Smart Nursery for Smart Cities: Infant Sound Classification Based on Novel Features and Support Vector Classifier

Ayyah Abdulhafith Mahmoud
Computer Engineering Department,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
e-mail: 201503144@pmu.edu.sa

Intessar Nasser A Alawadh
Department of Computer Engineering,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
e-mail: 201403226@pmu.edu.sa

Ghazanfar Latif
Department of Computer Science,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
e-mail: glatif@pmu.edu.sa

Jaafar Alghazo
Department of Computer Engineering,
Prince Mohammad bin Fahd University,
Al-Khobar, Saudi Arabia
e-mail: jghazo@pmu.edu.sa

***Abstract*—In the age of smart cities, it is envisioned that most processes within the smart city context will be smart and automated. This includes smart houses, smart kitchens, etc. within this context, a need will arise for Smart nursery rooms. Within the smart nursery concept, the infant needs will need to be fulfilled automatically, in addition, to infant monitoring and safety. The motivation of this work is to design a smart cradle system for a smart nursery room that automates the functions of the cradle based on the infant's sounds. Therefore, in this paper, we propose an infant sound classification technique based on the Support Vector Classifier (SVC) with Radial Basis Function (RBF) kernel using 18 extracted features of infant sounds. The proposed technique has been compared with two SVC kernel function, linear, and poly, as well as other classification algorithms including Decision Tree, Random Forest, and Gaussian Naive Bayes. As a result of comparing the confusion matrix, recall, F1 Score, accuracies, and precision values of various applied machine learning algorithms over-extracted features. SVC using RBF kernel function was found to be the most efficient model with an average accuracy of more than 96%. The proposed system outperforms all other systems proposed in the previous literature.**

Keywords-Baby crying Detection; Smart Nursery; Smart Cities; Support Vector Classifier; Decision Tree; Random Forest

I. INTRODUCTION

In the age of Internet of Things (IoT), Big Data, Cloud Computing, Artificial Intelligence, Robotics and other new technologies, the concept of SMART CITIES is gaining more interest. Smart Cities are cities visualized to be fully automated to serve the residents. Simple tasks such as cooking, taking out the trash, cleaning, and others will all be automated. Part of the smart cities visualization are smart houses in which each room in the house and even the surroundings of the house including safety and irrigation is automated. Family houses that include a nursery room will require the nursery room to be smart as well. The aspects of smart rooms especially a nursery room is very complex and a lot of research has to be done in automating the processes that monitor, serve and ensure the safety of infants. Integrating the new technologies in order to design systems that automate and provide services to humans is an area of research that spans over multiple disciplines. Smart Cities [1], Smart Medicine [2], Smart Education [3], and other disciplines are all jumping on the wagon of IoT, machines learning, AI and new technologies in proposing state of the art systems to better human life.

In this paper, we propose the use of machine learning to design a smart cradle for a smart nursery that will be part of the smart cities. Nowadays, parents go out of their way to check whether their babies are well or not especially during late night hours. For instance, an employed parent might be required to complete some job-related tasks, or he/she might need to do some household Chores while their baby is sleeping. Therefore, having a smart cradle system in such situations is an essential need to provide safety for the infant and peace of mind for the parents. The aim of this research is to design a smart cradle system within a smart nursery for smart cities based on machine algorithms to detect whether the infant is sleeping or crying and take the necessary automated actions.

In the past, in order to detect whether a baby is crying or not, people used baby monitors to detect the baby's sound frequency characteristics. This monitor performs the signal amplifications, timing, shaping and filtering when there is an input; and during the testing the output as a result, will prove if the detected sound is a cry or not. However, with the rapid development of technology, Machine Learning is utilized nowadays to perform the automatic detection and classification of baby sounds including crying.

Machine learning (ML) is when the computer systems study the algorithms and arithmetic models in order to implement a particular assignment efficiently by using inference and patterns instead of being directed by a specific instruction [4]. In ML, there are many ways in which the machine can learn. It could be Supervised, Unsupervised, and Reinforcement Learning. When the machine is

978-1-7281-6788-6/20/$31.00 ©2020 IEEE

supervised, it means it is learning from a supervisor that act as a teacher to provide labeled data to be helpful in training. Unlike Unsupervised learning, the machine is fed with unknown dataset and it identifies the pattern of its data by plotting them into the x-y axis to see in the clusters to identify features. In the Reinforcement learning, the learning is working based on the principle of feedback. For instance, we give the system unlabeled data to identify its characteristics as an output; when the system gives a wrong output, then we give a negative feedback to the machine by providing the accurate output. Therefore, the machine will learn from the feedback that will be provided to it. However, Supervised learning, includes classification and regression as two main categories [5]. Feature based Machine learning as well as deep learning is an area of research where various input data formats (voice, images, waveforms, etc.) are used with the intent of increasing the automatic detection and classification and reduce human intervention [6-8].

