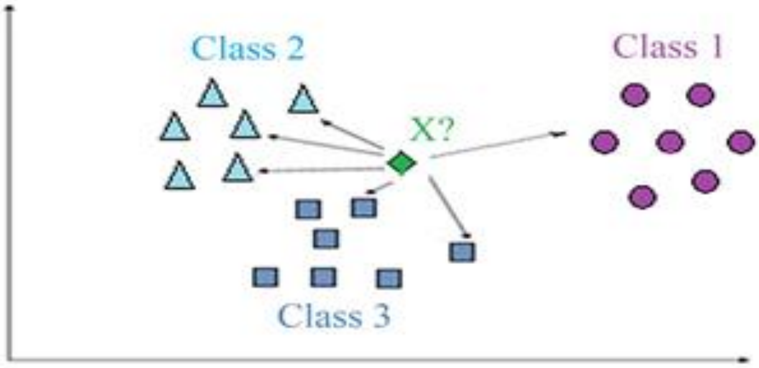

Fig. 3. A simple pictorial overview of the K-Nearest Neighbor (KNN) algorithm.

Random Forest: An RF classifier is made up of several trees, each of which is developed using some sort of randomization. Each tree's leaf nodes are given labels based on estimates of the posterior distribution across the picture classes. Each internal node has a test that divides the classification space of data into the best possible chunks. By transmitting a signal down each tree and summing the distributions of the reached leaves, a signal is categorized. Randomness can be introduced into the training process twice: while subsampling the training data to construct each tree using a different subset, and when choosing the node tests. Additionally, RF is particularly user-friendly in that it only requires two parameters, which are typically not very sensitive to one another. Due to its adaptability, RF is used in a wide range of industries, which contributes to their appeal. A set of criteria arranged in a hierarchical structure is referred to as a decision tree. It is a predictive model in which an instance is classified by traveling up the tree's root until it reaches a leaf, which corresponds to a class label, and then follows the route of satisfied conditions there. Simple categorization rules can be created from a DT.

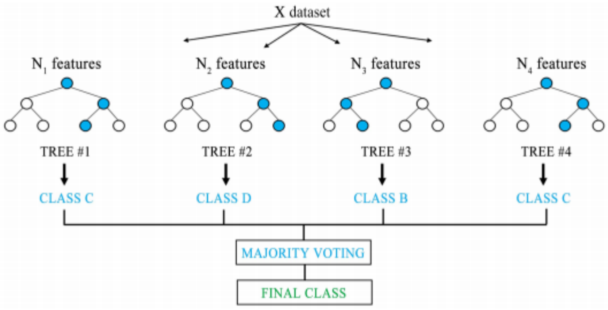

Fig. 4. A pictorial overview of the random forest (RF) algorithm.

Results and Conclusion:

Our algorithm code is written in Python. In order to compare the model results; we establish SVM models like linear kernel, polynomial kernel, KNN, Gaussian method and Random forest method. These models are compared to test the effectiveness. The metrics that are measured are accuracy, precision, average precision recall score, sensitivity, specificity and F1 score.

Table 1 - An example of a table.

Classification	***Accuracy***	***Precision***	***APS***	***Sensitivity***	***Specificity***	***F1 score***
Linear	72.09%	77.78%	0.6435	63.64%	80.95%	0.739
Polynomial	51.16%	55.46%	0.4884	100.00%	0.00%	0
Gaussian	88.37%	86.96%	0.8412	90.91%	85.71%	0.878
KNN	74.42%	78.95%	0.6664	68.18%	80.95%	0.756
Random forest	94.70%	87.50%	0.8818	95.45%	85.71%	0.9

The above table gives all the details about the metrices obtained for different classification methods. Random forest method gives the maximum accuracy of 94.7% with a precision of 87.50%, average precision recall score of 0.8818, Sensitivity of 95.45%, Specificity of 85.71% and F1 score of 0.9. Considering the matrices of the classification method it is seen that the Random Forest method works best to classify the laugh and cry audio of an infant.

