INFANTILE VOCAL DECODER

A PROJECT REPORT

submitted By

NANDU R NAIR

TVE18MCA035

to

the APJ Abdul Kalam Technological University

in partial fullfilment of the requirements for the award of the degree

of

Master of Computer Applications

Department of Computer Applications

College of Engineering

Trivandrum-695016

JULY 2021

Declaration

I undersigned hereby declare that the project report titled **"INFANTILE VOCAL DECODER"** submitted for partial fulfillment of the requirements for the award of degree of Master of Computer Applications of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by me under supervision of Smt. Pooja J P, Asst.Professor. This submission represents my ideas in my words and where ideas or words of others have been included. I have adequately and accurately cited and referenced the original sources. I also declare that I have adhered to ethics of academic honesty and integrity as directed in the ethics policy of the college and have not misrepresented or fabricated any data or idea or fact or source in my submission. I understand that any violation of the above will be a cause for disciplinary action by the Institute and/or University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title.

Place : Trivandrum **Nandu R Nair**

Date : 30/06/2021

DEPARTMENT OF COMPUTER APPLICATIONS

COLLEGE OF ENGINEERING
TRIVANDRUM

CERTIFICATE

This is to certify that the report entitled **INFANTILE VOCAL DECODER** submitted by **Nandu R Nair** to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the Degree of Master of Computer Applications is a bonafide record of the project work carried out by him under my guidance and supervision. This report in any form has not been submitted to any University or Institute for any purpose.

Internal Supervisor External Supervisor

Head of the Dept

Acknowledgement

First and for most I thank **GOD** almighty and to my parents for the success of this project. I owe a sincere gratitude and heart full thanks to everyone who shared their precious time and knowledge for the successful completion of my project.

I am extremely thankful to **Dr Jiji C V**, Principal, College of Engineering Trivandrum for providing me with the best facilities and atmosphere which was necessary for the successful completion of this project.

I am extremely grateful to **Dr. Sabitha S**, HOD, Dept of Computer Applications, for providing me with best facilities and atmosphere for the creative work guidance and encouragement.

I express our sincere thanks to **Smt. Pooja J P**, Asst. Professor, Department of Computer Applications, College of Engineering Trivandrum for her valuable guidance, support and advice that aided in the successful completion of my project.

I profusely thank other Asst. Professors in the department and all other staffs of CET, for their guidance and inspirations throughout my course of study.

I owe my thanks to my friends and all others who have directly or indirectly helped me in the successful completion of this project. No words can express my humble gratitude to my beloved parents and relatives who have been guiding me in all walks of my journey.

Nandu R Nair

Abstract

Cry is a form of communication for children to express their feeling. Baby's cry can be characterized according to its natural periodic tone and the change of voice. It has a base frequency (pitch) in range 250Hz to 600Hz. Through their baby's cries detection, parents can monitor their baby remotely only in important condition. This study of sound recognition has two main processes, the first process is feature extraction and the second process is classification or determining the sound pattern. In the Mel Frequency Cepstral Coefficient method, the analysis of changes in pre-emphasis, numbers of filter bank and numbers of cepstral are conducted. The selection of the filter bank value which applied must be greater than the cepstral value which applied.Cepstral values is adjusted to get the better accuracy. The highest percentage of accuracy is 90 percent when this system uses 8 as the cepstral value and 3 as the nearest neighbor value, and all rules are considered the best value based on the test results. The use of Mel Frequency Cepstral Coefficient as feature extraction method and Random Forest classification can be implemented to detect the baby is crying or not so that it can be applied as a solution for parents to monitor their children remotely only in certain condition.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The development of children up to five year old's is the basis of subsequent developments where in this phase the development of language, emotional, intelligence, and creativity occurs quickly, so they need to be monitored by theirs parents. Violence against children occurs at the age of 0-18 years where the highest rate occurs in children aged less than 2 years, on the data of 55 percent perpetrators of violence against children are family members and caregivers.One of the many technologies currently able to handle the problem of supervising room activities is Closed Circuit Television (CCTV) but a surveillance system that refers to children still need to be developed.

Crying is a form of communication for children to express their feeling.Studies have shown that newborns are crying differently depending on their need: hunger, tiredness, discomfort, eructation, pain, and so on.This system able to distinguish between different types of newborn cries and consequently estimate the baby's need by using the sounds, which includes :

- "Neh" - Baby feels hungry.
- "Eairh/Eargh" - Baby cries due to some gastric problems.
- "Owh/Oah" - Baby feels Sleepy or tired.
- "Eh" - Baby is unable to burp.
- "Heh" - Baby feels discomfort.

Baby's cry can be characterized according to its natural periodic tone and the change of voice. It has a base frequency (pitch) in range 250Hz to 600Hz. This study of sound recognition has two main processes, the first process is feature extraction and the second process is classification

or determining the sound pattern.Mel Frequency Cepstral Coefficient using Mel frequency filter bank. Based on this background, the use of Mel Frequency Cepstral Coefficient as feature extraction method must be set correctly so all numbers in pre-emphasis coefficient, numbers of filter bank and numbers of cepstral need to be analyzed. As well as the use of Random Forest to classify the baby is crying or not so that it can be applied as a solution for parents to monitor their children remotely only in certain condition. I analyzed the different cry signals by using time frequency analysis. A Hamming window with length 256 was used, the overlap was 128 and a 512 point fast Fourier transform (FFT) was used for calculating the STFT It is obvious that different catalogs of cry signals have different waveform and spectrum characteristics.

