AUTOMATIC CRYING EVENT DETECTION METHOD IN INDOOR HOME ENVIRONMENT

A THESIS

Submitted by

KARINKI MANIKANTA
(CB.EN.P2CEN17016)

in partial fulfillment for the award of the degree of

MASTER OF TECHNOLOGY
IN
COMPUTATIONAL ENGINEERING AND NETWORKING

Center for Computational Engineering and Networking

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

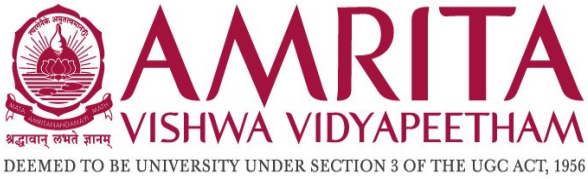
COIMBATORE - 641 112 (INDIA)

June - 2019

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112



BONAFIDE CERTIFICATE

This is to certify that the thesis entitled **" AUTOMATIC CRYING EVENT DETECTION METHOD IN INDOOR HOME ENVIRONMENT"** submitted by **KARINKI MANIKANTA (Register Number-CB.EN.P2CEN17016),** for the award of the **Degree of Master of Technology** in the **"COMPUTATIONAL ENGINEERING AND NETWORKING"** is a bonafide record of the work carried out by him under my guidance and supervision at Amrita School of Engineering, Coimbatore.

Dr. M.S. Manikandan	**Dr. K.P.Soman**	**Dr. K.P.Soman**
Project Guide	Project Co-guide	Professor and Head
Assistant Professor,IIT-BBS	Professor and Head,CEN	CEN

Submitted for the university examination held on ... ...

INTERNAL EXAMINER **EXTERNAL EXAMINER**

AMRITA SCHOOL OF ENGINEERING

AMRITA VISHWA VIDYAPEETHAM

COIMBATORE - 641 112

DECLARATION

I, **KARINKI MANIKANTA (CB.EN.P2CEN17016),** hereby declare that this thesis entitled **"AUTOMATIC CRYING EVENT DETECTION METHOD IN INDOOR HOME ENVIRONMENT"**, is the record of the original work done by me under the guidance of **Dr.M. S. Manikandan** Assistant professor for School of Electrical Sciences, IIT-BBS and **Dr. K.P. Soman**, Professor and Head, Centre for Computational Engineering and Networking, Amrita School of Engineering, Coimbatore. To the best of my knowledge this work has not formed the basis for the award of any degree/diploma/ associateship/fellowship/or a similar award to any candidate in any University.

Place: **Signature of the Student**

Date:

COUNTERSIGNED

Dr. K.P.Soman

Professor and Head

Center for Computational Engineering and Networking

Contents

Acknowledgement

First of all, I thank Almighty for the immeasurable blessings for smooth completion of my project. I am immensely pleased to express my sincere obligation to my project guide **Dr. M. S. Manikandan** Assistant Professor, School of Electrical Sciences, IIT-BBS. It is my genuine pleasure to thank my project co-guide **Dr. K.P. Soman**, Professor, Computational Engineering and Networking, Amrita Vishwa Vidyapeetham, Coimbatore, for her valuable guidance, dedication and encouragement for the successful completion of this project.

I am grateful to our Head, Dr.K.P.Soman and the Project Co-ordinator, Dr. E.A. Gopalakrishnan and the entire staff of CEN for their timely cooperation.

I avail this opportunity to thank my friends for their whole-hearted support during the project. I am greatly indebted to my loving parents and brother for being the motivating forces behind the completion of this dissertation. Also I am grateful for their invaluable help, moral support and encouragement throughout the course of study.

List of Figures

List of Tables

List of Abbreviations

ANN	Artificial Neural Network
BSCD	Baby Cry Sound Detection
BML	Boosting mixture learning
EM	Electromagnetic
FL	Frame Length
FFNN	Feed-forward Neural Network
GMM-UBM	Gassuian mixture model-univeral background model
HCBC	Hand crafted baby cry
LR	Logistic regression
MFCC	Mel-frequency Cepstral Coefficients
MFDWCs	Mel filter-bank discrete wavelet coefficients
ML	Machine Learning
MC-SVM	Multi-class Support Vectors Machines
OLS	Orthogonal least square
PCA	Principal components analysis
RNN	Recurrent Neural Networks
SVM	Support Vector Machine
TDNNS	Time-delay Neural networks
1D-CNN	One-dimensional Convolutional Neural Networks
2D-CNN	Two-dimensional Convolutional Neural Networks

Abstract

Effective automatic baby cry sound detection plays a significant role in many smart baby condition monitoring applications. In this paper, we present deep learning based effective baby cry sound detection (BCSD) method under different kinds of background sounds in indoor environments. We investigated the performance of the three BCSD methods by making use of mel-frequency cepstral coefficients (MFCC) and machine classifiers such as multi-class support vectors machines(MC-SVM), feed-forward neural networks (FFNNs), and one-dimensional convolutional neural networks (1D-CNN). We created the baby crying sounds for both training and testing of three models. In this study, we looked into the results of three methods under different frame lengths including 100 milli seconds, 250 milli seconds and 500 milli seconds. Evaluation results showed that the 1D-CNN with frame length of 500 ms provides promising results as compared to that of the frame lengths 100 milli seconds and 250 milli seconds. For frame length of 500 milli seconds, the 1D- CNN-based, FFNN-based and SVM-based BCSD method had F1-score of 98.86%, 98.46% and 97.97%, respectively for detecting cry sounds. The 1D-CNN based BCSD method had class-wise F1-score of above 98%. Results showed that three BCSD methods have promising results for the same test sound database including air-conditioner, fan, speech and music sounds.

Index Terms: Cry sound recognition, audio classification, sup- port vector machines, feed-forward neural networks, and 1D convolutional neural networks.

Chapter 1

Introduction

In everyones life, cry is the most important sign of life that is observed shortly after a babys live birth [1] - [5]. The cry is a multi- modal and non-static behavior that it contains a lot of information [1]. A detailed acoustic analysis was also carried out of measure and compare the acoustical characteristics of infant cry signals and have demonstrated the diagnostic potential of cry signals for pathological conditions [1]. Furthermore, some of the researchers have studied on the use of cry signals for identifying the structural problems such as cleft lip and invisible defects. In addition with aforementioned cry pathological analysis, baby cry sound recognition enables timely notification and allows parents to remotely monitor when their baby is crying. Thus, an automated recognition of crying sound patterns has become a most popular research topic in developing smart baby monitoring systems and analyzing

various patterns of crying sounds in different contexts such as hunger, sleepiness, pain and so on. Nowadays, increasing penetration of Smartphones and their sensing, computing and communication facilities have led to their use for monitoring different kinds of sound patterns daily routine and remote monitoring assistance [3][25]. Prior

works are often restricted to recognition of patterns of crying sounds for early diagnosis and treatment of newborns. Some of the cry sound detection methods were developed for cloud computing platform where an users device sends the audio to a centralized cloud for per- forming recognition task. The user-cloud computing demands higher bandwidth utilization costs and power consumption due to the continuous transmission of the recorded audio from a place of monitoring. Some of the past works focused on automatic detection of infant cry under different background sound environments or controlled settings. In this paper, we are presenting deep learning based effective baby cry sound detection (BCSD) method under different kinds of background sounds in indoor home environments. The BCSD methods are based on the mel-frequency cepstral coefficients (MFCC) and machine classifiers such as multi-class support vectors machines (MC-SVM), fully connected feed-forward neural networks (FFNNs), and one-dimensional convolutional neural networks (1D-CNN). The remaining sections of this paper is organized as follows. Chapter II Literature Survey. chapter III presents cry sound recognition methods. Chapter IV shows the results of three methods. Finally, conclusions are drawn in Chapter V.