The motivation of this research is to eventually design a smart nursery room for smart houses in smart cities. However, due to the complexity of the problem and design, we will start in this paper by designing a smart cradle for the smart nursery room. The aim of this paper is designing a smart automated cradle that can detect and classify infant sound and take the necessary action based on the classification output. In this paper, the focus will be on the classification of baby environmental sounds into 4 different classes.

The rest of the paper is organized as follows: section 2 details the literature review, section 3 shows the details of the dataset used in this research, section 4 explains the proposed method, section 5 shows all the results with comments and section 6 concludes the research. Section 7 lists all the references used in this study.

II. Literature Review

In [11], the authors introduced an auto noise detection system to classify and identity how infants sound when they are crying. The system is based on real recorded data in the Neonatal Intensive Care Unit (NICU) such as vigorous or whimpering crying. Linear Predictive coding (LPC) and Welch's Method are used to draw out insubstantial characteristic to identify the noises (infant's crying). To classify the noises, there are other features such as; the maximum power spectral density, and the standard deviation of the frequency signal. The database in this project is recorded in a hospital including infants' crying and other sounds of machines or knocking. All those noises are uploaded in the NICU. Tampa General Hospital is where the recording was placed via GoPro Hero 3 plus. VLC media player is used to observe the audio. The reported results reached an average accuracy of almost 90% in identifying different types of crying. Least Squares Support Vector Machine (LS-SVM), K-nearest Neighbors (KNN), and Random Forests (RF) were used in this work.

In [10], the authors presented a deduction system to categorize and classify different types of baby crying. This system is programmed to help both families and hospitals to take care of infants' pain, feeling sleepy, or hungry. The dataset of this experiment contains 138 sleep cries, 176 pain cries, and 176 hunger cries. There are fifteen extracted features from the dataset, which are designed to deduct the high frequency. Contrast, peak, spectral centroid, and pitch that are used to prepare the SVM classifier, verifying each audio via the Directed Acyclic Graph-SVM (DAG-SVM) algorithm. The infants' cries are classified using graph support vector machine and got an average accuracy of 92.17%.

In [11], the authors classify the differences between different infants' cries via integration of SVM and Orthogonal Least Square (OLS) using (OLS-SVM) and OLS with Multilayer Perceptron (MLP) classifier (OLS-MLP). The dataset and signals are gathered and collected from infants of ages ranging from a day old to seven months old. Mel Frequency Cepstrum Coefficient (MFCC) analysis were used to embed the information which were extracted in the cry detector. The extracted features were chosen based on its error reduction ratio (ERR) using OLS to distinguish between normal and asphyxiated cries, SVM and MLP are used as classifiers. A high accuracy 94.34% is achieved using the OLS-SVM in comparison to OLS-MLP.

In [12], the authors proposed a system to recognize the status of an infant baby whether he is hungry, sleepy, need to burp, uncomfortable, or in pain. Many classification techniques have been utilized in this approach such as SVM, Hidden Markov model, RF, MLP, Reservoir Network, and Gaussian Mixture model. The main goal of the method is to assist inexperienced parents and trainee pediatricians to diagnose the baby status. The datasets contain 1070 recordings that were taken in Hospital of the University of Damascus.

In order to detect acoustic event, in [13], the authors combined bottleneck features (BN) with acoustic features based on random forest regression classification in order to have an audio event detection system. To evaluate the proposed system, ITC-Irst and UPC-TALP databases were utilized which have a variety of acoustic events. The dataset includes multi audio events in meeting rooms environments such as chair moving, laughing, phone ringing, coughing, keyboard typing and unknown. The benefit of discriminative and low-dimensional bottleneck features was reported with a comparative 5.51% and 5.33% reductions in error rates.

In [14], the authors proposed a system for baby cry prediction and reasons. For the classification, novel Neuro-fuzzy techniques were applied with the Perceptual Linear Prediction for recognizing the baby cries. The results of Neuro-fuzzy classification achieved the most accurate performance compared with other classification algorithms.

