AN AUTOMATIC INFANT CRY SPEECH RECOGNITION USING ARTIFICIAL NEURAL NETWORK

Satrughan Kumar Singh*[1], Muniyan Sundararajan[2] and Jainath Yadav[3]

[1,3]Department of Computer Science, Central University of South Bihar, Gaya– 824236, Bihar, India

[2]Department of Mathematics & Computer Science, Mizoram University, Aizawl– 796004, Mizoram, India

*[1]satrughanksingh@yahoo.com

ABSTRACT

Infant cry speech recognition is one of the important and emerging research areas for medical science. The machine learning technology provides an automatic classification of diverse physical and physiological situations of the newborn or infant baby by their cry signal. In this study, we present a proposed method for automatic infant cry recognition & detection based on artificial neural network (ANN) through a soft computing system developed in MATLAB. We have developed an infant cry recognition system whose main task is to recognize infant cries into different classes. The clustering algorithm is used for splitting a given anonymous dataset into a fixed number of the clusters. We have considered two features, namely MFCC and LPCC which are extracted from an audio sample of infant cries and they are fed into recognition module. The overall average accuracy of the proposed ANN classifier is 94.77% approximately. In the proposed system, the experimental results have shown that ANN provides good accuracy for automatic infant cry detection. The evaluation revealed that ANN classifier achieved effective and good accuracy on the infant cries dataset.

Keywords: *Cry datasets, neural networks, computational audio processing, MATLAB, infant cry detection*

1. INTRODUCTION

The newborn birth cry of infant cries is important information about the health of an infant. Crying is the first sound the baby makes when he enters the world outside of his mother's womb, which is a very positive sign of a new healthy life [1]. Infant crying is the first sound of communication after immediate birth. The acoustic characteristic is extracted from the crying wave by a process called acoustic analysis. It is considered as the important criteria in determining the Apgar score as the newborn grows, the newborn infant cry acoustics changes with the integration of the vocal tract system. Newborn or infant are found to produce many sounds which reflect the learning their spoken communication. So that collected crying signals randomly from different databases. The Authors has analyzed seven different types of infant cry. When your baby cries, we think that he must be crying because of hunger, then the caretaker has to feed him. So in this state of hunger, feeding before or immediately after crying will be a signifier. Some hunger signs are smacking, rooting, and putting their hands to their mouth. If the baby cries due to colic or other problems related to colic, then the caretaker has to take some immediate measures. So this kind of state is called stomach pain. In a state of grief, when the baby comes due to sadness and sadness, the caretaker needs to pay special attention to the baby so that the baby feels a safe environment. Often there are many reasons for this type of situation such as feeling very uncomfortable with wet diapers, dirty diapers, need for sleep and feeling too cold or too hot. In the case of dirty diapers, some babies tell you immediately through hints and get an idea when they need to change dirty diapers now. In such a situation, such a condition can be checked very easily. In need of sleep, when the baby feels that he needs to sleep now and should be able to go from sleep to sleep, which is more difficult for him to do than we realize. Especially when babies are tired, they may cry and fuss instead of moving their legs in this type of position. In a very cold or very hot situation, if your baby feels too cold or too hot, you take her clothes off to change her diaper or clean the bottom of her feet with a cold wipe. , then in such a situation the baby can oppose you by stopping. In infant crying, colic is also a type of crying. In colic, the baby cries for a long time once or twice a day. It is usually comforting. It also works normal between the cries of the baby. In case of colic, the baby does not get sick, but also has enough to eat and does not feel hungry. If your baby is healthy and cries a lot, he may have a colic. On the other hand, if your baby is frustrated and embarrassed, he may refuse your efforts to pacify him. He may even signal by clenching his fists or pulling his knees up or bending his back. Caffeine is a type of stimulating baby cry that can lead to falling asleep or crying. Breastfeeding mothers should be very careful and very limited in their caffeine intake. When a baby cries when needed, the baby will need lots of hugs and reassurance to comfort him. Trying to hold the baby sling close to you can help the baby waving and singing.