1.1 Motivation

- Automatic baby crying detection can play a major role in both remote baby monitoring and evaluating the awareness of baby under different health condition.
- The development of robust baby crying detection method is still a challenging

task under various mixture background sounds.

- Modern IOT enabled monitoring system highly demands energy efficient method to detect crying events in real time.

1.2 Objectives

- Main focus of thesis is develop automatic crying detection method for smart baby care monitoring application.
- Exploring signal processing techniques machine learning classifier for improving robustness of baby crying detection under different kinds of indoor background noises such as fan-ac, music, speech.
- creating a crying sound database and indoore background database for developing a deep learning based crying detection method.
- Testing and validating the proposed crying detection method with recorded signal.

1.3 Organization

The rest of the thesis organized as follows. chapter 2 presents the literature reviews related to this work. chapter 3 present cry sound detection methods using the MFCC and spectrogram as a features using machine learning and deep learning based classifier. chapter 4 presents the evaluation results on audio database. Finally, conclusion and future directions are given in chapter 5.

Chapter 2

Literature Survey

In this chapter, we described the summary of each technique, various features for audio event recognition, various classifier, real time application and motivation behind this thesis work.

2.1 Conventional Machine Learning Based Methods

In this section, we described the brief overview of AER schemes based on conventional machine learning classifiers such as KNN, GMM, SVM, HMM, ANN from literature. In [1], Y.Kheddache et al., proposed a method to classify the cry samples like new born baby, health, pathological cry sound using PNN classifier. The techniques was evaluated for 2 classes sound using MFCC as feature. The results showed that 88.17% pretern cries and 82% full term infants. In [2], C. Y. Chang, et al., presented crying sound recognition based on MFCC feature and SVM classifier. The method was evaluated for 3 class sounds which recorded by different kinds of crying sounds. The results showed that the accuracy 92.12%. In [3], A. Sharma et al., proposed audio sound clas-

sification using MFCC, Spectrogram, histrogram features and Support Vector Machine and GMM classifier. The results showed that detection rate 97.18%. In [4], H.F. Alaie. et al., presented a omparative study using various classifier (MLP, PNN, SVN, GMM) for crying sound classification. These schemes where evaluated healthy infants and sick infants sound with MFCC, Spectrogram and histogram features. The results showed that overall the 2nd and 3rd adaptation methods mean and variance vector highest accuracy, sensitivity, specificity. In [13], V. Bhagatpati, et al., proposed KNN based method to classify using LFCC, MFCC features. The results showed that th classification accuracy of LFCC more than others. In [5], M.M. Jam, et al., presented the Mel filter bank discrete wavlet coefficients feature based on correct recognition rate scheme with ANN, MLP classifier. The results showed that the high correct recognition rate 93.2%. In [6], R. Torres, et al., proposed machine learning and deep learning technique (SVDD, CNN) to classify the baby and non baby crying sound. The scheme was evaluated for 195 cry sound by using MFCC and HCBC features. The results demonstrated that CNN based audio classifier performance better than the SVDD classifier. In [14], A. Osmani, et al., proposed method for crying sound classification based on KNN, Decision Tree classifier. The method was evaluated for 5 class sound with spectogram features. The results showed that the discomfort class accuracy better than other class. In [15], I.A. Bgnicg et al., proposed i-vector,GMM classifier as a classifier. The method is evaluated for 127 baby cryies of 6 class sound with MFCC feature. The results showed the GMM accuracy 50.6% and i-vector Accuracy 58%. In [16], I.A. Banica, et al., presented a method sound classification using GMM, i-vector classifica-

tion. The method was evaluated crying audio using MFCC feature. The results showed the Accuracy of 70%. In [17], W.S.Limantoro, et al., showed that MFCC as a feature and KNN classifier. The method evaluated on the 138 sound recording (Eairth, Heh, Nen,Owh) proposed features with an classifier. The results showed average accuracy 96.57%. In [7], S. Tejaswini, et al., proposed a method to classify the crying sounds using Support vector machine classifier . The technique was evaluated for 3 class sound using MFCC feature. The results showed that the classification accuracy of hunger and discomfort more than pain and hunger. In [8], Y.Lavner, et al., proposed the logistic regression and CNN based detection classifies to cry,opendoor,talking in home environment sound. The resluts showed that the false positive rates of the CNN classifier are lower than the corresponding rates of the logistic regression. In [18], Abou-Abbas, et al., proposed method to recognition rate crying sound using HMM as classifier. The method was evaluated on cry sound class with MFCC feature. The results showed that the IMF3+IMF4+IMF5 accuracy rate 86.69%. In [19], C.Y.Chang, et al., proposed cry, pain and hunger sound recognition using Zero crossing rate (ZCR), mean, MFCC, LPCC as feature. The results showed recognition accuracy 92%, 85.4%, 83.8% and 77% for infants born with in 2 weeks, 1 month, 2 month, 4 months respectively. In [9] R. Sahak, et al.,presented machine learning based healthy and asphyxia classification. The method was evaluated on MFCC feature. The results showed 93.86%classification accuracy. In [10] O. F. Reyes Galaviz, et al., proposed FFNN based cry detection method . The scheme was evaluated mexican-cuban datasets with MFCC feature. The results showed 50 feature cuban dataset recogition 91.07%. In [11] J.O. Garcia, et al.,

proposed feed forward neural network with using MFCC feature. The results showed that accuracy 97.43%. In [12], M. Petroni, et al., presented the a comparative study using verious classifier FFNN, TDNN, RNN, CC in the cry sounds (angry, pain, fear) sound. The results showed that the FFNN is highest correct calculation 77.9%. In [21], Zhang et al., presented deep learning(2D-CNN) based audio recognition. The method was evaluated on two types of standard datasets(RWCP, NOISEX-92) with spectogram as feature. The results showed the 93% classification accracy.

