International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

Vulnerable Children Model to Analyse Speech Detects Anxiety and Depression in Early Childhood using Machine Learning Algorithm

Bharath M[1], Anurag Umashankar[2], Prapulrag S[3]

[1,2,3]UG Student Department of Computer Science and Engineering, SJB Institute of Technology, Bengaluru India - 560060

Abstract *Childhood anxiety and depression often go undiagnosed. If left untreated se conditions, collectively known as internalizing disorders, are associated with long-term negative outcomes including substance abuse and increased risk for suicide. This project is a new approach to identify children with internalised disorders using a speech. We have planned to implement machine learning analysis of audio data from task can be used to identify children with an internalizing disorder. speech features most discriminative of internalizing disorder are analyzed in detail, showing that affected children exhibit especially voice at low pitch, with repeating speech inflections and content, and response with high pitch voice surprising stimulative to controls.*

Key Words: Childwood anxiety, Depression, Speech, Machine learning algorithm, learning analysis, stimulative control.

1. INTRODUCTION

Anxiety and depression can emerge in children at its youngest age as possible. National Institute of Mental Health estimates that at least 3.3% of children have had episodes of severe depression but symptoms are often overlooked until children can more clearly express ir discomfort given abstract emotions involved, and communicate ir impairment with help-seeking adults. Current standard with gold diagnostic assessment in children is to conduct a 50-90 minute semi-structured interview with a trained clinician and primary caregiver. Limitations such as waiting lists and insurance burden may slow assessment process, and poor report given by parents about child's emotions may prevent many children from receiving appropriate referrals and diagnoses. Many new tools that can feasibly and objectively screen children for se disorders, practise of pediatric visits would support surrounding adults in understanding intensity of ir child's distress, and to would help children to overcome ir problems.

Applications:

1.Healthcare

2.Education

1.1 Problem Formulation

Standard diagnostic assessment in young children is to conduct interview with a trained clinician and guardians or parents.

Limitations of standard diagnostic are:

1. Report provided by clinicians are inaccurate cause parental data and child data might be false
2. Process of garing data takes long duration.

1.2 Existring System:

Mood induction tasks have been increasingly used in research contexts to "press" for anxious, frustrating, joyful, or saddening affect. A child's behavioral and physiological response to tasks is recorded using a variety of technologies (i.e., video-recorded and coded behaviors and directly measured cortisol, heart rate variability, electrodermal activity, movement), manually processed, and studied in relation to ory-driven expectations. Trier-Social Stress Task (TSST) is one such mood induction task meant to induce performance anxiety by having a participant give a short, improvisational speech to a confederate audience pretending to be thoroughly bored and critical. Behavioral coding and physiological measures have not only been associated with task affect, but also with mental illness more generally.

Disadvantages of existing methods.

1. Prediction accuracy is less and suits for group of 3 to 6 years age, also data given by parents and child may not be correct
2. It take more time to analysis of data

1.3. Proposed System:

We employ speech analysis of child voice recordings during a 3-minute speech task and machine learning to detect clinically-derived anxiety and depressive diagnoses in children between ages of 3 and 7 years old.

Advantages of proposed system:

1. Cost effective.
2. Accuracy of detecting depression.

3. Less time computation

Objectives of Project

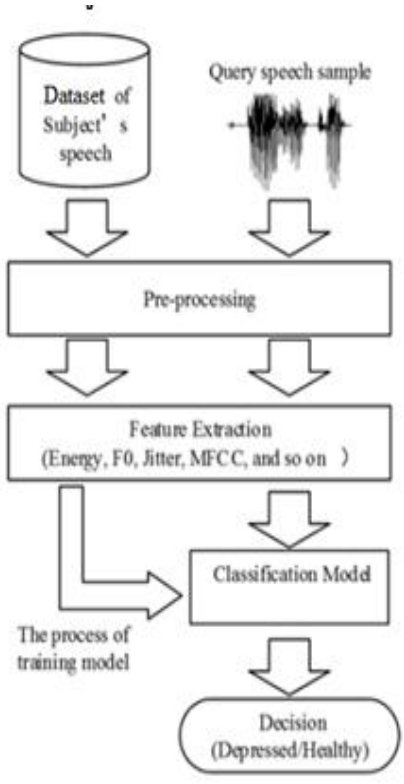
1. Accurately identify young children with internalizing disorders using a 3-minute speech task.
2. Improve accuracy of detecting depression in young children using machine learning approaches.
3. Build an automated, fast, low cost system to detect depression in young children.

