A Machine Learning Approach to Classify Biomedical Acoustic Features for Baby Cries

[#]Gaurav Aggarwal, [†]Kavita Jhajharia, [#]Jaweria Izhar, [‡,§]Manoj Kumar, and [§]Laith Abualigah

Abstract: Communication is imperative for living beings for exchanging information. But for newborns, the only way of communicating with the world is through crying, and it is the only medium through which caregivers can know about the needs of their children. Timely addressing baby cries is very important so that the child is relieved at the earliest. It has been a challenge, especially for new parents. The literature says newborn babies use The Dustan Baby Language to communicate. According to this language, there are five words to understand a baby's needs, which are "Neh" (hungry), "Eh" (burp is needed), "Owh/Oah" (fatigue), "Eair/Eargghh" (cramps), "Heh" (feel hot or wet, physical discomfort). This research aims to develop a model for recognizing baby cries and distinguishing between different kinds of baby cries. Here we more broadly focus on whether the infant is in pain due to hunger or discomfort. The study proposes a comparative approach using four classification models: random forest, support vector machine, logistic regression, and decision tree. These algorithms learn from the spectral features: chroma_stft, spectral_centroid, bandwidth, spectral_rolloff, mel-frequency cepstral coefficients, linear predictive coding, res, zero_crossing_rate extracted from the infant cry. The support vector machine model outperforms other classifiers for correctly classifying infant cries.

Key Words: Cry–Classification–SVM–Hunger–Discomfort–Spectral features–Acoustic features.

INTRODUCTION

The popularity of understanding Baby Cries is increasing, and this growing drift has divided into various domains of research and project work. The first trend is automatically identifying baby cry instances; basically, the baby cry instances are detected among various rare sound events.[1] The other existing direction is distinguishing a particular type of baby cry, such as pathological crying. Since crying is all a baby can do to express any discomfort, it seems that this multimodal signal carries a lot of information about him/her; hence, concealed information within a cry signal could clarify the infant's present psychological condition.[2] The third trend is the one on which we are working, and it is about understanding the baby's language. This implies understanding the underlying need of a baby as expressed through crying and was named after Pricilla Dunstan. He identified some patterns in her baby's cry and started the Dustan baby language program.[3,4] All these trends have been worked upon using various machine learning and deep learning techniques. This study uses several classification techniques to build models of different baby cries taken from the Donate Cry Corpus dataset. Learning to decode a newborn's cries has a long learning curve for new parents, adding to the stress and frustration of the sleep-deprived first few months, and face various difficulties in understanding their baby's needs.[5] These problems pose to be the primary motivation for building the classification models. Also, Pricilla Dunstan served as a motivation as she discovered the Dustan baby language, which is the baby's language that fascinated the most. The main objective of this research is to detect and classify the cry of a baby so that one can able to comprehend what the baby needs without any harassment.[6] This paper tries to build an automated tool that knows about the baby's basic needs such as hunger, discomfort, pain, and sleepiness by using his/her cry and helping the parent to act accordingly.[7] The pros and cons of baby cry prediction are also explained in Table 1.

The study proposes a comparative approach using four supervised learning models: random forest, support vector machine, logistic regression, and decision tree. These models train from the spectral features: chroma_stft, spectral_centroid, bandwidth, spectral_rolloff, Mel-frequency cepstral coefficients, linear predictive coding, res, zero_crossing_rate extracted from the infant cry using speech parametrization techniques. The random forest turns out to be the best among all the four classification algorithms.

PARTICIPANT'S DATASET

This dataset contains user-uploaded audio files in their original, unchecked, and unmodified forms of various infant cry recordings. It was uploaded through the Donate-a-cry Campaign mobile application. It is a collection of sounds in.WAV format with uniform bits and a sampling rate of 128 kbps and 8 kHz respectively. The duration of each sound is approximately 7 seconds. The dataset consists of 457 samples classified into two classes: 382 are classified as hungry and 75 as discomfort.[14]

Accepted for publication June 15, 2023.
From the #Department of Information Technology and Engineering, Amity University in Tashkent, Uzbekistan; †Department of Information Technology, Manipal University Jaipur, India; ‡Department of Engineering and Information Sciences, School of Computer Science, FEIS, University of Wollongong in Dubai, Dubai Knowledge Park, Dubai, UAE; and the §Department of Information Technology, Al al-Bayt University, Jordan and MEU Research Unit, Faculty of Information Technology, Middle East University, Amman 11831, Jordan.
Address correspondence and reprint requests to Gaurav Aggarwal, Department of Information Technology and Engineering, Amity university in Tashkent, Uzbekistan. E-mail: gaggarwal@amity.uz
Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

https://doi.org/10.1016/j.jvoice.2023.06.014

TABLE 1.
Pros and Cons of Our Project

PROS	CONS
1) A better understanding of the baby's needs	1) Less bonding between parents and infant
2) Almost 90% accurate result	2) Mechanically dependent for surety of voice
3) Early detection of any problem with infant	3) the Wrong prediction can lead to overfeeding
4) Ease of pressure for parents and pediatric	4) Time and resources could be a challenge

Speech Standardization

The goldwave noise reduction technique, which employs a frequency analysis method, was used to standardize speech recordings. The level of the recorded speech signal was first boosted to 100% using the maximize volume function in the Goldwave software. The software's several filter algorithms assisted in processing the signal-noise reduction filter function that was chosen. The Fast Fourier Transform size 12% and 90% scaling were used to maximize the overlapping signal window to 16x. Due to the few artifacts it introduced, the scaling value was not assumed to be 100%. Even with some noise present, the signal was unaffected. A clear and noise-free speech signal that can be used for feature extraction and classification was created using this method.