Table 2.1: Literature review summary of cry sound recognition

Ref./Author	**DataBase**	**Method**		**Performance**
		Features	Classifier	
[1]. Y. Kheddache, et al.,	3250 cry samples newborns, and includes healthy and pathologic cries	MFCC	PNN	88.71%-pretern cries, 82% full term infants
[2].C. Y. Chang, et al.,	176 Hunger cries, 138 sleepness cries, 176 pain cries	MFCC	SVM	Acc.92.17%
[3]. A. Sharma, et al.,,	9933 samples are training, 9665 sample validation set	MFCC, Spectrogram , Histogram	SVM, GMM	Detection rate 97.81% at the false alaram rate 4.17%
[4].H. F. Alaie, et al.	42 Healthy infants, 40 sick infants , 2 heart , 11 neurological , 18 respiratory, 4 blood, 4 others	MFCC.	MLP, PNN, SVM, GMM	Overall the 2nd and 3rd adaptation methods mean and variance vector highest Accuracy, Sensitivity, Specificity
[5].M. M. Jam, et al.,	Baby Chillanto Database.	MFDWC.	MLP,ANN	High correct Reg.rate 93.2%
[6]. R. Torres, et al.,	102 baby cry sound events 93 non baby cry	MFCC, HCBC	CNN,SVDD	CNN classifier is better than HCBC with 1% Improvement in Acc
[7].S. Tejaswini, et al.,	60 Healthy Babies, 826sec hunger, 536sec pain, 668sec discomfort	MFCC	SVM	Hunger and discomfort Acc.93.1%, Pain and Hunger Acc.90.27%, Discomfort and Pain Acc.71.29%
[8].Y. Lavner, et al.,	4096 samples of cry, opendoor,talking in home environment	MFCC	Logistic Regression, CNN	The false positive rates of the CNN classifier are lower than the corresponding rates of the Logistic regression
[9]. R. Sahak, et al.,	316 segments healthy, 284 asphyxia	MFCC	SVM	Classification accuracy 93.86%
[10].O.F. Reyes Galaviz, et al.,	Mexican-Cuban infant cry Database	MFCC	FNN	50 feature C.Babies Reg.91.07%, M.Babie Reg.100%, Mixed Reg. 92%
[11].J. O. Garcia, et al.,	253 samples Normal and 253 samples Pathological cry	MFCC	FNN	Acc.97.43%
[12].M.Petroni, et al.,	230 cry sounds(Angry,pain,fear)	MFCC	FFN, TDNN, RNN, CC	FFN is highest correct classification 77.9%
[13].V. Bhagatpati, et al.,	150 babies training ,40 babies are testing.	LFCC, MFCC.	K-NN	LFCC Avg.Classification Acc.91.02%, MFCC Avg.classcification Acc.85.76%
[14].A. Osmani, et al.,	59 Eh, 19 Eairh, 12 Heh, 131 Neh, 67 Owh, total duration 890.08(s)	Spectrogram	Bagged Tree,Boosted Tree, Subspace K-NN,Decision Tree, SVM	Discomfort class accuracy better than other class.
[15].I. A. Bgnicg, et al.,	127 babies(colic,erucation,discomfort,hunger,pain,tiredness)	MFCC	GMM, i-Vector	GMM-UBM Acc.50.6%, i-Vector Acc.58.0%
[16].I. A. Banica, et al.,	DBL DataBase	MFCC	GMM, i-Vector	Acc. 70%
[17].W. S. Limantoro, et al.,	139 sound recording and divided into four class:Eairh/Eh,Heh,Nen,Owh	MFCC	K-NN	Avg.ACC.96.57%
[18].L. Abou-Abbas, et al.,	200 cry signal, 66.5% of voiced cry sound,35.5% acoustic activaties	MFCC	HMM	IMF3+IMF4+IMF5 Acc.Rate 86.69%
[19].C. Y. Chang, el al.,	437 cries hunger, 431 pain, 392 sleepy	Average,Intensity,pitch, Root mean Square, ,ZCR,Vally,MFCC,LPCC	SVM	Reg. Acc. 92%, 85.4%, 83.8%, and 77% for infants born with in 2 weeks, 1 months,2 months, and 4 months respectively.

Chapter 3

Cry Sound Recognition Schemes

In this chapter, we presents cry sound recognition schemes using the MFCC and spectrogram as audio features with machine learning and deep learning based classifiers. In this study. Our objective is to develop effective and efficient cry sound recognition scheme for automatically recognition four sound classes including fan-ac, music, speech, crying by choosing the suitable frame audio frame size. Since each of the source can produce sounds with different duration which need to be considered in the creation of the sound models for achieving better recognition accuracy with processing time. In some of the sound recognition method like speech processing, selection of optimal speech frame size was well studied based on the glottal periods for processing and analysis of speech signals. But it is very difficult to have a complete study of the duration of each of the sounds that can be produced by different sources in real-life. Thus, we study the performance of the AER schemes under audio frame sizes of 100ms, 250ms, and 500ms. The each technique is discussed in detail in the next sections. The block diagram of the machine learning(ML) and Deep Learning(DL) classifiers based cry sound recognition schemes is shown in fig which consists of four major steps: pre-processing, feature

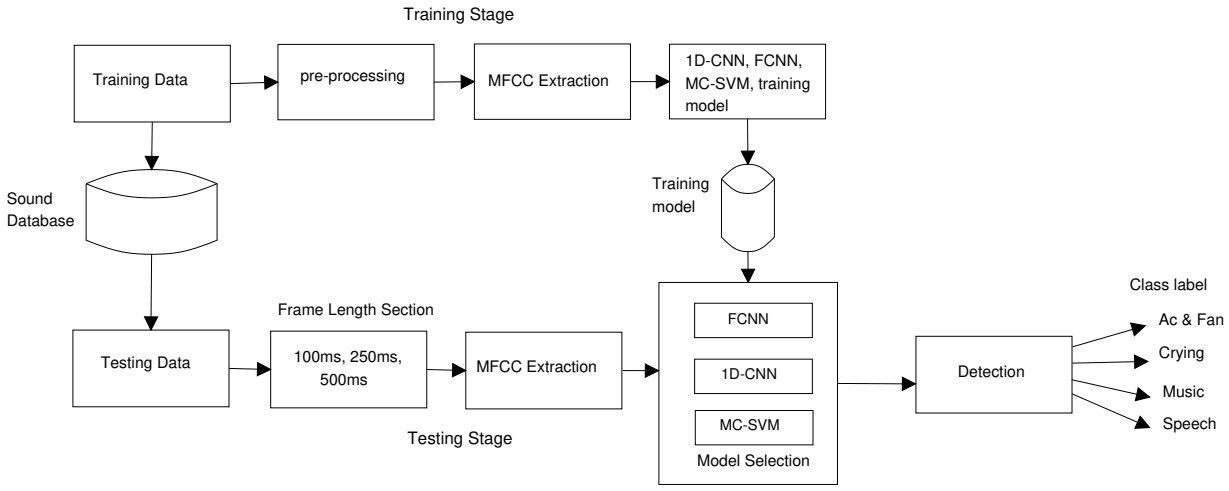


Figure 3.1: Blcok Diagram of the cry sound recognition methods using the machine learning classifier and MFCC feature

extraction, sound models and recognition. Each of the cry sound recognition steps is described in the next section.

3.1 Audio Feature

Audio feature should offer good representation of the audio signal. There are several steps to extract the features from a sound recording.

3.1.1 Preprocessing

In practical the recorded cry sound signals are often corrupted with low-frequency noise components such as microphone artifacts, recording instrument biasing and power-line interface which are generated by the sensors movements and electromagnetic (EM) since the audio sensor is exposed to the environments. Therefore, the audio recorded signal is sent through a high-pass filter with threshold frequency of 60 Hz. Then, the signal is split as 100 ms, 250 ms and 500 ms which are considered for performance

evaluation. The intensity of the cry sound varies due to the stochastic nature of cry sound production and the area of acoustic sensor from a sound source area which can be time fluctuating in the sound observation zone. In practice the sound source location can be unknown. Although the feature is not sensitivity to the amplitude level, the amplitude normalization is done on the zero mean signal formed by audio in order to restrict the microphone sensitivity changes.

3.1.2 MFCC Feature Extraction

In this study, we use the mel-frequency cepstral coefficients (MFCC) features for recognizing the cry sound signals. Each of the frame, we extract MFCC features for the cry, fan and air-conditioner, music, and speech signals. The feature origin is described below as described in [20]:

- **Pre-emphasis filter** It is used to intensify the very high frequencies which equilibrium the spectrum, as in general high frequencies have low amplitudes related to lower frequencies and improves the SNR ratio. Pre-emphasis filter is executed as

$$y[n] = x[n] - \alpha x[n-1], \tag{3.1}$$

  where α is fixed as 0.97.