In [15], the authors proposed an infant crying detection system based on modern features extraction using machine learning equipment. The research utilized Librosa library to calculate the features extraction parameters such as delta-MFCC, zero-crossing, MFCC and pitch of the signal in python program. The dataset includes crying sounds and non-crying sounds like people speech, baby laughing, whimpering and toy noise. By utilizing Zero-R rule, multilayer perceptron and random tree bagger classifications,

2.5% error rate was shown using hundred trees from random forest classifier.

III. EXPIREMENTAL DATASET

The aim of this paper is to design a smart cradle based on the detection and classification of baby sounds using novel feature set. The performance of the system is compared using five supervised classification algorithms. The dataset used in the paper consists of 432 recorded audio sound [16]. The dataset has four categories of sounds; 108 crying, 108 silence, 108 laughter, and 108 various noises like snoring sounds, dog barking, chirping Sparrow, thunder sound, car horns, and many other sounds of noise. The size of each audio file varies between ninety-five to eight hundred and eighty-two Kilobyte. There are three formats of the recorded audio files which are M4A, MP3 and MP4. Moreover, the duration of each audio file is about five seconds where the sample rate is about forty-four kilohertz.

IV. PROPOSED METHOD

The proposed method consists of three different phases. Phase 1 is the preprocessing of the stereo audio files consisting of infant sounds. Phase 2 consists of extracting novel features from the audio sounds that can be used as input for the classification phase. The final third phase is the classification.

A. Preprocessing the Dataset

As part of the preprocessing step, the dataset has been standardized by performing both scaling and centering. The standardizing has been done to give each variable equal chance to influence the model. This has been done by calculating the mean and standard deviation of each column in the dataset then substitute each value inside the dataset by the following:

$$X = \frac{x - \bar{x}}{sd}$$

This Scaled dataset therefore will have zero mean and one standard deviation. The preceding procedure is essential and has a high impact on the performance of the classifiers later on especially in Support Vector Machine Classifier (SVC), since its kernel functions (RBF, Linear) assume that the dataset is centered around zero.

B. Feature Extraction

In order to extract the different features of 5-second-long audio files. Each audio file has been represented as an audio time series by firstly taking the average of the left and right channel to convert from stereo audio file into a mono audio file. Then, the monophonic signal has been resampled at a sampling rate of 44100Hz [17].

After representing each monophonic audio file into an array of 220500 row and one column, the signal has been divided into multiple overlapping frames of length 2048 after being padded in reflect mode. A total of 431 frames were generated. The root mean square (rms) value of each frame is computed, which represent the average power of a signal, using the following equation:

$$rms_l = \sqrt{\frac{1}{N}\sum_{n=0}^{N-1} |(\mathrm{x}_l[\mathrm{n}])|^2}$$

where N represent the size of the frames.

Moreover, Zero Crossing Rate (ZCR), which can be used to measure the nosiness in a signal, is computed for each frame in the audio signal as well [18]. However, this time the audio signal has been padded by the value of their edges before dividing it into frames of the same size and finding the rate at which the signal change signs from positive to negative and vice versa in each frame.

$$z_{l=} \frac{1}{2N}\sum_{n=0}^{N-1} |sgn_L[x[n]] - sgn_L[x[n-1]]|$$

where: $sgn_L[x[n]]$= 1 when x[n]≥ 0

$sgn_L[x[n]]$= 1 when x[n] ≤0

The following three spectral features, spectral roll-off, spectral centroid and spectral bandwidth of the monophonic audio are computed after finding its magnitude spectrum. In order to find the magnitude spectrum; the sequence representing the monophonic audio is padded by reflect mode then divided into multiple overlapping frames of size 2048. Since audio cannot be represented using a single spectrum, hope-length of 512 and Hanning window is used to generate a sequence of spectrums corresponding to the sequence of frames using short time Fourier transform (STFT) [19]. The STFT of each frame is computed using the following equation:

$$X_l[K] = \sum_{-N/2}^{\frac{N}{2}-1} w[n]x[n + lH]e^{-j\frac{2\pi Kn}{N}}$$

where L represent frame number, W is analysis window
H is hope-length, and N is Frame length

Now a sequence of magnitude spectrums $|X_l[K]|$ can be used to find the corresponding spectral roll-off of each frame, which represent the frequency below which a 90% of the total spectral energy lies.