2. LITERATURE REVIEW

Infant baby cry classification is a challenging task for machine learning approach in the next generation of artificial intelligence based computing research. There are many machine learning model and classifiers that are effectively to classify an infant baby cry sound classification. In this literature survey section, we presented the related work with classifier methods carried out for infant baby cry classification by many researchers are shown in **Table 1**.

Table 1: The recognition accuracy of infant cries using different models given in the literature

Ref. No.	Author(s)	Model(s) used	Accuracy (%)
[3]	Pai C-Y (2016)	kNN	76.47
[4]	Nanni L et al. (2010)	SFFS-SVM	93

[5]	Brahnam S et al. (2006)	SVM	87.8
[7]	Barajas-Montiel SE et al. (2006)	FSVM	97.83
[8]	Zamzami Z et al. (2015)	KNN	96
		SVM	94
[9]	Fotiadou E et al. (2014)	SVM	98
[10]	Pal P et al. (2006)	LPCC	64
		MFCC	74.2
[11]	Abdulaziz Y et al. (2010)	MFCC	76.2
		LPCC	57
[12]	Severini et al. (2019)	DNN (PRC-AUC)	87.28
[13]	Abou-Abbas L et al. (2017)	FFT-GMM	94.29
		EMD-HMM	92.16
[14]	Alaie HF et al. (2016)	MLP	91.68
[15]	Naithani G et al. (2018)	HMM	89.2
[16]	Mohammed YA (2018)	CDHMM	96.1
		ANN	79
[17]	Ahmad SMBS et al. (2010)	HMM	96.1
[18]	Abou-Abbas L et al. (2015)	HMM	83.79
[19]	Jam MM (2009)	MLP ANN	88.3
[20]	Zabidi A (2010)	MLP-ANN	95.07
[21]	Brahnam S et al. (2007)	NNSOA	90.20
		SVM	82.35
		PCA	80.35
		LDA	76.96
		SVM	94

3. METHODOLOGY

In this part, we describe the process and modules for the proposed system. The block diagram of the proposed model for training and testing datasets in infant cry detection system is presented in **Figure 1**.

 ISSN-2581-8953

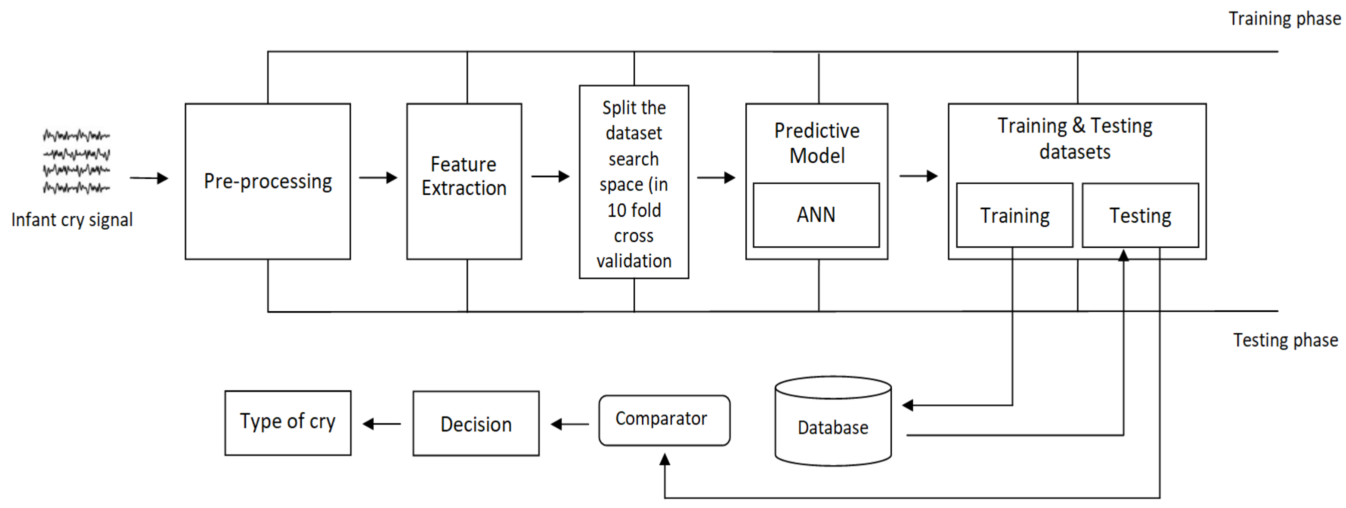

Figure 1: Block diagram of proposed model for training & testing datasets in infant cry detection system