Chapter 2

Problem Defnition and Motivation

Since children in distress will almost always cry, detecting their screams could trigger an alert. The problem is that to be effective the application needs to be precise. Now only pediatrician's can understand why babies are crying. Parents or caretakers may not be aware of babies needs according to their cries. One of the many technologies currently able to handle the problem of supervising room activities is Closed Circuit Television (CCTV) but a surveillance system that refers to children still need to be developed.

A baby's cry is an important signal that they need care. The aim is to develop a system that detects the actual reasons behind babies cries. The key Features are:

- Baby's cry can be characterized according to its natural periodic tone and the change of voice based on the frequency (pitch) range in between 250Hz to 600Hz
- Study of sound recognition has two main processes, the first process is feature extraction and the second process is classification or determining the sound pattern

The development of children up to two year olds is the basis of subsequent developments where in this phase the development of language, emotional, intelligence, and creativity occurs quickly, so they need to be monitored by theirs parents.Only experienced parents, caregivers, doctors, and nurses understand the cries based on their experiences.Young parents have trouble calming down their babies because all cry signals sound the same to them.

Chapter 3

Literature Review

Cry is a form of communication for children to express their feelings.Through their baby's cries detection, parents can monitor their baby.This system able to distinguish between different types of newborn cries and consequently estimate the baby's need by using the sounds.Here, I take some of the papers related to Infantile Vocal Decoder using various methods,

- Lichuan Liu Audio features of infant cry signals were obtained in time and frequency domains. Features extracted from audio feature space include : linear predictive coding (LPC), linear predictive cepstral coefficients (LPCC), Bark frequency cepstral coefficients (BFCC), Mel frequency cepstral coefficients (MFCC).

- Dewi, S. P, Prasasti, A. L Irawan, B(An Automatic Infant's Cry Detection Using Linear Frequency Cepstral Coefficients (LFCC), 2014) The average recognition rate performance analysis using LFCC in infant's cry detection about 91.58 percent and then when using MFCC the average recognition rate performance about 78.8 percent. in the other scenario, the average performance of classification accuracy using FFNN classifier about 91.02 percent when using LFCC and 85.76 percent when using MFCC. In this scenario, if the number of training samples is high in the end performance of classification is small

3.1 Using Mel-frequency cepstral coefficients

The mel frequency cepstrum is described by mel frequency cepstral coefficients (MFCC). Mel frequency cepstrum is a representation of a sound's short-time power spectrum based on a linear cosine transform of a log spectrum on a non-linear mel scale of frequency in sound

processing.MFCC employs the Mel-scale bank filter, which is a triangle logarithmic band-pass filter. It results in a higher frequency filter with a larger bandwidth. After the outputs of the previous step are translated back into the time domain so that the signal can be displayed well, the MFCC process ends with Discrete Cosine Transformation (DCT).These findings are organised into a row of an audio vector called Mel Frequency Cepstral Coefficient.

3.2 Using Random Forest

A Random Forest is an ensemble technique that uses several decision trees and a technique called Bootstrap and Aggregation, sometimes known as bagging, to solve both regression and classification problems. Rather of depending on individual decision trees, the main idea is to aggregate numerous decision trees to determine the final outcome.As a fundamental learning model, Random Forest uses several decision trees. Row and feature sampling are done at random from the dataset, resulting in sample datasets for each model. Bootstrap is the name of this section. Experiments have shown that they are powerful on infant cry classification.

Chapter 4

Requirement Analysis

4.1 Purpose

There are numerous studies currently being conducted on the detection of newborn crying for a variety of objectives. The correct interpretation of a crying baby is important for the medical goal of ensuring that the caregiver knows how to properly treat the baby.Because cry signals contain information about a baby's health and can be interpreted to some extent by experienced parents and professionals, recognising and analysing an infant's cry is not only conceivable, but also has significant medical and societal implications.I gather and analyse audio aspects of infant cry signals in the temporal and frequency domains in this project. For cry language identification, and I can classify provided cry signals into specific cry meanings based on the relevant properties.Machine Learning is used to create the "Infantile vocal decoder." The technique assists parents in determining the cause of their child's weeping.

4.2 Overall Description

The cry of a newborn is a crucial indicator that they require attention. The goal is to create a system that can recognise the true causes of a baby's crying.By assessing the properties of audible cry components detected in a realistic clinical environment, a newborn cry-based diagnostic system intends to achieve preliminary screening of infant diseases.The feature extraction procedure and the classification or determining the sound pattern are the two key operations in the infantile vocal decoder. Mel-spectrogram as a feature extraction method and Random Forest as a classification method can be used to determine whether or not the baby is sobbing.As a

result, it can be used as a solution for parents to remotely watch their children only in limited circumstances.