Frame and Windowing

Dividing the signals into overlapping/non-overlapping frames of short-time windows of 25 ms.

Hamming window in Equation 1 is used

$$w[n] = a - b\cos\left(\frac{2\pi n}{N-1}\right) \text{where} \frac{N-1}{2} \le n \le \frac{N+1}{2} \tag{1}$$

with a = 0.54 and b = 1-a = 1–0.54 = 0.46.

METHODOLOGY

Speech Parameterization

Speech parametrization has done on the dataset and obtained eight spectral features, namely, chroma_stft, spectral_centroid, spectral_bandwidth, spectral_rolloff, mel-frequency cepstral coefficients, LPC (Linear Predictive Coding), res, and zero_crossing_rate. These features were extracted using the LibROSA Library, a Python package for music and audio analysis. This step results in forming a.CSV file containing the features in numerical and labeled data to apply the machine learning models.[8,9] The feature description is mentioned in Table 2.

Classification Models

Logistic regression: not a regression algorithm, but a classification one. On the basis of a predetermined set of the independent variable, it is used to determine discrete values (Boolean values like 0/1, yes/no, and true/false) (s). It basically involves fitting data to the logistic regression function to forecast the likelihood that an event will occur. Thus, it is often referred to as logit regression. Because it forecasts probability, its output values range from 0 to 1.[17–19] The graph of the sigmoid function is shown in Figure 1.

Random Forest

An acronym for a collection of decision trees is random forest. The "Forest" in the random forest is a collection of decision trees. Each tree provides a categorization for a new item based on the attributes, and we can observe the tree "votes" for that classification. The categorization with the highest votes is chosen by the forest (out of all the trees in the forest).[20,21]

Support Vector Machine (epSVM)

It is a supervised machine learning technique used primarily for classification issues, while it can also be used for regression. The SVM algorithm plots each data point as a point in an n-dimensional space, where n is the number of features and each feature's value is a specific coordinate's value. The next step in classification is to identify the hyperplane that effectively distinguishes the two classes.[23,24]

Decision Trees

It is a specific kind of supervised learning algorithm that is frequently applied to classification issues. Interestingly, it functions for both continuous and categorical dependent variables. We divide the population into two or more homogenous sets using this approach. To create as many separate groups as feasible, this is done based on the most important characteristics/independent variables.[22]

10-Fold Cross Validation

After training the model, it is not necessary that the model perform well with test data. The pretrained model should give better accuracy with appropriate precision and recall. For this, the validation of the model is required. 10 folds cross-validation technique is widely used to validate and test the effectiveness of the classification models. If the dataset is small this technique gives very effective results.

Once we are done training our model, we cannot just assume that it will work well on data that it has not seen before. We need to be sure of the accuracy of the predictions that our model makes. For this, we need to validate our model. Cross-validation is one of the techniques used

TABLE 2.
Description and Parameter Specifications of Speech Features

S no.	Speech features	Parameters specifications	Description
1	Chroma_stft	Libros. feature.chroma_stft(y = None, sr=22050, S=None, norm=inf, n_fft=2048, hop_length=512, tuning=None, **kwargs)	Applied to create chromagrams from waveforms or power spectrograms
2	Spectral_centroid	Libros. feature.spectral_centroid(y = None, sr=22050, S=None, n_fft=2048, hop_length=512, freq=None)	Used in spectral centroid computation. An average (centroid) is taken from each frame of a magnitude spectrogram after each frame has been normalized three times and treated as a distribution over frequency bins.
3	Spectral_bandwidth	Libros.feature.spectral_bandwidth(y = None, sr=22050, S=None, n_fft=2048, hop_length=512, freq=None, centroid=None, norm=True, $P = 2$)	Used to compute pth-order spectral bandwidth.
4	Spectral_rolloff	Libros.feature.spectral_rolloff(y = None, sr=22050, S=None, n_fft=2048, hop_length=512, freq=None, roll_percent=0.85)	Used to compute roll-off frequency.
5	MFCC	Libros.feature.mfcc(y = None, sr=22050, S=None, n_mfcc=20, dct_type=2, norm='ortho', **kwargs)	Used to compute Mel-frequency cepstral coefficients (MFCCs).
6	LPC	Libros.core.lpc(y, order)	Used to compute Linear Prediction Coefficients via Burg's method. This function applies Burg's method to estimate coefficients of a linear filter on y of order.
7	RMS	Libros.feature.rms(y = None, S=None, frame_length=2048, hop_length=512, center=True, pad_mode='reflect')	Used to compute root-mean-square (RMS) value for each frame, either from the audio samples y or from a spectrogram S.
8	Zero_crossing_rate	Libros.feature.zero_crossing_rate(y, frame_length=2048, hop_length=512, center=True, **kwargs)	Used to compute the zero-crossing rate of an audio time series.