2. Methodology for Proposed Project

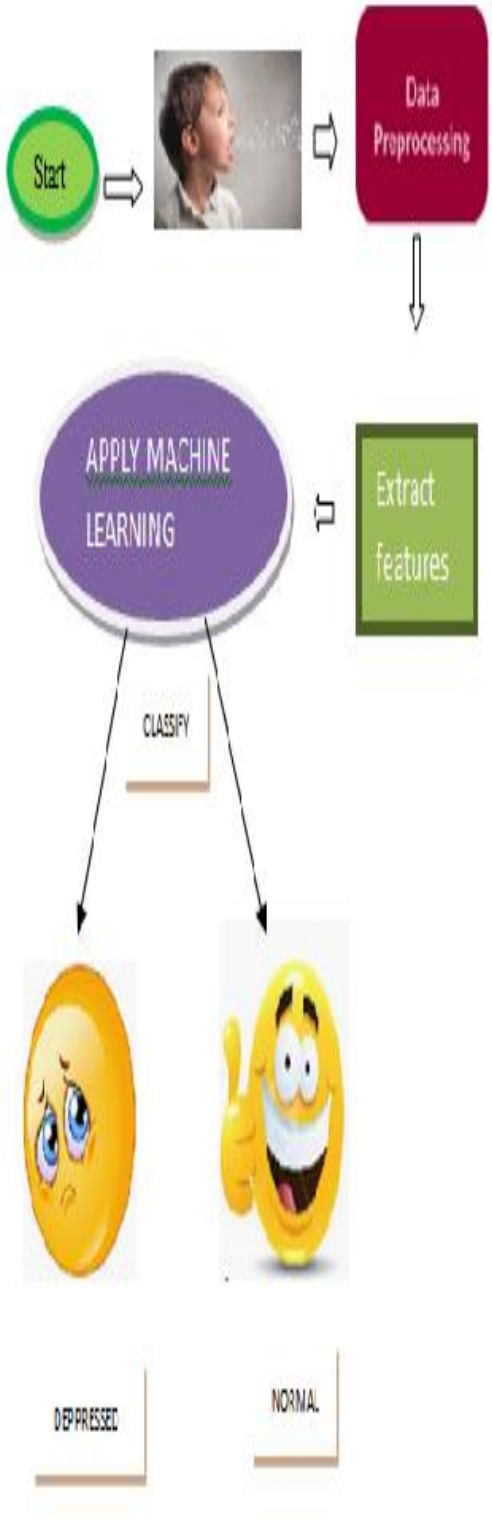
Binary classification models such as **Logistic Regression, Linear Kernel, A random forest:**

3. Architecture of Project

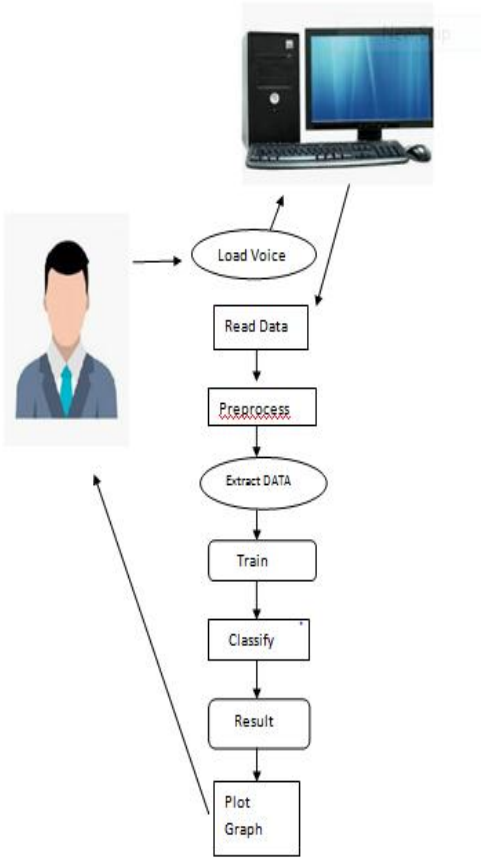
We are collecting child audio dataset using VAD n will perform preprocessing, later will extract features from audio, with help of machine learning model will classify m as depressed or normal.