The working process is based on training & testing phase in supervised pattern classification. Audio signal decomposition accepts the input data of audio infant cry signal. The original audio signals of infant cries are converted from the time domain to another domain viz. frequency or ceptral domain for better characterizes it. This method receives the cry signal and draws out important information within each dataset frame to form a feature vector sets. The learning system for training infant cry signals provides a mechanisms of acoustic model generation. Training data is used to estimate the weights of the neural network [6].

3.1. Mathematical Background for Cry Signal Analysis

Sometimes signal decomposition also called forepart module in any audible analysis procedure. It is the first and initial step to develop the system of recognition. Signal decomposition facilitates the extraction and separation of audio signal components from the composite signal, which should ideally be related to semantic units. Statistical program can be used such a model, which based on the training corpus, for analyzing new, unknown texts. Training facilitates a learning system for creating an acoustic training dataset. The training model is used to decide the cry class: hungry, pain, sleeping and so on. The task of training module is pre-processing of cries datasets and feature extraction with validation establishment. Hidden markov model (HMM) chain contains all the possible states of a system and the probability of transiting from one state to another.

$$P(X_{n+1}) = \frac{x \mid X_1 = x_1, X_2 = x_1, \ldots, X_n = x_n}{\text{can be ignored}} = P(X_{n+1} = x \mid X_n = x_n)$$

This model is easy to use and handle. However, in many machine learning systems, not internal states. Some works treat them as latent factors for the inputs. An internal state will be {H or S}. But someone can get some relative hints from what observed. The probability of observing given internal states is emission

 ISSN-2581-8953

probability, and transiting from one internal state to another is transition probability. Gaussian mixture model (GMM) is a famous and oldest clustering method in which each cluster is modeled using Gaussian distribution. This model facilitates a very flexible and probabilistic approach on data modeling having hard assignments into clusters like k-means and hierarchical clustering. Each data points are generated by Gaussian distribution.

3.2. Datasets Collection for Infant Baby Cries

A digital high quality sound recorder and microphone placed 5 cm from the infant baby's mouth with the help of nurse or parents and connected to laptop. A datasets of total 927 infant cry signals were recorded. These cries signal samples are recorded with sampling rate of 16 KHz, resolution of 16 bits, and duration of 30 seconds. These cry signals were categorized separately with different cry types. The cry types are categorized as sad, unhappy, dirty, need sleep, too cold too hot, colic, hunger and caffeine. Out of 927 cry signal, sad contains 86, unhappy contains 112, 146 of dirty, need sleep contains 98, 118 of too cold too hot, 132 to colic, hunger contains 129 and 106 of caffeine cry type.

3.3. Pre-Processing For Infant Cry Signals

The data pre-processing phase is the initial phase in any machine learning projects, leading to better results from the applied model. **Figure 2** shows the block diagram of process of pre-processing steps. It is required for voice activity detection and noise reduction. The proper dimensions for applying filter to remove unordered impurities and unwanted noise from signal.