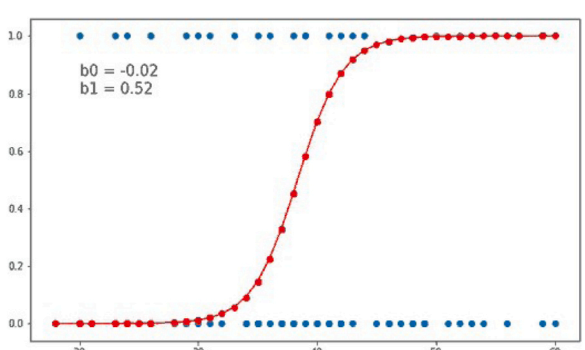

FIGURE 1. Graph of Sigmoid function.

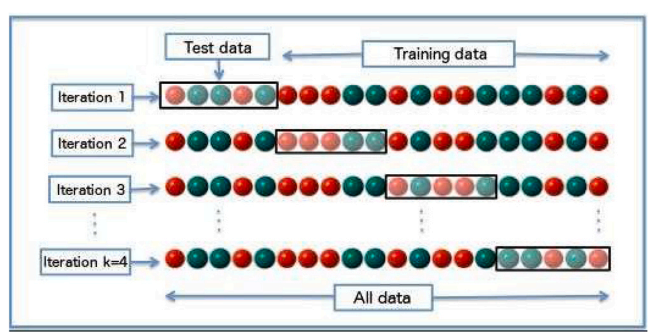

FIGURE 2. 10-fold cross validation.

to test the effectiveness of machine learning models, it is also a procedure to re-sample used to evaluate a model if we have a limited dataset.[25] In comparison to other methods, 10-Fold is well-liked and simple to comprehend, and it typically produces a less biased model. Since it makes sure that every observation from the original dataset has a fair chance of appearing in both the training and test sets. This method follows the below steps. The Block diagram of 10-fold cross-validation is shown in Figure 2.

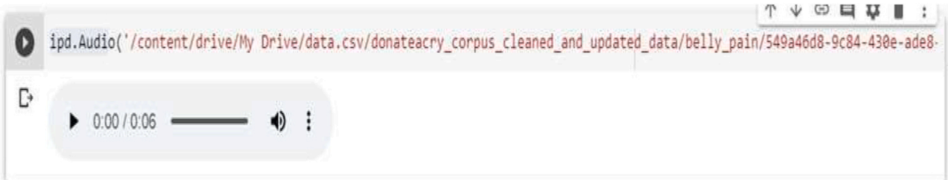

FIGURE 3. Reading/listening to the audio files.

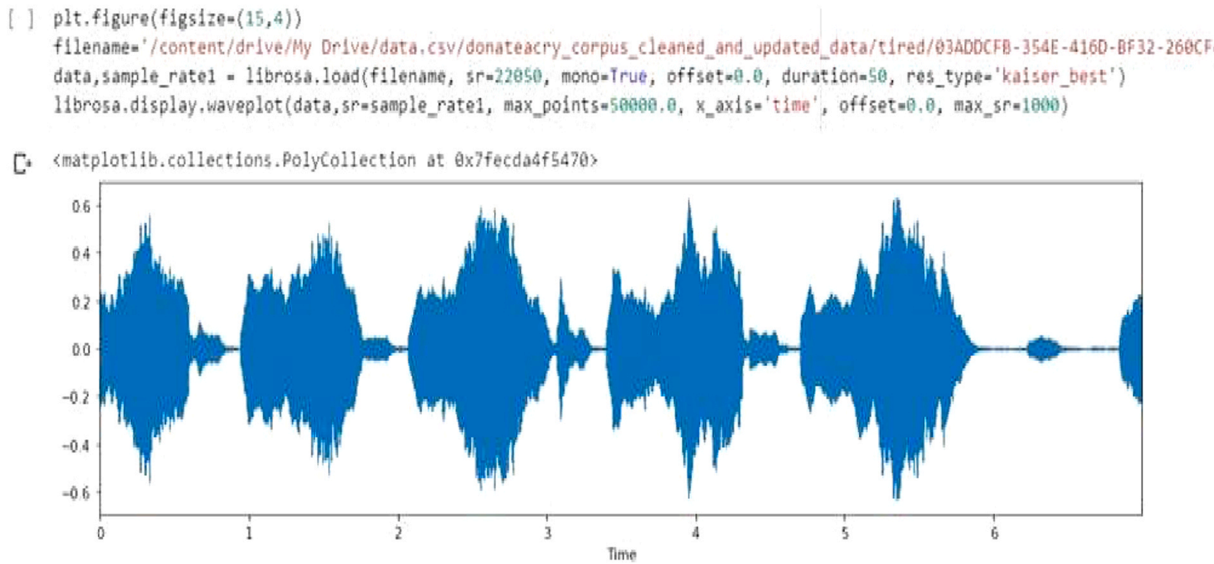

FIGURE 4. Plotting waveform.