4.2.1 Product Functions

- Feature extraction using MFCC
- Classification or determining the sound pattern
- Training and testing of model
- Develope a user interface
- connect the UI with the model using Flask.

4.2.2 Hardware Requirements

- Processor : Intel Core i3
- Storage : 512 GB Hard Disk space
- Memory : 4 GB RAM

4.2.3 Software Requirements

- Operating System : Linux/Windows
- Platform : Python
- Librarie used : pandas, matplotlib, numpy, sklearn,Librosa

4.3 Functional Requirements

The functional requirements includes all the activities or processes that should be achieved by the proposed system. It includes

- **Librosa:** Librosa is a robust Python package for working with and analysing audio. It's the first step toward working with audio data at scale for a variety of applications, from identifying a person's voice to extracting personal features from audio.It provides the necessary

building elements for the development of music information retrieval systems. Librosa aids in the visualisation of audio signals as well as feature extractions utilising various signal processing techniques.

- **sklearn:** sk learn (formerly sci-kit learn and sometimes called sk learn) is a machine learning library can be used in python programming language. By using this library, we can implement various regression, classification and clustering algorithms such as random forest, support vector machine, k-means and DBSCAN. And the sk learn library is built in a way that it can work with various scientific and numeric libraries of python such as scipy and numpy.
- **NumPy:** NumPy is a Python package for array processing. It includes a high-performance multidimensional array object as well as utilities for manipulating them. It is the most important Python package for scientific computing.NumPy can be used as a multi-dimensional container of generic data in addition to its apparent scientific applications. Numpy allows arbitrary data types to be created, allowing NumPy to connect with a wide range of databases cleanly and quickly.

4.4 Non Functional Requirements

4.4.1 Performance Requirements

- Accuracy : Accuracy in functioning and the nature of user-friendly should be maintained by the system.
- Speed : The system must be capable of offering speed.
- Low cost: This system is very cheap to implement and is also user-friendly.
- Less Time consuming: It uses very less time comparing to the existing sysytem .
- User Friendly: This proposed system is highly user friendly they enables to create a good environment.

4.4.2 Quality Requirements

- Scalability : The software will meet all of the functional requirements.

- Maintainability : The system should be maintainable. It should keep backups to atone for system failures, and should log its activities periodically.
- Reliability : The acceptable threshold for down-time should be large as possible. i.e. mean time between failures should be large as possible. And if the system is broken, time required to get the system backup again should be minimum.
- Availability: This system is easily available as the core equiments in building the sofware is easily obtained.
- High- Functionality: This system is highly functional in all environment since, They are highly adaptable.

Chapter 5

Design And Implementation

The proposed system is used to classify baby cry by using a pre-trained model. The model is trained using Random Forest upon the features extracted by MFCC..

5.1 Overall Design

The suggested system is built on a client-server model. That is to say, the Infantile vocal decoder has both a client and a server component. The user inputs the voice signal into the client portion, which must be assessed. The client receives the evaluated result after passing the input to the server. The client side is constructed with HTML and Python, while the server side is written in Python Flask.

5.1.1 System Design

The system is web based. The input is taken from the user through a web page and the input is passed to the python program running in the server side. The server program perform tasks such as pre processing and feature extraction on the input data. The results of these processes are used to evaluate the input using the pre trained model.

The model is created using The data sourced from a corpus entitled the "Donate-a-Cry Corpus".This corpus, a compilation of audio recordings of babies crying, appears to have originally been collected as part of a campaign entitled the "Donate-a-Cry Campaign" in Sweden in 2015. More information about this campaign cannot be ascertained as the listed campaign website in the github repository has been converted into a website for Japanese escorts. Further work on

this dataset, including removing background noise, standardizing sample rates, and eliminating outliers, was completed by researchers at the Royal Institute of Technology of Sweden in 2018.

5.1.2 Methodology

There are two parts in this project. The first part is the creation of the model and the second one is the creation of user program which will work with the pre-trained model.

The main process of the Infantile vocal decoder is the creation of the trained model. The major steps in the model creation are Model Building, Loading Dataset, Feature extraction, Data split,Classification using Random Forest. The major steps in the model creation are mentioned below.