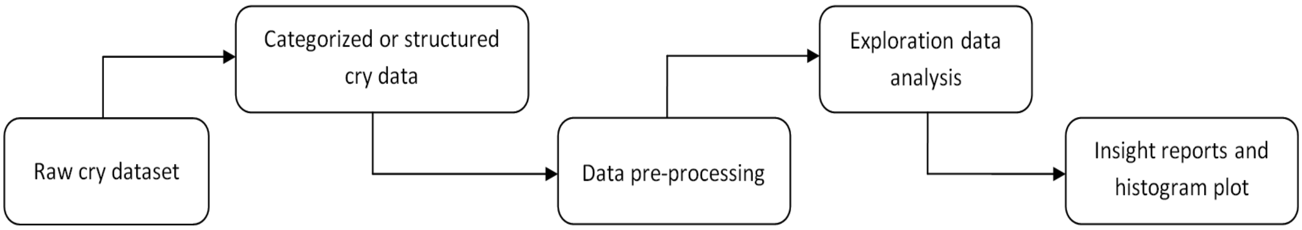

Figure 2: Process of pre-processing

3.4. Features Extraction

The feature selection method can be used to choose important features and to remove inconsistencies and redundancies from these pre-processed image datasets [2]. Features are selected by the feature extraction process by extracting some important features from the pattern of the infant cry audio signal. These facilities will be selected on a special basis which will effectively discriminate between various causes like pain, hunger, crying etc. Once the complete feature extraction step is completed, a feature vector file is generated, which is handed over to the classifier for processing in the next step. MFCC, Pitch, and LPCC are major issue for planning to use in infant cry detection. MFCCs are derived to represent the acoustic characteristics of the audio signal. With the help of MFCC, the short term power spectrum of an audio infant cry signal is shown. Mel Frequency Cepstral Coefficient (MFCC) is an audio feature commonly used in speech recognition systems. While calculating MFCC we will make audio signal into short frames. Then we take power spectrum which helps in identifying the different frequencies present in audio. Then we take logarithm of that value which decrease the difference in loudness and makes easy to compare. MFCCs are coefficients that describe the mel frequency cepstrum. Pitch is an important attribute of voiced speech that contains particular information. Initially input voice push bottom is pressed and the test cry signal is loaded. It is based on splitting the audio segment into small frames. Linear predictive coding is a tool used mostly in audio signal processing and speech processing. It is one of the most powerful speech analysis techniques LPCC is linear predicted ceptral coefficients in the cepstrum domain analysis where an inputted speech sample with a linear combination of the past speech samples for finding approximation, autocorrelation and prediction on current time-domain sample. It specifically assesses and estimates the response of the human auditory system and at the same time also demonstrates good performance in various applications. The selected feature set should accurately capture the information about the health condition of the baby. Cry signal can be described by their features within three common domains, namely time domain analysis, frequency domain analysis, and ceptral domain analysis.

3.5. Proposed Model for Classifying Infant Cries Using Acoustic Analysis

In this study, the classification step involves ANN classifiers to classifying infant cry types with multi-layered architectural framework of input layer, hidden layer, and output layer by applying 10-fold matrix operation for training and testing cry datasets with cross validation. These five layers (one input layer, three hidden layers and one output layer) have 13, 20, 7, 20 and 13 neurons, respectively. Artificial neural networks are matured to work best on large datasets. It robustly implements non-linear mappings that are used in many applications. In this model, the sound source is to be located at larynx, and the vocal folds with different modes of vibration are represented by three types of infant cry, namely phonation, hyperphonation and dysphonation. In phonation cry type, the vocal folds vibrate completely at 0-F range of 250-700 Hz (approximately). Conversely, hyperphonation cry type results the vocal tract from vibration with a 0-F range of 1000-2000 Hz (approximately). In hyperphonation cry type, only a thin portion of the vocal ligament is involved. An aperiodic and periodic source of cry sound occurs during the generation of vocal folds and tracts. However, vocal folds and tracts are modulated by the turbulence sound. The resultant filtering process gives the positions which are dependent upon shape of vocal folds and tracts. In a higher fundamental frequency, it is expected to result in increased pressure that may be influenced by the contraction of the musculature. An infant cry sequence consists of a series of long expiratory cry tone or sound with intervals. In the state of newborn respiratory system, it is expected that the duration od sound interval of the infant cry units are dependent. It is also expected that a diminished vital capacity results lower and short intensity infant cry units with a relative small tidal volume.