The sequence of magnitude spectrums $|X_l[K]|$ can also be used to find the corresponding spectral centroid of each frame, which is used to determine the frequency at which the energy of the spectrum is centered or at which the sound is dispersed. The spectral centroid of each frame is computed by weight averaging the corresponding magnitude spectrum by the center frequency of each bin.

$$C_l = \frac{\sum_{k=0}^{N-1} K|X_l[K]|}{\sum_{k=0}^{N-1} |X_l[K]|}$$

The spectral bandwidth which is a sequence of frequency bandwidth of each frame is computed using the magnitude spectrum $|X_l[K]|$,the spectral centroid C_l , and center frequency of each bin. The following equation has is used to calculate the spectral bandwidth of each frame.

$$B_l = \sqrt{\sum_{K=0}^{N-1} |X_l[K]|\ (\mathrm{K}-C_l)^2}$$

Moreover, the Mel-frequency cepstral coefficients (MFCCs) of the monophonic signal is computed as well by mapping the magnitude spectrum of each frame $|X_l[K]|$ to mel_scale_spectrum using filter banks [20]. Then applying log function to convert the magnitude spectrum to Decibel unit (db). Finally, the Discrete Cosine Transform DCT type 2 with orth normalization mode is applied to find 13 mel-cepstrum of each frame. The Following equation is used to calculate (MFCCs) of each frame.

$$\tilde{|X_l[K]|}=\frac{1}{M}\sum_{k=i_s}^{ie}|X_l[K]|*M_i[K]$$
$$\tilde{|X_l[K]|}_{-}db = 10 * log_{10}|X_l[K]|$$
$$13 - mfcc_l = DCT(\tilde{|X_l[K]|}_{-}db)$$

where i is the filter function i = 0,1, 2, …,12.

C. Classification

To classify the dataset into 4 different classes, five different supervised multiclass classifiers are applied including K-Nearest Neighbor, Decision Tree, Support Vector Machines, Random Forest, Gaussian Naïve Bayes [21]. The confusion matrix, recall, f1_score, accuracy, and precision of these classifiers are compared to evaluate their performance and generate the best model. Before applying the classifiers, the dataset has been split into 75% training and 25% testing K-Nearest Neighbor classifier using Brute Force algorithm is applied to find the 5 nearest samples in the dataset using Euclidean distance values and classify based on the label or the class of these 5 neighbors. Moreover, KNN based on Ball-tree Algorithm is applied as well.

Decision Tree Classifier uses the concept of recursively partitioning the dataset into two classes based on certain conditions until a pure leave node is reached, which contains samples that have the same label or belong to the same class. However, the choice of the splitting conditions is done based on the node information gain value, which is a measure of the node impurity. The information gain can be computed using two different criteria Gini and entropy which is used to compare the performance of the Decision Tree Classifier.

The Random Forest Classifier performs classification by taking the majority vote of the different results generated by 100 different decision trees. These decision trees are constructed by different splitting conditions. In the same manner the splitting of each tree into nodes is done based on Gini and entropy.

Support Vector Classifier (SVC) draws a hyperplane to split the dataset into multiple classes using the distance margin or distance between two support vectors. In most situation, getting the optimal hyperplane requires transformation to higher dimensions, therefore, different kernel functions are used. Three kernel function linear, polynomial and radial basis function (RBF) are applied [22].

Another classifier which has been applied is Gaussian Naïve Bayes classifier. This classifier works based on Bayes' theorem with Naive assumption of independency between features. During the training part, the model for each class is found by computing the Gaussian for each feature in each class, then assume the independency of these features by multiplying the Gaussian for the features in the same class. In the testing part, the probability of each feature is computed under each model using the following equation.

$$P(f_i|C_i) = \frac{1}{\sqrt{2\pi\sigma^2{}_{f_i.c_i}}} exp - (\frac{(f_{i-\mu_{f_i.c_i}})^2}{\sigma^2{}_{f_i.c_i}})$$

After that, these probabilities are combined together using Bayes Rules before finally finding to which class the sample belongs.

V. Experimental Results

Table I illustrates the performance of the five selected supervised multiclass classifiers. Overall, it can be seen that SVC specifically RBF kernel Achieves the highest accuracy among all classifiers with an average accuracy of 96.3%. Other parameters of the SVC using RBF kernel achieved the highest results as well with Recall 96.3%, F1 score 96.23%, and a precision of 96.61%. This high performance can be proven by the diagonal values of its confusion matrix table 2, which shows that almost 3 classes out of 4 has been predicted completely right by the generated model. This high performance is a subsequent effect of the scaling and centering which is performed in the preprocessing step. This is mainly because of the fact that RBF assumes its dataset is centered around zero.