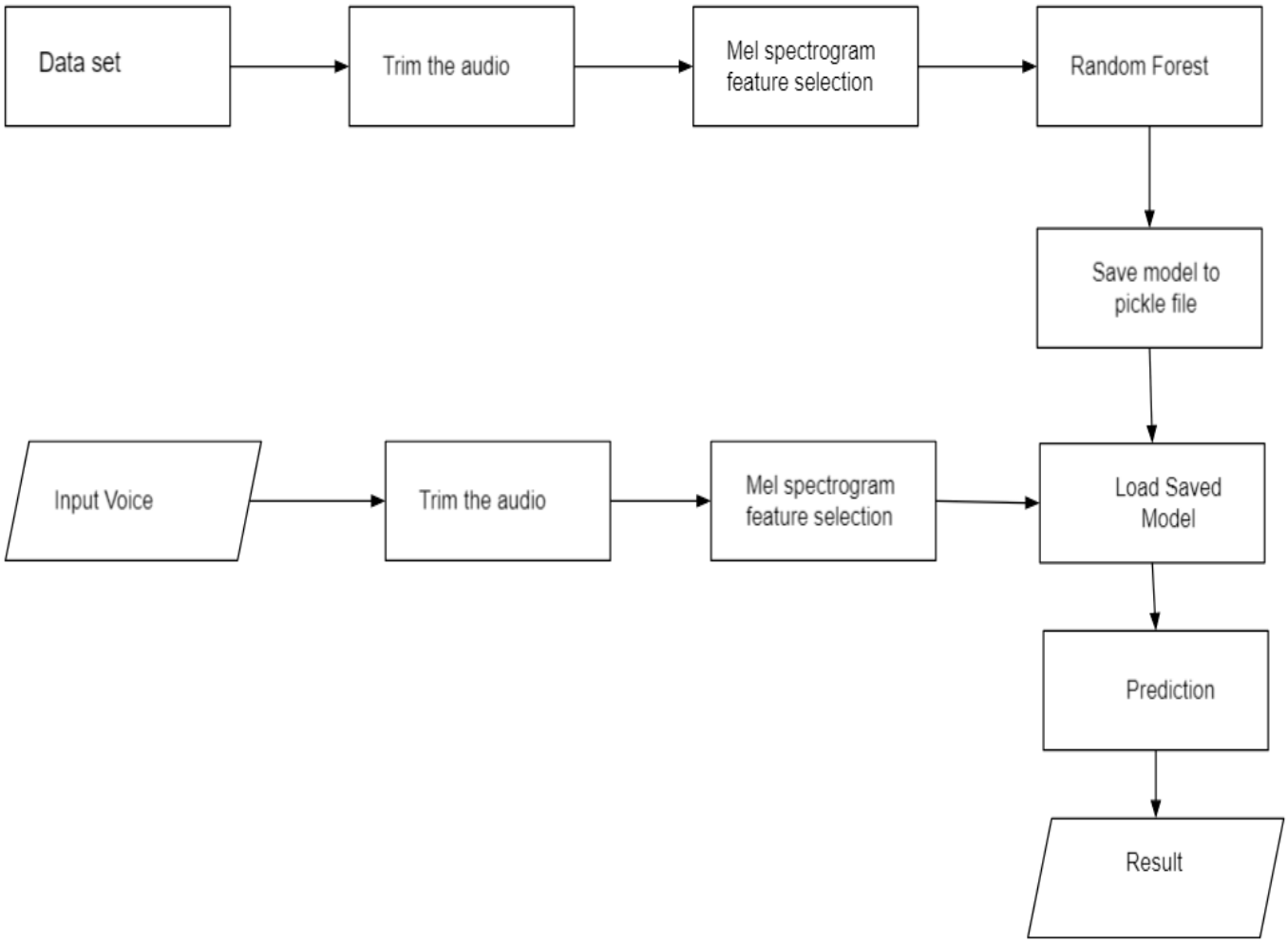

Figure 5.1: Architecture of model Creation

- **Model Building:**The entire work is done on Jupiter notebook and with keras backend API. Python functions was written to produce the for mentioned features from the sounds, the file paths in the baby cry array was used to extract the sound features by calling each of the functions. The first thing that was done is importing the necessary modules.

- **Loading Dataset:** Extraction of sounds paths into an array, the downloaded data was kept in different folders. The file paths were specified in the python program. The sounds were classified into two arrays in the code which was the baby cry array and the non-baby cry array. The baby cry array entails sound from the baby cry folder while the non-baby cry array consists of sounds from the baby laugh, baby silence and noise. And other domestic sounds.The data I use for this analysis is sourced from a corpus entitled the "Donate-a-Cry Corpus".

  To extract features from this data, I took the Fourier transform of each sound file, normalizing the frequencies produced. This allowed us to eliminate the time dimension and analyse if the frequencies in a baby's cry contain the requisite information to differentiate the reasons for crying.

- **FEATURE EXTRACTION:** To extract the useful features from sound data, I am using Librosa library. It provides several methods to extract different features from the sound clips.I am using the method melspectrogram. It Compute a Mel-scaled power spectrogram.

  In sound processing, the mel-frequency cepstrum is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

- **COMPUTING MEL-SCALED SPECTROGRAM** To extract the useful features from sound data, I m using Librosa library. It provides several methods to extract different features from the sound clips and I use the method melspectrogram. It Compute a Mel-scaled power spectrogram.

  In sound processing, the mel-frequency cepstrum is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency.

  A spectrogram is calculated by computing the fast Fourier transform (FFT) over a series of overlapping windows extracted from the original signal. The process of dividing the

signal in short term sequences of fixed size and applying FFT on those independently is called Short-time Fourier transform (STFT). The spectrogram is then calculated as the (typically squared) complex magnitude of the STFT. Fourier Transform is a function that gets a signal in the time domain as input, and outputs its decomposition into frequencies.