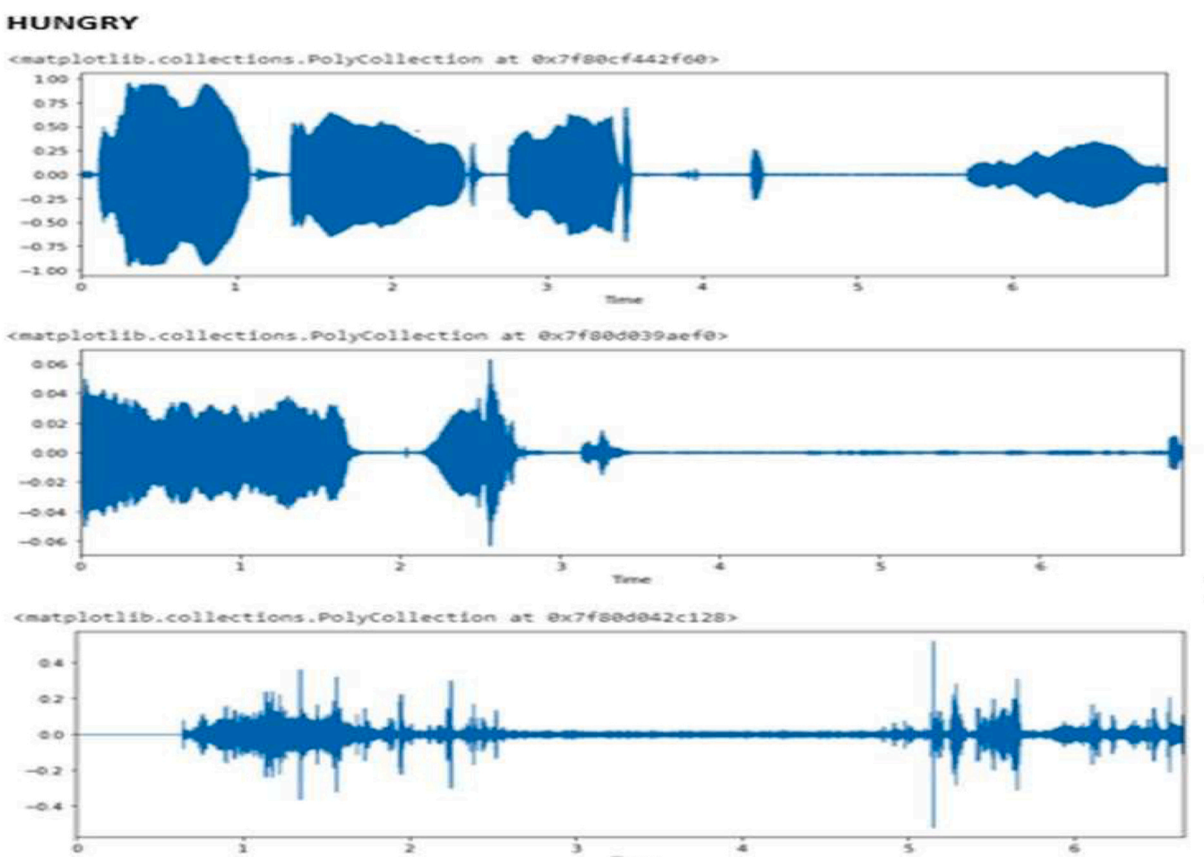

FIGURE 5. Plotting hungry waveform.