4. RESULTS & DISCUSSIONS

For implementing proposed ANN based soft computing system, we have considered MATLAB programming language. It uses confidence measures to detect and classify cries through 60% split, 70% split, 80% split and 10-fold Cross-Validation (CV) on the infant cry datasets. In this study, we categorized cry types in 8 groups such as sad, unhappy, diaper (containing three sub-groups with dirty, need sleep and too cold too hot), colic, hunger and caffeine. **Table 2** shows the average accuracy of proposed ANN classifier through different split and 10-fold CV.

Table 2: Average accuracy (in %) of infant cry detection through different splits and 10-fold CV using proposed ANN classifier

Cry type		60% split	70% split	80% split	10-fold CV
Sad		93.33	94.62	96.02	97.83
Unhappy		90.54	91.09	93.44	94.68
Diaper	**Dirty**	92.66	93.53	95.78	96.32
	Need sleep	94.15	95.62	96.87	97.76
	Too cold too hot	89.45	91.87	93.12	94.11
Colic		93.45	94.72	96.23	98.52
Hunger		94.03	95.66	97.24	98.21
Caffeine		94.69	94.92	95.64	96.64
Average accuracy		92.79	94.00	95.54	96.76

From the **Table 2**, we observed that the accuracy of 10-fold CV is better than 60% split, 70% split and 80% split for infant cry detection. The higher accuracy is observed for 10-fold CV is 98.52 (for colic type). The average detection accuracy of the ANN classifier through 60% split, 70% split, 80% split and 10-fold CV are 92.79%, 94.00%, 95.54% and 96.76%, respectively. Finally, an overall average accuracy of the proposed ANN classifier is 94.77% approximately.

5. CONCLUSIONS

The decisions for classifying infant cries are quite difficult. In this study, the performance of ANN classifier is accessed that recognize the automatic infant cry types by applying machine learning algorithms. As we know, automatic infant cry recognition is a powerful tool to understand the needs of the infant and their emotional condition. We have looked at different papers and noticed that different authors have also used different methodologies and provided different accuracy. So we can say that methodology and accuracy both are complementary to each other and if methodology is changed then accuracy will also change and the same conclusion we have obtained from different paper of study. In this proposed work, we observe that an overall accuracy of the recognition with different splits (60%, 70% and 80% split) and 10-fold CV achieved by ANN is 94.77% approximately. Thus, the proposed ANN classifier model facilitate an efficient and reliable option which automatically alert parents when infant babies are being left alone either in vehicles or apartments or offices or home. It has been evaluated effectively on a collected infant speech corpus. Thus, this system can be used by guardians and baby care takers. This system also gives the suggestion so that user can take necessary action.

REFERENCES

1) Alaie Hf, Tadj C. Cry-Based Classification Of Healthy And Sick Infants Using Adapted Boosting Mixture Learning Method For Gaussian Mixture Models. Modelling And Simulation In Engineering. 2012; 983147:1-10. Doi:10.1155/2012/983147

2) Singh Sk, Yadav J. Machine Learning & Image Processing For Hand Written Digits And Alphabets Recognition From Document Image Through Matlab Simulation. Iop Conf. Series: Materials Science And Engineering. 2021;1084(2021)012021:1-8. Doi:10.1088/1757-899x/1084/1/012021

3) Pai C-Y. Automatic Pain Assessment From Infants' Crying Sounds [Thesis]. University Of South Florida; 2016.

4) Nanni L Et Al. A Local Approach Based On A Local Binary Patterns Variant Texture Descriptor For Classifying Pain States. Expert Systems With Applications. 2010;37(12):7888–7894.

5) Brahnam S Et Al. Machine Recognition And Representation Of Neonatal Facial Displays Of Acute Pain. Artificial Intelligence In Medicine. 2006;36(3):211-222. Doi:10.1016/J.Artmed.2004.12.003

6) Yadav J, Rao Ks. Prosodic Mapping Using Neural Networks For Emotion Conversion In Hindi Language. Circuits, Systems, And Signal Processing. 2016;35:139–162.

7) Barajas-Montiel Se, Reyes-García Ca. Fuzzy Support Vector Machines For Automatic Infant Cry Recognition. Intelligent Computing In Signal Processing And Pattern Recognition, Part Of The Lecture Notes In Control And Information Sciences Book Series (Lncis)2006;345:876–881.