- After pre-emphasis, window function such as Hamming window is applied to each audio frame the Hamming window function is defined as

$$h(n) = 0.54 - 0.46 \ cos\left(\frac{2\pi n}{N-1}\right) \tag{3.2}$$

where, $0 \leq n \leq N-1$, where N represents the length of audio frame. The window function $h[n]$ is multiplied with the filtered signal $y[n]$.

- **Fourier Spectrum**:The magnitude spectrum is computed by taking the fast Fourier transform (FFT) of the windowed audio frame z[n] that corresponds to different energy distribution over frequencies.

- **Mel-Frequency Spectrum**: The magnitude spectrum is procreated by using a set of 26 triangular band-pass filters which has the places at regular intervals on the Mel-scale, concerned to the linear frequency f by the following equation:

$$M(f) = 1125 \ \ln\left(1 + \frac{f}{700}\right). \tag{3.3}$$

The Mel-frequency is relative to logarithm of linear frequency that manifests the same audio effects in the peoples subjective perception.

- **MFCC**: As, filter bank coefficients are highly corresponding to the discrete cosine transform (DCT). DCT is applied to decor-relate the filter bank coefficients and the coefficients are calculated as

$$C_n = \sum_{k=1}^{k} (\log \ S_k) \left[n \left(k - \frac{1}{2} \right) \frac{\pi}{K} \right] \tag{3.4}$$

where, n = 1, 2....., K and K = 26, number of triangular band-pass filters, Sk energy output of the Kth triangular band-pass filter.

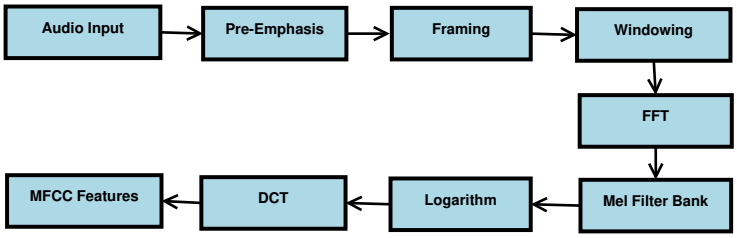


Figure 3.2: The General Block Diagram Of MFCC

Thus above feature extraction step results in 26 coefficients. In this process, we used the lower 12-13 of 26 coefficients for each frame. The parameters which are used for extracting the MFCC are: pre-emphasis as 0.97, number of filters 26, lower band edge as 0 Hz, higher band edge as 8000 Hz, and cepstral coefficients of 13 for every audio frame.

3.1.3 Spectrogram Image Feature (SIF)

A Spectrogram is a visual representation of the intensity in the spectrum of frequencies, of a sound, that varies with time[23]. Spectrogram is generated from an audio signal using Short Time Fourier Transform (STFT) [21]. In forming the spectrogram image, discrete Fourier transform (DFT) is applied to the windowed signal as [22]:

$$X(k,t) = \sum_{n=0}^{N-1} x(n)w(n)exp(-\frac{j2\pi kn}{N}) \tag{3.5}$$

where hanning window function w(n) is defined as [**?**]:

$$w(n) = 0.5 - 0.5\ cos\left(\frac{2\pi n}{N-1}\right) \tag{3.6}$$

$0 \leq n \leq N-1$, where N is the length of the window, x(n) is the time-domain signal, X(k,t) is the k^{th} harmonic corresponding to the frequency $f(k) = kf_s/N$ for the t^{th} frame, f_s is the sampling rate.

The STFT of an acoustic event is computed using Hanning window of size 100ms with 50% overlap and 16000 Hz sampling rate. This gives the spectrum of complex values (X(k, t)). The magnitude of STFT yields linear spectrogram and log of linear spectrogram yields logarithmic spectrogram S(k, t) as [23]:

$$S(k,t) = log(|X(k,t)|) \tag{3.7}$$

where k is frequency bin and t is the time frame. The log operation reduces the dynamic range of spectrogram energies and enhances the spectral components belonging to an acoustic event. The same procedure is followed for the generation of spectrogram image for 250 ms and 500 ms frame length. The spectrogram image is obtained with dimension 128×128 using above procedure. The spectrogram image of audio events are shown in Fig. 3.3.

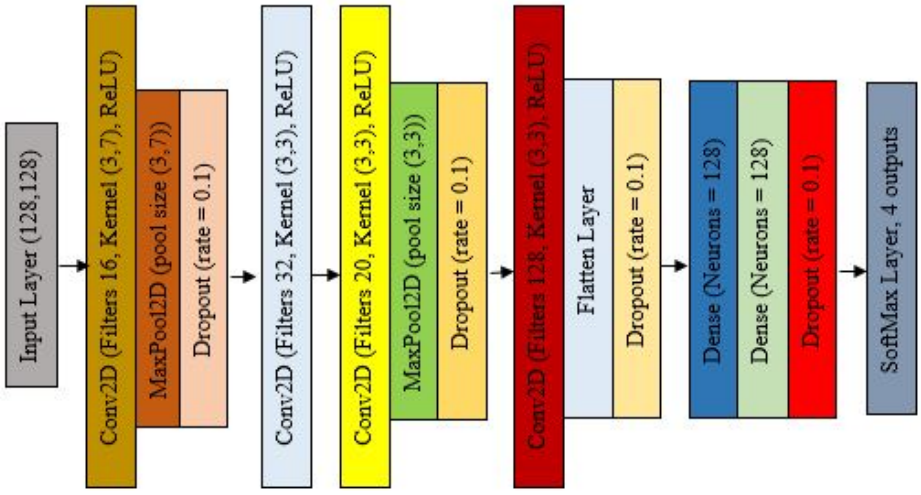


Figure 3.3: 2D CNN based cry sound recognition scheme

Table 3.1: Specification of 2D-CNN for Cry sound Recognition

Parameters	Frame Length		
	100 ms	250 ms	500 ms
Audio Format	wav	wav	wav
Channel	mono	mono	mono
Bit depth	16-bit	16-bit	16-bit
Class	4	4	4
Training Duration	192min	192min	192min
Testing Duration	48 min	48 min	48min
Image Width	128	128	128
Image Heigth	128	128	128
Sampling rate	16000 Hz	16000 Hz	16000 Hz
Feature	Spectrogram	Spectrogram	Spectrogram
Convolution layer	4	4	4
Maxpooling layer	2	2	2
Dropout layer	4	4	4
Activation Function	Relu, Softmax	Relu, Softmax	Relu, Softmax
Batch size	21	21	21
Learning Rate	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam
Epochs	30	30	30

3.2 Machine learning Classifier

In this study we evaluated the performance of the machine learning based classifiers such as multi-class support vector machines(MC-SVM) under different audio frame size 100, 250 and 500 ms for automatically recognizing the 4 sound classes using MFCCs audio feature.

3.2.1 Multi-class Support Vector Machine

For trained information (xi, yi) i = 1, 2, ...N , the optimization of SVM is equated as follows:

$$\arg\min_{w,\xi,b}\{\frac{1}{2}\|\mathrm{w}\|^2 + \mathrm{C}\sum_{\mathrm{i}=1}^{\mathrm{N}}\xi_{\mathrm{i}}\} \tag{3.8}$$

subject to the constraints

$$y_i(w \cdot x_i - b) \geq 1 - \xi_i, \qquad \xi_i \geq 0 \tag{3.9}$$

where ξ_i is non-negative slack variable. It is evaluated using a Lagranzian formulation of the problem, producing the multipliers i and decision function:

$$f(x) = sgn\left(\sum_{i=0}^{N} y_i\alpha_i x \cdot x_i + b\right) \tag{3.10}$$

Where, N represents the no. of training specimens and x denotes a feature vector. Non-linear kernel function K (xi, xj) is used to substitute the dot products x and xi , with the influence of showing the data into a higher-dimensional linearly separable

space. The decision function is defined as:

$$f(x) = sgn\left(\sum_{i=0}^{N-1} y_i \alpha_i K(x, x_i) + b\right) \tag{3.11}$$

In this study, the Gaussian radial kernel used and is defined as $K_R(x, x_i) = exp\left(-\gamma |x - x_i|^2\right)$. The specifications of the MC-SVM based Cry sound recognition scheme are mentioned in Table 3.2.