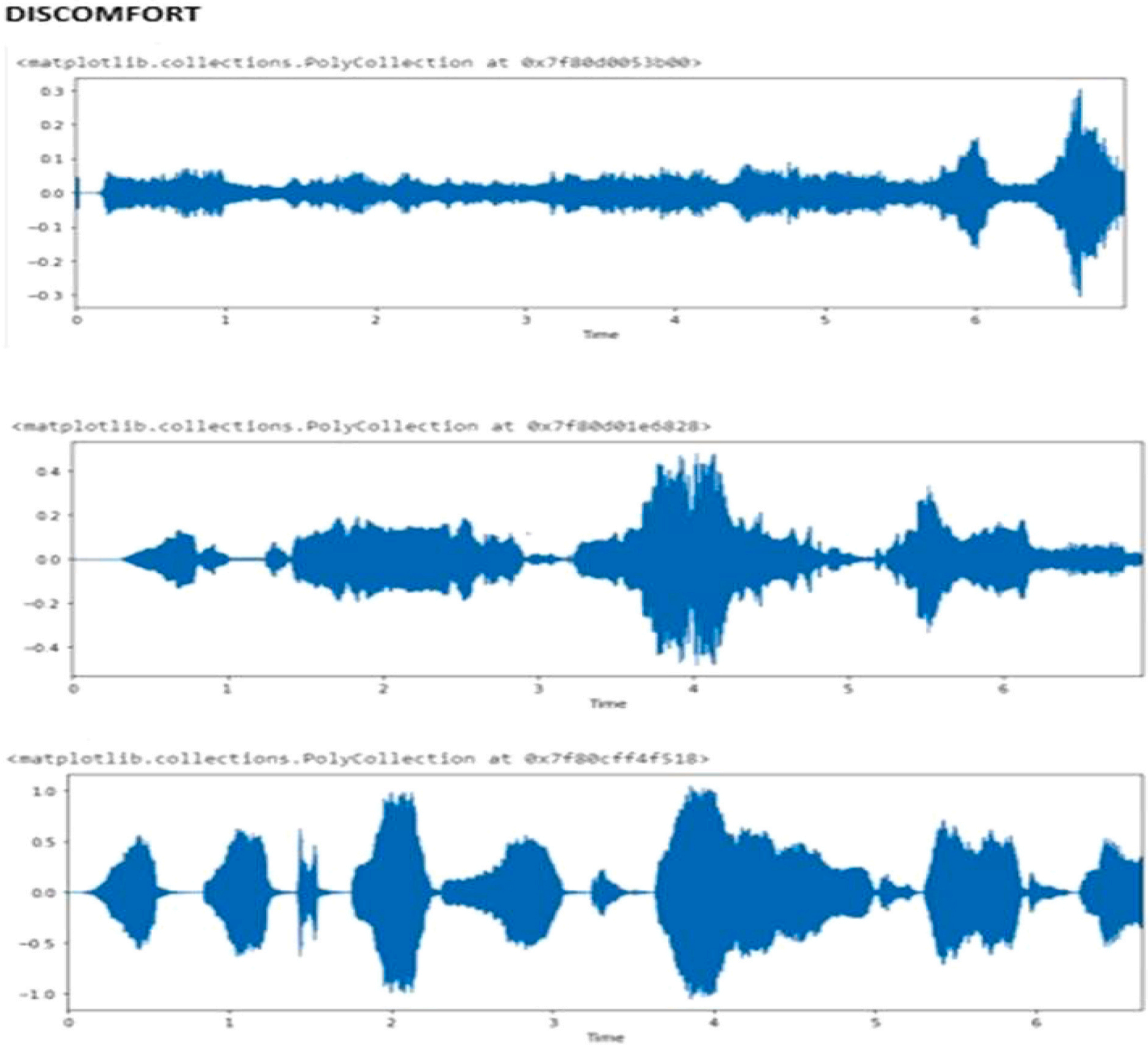

FIGURE 6. Plotting discomfort waveform.

```
[1] import librosa
    from librosa import feature
    from librosa import core
    import numpy as np

    f_list_1 = [
        feature.chroma_stft,
        feature.spectral_centroid,
        feature.spectral_bandwidth,
        feature.spectral_rolloff,
        feature.mfcc,
        core.lpc
    ]

    f_list_2 = [
        feature.rms,
        feature.zero_crossing_rate

    ]

    def get_feature_vector(y,sr):
      featvect_1 = [ np.mean(funct(y,sr)) for funct in f_list_1]
      featvect_2 = [ np.mean(funct(y)) for funct in f_list_2]


      feature_vector =   featvect_1 + featvect_2
      return feature_vector
```

FIGURE 7. Feature extraction.

RESULTS

Speech Parameterization

Speech parametrization has been done using the Python library to extract the speech features of the infant's cries. Speech features such as chroma_stft, spectral_centroid, bandwidth, spectral_rolloff, mel-frequency cepstral coefficients, linear predictive coding, res, zero_crossing_rate extracted from the speech signal and converted into a.csv file. The step-by-step process of speech parameterization has shown below: Figures 3 and 4.

Figures 5 and 6 has shown the speech waveform of a hungry child and the speech waveform of s discomfort child respectively.

Speech parameterization has applied and extracted eight acoustic features.[10–15] The features are extracted as the mean value for each sample and then all the features and samples are combined into a vector [no. of samples] [no. of features = 8] as shown in Figures 7 and 8.[16]

```
[ ] audiosfeat1 = []
    for file in audio1f:
      '''
      y is the time series array of the audio file, a 1D np.ndarray
      sr is the sampling rate, a number
      '''
      y,sr = librosa.load(file,sr=None)
      feature_vector = get_feature_vector(y, sr)
      audiosfeat1.append(feature_vector)
      print('@', end= " ")

@ @ @ @ @ @ @ @ @ @ @ @ @ @ @ @

print(audiosfeat1)

[[0.2763468, 1212.645457849391, 1013.8412480194755, 2392.7201704545455, -21.87616,
```

FIGURE 8. Creation of a list of features.

```
import csv

o1 = 'tired.csv'
o2 = 'hungry.csv'

header =[
     'chroma_stft',
     'spectral_centroid',
     'spectral_bandwidth',
     'spectral_rolloff',
     'mfcc',
     'lpc',
     'rms',
     'zero_crossing_rate'
]
with open(o1,'+w') as f:
  csv_writer = csv.writer(f, delimiter = ',')
  csv_writer.writerow(header)
  csv_writer.writerows(audiosfeat1)

with open(o2,'+w') as f:
  csv_writer = csv.writer(f, delimiter = ',')
  csv_writer.writerow(header)
  csv_writer.writerows(audiosfeat2)
print('a')

a
```

FIGURE 9. CSV files for the combined features.

Figure 9 shows the CSV (comma-separated values) files for the features combined to obtain the final numerical dataset.

The final dataset is shown in Figure 10.

Data Standardization

Reading the dataset and saving it into X and the result values into y in a list format and standardizing the values. Figure 11: reading, storing, and standardizing the data.