The Mel Scale, mathematically speaking, is the result of some non-linear transformation of the frequency scale. Mel Scale is constructed such that sounds of equal distance from each other on the Mel Scale, also "sound" to humans as they are equal in distance from one another. signal is a variation in a certain quantity over time. For audio, the quantity that varies is air pressure. I can take samples of the air pressure over time. The rate at which I sample the data can vary, but is most commonly 44.1kHz, or 44,100 samples per second. What I have captured is a waveform for the signal, and this can be interpreted, modified, and analyzed with computer software.

- **Data split:**The data is split into 80% for training 10% validation and 10% for testing using sklearn python. Here 80% data is given for training the model 10% data is used to validate the trained model and after hyper parameter tuning then test the best selected trained model with 10% test set.

- **Classification using Random Forest:**Classification performed through the Random Forest. Random Forests is an extension of machine learning classifier which includes the bagging to improve the performance of Decision Tree. Random Forests are weighted combinations of tree classifiers that use a random selection of attributes to build the decision taken at each node. Random forests are built by combining through averaging the predictions of several decision trees, each one trained in isolation. It combines tree predictors, and trees are depended on a random vector which is independently sampled. The distribution of all trees is the same. Random Forests splits nodes using the best among of a predictor subset that are randomly chosen from the node itself, instead of splitting nodes based on the variables. The time complexity of the worst case of learning with Random Forests is O(M (dn log n)) , where M is the number of growing trees, n is the number of instances, and d is the data dimension.

  The second part of the project is to build the user interface. The user interface is build using HTML and Python. This is the part of project which deals with the user. The input text is fed into the server through this. And the results returned are also displayed in the

user program. The interface is built in a way such that it is easy and understandable for the person who uses it. For that I uses responsible HTML designs which uses Bootsstrap, CSS and JavaScript also to provide the better user experience. Python Flask is also used for the development of user interface.

5.2 Data Flow Diagram

DFD is one of the graphical representation techniques used in a project to show the flow of the data through a project. DFD helps us to obtain an idea about the input, output, and process involved. The things absent in a DFD are control flow, decision rules, and loops. It can be described as a representation of functions, processes that capture, manipulate, store, and distribute data between a system and the surrounding and between the components of the system. The visual representation helps for good communication.

In level 0 the basic data flow of the application is showcased. It does not show the flow of data much deeper. It will be evaluated in the higher levels of Data Flow Diagram. The Data Flow Diagram of Infantile Vocal Decoder is shown below.

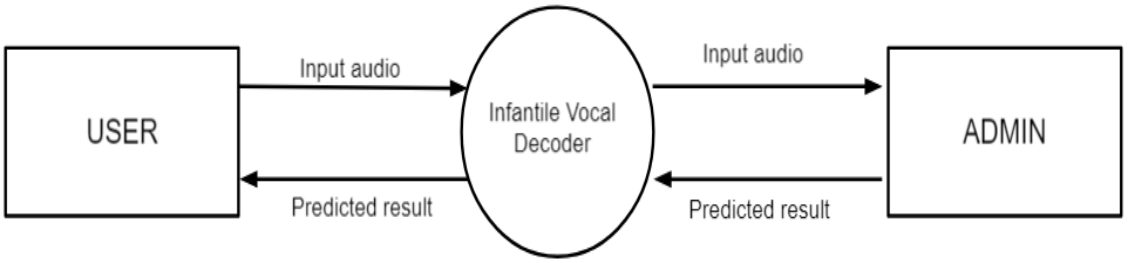

Figure 5.2: Level 0 DFD

The diagram shows Level 0 Data flow diagram of the Infantile Vocal Decoder. As the diagram indicates there is a user part and an admin one. The input of the project is the audio file of child cry by the user. Then the audio is passed to the admin part for the evaluation. The evaluation of Infantile vocal Decoder is occurred in the admin side. The Prediction of the evaluation is passed back to the user through the application. This is how the data flows through the application. Since there is no database in the application, the data is not stored anywhere. The data is lost after the evaluation.