Table 3.2: Specification of Multi-class SVM Classifier.

Parameters	**Frame Length(FL)**		
	100ms	**250ms**	**500ms**
Audio Format	.wav	.wav	.wav
Channel	mono	mono	mono
Bit depth	16 bit	16 bit	16 bit
Class	4	4	4
Training Duration	384 min	384 min	384 min
Testing Duration	96 min	96 min	96 min
sampling rate	16kHz	16kHz	16kHz
Feature	MFCC	MFCC	MFCC
N0.of samples	1600	4000	8000
NFFT	2048	4096	8192
kernel	RBF	RBF	RBF
penalty parameter C	1	1	1
Gamma	0.0769	0.0769	0.0769
Degree	3	3	3
cache size	200	200	200
Tolerance	0.001	0.001	0.001
No .of Neurons	1600	1600	1600
Batch Size	128	128	128
Learning Rate	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam
Epochs	Infinity	Infinity	Infinity
Training time	42 Hour	5 Hour	1 Hour
Trained Model Size	28.9 MB	10.8 MB	5.0 MB

3.2.2 Feed-Forward Neural Network Classifier

The feed- forward neural network (FFNN) contains three layers. They are input, output and hidden layer. Every layer obtain its input from a preceding layer and then evaluates and transforms data. The translated data send to the immediate layer. Every layer joining has its own weights. The block diagram of feed forward neural network is shown Fig. 3.4. The non-linear transformation is defined as:

$$h^i = f(W^i h^{i-1} + b^i), \quad for \ 0 \leq i \leq L \tag{3.12}$$

where h0 corresponds to the input x. Wi, bi are weight matrices and bias vectors for the ith layer, and the desired output of final layer is hL. The non-linearity f is usually a sigmoid or tanh. Nevertheless, the ReLU activation function is used for faster convergence of models and ReLU is defined as f (x) = max(0, x), which is more simpler than the tanh or sigmoid activation functions as they requires piece-wise linear operator than that of a fixed point sigmoid or tanh. The FFNN based cry sound detection identifications are listed in Table 3.3.

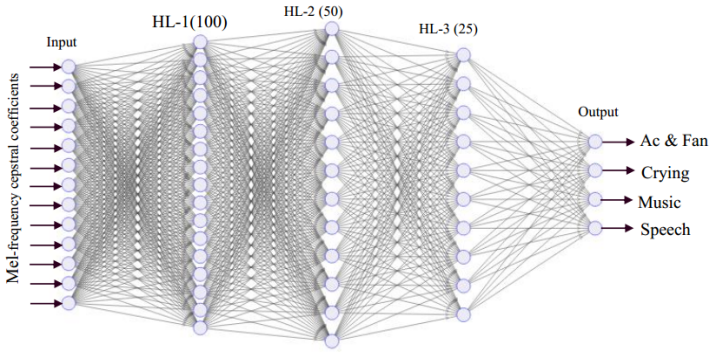


Figure 3.4: Block Diagram of the Fully Connected FFNN cry sound recognition

Table 3.3: Specification of FFNN for Cry Sound Recognition

Parameters	**Frame Length(FL)**		
	100ms	**250ms**	**500ms**
Audio Format	Wav	Wav	Wav
Channel	Mono	Mono	Mono
Bit depth	16 bit PCM	16 bit PCM	16 bit PCM
Class	4	4	4
Training duration	384 min	384 min	384 min
Testing duration	96 min	96 min	96 min
Sampling Rate	16 kHz	16 kHz	16 kHz
Feature	MFCC	MFCC	MFCC
Number of samples	1600	4000	8000
NFFT	2048	4096	8192
Hidden layers	3	3	3
Activation Function	softmax, Relu	softmax, Relu	softmax, Relu
Fold	5	5	5
No .of neurons	1600	1600	1600
Batch Size	128	128	128
Learning rate	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam
Epochs	30	30	30
No .of Parameters	7829	7829	7829
Training time	11.61 Sec	3.79 Sec	1.93 Sec
Trained model size	127.1 kB	127.1 kB	127.1 kB

3.2.3 One-dimensional CNN Classifier for Cry Sound Recognition

Fig. 3.5 illustrates a block diagram of the one-dimensional CNN based cry sound recognition. The specifications of the ID-CNN architecture are summarized in Table 3.4.In this work, we performed study to choose optimal no. of layers and the parameters were varied for achieving the better performance. In this study, the proposed 1D-CNN architecture contains of four convolutions, one drop out, two max-pooling, two fully connected layers and one flatten. The filter shifts are defined as one and two for convo-

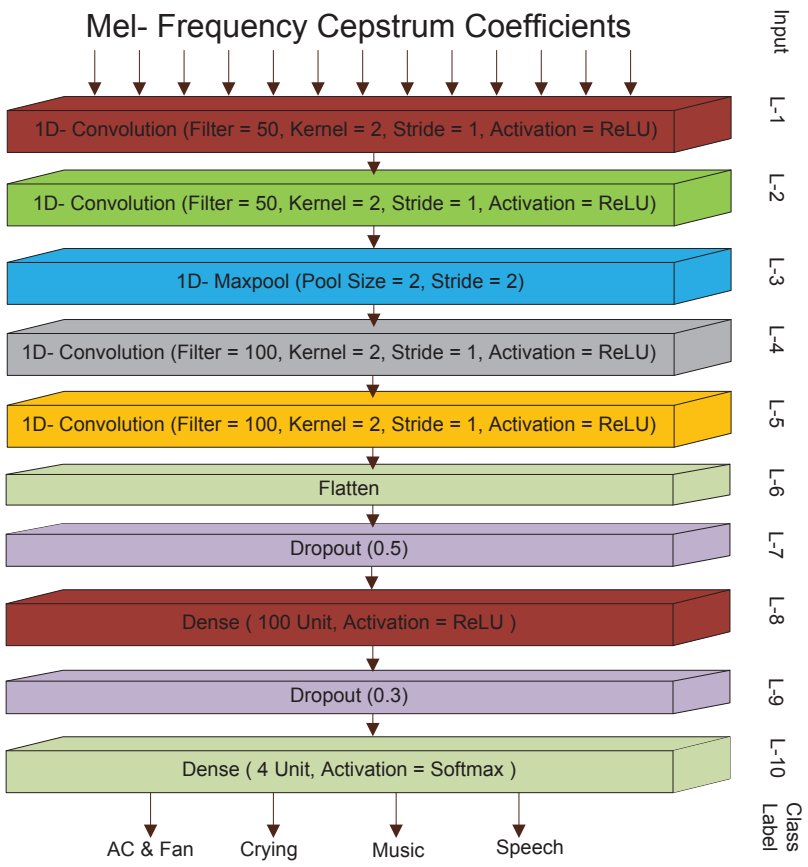


Figure 3.5: One-dimensional CNN architecture for cry sound recognition

lution and max-pooling, respectively. The max- pooling operation minimizes the size of feature maps and also holds the crucial and substantial features of the audio feature. The flatten layer converts Two-Dimensional output into One-Dimensional output. To restrict the model over-fitting, before each dense layer, a dropout layer is used. In the last stage of process, the fully connected layer used to interlink the neurons in the last layers. The stating of the one-dimensional CNN based cry sound detection is listed in Table 3.4.