64	0.304779	1512.417	1047.674	2865.945	-16.9128	4.14E-05	0.055765	0.258758	discomfort
65	0.392436	1231.275	952.0748	2311.506	-18.2629	0.000163	0.043511	0.223957	discomfort
66	0.222782	872.265	677.4502	1317.302	-10.3984	0.044732	0.297763	0.1547	discomfort
67	0.335631	1278.939	890.1866	2273.97	-17.1001	0.000249	0.041245	0.227965	discomfort
68	0.36917	1657.967	774.4191	2518.801	-18.785	0.491004	0.11019	0.372983	discomfort
69	0.303567	1250.885	819.8965	2206.036	-21.0023	0.239472	0.050132	0.232514	discomfort
70	0.409235	967.8202	899.018	1850.284	-26.0251	0.000323	0.008983	0.201993	discomfort
71	0.321983	1875.677	894.0634	2894.093	-17.1105	0.436312	0.063087	0.454572	discomfort
72	0.426017	1303.955	1004.728	2569.954	-13.9024	0.456252	0.035527	0.221088	discomfort
73	0.400473	1152.034	1106.183	2570.241	-26.3593	0.004113	0.001467	0.129923	discomfort
74	0.449821	1597.251	968.0963	2637.678	-27.6522	0.000143	0.005405	0.336239	discomfort
75	0.53989	1724.884	888.862	2734.556	-28.5354	0.045359	0.00493	0.363602	discomfort
76	0.392775	1703.538	830.6463	2655.483	-20.6381	0.836218	0.091974	0.377994	discomfort
77	0.460996	1381.156	1022.757	2590.345	-31.8632	0.006148	0.000853	0.257933	hungry
78	0.504262	1101.828	926.5277	2220.667	-25.6908	0.043601	0.006271	0.151937	hungry
79	0.256459	1289.919	856.9058	2278.693	-8.7733	0.024393	0.194313	0.227384	hungry
80	0.293549	1846.934	859.9338	2886.898	-13.4996	0.634889	0.111819	0.384707	hungry
81	0.491528	1469.59	913.9718	2469.141	-19.0262	4.43E-05	0.061903	0.293279	hungry
82	0.247078	1068.402	864.0536	1998.118	-15.5942	1.87E-05	0.172394	0.155571	hungry
83	0.414935	1334.69	906.8851	2366.356	-29.1677	0.019229	0.004806	0.235329	hungry
84	0.354079	1409.765	891.633	2393.075	-17.5779	2.27E-05	0.084315	0.280256	hungry
85	0.409208	1568.892	928.9265	2615.092	-18.0507	2.36E-05	0.053283	0.318302	hungry

FIGURE 10. Final dataset of infant cries.

```
array([[ 2.76e-01,  1.21e+03,  1.01e+03,  2.39e+03, -2.19e+01,  4.29e-04,
         2.70e-02,  1.84e-01],
       [ 3.85e-01,  1.28e+03,  9.87e+02,  2.42e+03, -2.62e+01,  1.03e-04,
         6.61e-03,  2.07e-01],
       [ 3.99e-01,  1.27e+03,  9.67e+02,  2.42e+03, -2.33e+01,  7.14e-04,
         1.37e-02,  2.15e-01],
       [ 2.94e-01,  1.33e+03,  7.79e+02,  2.25e+03, -2.47e+01,  1.76e-01,
         2.01e-02,  2.48e-01],
       [ 3.30e-01,  1.05e+03,  8.74e+02,  1.97e+03, -2.35e+01,  2.63e-05,
         2.92e-02,  1.58e-01]])
```

```
array(['discomfort', 'discomfort', 'discomfort', 'discomfort',
       'discomfort'], dtype=object)
```

```
array([[-0.91, -0.56,  1.26, ..., -0.56, -0.72, -0.86],
       [ 0.36, -0.33,  1.01, ..., -0.56, -1.04, -0.59],
       [ 0.51, -0.36,  0.81, ..., -0.56, -0.93, -0.49],
       ...,
       [-1.35, -0.89, -0.32, ..., -0.56, -0.31, -0.95],
       [-0.66, -0.4 ,  0.09, ..., -0.56,  0.37, -0.56],
       [-0.99, -0.89, -0.47, ..., -0.56, -0.35, -0.72]])
```

FIGURE 11. Reading, storing, and standardizing the data.

Classification Accuracy

Four classification models are applied to accurately classify the baby's cries into hunger and discomfort. The dataset consists of 457 samples classified into two classes: 282 are classified as hungry and 175 as discomfort. As the dataset is relatively inclined toward hunger, a 10-folds cross-validation approach is applied to test the effectiveness of machine learning models. Librosa library has been used for speech

TABLE 3.
In Spectrogram Image Analysis, SVM Performs Better Than Similar Classifiers for Identifying Infant Cry Types

Models	Metric	Discomfort	Hungry
Random forest	Accuracy	94.32%	94.03%
	Specificity	94.51%	94.17%
	Sensitivity	94.22%	93.69%
Logistic regression	Accuracy	93.81%	93.75%
	Specificity	93.88%	94.09%
	Sensitivity	93.64%	93.57%
Decision tree	Specificity	92.28%	92.32%
	Accuracy	92.11%	92.22%
	Sensitivity	91.93%	92.02%
SVM	Accuracy	97.34%	97.12%
	Specificity	97.56%	97.37%
	Sensitivity	97.06%	96.86%

SVM, support vector machine.

TABLE 4.
Studies Conducted With Similar Features and Classification Models

Method	Dataset	Features	Classifier	Accuracy (%)
Bănică et al[12]	Dunstan baby language	MFCC	GMM	81.80
Kheddache et al	Self-recorded	MFCC	PNN	60.00
Kheddache et al	Self-recorded	Tglide, RFsdys, TRP, TUP, F0glide,F0	PNN	72.00
Bano et al	Self-recorded	Pitch, energy, and mel frequency cepstral coefficients (MFCCs)	*k*-NN	86.00
Maghfira et al	Dunstan baby language	Spectrogram	CNN-recurrent neural network (RNN)	94.97
Franti et al	Dunstan baby language	Spectrogram	CNN	89.00
Bano et al	Dunstan baby language	Pitch, energy, and MFCCs	*k*-NN	85.00
Abou-Abbas et al	Self-recorded	MFCC	HMM	83.79
Kabir et al	Baby chillanto database	Spectrogram	CNN+SVM	91.10
Sharma et al	Donate a cry corpus	Frequency, entropy, spectral	GMM	81.27
Anders et al	Internet	Spectrogram	CNN	72.00

parameterization. Eight speech features were used in the classification models to predict the accuracy. Among logistic regression, seep support vector machine, decision tree and random forest. SVM classification model has outperformed all the other classifiers with an accuracy of 97.34%. In this study, two types of cries are classified. The classification accuracies for all the classifiers has been shown in Table 3.