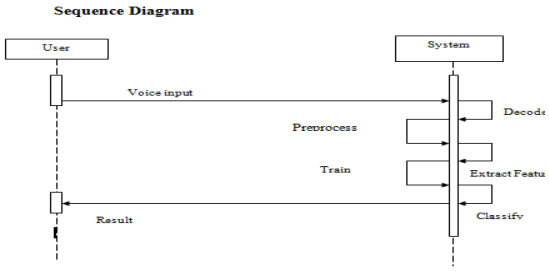
1. **Flow chart:**



2. Use Case Diagram:

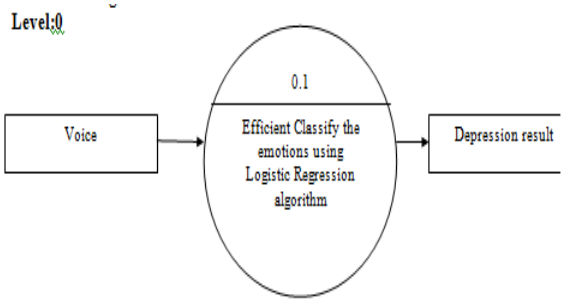


3. Sequence Diagram

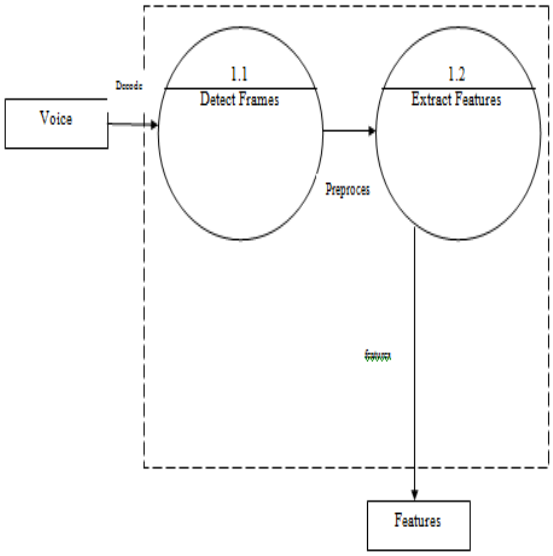


sequence diagram will determine users states as active/ inactive

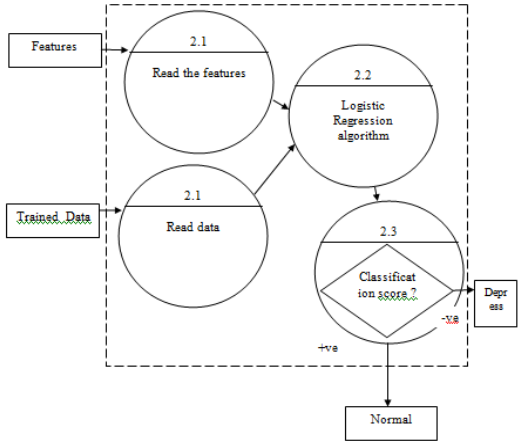
4. Data Flow Diagram:



Level: 1



Level 2:



5. Requirements

Software:

- OS: Windows 7.
- Coding Language: Python3.6
- Tools: Python IDLE

Hardware:

- System: Pentium i3.
- Hard Disk: 120 GB.
- Monitor: 15'' LED
- Input Devices: Keyboard, Mouse
- Ram: 4 GB

6. Motivation/ Scope of Project

With NMHP and its implementation arm, District Mental Health Programme (DMHP) has a greater scope for strengthening and scaling up of services for depression through focused and dedicated interventions. Screening and treatment for depression using simple tools and standard protocols can be implemented at various levels of health care, workplace and educational settings. Accessibility to treatment for depression can be enhanced through provision of services and ensuring continuous supply of basic drugs at primary health care settings.

Implementation

A. Clinical Measures

Speech Task is an adapted version of Trier Social Stress Task for children (TSST-C), which has been shown to induce anxiety in children 7 and older. This task, which was conducted during home visit, is standardized, and all research assistants were trained to carry out task according to protocol including displaying flat affect through duration of task. In Speech Task, participants are instructed to prepare and give a three-minute speech and are told that y will be judged based on how interesting it is. y are given three minutes to prepare, and n begin ir three-minute speech. A buzzer is used to interrupt participant's speech with 90 and 30 seconds remaining in task. At each interruption, experimenter informs participant of time remaining in task using a standardized script. experimenter responds to participant questions as necessary. Each speech was recorded using a standard video camera, truncated to include just three-minute speech task, and audio was extracted for furr analysis.