Table 3.4: Specification of 1D-CNN Based cry sound recognition Scheme For Different Frame Length

layers(Type)	**Frame Length**					
	100ms		**250ms**		**500ms**	
	Shape	**Parameters**	**Shape**	**Parameters**	**Shape**	**Parameters**
Conv1D	(12, 60)	180	(12, 60)	180	(12, 60)	180
Conv1D	(11, 50)	5050	(11, 50)	5050	(11, 50)	5050
Maxpooling	(5, 500)	0	(5, 500)	0	(5, 500)	0
Conv1D	(4, 100)	10100	(4, 100)	10100	(4, 100)	10100
Conv1D	(3, 100)	20100	(3, 100)	20100	(3, 100)	20100
Flatten	300	0	300	0	300	0
Dropout	300	0	300	0	300	0
Dense	100	30100	100	30100	100	30100
Dropout	100	0	100	0	100	0
Dense	4	404	4	404	4	404

Table 3.5: Specifications of 1D-CNN for Cry Sound Recognition

Parameters	**Frame Length(FL)**		
	100ms	**250ms**	**500ms**
Audio Format	.Wav	.Wav	.Wav
Channel	Mono	Mono	Mono
Bit depth	16 bit PCM	16 bit PCM	16 bit PCM
Class	4	4	4
Training duration	384 min	384 min	384 min
Testing duration	96 min	96 min	96 min
Sampling Rate	16 kHz	16 kHz	16 kHz
Feature	MFCC	MFCC	MFCC
Number of samples	1600	4000	8000
NFFT	2048	4096	8192
Convolution layer	4	4	4
Dropout	2	2	2
Max pooling layer	1	1	1
Activation Function	Softmax, Relu	Softmax, Relu	Softmax, Relu
Fold	5	5	5
No .of neurons	1600	1600	1600
Batch Size	128	128	128
learning rate	0.001	0.001	0.001
optimizer	Adam	Adam	Adam
epochs	30	30	30
No .of Parameters	65904	65904	65904
Training time	52 min	23 min	10 min
Trained model size	843.0 kB	843.0 kB	843.0 kB

Chapter 4

Results and discussion

In this chapter, we evaluate the performance of the deep learning classifiers and machine learning classifier based four Cry sound recognition schemes using a wide variety of audio recorded using different kinds of audio recording devices.

4.1 Experimental Setup

The description of experimental setup used in this thesis work is given below:

Hardware : Descktop PC

RAM : 8 GB

OS Type : 64 bit

Processor : Intel Xeon(R) CPU E5-2620 v3 @ 2.4GHz x12

Graphic Card: NVIDIA Quadro K-2200/PCIe/SSE2

Softwares:

Os: Ubuntu 16.04 LTS

Anaconda Python (Python 3.6.5, JupyterNotebook)

Deep Learning Framework: Tensorflow (1.12.0), keras (2.2.0)

Machine leaning Framework: Scikit-learn

Audio Features library: python speech feature

Audio analysis package: LibROSA

Matlab 2015b

Audacity, wavesurfer

4.2 Description of Validation Database

There is no much freely-available cry sound databases in the literature. In the past published works, the cry sound databases were created for performance evaluation but not available for public access. Since the major objective of this paper is to evaluate the performance of the machine learning based baby crying sound detection methods under different indoor background sounds, we created the sound database including cry, fan and air-conditioner, speech, and music that are described in Table 4.1. We used EVISTR digital voice recorder and H1n Handy recorder for recording audio signals under various home conditions. The signals are converted to digital signals at rate of 44.1 kHz sampling and 16-bit resolution. In addition to our audio signals, we also gathered from the public multimedia website (like freesound.org, YouTube, GTZAN library, soundsnap.com and pond5.com).The duration of the audio is about 8 hours in total. For training and test purposes, the audio recorded signal is resampled to 16 kHz.

Table 4.1: Description of Validation Database

Title	Class	Contribution		
		Duration(hrs)	Self	Internet
Ac & Fan	Split ac, Central ac, Ceiling fan, Pedestal fan, Wall fan, Table fan.	2	100%	-
Crying	Crying	2	-	crying web sites
Music	Classical, country, Disco, Hiphop, Jazz, Metal, pop, Raggae, Rock, Tv news, tv programs, Movies.	2	90%	10%(GIZAN)
Speech	Male, Female, children.	2	100%	-

4.3 Performance Evaluation

The performance of cry sound recognition methods are analysis by using benchmark metrics such as Recall (sensitivity), Precision and F1-score and overall accuracy.

4.3.1 Confusion Matrix

A confusion matrix best represents the outputs of predctive analysis algorithms[24]. In this work evaluated the benchmark metrics of multiclass problem the general confusion matrix. It contains some specific terminology which are briefly explained below.

- True Positive(TP): is when the predicted output and the true output are both positive. In the figure the diagonal elements of confusion matrix are TP (i.e., TPF (for Fan-Ac)), (TPC (for the crying class) ete)

- False Positive(FP) is when the predicted output is positive while the true output

is negative. The total number of Fpfor class is sum of values in corresponding column except TP in that column. For example FP for the case of FAN-AC(FA) class is given as:

$$FP_{FA} = ER_{CFA} + ER_{SFA} + ER_{MFA} \tag{4.1}$$

- False Negative(FN) is when the predicted output is negative while the true output is positive. The total number of FN for a class is sum of values in corresponding row except TP in that row. For example FN for the case of FAN-AC(FA) class is given as:

$$FN_{FA} = ER_{FAC} + ER_{FAS} + ER_{FAM} \tag{4.2}$$

- True Negative (TN) is when the predicted output and the true are both negative. The total number of TN for a specific class is sum of all columns and rows excluding that classs column and row. For example TN for the case of FAN-AC (FA) class is given as:

$$TN_{FA} = TotalPrecdiction - (TP_{FA} + FP_{FA} + FN_{FA}) \tag{4.3}$$

- The total number of test sample of any class would be sum of corresponding row(i.e., TP+FN for that class)

The performance of a classifier is evaluated using certain performance parameters as discussed below:

1. Accuracy:It is the ratio between correctly predicted outcomes and sum of all

predictions. It shows, overall how often is the classifier correct.

$$OverallAccuracy(OA) = \frac{AllTPs}{TotalMatrixSum} \tag{4.4}$$

2. Precision: It is defined as out of all the classes, how much model predicted correctly. It should be high as possible.

$$Precision(Pr) = \frac{TP}{TP + FP} \tag{4.5}$$

3. Recall(Sensitivity): It is defined as out of all the positive classes, how much model predicted correctly. It should be high as possible.

$$Recall(Re) = Sensitivity(Se) = \frac{TP}{TP + FN} \tag{4.6}$$

4. F1-score: It is difficult to compare two models with low precision and high recall or vice versa. So to make them compareble, F1-score is used. F1-scorehelps to measure Recall and Precision at the same time. It uses Harmonic mean in place of arithmetic mean by pubishing the extreme values more.

$$F1 - Score(F1) = \frac{2Re * Pr}{Re + Pr} \tag{4.7}$$

4.4 Performance Comparison

In this section, we discussed the performance comparison of each method used in this thesis work. First, we discussed the performance of three classifier for audio event recognition using MFCC as feature.then, we discussed the performance of 2D-CNN classifier using spectroram image as feature for audio event recognition. Finally, we compare the all four AER schemes used in thesis work for audio event recognition.