CONCLUSION AND FUTURE SCOPE

The research investigated the use of speech signal features for classifying infant cries. To build a robust classification model, different speech features were extracted from the infant's cries. A total of 457 speech samples were used in the study. Feature 1 is the extraction of spectral, root-mean-square, and zero_crossing_rate features obtained. Feature 2 is the mel-frequency cepstral coefficients, which are calculated over the total number of frames. Feature three is the LPC cepstral coefficients which depend on the past samples. Four supervised learning models are used to test the accuracy: logistic regression, decision tree, SVM, and random forest. It has been observed that the random forest classification algorithm outperformed other classification models with an accuracy of 97.34%. In the next phase, the SVM model will also be introduced in this research to get better results. Several research has been taken place with similar techniques to correctly classify infant cry into multiple categories. Some of the renounced studies are tabulated in Table 4.

Few models were able to forecast discomfort when applied without K-fold cross validation. When the models were applied using K fold cross validation (value of K being 10), we were able to predict the accuracy of the Discomfort class using simple decision trees and random forest, but even then, the prediction rate for discomfort was quite low compared to that of Hungry. More of the discomfort samples could be predicted when the models were used with different ratios of the class samples. The random forest model performed best and produced the highest true predictions when we divided our dataset into equal portions for hungry and discomfort, respectively.

In future work, we are planning to apply the same methodology to a mature and larger dataset. Also, add more precise and variant features in the feature extraction phase. In speech parameterization, many handcrafted features such as wavelet-based, energy-based, and energy-based parameterization techniques will also be included in the upcoming work.[26–30] These techniques are well advanced and may increase the reliability of the models with five types of infant cries. Classifying the infant cry into more than two classes as per universal theory that is, into five classes, namely belly pain, hungry, discomfort, sleepiness, and burping. And pairing this classifying model with an app to make it easily and freely available to everyone.

DISCUSSION

Identifying the type of cries using machine learning techniques are quite helpful for caregivers in many ways such as: a) Early Detection: By using machine learning models to categorize screams as signs of hunger or discomfort, caregivers can be alerted to their infants' needs before they become serious problems. This can facilitate rapid response and proper care, enhancing the infant's general well-being. b) Improved communication: infants, particularly in the early stages, may find it difficult to adequately convey their demands. By offering insights into the demands underlying the scream, machine learning models can operate as a tool to close the communication gap between caregivers and newborns, enabling caregivers to comprehend, and meet those needs more effectively. c) Personalized care: algorithms that use machine learning can adapt to the specific traits of each newborn by learning from their individual cry's patterns. The quality of care given is improved by allowing carers to customize their caregiving tactics based on the unique requirements and cues of their newborns. d) support for new parents: it can be difficult for new parents or caregivers to decipher the significance of an infant's cries if they lack previous experience. To help parents, comprehend various cry kinds, potential reasons, and suitable solutions, machine learning models can be a useful resource.

Machine learning techniques offer technological feasibility for infant cry classification due to the following reasons: a) The capacity to gather enormous datasets of recordings of infant cries has been made possible by technological improvements and an increase in the use of digital gadgets. With the use of these datasets, machine learning models may be trained to make precise classifications. b) Effective algorithms: machine learning algorithms have shown promising results in a variety of signal processing applications, including cry categorization. Because of these algorithms' effective processing and analysis of cry signals, technological deployment is possible. c) Real-time processing: machine learning models can be implemented on hardware or specialized software with enough processing power to enable real-time analysis of cry signals. This improves the technology's practicality and utility by enabling caregivers to get quick feedback and replies based on the classification results.

Funding: No funding.

CRediT authorship contribution statement

GA: Conceptualization, Methodology, experiments; **A, KJ**: Literature Study, **GA, KJ**: Data creation, Experiments; **KJ, GA, JI**: Classification; **KJ**: Validation; **GA, JI**: Analysis and validation; **JI**: formal analysis; **KJ, GA**: investigation; **GA**: resources; **KJ**: writing—original draft preparation; **GA, KJ, JI**: writing—review and editing; **JI**: visualization.

DATA AVAILABILITY STATEMENT

The data for baby cry is available online from the Kaggle website.

DECLARATION OF COMPETING INTEREST

It has been declared that there is no conflict of interest for any author in the manuscript.