On the other hand, SVC with poly kernel functions, Decision Tree, and Gaussian Naive Bayes models is generating an accuracy, recall, f1_score and precision below 90%. However, taking the prediction generated by multiple decision trees using random forest classifier contributes on raising the accuracy by approximately 4% as compared to Decision Trees. Another enhancement which can applied to improve the performance of decision tree can be obtained by using entropy impurity to compute information gain. This improvement is not only shown in the model accuracy, but also on its recall, f1_score, precisions, and obviously its confusion matrix. The Gaussian Naïve Bayes models has the least accuracy and precision with results of 87.96% and 89.24%, respectively. This is expected because Gaussian Naive Bayes assumes the undependability between features, which is not always right such as the case of spectral centroid and spectral bandwidth features.

TABLE I. Comparison of Five Different Classifiers in Terms of Accuracy, Recall, F1 Score and Precisian

Multiclass Classifier		Accuracy	Recall	F1 Score	Precision
SVC	Linear	92.59%	92.59%	92.33%	93. 96%
	RBF	96.30%	96.30%	96.23%	96.61%
	Poly	88.89%	88.89%	88.19%	90.96%
Decision Tree	Gini	87.96%	87.96 %	87.98%	89.63%
	Entropy	89.81%	89.81%	89.57%	90.05%
Random Forest	Gini	92.59%	92.59%	92.55%	92.99%
	Entropy	92.59%	92.59%	92.50%	93.51%
KNN	Brute	91.67%	91.67%	91.24%	93.13%
	Ball Tree	91.67%	91.67%	91.24%	93.13%
Gaussian Naïve Bayes		87.96%	87.96%	88.09%	89.24%

Another observation showing that entropy and gini results in random forest are almost generating the same accuracy, and recall with slight change in precision and consequently f1_score. In the same manner, KNN using two different algorithms, which are brute and ball tree. They generate identical accuracy, recall, f1_score, precision, and confusion matrixes despite the advantage ball tree has over brute in terms of predicting time complexity.

TABLE II. COMPARISON OF FIVE DIFFERENT CLASSIFIERS IN TERMS OF THEIR CONFUSION MATRIXES

Classifier (Kernel)	Confusion Matrix	Classifier (Kernel)	Confusion Matrix
Decision Tree (Gini)	26 0 1 0 0 27 0 0 7 0 19 1 3 0 1 23	Decision Tree (Entropy)	26 0 1 0 0 27 0 0 6 0 19 2 0 0 2 25
Random Forest (Gini)	25 0 2 0 0 27 0 0 6 0 21 0 0 0 0 27	Random Forest (Entropy)	26 0 1 0 0 27 0 0 7 0 20 0 0 0 0 27
KNN (Brute)	27 0 0 0 0 27 0 0 7 0 18 2 0 0 0 27	KNN (Ball-Tree)	27 0 0 0 0 27 0 0 7 0 18 2 0 0 0 27
SVC (Linear)	27 0 0 0 0 27 0 0 7 0 19 1 0 0 0 27	SVC (Poly)	27 0 0 0 0 27 0 0 3 0 23 1 0 0 0 27
SVC (RBF)	26 0 0 1 0 27 0 0 4 0 16 7 0 0 0 27	Gaussian Naive Bayes	24 0 3 0 0 27 0 0 8 0 19 0 2 0 0 25

VI. CONCLUSION

Designing a full smart nursery for smart cities is a complex task and will require research in machine learning, AI, smart systems, etc. In this paper, we propose the design of one component in the smart nursery that of a smart cradle using machine learning and novel features. The smart cradle is intended to help parents ensure the safety and comfort of their infants. The complete smart cradle design is envisioned to contain features of automatically switching the cradle on in certain situations, notify parents in other situation, and other smart functionalities. All these decisions and which action to take depends on the detection and classification of the baby sounds.

For the detection and classification of the baby sounds, we proposed in this paper the use of novel features and through experiments determined that SVC using RBF is the classifier of choice in the system that achieved the most accurate results among other classifiers. 18 extracted features were used. The system is compared with four other supervised classifiers, on the dataset leads to the conclusion that illustrates the SVC with RBF kernel achieved the best accuracies of more than 96% using proposed features.

For future work, we intend to continue the design of the smart nursery with all its components as well as continue the work on the smart cradle to achieve better accuracy with classification of more than the 4 categories of sounds in the dataset used.

REFERENCES

[1] Visvizi, A., & Lytras, M. D. (2020). Sustainable Smart Cities and Smart Villages Research: Rethinking Security, Safety, Well-being, and Happiness.