4.4.1 Cry sound recognition with MFCC

In the first stage of our thesis work we used MFCC feature to represent the audio and three classifier based on machine learning and deep learning such as MC-SVM, FFNN and 1D-CNN to detect audio events. While training these models we observed that FFNN took less time training comparison to MC-SVM and 1D-CNN model. It is also observed that the trained model size of 1D-CNN is very less contrast to MC-SVM and FFNN models because the number of learning parameters in 1D-CNN is less than other models. The method is evaluated on 1 hour 36 min 0f 4 class audio data set against different frame length. For each frame length , the confusion matrix of three AER schemes are summarized in table and their overall performance are summarized in table. The results shows that the FFNN and 1D-CNN based AER scheme had the F1-score values 98.38% and 98.77% for the audio frame size of 500ms whereas MC-SVM based AER scheme had an overall accuracy 97.87%. The 1D-CNN based AER scheme had class wise accuracy is greater than 90%. the computational analysis results show that the prediction time 1D-CNN based scheme is faster than the FFNN based cry sound recognition scheme.

4.4.2 Cry sound recognition with Spectrogram Image Feature

In second stage of our thesis work we used spectogram image feature to represent audio signal and image classifier 2D-CNN to detect the events. The method is evaluated of 96 min of 4 class audio data set against frame length. For each frame length, the confusion matrix of 2D-CNN based AER scheme are summarized in table and the

overall performance are summarized in table. The results shows that 2D-CNN based AER scheme had the f1-score value of 98.84% and overall accuracy 98.98% for audio frame size of 500ms. In this scheme the class-wise accuracy for 500ms frame size is greater than 100ms and 250ms frame size. It is also observed that the recognition rate of speech, crying, fan-ac, music are mor than 95%.

Table 4.2: Confusion Matrix For Three ML based BCSD Methods for Frame Length(FL) Of 100 ms, 250 ms, and 500 ms

	1D-CNN				**FFNN**				**SVM**			
Frame Length (FL) = 100ms												
	A&F	**C**	**M**	**S**	**A&F**	**C**	**M**	**S**	**A&F**	**C**	**M**	**S**
AC& FAN	**144098**	3	27	6	**143914**	10	172	38	**144092**	0	33	9
CRY (C)	1	**139244**	2114	2218	1	**139933**	2089	1554	16	**138353**	2480	2728
MUSIC (M)	44	1602	**137947**	4743	42	1976	**138531**	3797	184	1809	**138197**	4146
SPEECH (S)	12	1489	2298	**140125**	24	2297	3217	**138386**	74	2459	2897	**138494**
Frame Length (FL) = 250ms												
	A&F	**C**	**M**	**S**	**A&F**	**C**	**M**	**S**	**A&F**	**C**	**M**	**S**
AC&FAN	**57803**	0	4	0	**57800**	0	7	0	**57802**	0	5	0
CRY	5	**56117**	695	612	0	**56271**	731	427	8	**55760**	751	910
MUSIC	15	275	**56102**	1248	7	504	**55768**	1361	53	576	**55398**	1613
SPEECH	6	377	824	**56288**	11	697	934	**55853**	22	740	1057	**55676**
Frame Length (FL) = 500ms												
	A&F	**C**	**M**	**S**	**A&F**	**C**	**M**	**S**	**A&F**	**C**	**M**	**S**
AC&FAN	**28769**	0	1	1	**28766**	0	1	4	**28761**	0	6	4
CRY	1	**28564**	171	196	0	**28456**	198	278	8	**28243**	272	409
MUSIC	1	141	**28464**	380	3	220	**28123**	640	20	213	**27992**	761
SPEECH	6	142	354	**27980**	5	189	312	**27976**	13	254	487	**27728**

Table 4.3: Comparison of performance of different classifiers with different frame lengths.

		1D-CNN			FFNN			SVM		
FL	**Class**	**PR**	**RR**	**F1**	**PR**	**RR**	**F1**	**PR**	**RR**	**F1**
100ms	**AC & FAN**	99.96	99.97	99.96	99.95	99.84	99.89	99.81	99.97	99.88
	CRY	97.82	96.98	97.39	97.03	97.46	97.24	97	96.36	96.67
	MUSIC	96.38	95.87	96.37	96.19	95.97	96.07	96.23	95.74	95.98
	SPEECH	95.26	97.36	96.29	96.25	96.15	96.19	95.26	96.22	95.69
	Avg.	97.35	97.54	97.5	97.35	97.35	97.34	97.07	97.07	97.05
250ms	**AC & FAN**	99.95	99.99	99.96	99.96	99.98	99.96	99.85	99.99	99.91
	CRY	98.85	97.71	98.27	97.91	97.98	97.94	97.69	97.09	97.99
	MUSIC	97.35	97.71	97.52	97.08	96.75	96.91	96.83	96.11	96.46
	SPEECH	96.8	97.9	97.34	96.89	97.14	97.01	95.66	96.83	96.24
	Avg.	98.23	98.32	98.27	97.96	97.96	97.95	97.5	97.5	97.65
500ms	**AC & FAN**	99.97	99.99	99.97	99.97	99.98	99.97	99.85	99.96	99.9
	CRY	99.01	98.72	98.86	98.58	98.35	98.46	98.34	97.61	97.97
	MUSIC	98.18	98.19	98.18	98.21	97.02	97.61	97.33	96.57	96.94
	SPEECH	97.97	98.23	98.09	96.8	98.22	97.5	95.93	97.35	96.66
	Avg.	98.78	98.78	98.77	98.39	98.39	98.38	97.86	97.87	97.86

Table 4.4: Confusion Matrix for 2D-CNN based cry sound recognition schemes for FL of 100 ms, 250ms, and 500ms using Spectrogram as a feature

	FL = 100 ms			
	Crying	Ac & Fan	Music	Speech
Crying	14282	0	66	52
Ac & Fan	0	14398	0	2
Music	65	1	14245	89
Speech	530	6	373	13491
	FL =250 ms			
	Crying	Ac & Fan	Music	Speech
Crying	5672	0	7	81
Ac & Fan	0	5760	0	0
Music	23	0	5655	82
Speech	34	4	39	5683
	FL = 500 ms			
	Crying	Ac & Fan	Music	Speech
Crying	2856	0	8	16
Ac & Fan	0	2880	0	0
Music	13	0	2848	19
Speech	39	1	36	2804

Table 4.5: Performance of 2D-CNN scheme using spectrogram feature

Frame Length	Class	PR	Re	F1
100 ms	Crying	95.99	99.19	97.55
	Ac & Fan	99.95	99.98	99.96
	Music	97.01	98.92	97.96
	Speech	98.95	93.68	96.24
	avg/total	97.97	97.94	97.92
250 ms	Crying	99	98.47	98.73
	Ac & Fan	99.93	100	99.96
	Music	99.19	98.17	98.67
	Speech	97.21	98.66	97.92
	avg/total	98.83	98.82	98.82
500 ms	Crying	98.21	99.16	98.68
	Ac & Fan	99.96	100	99.97
	Music	98.47	98.88	98.67
	Speech	98.76	97.36	98.05
	avg/total	98.85	98.85	98.84

Table 4.6: The overall accuracy of all the classifier against different frame length

FL	SVM	FFNN	1D-CNN
100 ms	97.07	97.35	97.47
250 ms	97.51	97.96	98.23
500 ms	97.87	98.39	98.78