5.3 Screenshots of user interface

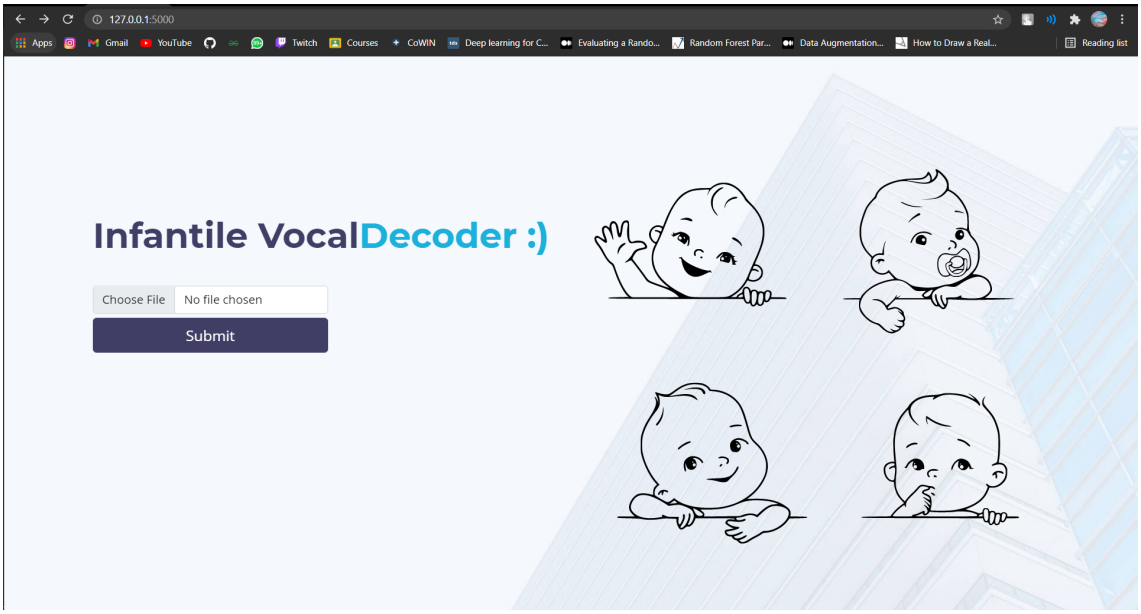

Figure 5.3: input

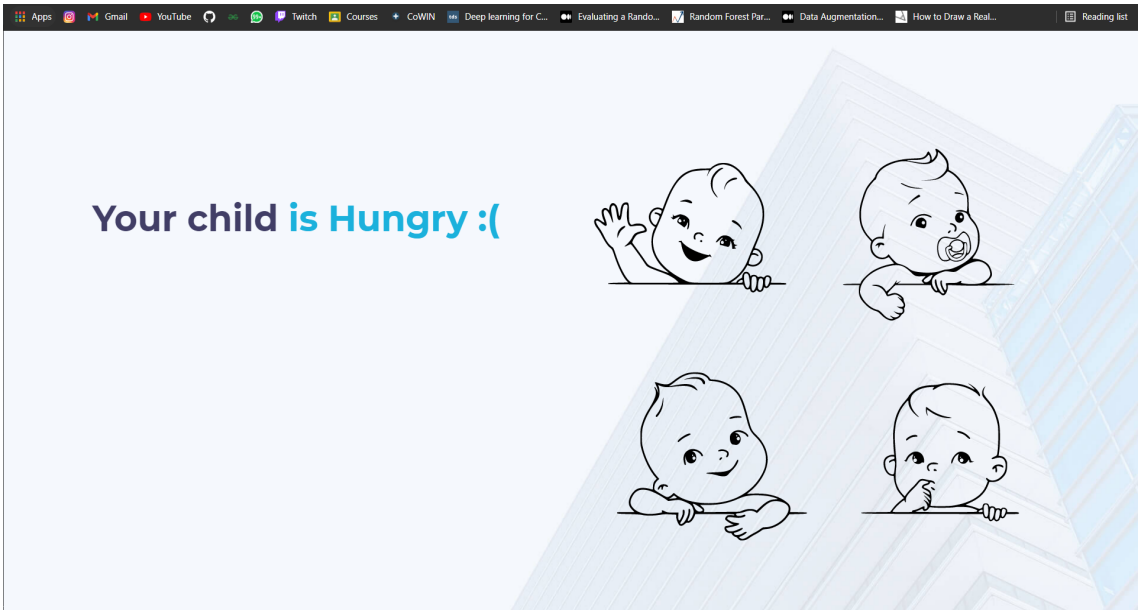

Figure 5.4: output

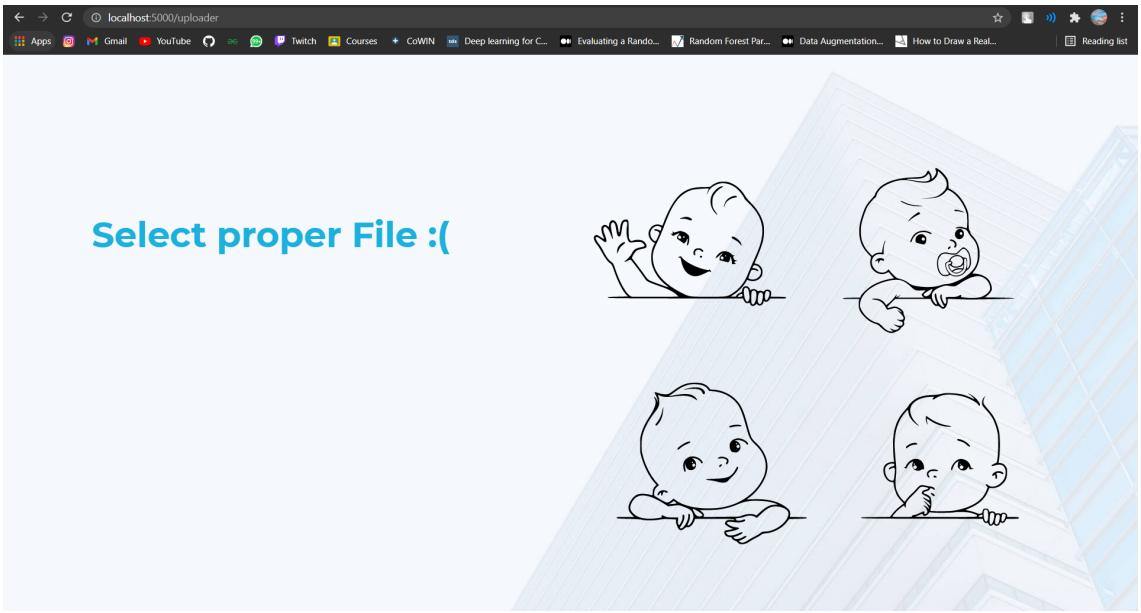

Figure 5.5: output2

Chapter 6

Coding

6.1 Algorithm for Creating the model:

1: Split the data set into training data set and testing dataset. 80% of the dataset is used for training and the remaining 20% is used for testing to obtain better result.