8) Zamzami G Et Al. Pain Assessment In Infants: Towards Spotting Pain Expression Based On Infants' Facial Strain. In: Proceedings Of The 11th Ieee International Conference And Workshops On Automatic Face And Gesture Recognition (Fg); 4-8 May 2015; Ljubljana. Slovenia: Ieee; 2015. Doi:10.1109/Fg.2015.7284857

9) Fotiadou E Et Al. Video-Based Facial Discomfort Analysis For Infants. Visual Information Processing And Communication. 2014;9029. Doi:10.1117/12.2037661

10) Pal P Et Al. Emotion Detection From Infant Facial Expressions And Cries. In: Proceedings Of The International Conference On Acoustics Speech And Signal Processing Proceedings; 14-19 May 2006; Toulouse. France: Ieee. Doi:10.1109/Icassp.2006.1660444

11) Abdulaziz Y Et Al. Infant Cry Recognition System: A Comparison Of System Performance Based On Mel Frequency And Linear Prediction Cepstral Coefficients. In: Proceedings Of The International Conference On Information Retrieval & Knowledge Management (Camp); 17-18 March 2010; Shah Alam. Malaysia: Ieee. Doi: 10.1109/Infrkm.2010.5466907

12) Severini M Et Al. Automatic Detection Of Cry Sounds In Neonatal Intensive Care Units By Using Deep Learning And Acoustic Scene Simulation. Advanced Information Sensing And Learning Technologies For Data-Centric Smart Health Applications, Ieee. 2019;7:51982-51982. Doi:10.1109/Access.2019.2911427

13) Abou-Abbas L Et Al. A Fully Automated Approach For Baby Cry Signal Segmentation And Boundary Detection Of Expiratory And Inspiratory Episodes. The Journal Of The Acoustical Society Of America. 2017;142(3):1318-1331. Doi:10.1121/1.5001491

14) Alaie Hf Et Al. Cry-Based Infant Pathology Classification Using Gmms. Speech Communication. 2016;77:28–52.

15) Naithani G Et Al. Automatic Segmentation Of Infant Cry Signals Using Hidden Markov Models. Eurasip Journal On Audio, Speech, And Music Processing. 2018;2018(1):1-14. Doi:10.1186/S13636-018-0124-X

16) Mohammed Ya. A Comparative Study Between Artificial Intelligence Techniques In An Automatic Infant's Pain Cry Identification System, In: Proceedings Of The 1st Annual International Conference On Information And Sciences; 20-21 November 2018; Fallujah. Iraq: Ieee. Doi:10.1109/Aicis.2018.00019

17) Ahmad Smbs Et Al. An Effective Automatic Infant Cry Classification System Based On Continuous Hidden Markov Model Technique Of Fundamental Mel-Frequencies In Acoustic Features. In: Competition/University Tenaga Nasional/College Of Information Technology, Dept. System And Networking; August 2010; Uniten. Malaysia.

18) Abou-Abbas L Et Al. Automatic Detection Of The Expiratory And Inspiratory Phases In Newborn Cry Signals. Biomedical Signal Processing And Control. 2015;19:35-43.

19) Jam Mm, Sadjedi H. A System For Detecting Of Infants With Pain From Normal Infants Based On Multi-Band Spectral Entropy By Infant's Cry Analysis. In: Proceedings Of The Second International

Conference On Computer And Electrical Engineering; 28-30 December 2009; Dubai. United Arab Emirates: Ieee; 2009. Doi:10.1109/Iccee.2009.164

20) Zabidi A Et Al. Binary Particle Swarm Optimization For Selection Of Features In The Recognition Of Infants Cries With Asphyxia. In: Proceedings Of 7th International Colloquium On Signal Processing And Its Applications; 4-6 March 2011; Penang. Malaysia: Ieee; 2011. Doi:10.1109/Cspa.2011.5759886

21) Brahnam S Et Al. Machine Assessment Of Neonatal Facial Expressions Of Acute Pain. Decision Support Systems. 2007; 43(4):1242-1254. Special Issue Clusters.