B. Audio Data Processing

Audio data from speech task were sampled at 48 kHz and processed via a voice activity detector that discriminates instances of speaking from background noise. VAD operates on signal energy and has been designed to have a high sensitivity towards speech. Speech epochs were identified when energy within a sliding window was above baseline noise. Identified raw speech epochs, which included full sentences, phrases, phonemes, and high energy noise, were smood using a median filter with window length of 0.21 seconds. This ensured that natural pauses in speech were contained within a single speech epoch and that short-duration phonemes and noise were removed. Due to realities of collecting data from children in home, many recordings had low signal-to-noise ratios and were corrupted by significant harmonic background noise. Thus, each audio file and its detected speech epochs were screened manually for quality.

C. Extraction of Audio Features

To characterize ability of proposed approach for identifying children with an internalizing disorder, we first partitioned each three-minute speech task into three phases, boundaries of which were defined by buzzer interruptions inherent to task. To parameterize audio signal within each phase, we computed following features for each speech epoch: speech epoch duration, zero crossing rate (ZCR) of audio signal, Mel frequency cepstral coefficients (MFCC), dominant frequency, mean frequency, perceptual spectral centroid (PSC), spectral flatness, skew and kurtosis of power spectral density (PSD), ZCR of z-score of PSD (ZCR zPSD) for all speech epochs, first, second, and third formants, and percentage of signal energy above 200 to 2000 Hz. We also extracted mean, median, standard deviation, maximum, and minimum ZCR zPSD from sliding windows in time and frequency domains within each speech epoch. Descriptive statistics (i.e., mean, standard deviation, median, maximum, minimum) were computed for each feature within each phase.

Procedure Feature

Extraction(audio files):

INPUT: Audio files

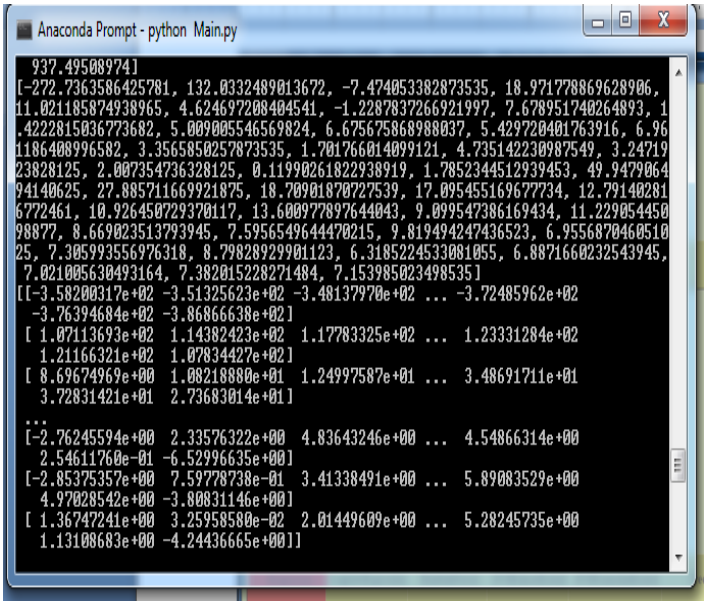
OUTPUT: Extract MFCC core features

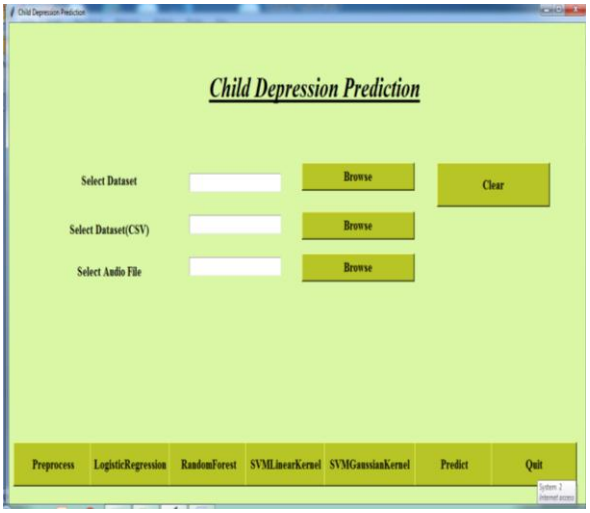
Steps:

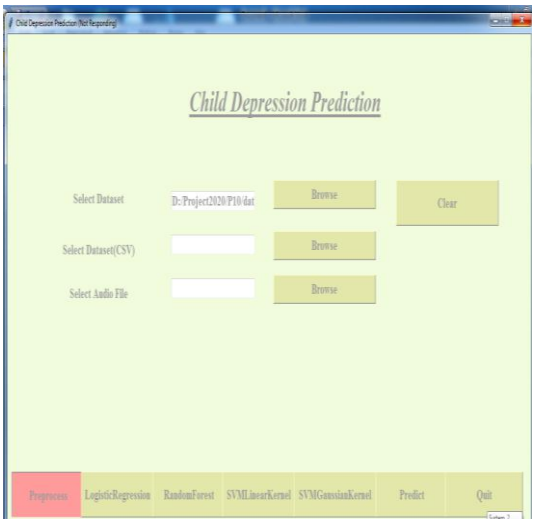
1. Read dataset files

2. For each file extract core features from file using librosa library

3. Write extracted features in corresponding csv file.







D. Model Development and Analysis

Binary classification models relating audio signal features from each phase to internalizing disorder determined via K-SADS-PL with clinical consensus were trained using a supervised learning approach on data classified as High Quality. Classifier performance was established using leave-one-subject-out (LOSO) cross validation. In this approach, data from all but one participant (N=42) were partitioned into a training dataset and converted to z-scores prior to performing Davies-Bouldin Index based feature selection. This yields eight features with zero mean and unit variance that best discriminate between diagnostic groups. Thus, 42 observations of se eight features were used to train binary classification models for predicting internalizing diagnosis.

Same eight features were extracted, converted to z-scores based on parameters (e.g. mean, variance) from training set, and used as input to model for predicting diagnosis of one remaining test subject. score threshold to determine diagnosis was set for each iteration using Number Needed to Misdiagnosis criteria based on ROC curve of training data. This measure, which is maximized to find appropriate threshold, is an estimate of number of patients who need to be tested in order for one to be misdiagnosed. This process was repeated 42 times until diagnosis of each subject had been predicted.

INPUT: Dataset

OUTPUT: Classified results

1. **Logistic regression – LR:**

Steps:

1. Read dataset

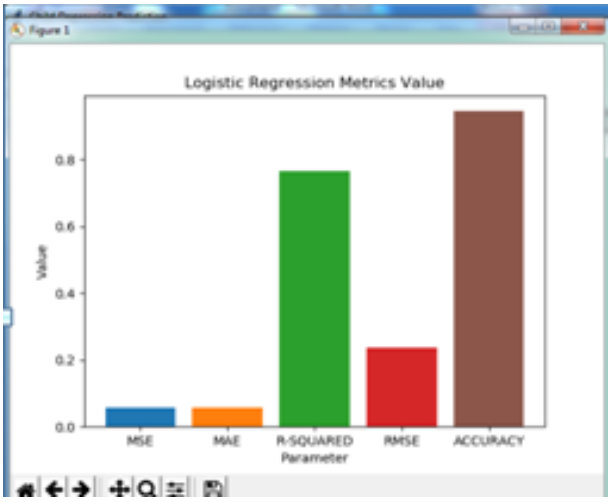
2. Split Dataset into train and test data

3. Use train data to train logistic regression model

4. Predict result using sigmod function

5. Return predicted class.

6. Calculate accuracy measures.



2. **Support vector machine with a linear kernel – SL:**

Steps:

1. Read dataset

2. Split Dataset into train and test data

3. Use train data to train SVMmodel

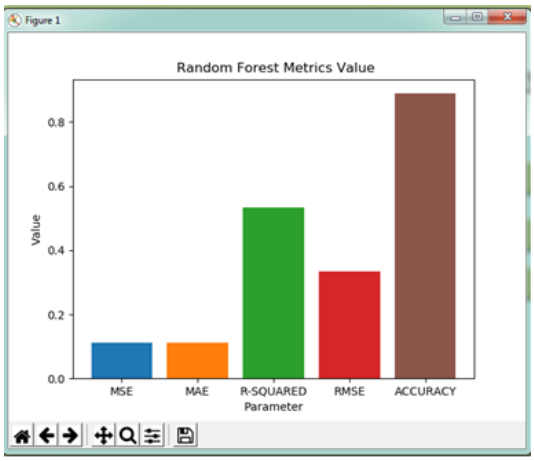
4. Return predicted class.