Table 4.7: The overall accuracy of 2D-CNN in different frame length

FL	2D-CNN
100 ms	97.94
250 ms	98.83
500 ms	98.85

Chapter 5

Conclusion

In this thesis, we presented for cry sound recognition scheme using the MFCC and Spectogram as audio features with machine learning and deep learning based classifiers such as multi-class SVM, FFNN, 1D-CNN, and 2D-CNN for recognizing 4 sound classes. We created audio database including speech(mail, female, children), crying, music(tv shows, laptap sounds, movies, news), fan-ac(table fan, round fan, split ac,central ac, ceiling fan, wall fan) sounds by considering different recording devices such as commercials hand-held audio recorders, smart phones, laptop PC and different recording home environments. In the first stage of our thesis work, we used MFCC as feature and MC-SVM, FFNN, 1D-CNN as a classifiers to classify the four class audio events. The results showed that 1D-CNN outperforms other schemes for audio frame length 500ms. In the final stage of our thesis work, we used spectrogram image as feature and 2D-CNN as a classifier to classify the four class audio events. The results showed that recognition rate for 500 ms frame length based 2D-CNN classifier is more than 100 ms and 250ms. We also observed that MFCC feature with 1d-CNN classifier gives more recognition accuracy than spectrogram feature based on 2D-CNN classifier. The recog-

nition of speech and fan-ac sounds with more than 95% accuracy are the state-of-art of our thesis work. For the future work, collection of more audio database for testing of our methods. There is also possibility improving the performance of models by applying post processing method and reduce the computational complexity in predition time by studying the hyper-parameter and algorithms of classifier. The frame wise multi-label classification is the future research in this area.

References

1. Y. Kheddache, C. Tadj, Identification of diseases in newborns using advanced acoustic features of cry signals, Biomedical Signal Processing and Control. 50, pp. 35-44, Jan. 2019.

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

3. A. Sharma, S. Kaul, Two-stage supervised learning-based method to detect screams and cries in urban environments, IEEE/ACM Trans. on Audio, Speech, and Language Processing, Vol. 24, pp.290-299, Feb. 2016.

4. H. F. Alaie, L. Abou-Abbas, C. Tadj, Cry-based infant pathology classification using GMMs, Speech communication.77, pp. 28-52. Dec. 2015.

5. M. M. Jam, H. Sadjedi, Wavelet-based automatic cry recognition system for detecting infants with hearing-loss from normal infants, The Journal. of Engineering, Vol. 11, pp. 63-64, Sep. 2013.

6. R. Torres, D. Battaglino, L. Lepauloux, Baby cry sound detection: A comparison

of hand crafted features and deep learning approach, in Proc. Int. Conf. on Engineering Applications of Neural Networks, pp. 168-179. Springer, Cham, Aug. 2017

7. S. Tejaswini, N.Sriraam, G. C. M.Pradeep, Recognition of infant cries using wavelet derived mel frequency feature with SVM classification, in Proc. Int Conf IEEE Circuits, Controls, Communications and Computing, pp. 1-4, Oct. 2016.

8. Y. Lavner, R. Cohen, D.Ruinskiy, H. IJzerman, Baby cry detection in domestic environment using deep learning, in Proc Int. Conf. IEEE the Science of Electrical Engineering,pp. 1-5, Nov. 2016.

9. R. Sahak, W. Mansor, L. Y. Khuan, A. Zabidi, A. I. M. Yassin, "Detection of asphyxia from infant cry using support vector machine and multilayer perceptron integrated with Orthogonal Least Square," In Proc. IEEE-EMBS Int. Conf. Biomedical and Health Informatics pp. 906-909, Jan. 2012.

10. O. F. Reyes-Galaviz, S. D. Cano-Ortiz, C. A. Reyes-Garca, Evolutionary-neural system to classify infant cry units for pathologies identification in recently born babies, in Proc. 7th Int. Conf. IEEE Artificial Intelligence, pp. 330-335, Oct. 2008.

11. J. O. Garcia, C. A. R. Garca, Acoustic features analysis for recognition of normal and hypoacustic infant cry based on neural networks, In Int. Work-Conf. on Artificial Neural Networks, pp. 615-622.Springer, Berlin, Heidelberg. June. 2003.

12. M. Petroni, A. S. Malowany, C. C. Johnston, B. J. Stevens, Classification of infant cry vocalizations using artificial neural networks, in Proc. Int. Conf. IEEE Acoustics, Speech, and Signal Processing, Vol. 5, pp. 3475-3478, May. 1995.

13. V. Bhagatpatil, V. M. Sardar, An automatic infants cry detection using linear frequency cepstrum coefficients, Int. Journal of Scientific and Engineering Research, Vol. 3, pp.1379-1383, 2014.

14. A. Osmani, M. Hamidi, A.Chibani, Machine Learning Approach for Infant Cry Interpretation, in Proc. 29th Int. Conf. IEEE Tools with Artificial Intelligence, pp. 182-186, Nov. 2017.

15. I. A. Bgnicg, H.Cucu, A. Buzo, D. Burileanu, C. Burileanu, Baby cry recognition in real-world conditions, in Proc. IEEE 39th Int. Conf. on Telecommunications and Signal Processing, pp. 315-318, June. 2016.

16. I. A. Banica, H. Cucu, A. Buzo, D. Burileanu, C. Burileanu, " Automatic methods for infant cry classification," In Proc. Int. Conf. on IEEE Communications, pp. 51-54, June. 2016.

17. W. S. Limantoro, C. Fatichah, U. L. Yuhana, Application development for recognizing type of infant's cry sound, in Proc. Int. Conf. IEEE Information and Communication Technology and Systems, pp. 157-161, Oct. 2016.

18. L. Abou-Abbas, L. Montazeri, C. Gargour, C. Tadj, On the use of EMD for automatic newborn cry segmentation, in Proc. Inte. Conf. IEEE Advances in

Biomedical Engineering, pp. 262-265, Sep. 2015.

19. C. Y. Chang, Y. C. Hsiao, S. T. Chen, " Application of incremental SVM learning for infant cries recognition," in Proc. 18th Int. Conf. on IEEE Network-Based Information Systems. pp. 607-610. Sep. 2015.

20. S. Soni, S. Dey and M. S. Manikandan, "Automatic Audio Event Recognition Schemes for Context-Aware Audio Computing Devices," 2019 7th Int. Conf. Digital Inform. Process. and Comm. (ICDIPC), Turkey, 2019, pp. 23-28.

21. H. Zhang, I. McLoughlin and Y. Song, "Robust sound event recognition using convolutional neural networks," in *Proc.IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)* pp. 559-563, 2015.

22. Sharan, Roneel V., and Tom J. Moir, "Acoustic event recognition using cochleagram image and convolutional neural networks," *Applied Acoustics* vol. 148, pp. 62-66., 2019.

23. Z. Ren et al., "Deep Scalogram Representations for Acoustic Scene Classification," *IEEE/CAA Journal of Automatica Sinica,* vol. 5, no. 3, pp. 662-669, 2018.

24. Chollet, Francois, "Deep learning with python," *Manning Publications Co.,* 2017.

25. M. S. Manikandan, K. P. Soman, " A novel method for detecting R-peaks in electrocardiogram (ECG) signal", Biomedical Signal Processing and Control. 2012 Mar 1;7(2):118-28.

Publication based on this research work

1. Karinki Manikanta, K.P. Soman , M. Sabarimalai Manikandan, *"Deep Learning Based Effective Baby Crying Recognition Method Under Indoor Background Sound Environments"* , Seventh International Conference on Innovations in Computer Science and Engineering(ICICSE - 2019), Springer.(submitted)