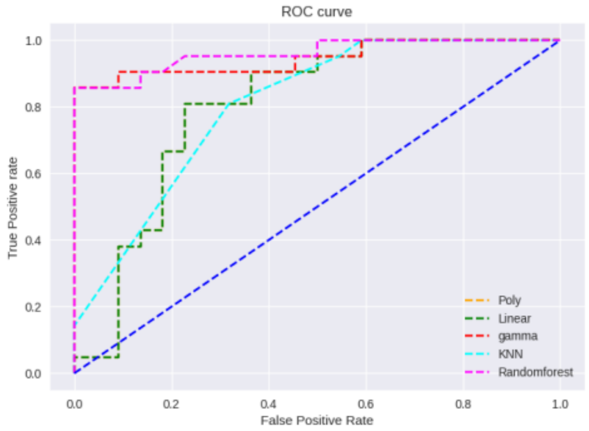

Fig. 5. ROC curve.

The above graph shows the ROC curve of the classification methods we have used and is observed that Random Forest method is the best classifier among all.

References

[1] Geangu, Elena, et al. "Contagious crying beyond the first days of life." Infant Behavior and Development 33.3 (2010): 279-288.

[2] Douglas, Pamela S., and Harriet Hiscock. "The unsettled baby: crying out for an integrated, multidisciplinary primary care approach." Medical Journal of Australia 193.9 (2010): 533-536.

[3] Zeifman, Debra M., and Ian St James-Roberts. "Parenting the crying infant." Current opinion in Psychology 15 (2017): 149-154.

[4] Orlandi, Silvia, et al. "Testing software tools for newborn cry analysis using synthetic signals." Biomedical Signal Processing and Control 37 (2017): 16-22.

[5] Liu, Lichuan, et al. "Infant cry language analysis and recognition: an experimental approach." IEEE/CAA Journal of Automatica Sinica 6.3 (2019): 778-788.

[6] Yoo, Hyunjoo, et al. "Acoustic correlates and adult perceptions of distress in infant speech-like vocalizations and cries." Frontiers in psychology 10 (2019): 1154.

[7] Dunstan, P. "Calm the crying: The Secret Baby Language." (2012): 16-17.

[8] Parga, Joanna J., et al. "Defining and distinguishing infant behavioral states using acoustic cry analysis: is colic painful?." Pediatric research 87.3 (2020): 576-580.

[9] Boukydis, CF Zachariah, and Barry M. Lester, eds. "Infant crying: Theoretical and research perspectives." (2012).

[10] Abdulaziz, Yousra, and Sharrifah Mumtazah Syed Ahmad. "Infant cry recognition system: A comparison of system performance based on mel frequency and linear prediction cepstral coefficients." 2010 International Conference on Information Retrieval & Knowledge Management (CAMP). IEEE, 2010.

[11] Jeyaraman, Saraswathy, et al. "A review: survey on automatic infant cry analysis and classification." Health and Technology 8 (2018): 391-404.

[12] Chang, Chuan-Yu, et al. "An efficient classification of neonates cry using extreme gradient boosting-assisted grouped-support-vector network." Journal of healthcare engineering 2021 (2021).

[13] Bănică, Ioana-Alina, et al. "Automatic methods for infant cry classification." 2016 International conference on communications (COMM). IEEE, 2016.

[14] Chang, Chuan-Yu, et al. "DAG-SVM based infant cry classification system using sequential forward floating feature selection." Multidimensional Systems and Signal Processing 28 (2017): 961-976.

[15] Rezaee, Khosro, et al. "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 (2023).

[16] Nimbarte, Nita, et al. "New Born Baby Cry Analysis and Classification." 2023 4th International Conference for Emerging Technology (INCET). IEEE, 2023.

[17] Ashwini, K., et al. "Deep learning assisted neonatal cry classification via support vector machine models." Frontiers in Public Health 9 (2021).

[18] Abbaskhah, Ahmad, Hamed Sedighi, and Hossein Marvi. "Infant cry classification by MFCC feature extraction with MLP and CNN structures." Biomedical Signal Processing and Control 86 (2023): 105261.

[19] Boateng, Ernest Yeboah, Joseph Otoo, and Daniel A. Abaye. "Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review." Journal of Data Analysis and Information Processing 8.4 (2020): 341-357.

[20] Boateng, Ernest Yeboah, Joseph Otoo, and Daniel A. Abaye. "Basic tenets of classification algorithms K-nearest-neighbor, support vector machine, random forest and neural network: a review." Journal of Data Analysis and Information Processing 8.4 (2020): 341-357.