REFERENCES

1. Mercuri M, Stack DM, Mantis I, et al. Maternal and infant touching behaviours during perturbed interactions: associations with maternal depressive symptomatology and infant crying. *Infant Behav Dev.* 2023;71:101821.
2. Muller I, Ghio D, Mobey J, et al. Parental perceptions and experiences of infant crying: a systematic review and synthesis of qualitative research. *J Adv Nurs.* 2023;79:403–417.
3. Crouch JL, Bridgett DJ, Milner JS, et al. Prolonged infant crying: caregiving quality and child physical abuse risk. *J Interpers Violence.* 2023;38:3298–3320.
4. Manigault AW, Sheinkopf SJ, Carter BS, et al. Acoustic cry characteristics in preterm infants and developmental and behavioral outcomes at 2 years of age. *JAMA Netw Open.* 2023;6:e2254151.
5. Cymet T, Sciuva D, Sciuva A. A cry to look at crying differently. *Med Rep Case Stud.* 2023;8:1–3.
6. Khalilzad Z, Tadj C. Using CCA-fused cepstral features in a deep learning-based cry diagnostic system for detecting an ensemble of pathologies in newborns. *Diagnostics.* 2023;13:879.
7. Cohen R, Ruinskiy D, Zickfeld J, et al. Baby cry detection: deep learning and classical approaches. In: Pedrycz W, Chen S-M, eds. *Development and Analysis of Deep Learning Architectures, Studies in Computational Intelligence.* Cham: Springer International Publishing; 2020:171–196. https://doi.org/10.1007/978-3-030-31764-5_7.
8. Sujatha K, Nalinashini G, Ganesan A, et al. *Internet of medical things for abnormality detection in infants using mobile phone app with cry signal analysis. Implementation of Smart Healthcare Systems using AI, IoT, and Blockchain.* Academic Press; 2023:169–191.

9. Lahmiri S, Tadj C, Gargour C, et al. Optimal tuning of support vector machines and k-NN algorithm by using Bayesian optimization for newborn cry signal diagnosis based on audio signal processing features. *Chaos Solitons Fract.* 2023;167:112972.
10. Vaishnavi V, Suveetha Dhanaselvam P. Neonatal cry signal prediction and classification via dense convolution neural network. *J Intell Fuzzy Syst.* 2022;42:6103–6116.
11. Matikolaie FS, Tadj C. On the use of long-term features in a newborn cry diagnostic system. *Biomed Signal Process Control.* 2020;59:101889.
12. Lahmiri S, Tadj C, Gargour C, et al. Characterization of infant healthy and pathological cry signals in cepstrum domain based on approximate entropy and correlation dimension. *Chaos Solitons Fract.* 2021;143:110639.
13. Lahmiri S, Tadj C, Gargour C. Biomedical diagnosis of infant cry signal based on analysis of cepstrum by deep feedforward artificial neural networks. *IEEE Instrum Meas Mag.* 2021;24:24–29.
14. Hariharan M, Sindhu R, Vijean V, et al. Improved binary dragonfly optimization algorithm and wavelet packet based non-linear features for infant cry classification. *Comput Methods Programs Biomed.* 2018;155:39–51.
15. Badreldine OM, Elbeheiry NA, Haroon ANM, et al. *Automatic diagnosis of asphyxia infant cry signals using wavelet based mel frequency cepstrum features. 2018 14th International Computer Engineering Conference ((ICENCO))*. IEEE; 2018:96–100.
16. Sarria-Paja M, Falk TH. Fusion of auditory inspired amplitude modulation spectrum and cepstral features for whispered and normal speech speaker verification. *Comput Speech Lang.* 2017;45:437–456.
17. Kheddache Y, Tadj C. *Newborn's Pathological Cry Identification System*. IEEE; 2012:1024–1029.
18. Hariharan M, Yaacob S, Awang SA. Pathological infant cry analysis using wavelet packet transform and probabilistic neural network. *Expert Syst Appl.* 2011;38:15377–15382.
19. Manfredi C, Bandini A, Melino D, et al. Automated detection and classification of basic shapes of newborn cry melody. *Biomed Signal Process Control.* 2018;45:174–181.
20. Mary L. *Extraction of Prosody for Automatic Speaker, Language, Emotion and Speech Recognition*. Springer; 2018.
21. Hu X, Yi W, Jiang L, et al. Classification of metaphase chromosomes using deep convolutional neural network. *J Comput Biol.* 2019;26:473–484.
22. Bushehri SF, Zarchi MS. An expert model for self-care problems classification using probabilistic neural network and feature selection approach. *Appl Soft Comput.* 2019;82:105545.
23. Bashiri A, Hosseinkhani R. Infant crying classification by using genetic algorithm and artificial neural network. *Acta Medica Iran.* 2020:531–539.
24. Mahmoud AA, Alawadh INA, Latif G, et al. *Smart nursery for smart cities: infant sound classification based on novel features and support vector classifier. 2020 7th international Conference on Electrical and Electronics Engineering*. IEEE; 2020:47–52.
25. Matikolaie FS, Kheddache Y, Tadj C. Automated newborn cry diagnostic system using machine learning approach. *Biomed Signal Process Control.* 2022;73:103434.
26. Guido RC. Wavelets behind the scenes: practical aspects, insights, and perspectives. *Phys Rep.* 2022;985:1–23.
27. Guido RC. A tutorial on signal energy and its applications. *Neurocomputing.* 2016;179:264–282.
28. Guido RC. A tutorial review on entropy-based handcrafted feature extraction for information fusion. *Inf Fusion.* 2018;41:161–175.
29. Guido RC. Enhancing teager energy operator based on a novel and appealing concept: signal mass. *J Frankl Inst.* 2019;356(4):2346–2352.
30. Guido RC. ZCR-aided Neurocomputing: a study with applications. *Knowl-based Syst.* 2016;105:248–269.