2: The training dataset is used for the preprocessing stage and the preprocessed data is further used to extract the features.The features is extracted using MFCC .

3: The model is created using Random Forest with the selected features.

4: The testing dataset is feeded into the created model and their results are noted down.

5: The result of testing dataset evaluated using the created model is then compared with the actual values of the testing dataset to evaluate the efficiency of the model.

6: Further tuning is performed upon the created model to improve the efficiency of the model.

6.2 Algorithm for web Application and Prediction:

1: Read the audio or child from the user through the user interface.

2: On button click the value in the web page is passed to the server program fror the evaluation of the audio.

3: From the server program, access the input voice and perform the preprocessing tasks on it.

4: The preprocessed voice is used to extract the required features from it using MFCC.

5: Using the pretrained model, I evaluate the input audio using the extracted MFCC features.

6: The Prediction is evaluated by the results of the model and the result is passed to the web page.

7: The Prediction is displayed in the web application.

Chapter 7

Testing and Implementation

7.1 Testing and various types of testing used.

Once a software is developed, the major activity is to test whether the actual results match with the experimental results. This process is called testing. It's used to make sure that the developed system is defect free. The main aim of testing is to find the errors and missing operations by executing the program. It also ensure that all of the objectivs of the project are met by the developer. The objective of testing is not only to evaluate the bugs in the created software but also finding the ways to improve the efficiency, usability and accuracy of it. It aims to measure the functionality, specification and performance of a software program. Tests are performed on the created software and their results are compared with the expected documentation. When there are too much errors occurred, debugging is performed. And the result after debugging is tested again to make sure that the software is error free. The major testing processes applied to this project are unit testing, integration testing and system testing. In unit testing, our aim is to test all individual units of the software. It makes sure that all of the units of the software works as it intended. In integration testing, the combined individual units are tested to check whether it met the intended function or not. It helps us to find out the faults that may arise when the units are combined. In system testing the entire software is tested to make sure that it satisfies all of the requirements. The tables shown below describes the testing process occurred during the development of this project "Infantile vocal Decoder". This defines the various steps took to create the project error free.

7.1.1 Unit Testing

Test Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	create the user interface	To load the web page with required fields	Same as expected	Pass
2	pre-processing	clean the dataset for feature extraction	same as expected	Pass
3	extract features from dataset	extract various features from dataset and store it in a csv file	csv file generated	Pass
4	training and testing of model	create the model and store it in a pickle file	pickle file generated	Pass
5	prediction	predict the result accurately	same as expected.	Pass
6	python server program	set up a python flask server to run the program	Same as expected	Pass

Table 7.1: Unit test cases and results

7.1.2 Integration Testing

Test Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	load the user interface from python	the user interface is loaded when I run the flask program	Same as expected	Pass
2	pass input audio from web page to server	To pass the input audio selected by the user to the python program to and receive it there.	Same as expected	Pass
3	Prediction	load the previously generated pickle file to the server and predict the result with it and extracted features.	Same as expected	Pass
4	display results	pass the result to web page and display it there	Same as expected	Pass

Table 7.2: Integration cases and result

7.1.3 System Testing

Text Cases and Result

Sl No	Procedures	Expected result	Actual result	Pass or Fail
1	to run python server	Server program executed successfully, hence the entire program worked without any crash	Same as expected	Pass
2	Prediction	allow user to input audio and output generated according to the input audio.	Same as expected	Pass

Table 7.3: System test cases and results

Chapter 8

Results and Discussion

The main aim of the project was to predict the reason behind the baby cry. And it is observed that the system performs all the functionalities as expected. By using this machine learning model the computer can evaluate an baby cry audio and predict the result

8.1 Advantages and Limitations

The proposed system is a machine learning model to evaluate the baby cry and predict the reason. The proposed system posses more advantages over the existing system. The proposed system save a huge amount of time. It will reduce the rush in hospitals to meet pediatricians.Like every other system, this system also have it's own disadvantages. But they are negligible while comparing with the advantages and they can be overcame in future.

8.1.1 Advantages

- Through this AI system , decoding of the infant become easily to parents.
- Detect the emotions of infants like hunger, pain , discomfort etc.
- Fastest and accurate prediction
- It will reduce the rush in hospitals to meet pediatricians.
- It hold information regarding their needs or health.