[2] Latif, G., Shankar, A., Alghazo, J. M., Kalyanasundaram, V., Boopathi, C. S., & Jaffar, M. A. (2019). I-CARES: advancing health diagnosis and medication through IoT. *Wireless Networks*, 1-15.

[3] Uskov, V. L., Howlett, R. J., & Jain, L. C. (Eds.). (2015). *Smart education and smart e-learning* (Vol. 41). Springer.

[4] Banda, J.M., Seneviratne, M., Hernandez-Boussard, T. and Shah, N.H., 2018. Advances in electronic phenotyping: from rule-based definitions to machine learning models. *Annual review of biomedical data science*, *1*, pp.53-68.

[5] Géron, A., 2019. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. *O'Reilly Media*.

[6] Latif, G., Iskandar, D. A., Alghazo, J. M., & Mohammad, N. (2018). Enhanced MR image classification using hybrid statistical and wavelets features. *IEEE Access*, *7*, 9634-9644.

[7] Al-Hmouz, A., Daqrouq, K., Al-Hmouz, R., & Alghazo, J. (2017). Feature Reduction Method for Speaker Identification Systems Using Particle Swarm Optimization. *International Journal of Engineering and Technology*, *9*(3).

[8] Latif, G., Khan, A. H., Butt, M. M., & Butt, O. (2017), IoT based Real-Time Voice Analysis and Smart Monitoring System for Sisabled People. *Asia Pacific Journal of Contemporary Education and Communication Technology (APJCECT)*, Volume 3(2), pp. 227-234.

[9] Pai, C.Y., 2016. Automatic pain assessment from infants' crying sounds, *Graduate Thesis, University of South Florida.*

[10] Chang, C.Y., Chang, C.W., Kathiravan, S., Lin, C. and Chen, S.T., 2017. DAG-SVM based infant cry classification system using sequential forward floating feature selection. *Multidimensional Systems and Signal Processing*, *28*(3), pp.961-976.

[11] Sahak, R., Mansor, W., Khuan, L.Y., Zabidi, A. and Yassin, A.I.M., 2012, January. Detection of asphyxia from infant cry using support vector machine and multilayer perceptron integrated with Orthogonal Least Square. In *Proceedings of 2012 IEEE-EMBS International Conference on Biomedical and Health Informatics* (pp. 906-909).

[12] Ntalampiras, S., 2015. Audio pattern recognition of baby crying sound events. *Journal of the Audio Engineering Society*, *63*(5), pp.358-369.

[13] Xia, X., Togneri, R., Sohel, F. and Huang, D., 2017, July. Random forest regression based acoustic event detection with bottleneck features. In *2017 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 157-162). IEEE.

[14] Srijiranon, K. and Eiamkanitchat, N., 2014, October. Application of neuro-fuzzy approaches to recognition and classification of infant cry. In *TENCON 2014-2014 IEEE Region 10 Conference* (pp. 1-6). IEEE.

[15] Bhattacharya, S., 2016. *Infant cry detection* (Doctoral dissertation, University of Miami).

[16] Piczak, K.J., 2015, October. ESC: Dataset for environmental sound classification. In *Proceedings of the 23rd ACM international conference on Multimedia* (pp. 1015-1018).

[17] McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E. and Nieto, O., 2015, July. librosa: Audio and music signal analysis in python. In *Proceedings of the 14th python in science conference* (Vol. 8).

[18] Bachu, R.G., Kopparthi, S., Adapa, B. and Barkana, B.D., 2010. Voiced/unvoiced decision for speech signals based on zero-crossing rate and energy. In *Advanced Techniques in Computing Sciences and Software Engineering* (pp. 279-282). Springer, Dordrecht.

[19] Liu, H., Li, L. and Ma, J., 2016. Rolling bearing fault diagnosis based on STFT-deep learning and sound signals. *Shock and Vibration*, *2016*.

[20] Palo, H.K., Chandra, M. and Mohanty, M.N., 2018. Recognition of human speech emotion using variants of Mel-Frequency cepstral coefficients. In *Advances in Systems, Control and Automation* (pp. 491-498). Springer, Singapore.

[21] Singh, A. and Lakshmiganthan, R., 2018. Impact of different data types on classifier performance of random forest, naive bayes, and k-nearest neighbors algorithms.

[22] Dogaru, R. and Dogaru, I., 2018, June. Optimized Super Fast Support Vector Classifiers Using Python and Acceleration of RBF Computations. In *2018 International Conference on Communications (COMM)* (pp. 193-196). IEEE.