5. Calculate accuracy measures

3. **Support vector machine with a gaussian kernel – SG**:

Steps:

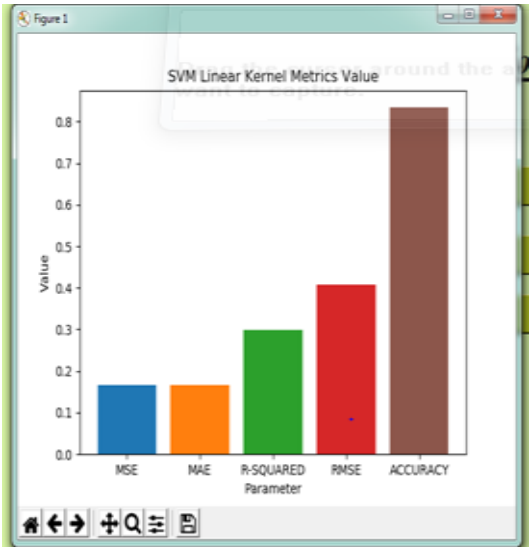
1. Read dataset
2. Split Dataset into train and test data
3. Use train data to train SVMmodel
4. predict result using test data
5. Return predicted class.
6. Calculate accuracy measures



4. **Random Forest – RF**

Steps:

1. Read dataset
2. Split Dataset into train and test data
3. Use train data to train RF model
4. predict result using test data
5. Return predicted class.
6. Calculate accuracy measures



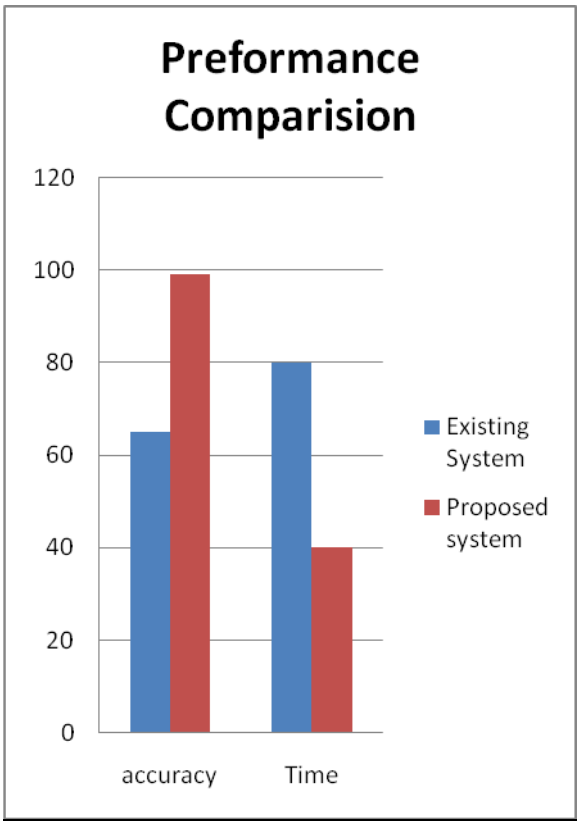
4. Test Cases

Test Case#	1
Test Name	User input file format
Test Description	To test user input file as dataset folder which contains audio .wav file
Input	Dataset folder
Expected Output	file should be read by program and print path of input folder on console
Actual Output	file is read and print contents accordingly
Test Result	Success

Test Case#	2
Test Name	User input format
Test Description	To test user input file as dataset folder which contains audio .wav file
Input	folder as null
Expected Output	It Should show alert Message enter valid input
Actual Output	Shown alert message
Test Result	Success

Test Case#	3
Test Name	Feature Extraction
Test Description	To test wher extracting important features
Input	.wav file
Expected Output	It extract important features
Actual Output	Its extracted important features
Test Result	Success

5. Performance Evaluation:



Conclusion:

Results presented herein suggest that a machine learning analysis of child speaking patterns during a short anxiety induction task is able to identify children with internalizing psychopathology. This statistical classification model outperforms clinical thresholds on parent-reported child symptoms collected with CBCL, indicating its potential as an objective screening tool in this population. A detailed analysis of audio features selected for this classification indicated that affected children exhibit low-pitch voices, with repetitive inflection and content and high-pitched response to surprising stimuli.

Future Enhancement:

In future we can use deep learning algorithm to increase efficiency of prediction result.

References

[1] T. E. Chansky and P. C. Kendall, "Social expectancies and self-perceptions in anxiety-disordered children," J. Anxiety Disord., Aug. 1997.

[2] H. L. Egger and A. Angold, "Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology," J. Child Psychol. Psychiatry, Apr. 2006.

With the guidance of Ragavendra K M, Assistant Professor in Computer science dept. SJBIT.