8.1.2 Limitations

- The current dataset is comparatively small. hence the results are neither too good nor too bad. It can be improved by improving dataset
- Further testing must be carried out in a noisy and not noisy place.
- Many other feature extraction method and classification method can be used to compare their performance.

8.2 Accuracy

Accuracy is defined as the percentage of correct predictions for the test data. It can be calculated easily by dividing the number of correct predictions by the number of total predictions.Classification Accuracy is what usually mean, when use the term accuracy. It is the ratio of number of correct predictions to the total number of input samples.The most commonly used metric to judge a model and is actually not a clear indicator of the performance. The worse happens when classes are imbalanced.The accuracy of Infantile Vocal Decoder is 91 percentage .

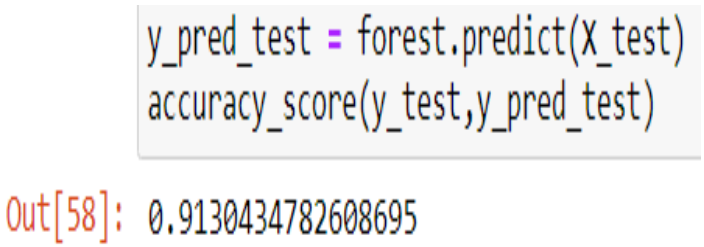

Figure 8.1: Accuracy

Chapter 9

Conclusion and Future Scope

Infantile vocal decoder uses Mel Frequency spectrogram as feature extraction method and Random Forest as classification method can detect the baby is crying or not.

I developed a method for pattern recognition of baby's emotions expressed in the baby's cries. The main goal of this study is to apply the proposed method to improve baby care in the baby's home. I successfully applied the proposed method for pattern recognition of baby's emotions. The average accuracies of emotion recognition were 89%. Tests on the pre-emphasis value, filter bank value, cepstral value, and Random forest have different calculation scenarios. The use of coefficients in the preemphasis does not give significant impact to improve the accuracy in the classification process but it affects the quality of the feature extraction results. The choice of filter bank value and cepstral value can affect the accuracy of the classification process even though it is not significant. The use of Random Forest value will affect accuracy during the classification process and so does the quality of the testing data.

From this study, I can conclude that the best results in the testing scenario are: The number of the filter bank value must be greater than the cepstral value and the cepstral values are adjusted to get the best performance in baby's cry detection

I can conclude that the use of the Mel Frequency Cepstral Coefficient method can be implemented in the baby crying detection system. The sound for training data and test data must be clean from noise to get the better accuracy, so it will produce good characteristic values. This analysis study is still not enough to overcome the noise in baby crying sound. It has to add any normalization process and the pre-processing process to reduce the noise prior the core process.

Bibliography

[1] Eduard Franti,Ioan Ispas,Monica Dascalu "Testing the Universal Baby Language Hypothesis - Automatic Infant Speech Recognition with CNNs",IEEE,2018.

[2] Bhagatpatil Varsharani V, V. M. Sardar," An Automatic Infants Cry Detection Using Linear Frequency Cepstral Coefficients(LFCC)" ,IJTEEE , 2014.

[3] Sita Purnama Dewi,Anggunmeka Luhur Prasasti, Budhi Irawan, "The Study of Baby Crying Analysis Using MFCC and LFCC in Different Classification Methods",IEEE,2019.

[4] Aomar Osmani , Massinissa Hamidi, Abdelghani Chibani,"Machine Learning Approach for Infant Cry Interpretation",IEEE,2018.

[5] Lichuan Liu,Wei Li Yang Li,Kevin Kuo," Infant cry signal detection, pattern extraction and recognition", IEEE,2019.

[6] V. V Bhagatpatil and P. V. M. Sardar, "An Automatic Infant's Cry Detection Using Linear Frequency Cepstrum Coefficients (LFCC)," vol. 5, no. 12, pp.

[7] N. Dave, "Feature Extraction Methods LPC , PLP and MFCC In Speech Recognition," Int. J. Adv. Res. Eng. Technol., vol. 1, no. Vi, pp. 1–5, 2013.

[8] G. V. I. S. Silva and D. S. Wickramasinghe, "Infant Cry Detection System with Automatic Soothing and Video Monitoring Functions," J. Eng. Technol. Open Univ. Sri Lanka, vol. 5, no. 1, pp. 36–53, 2017.

[9] R. M. Aarts, "Audio signal processing device," J. Acoust. Soc. Am., vol. 120, no. 6, p. 3445, 2006.

[10] S. S. Jagtap, P. K. Kadbe, and P. N. Arotale, "System propose for Be acquainted with newborn cry emotion using linear frequency cepstral coefficient," Int. Conf. Electr. Electron. Optim. Tech. ICEEOT 2016, pp. 238–242, 2016.

[11] P. K. Sari, K. Priandana, and A. Buono, "Perbandingan Sistem Perhitungan Suara Tepuk Tangan dengan Metode Berbasis Frekuensi dan Metode Berbasis Amplitudo Comparison of Applause Calculation Systems using FrequencyBased Method and Amplitude-Based Method," J. Ilmu Komput. Agri-Informatika, vol. 2 Nomor 1, pp. 29–37, 2013.