Giving Voice to Vulnerable Children: Machine Learning Analysis of Speech Detects Anxiety and Depression in Early Childhood

Mahesh TR[1], Vamsi Krishna G[2], Sathwik P[3], Ajith Chowdary V[4], Hemchand G[5]

[1]Professor, Dept. of computer science Engineering, SET-Jain University, Benguluru-Karnataka, India
[2-5]Student , Dept. of computer science Engineering, SET-Jain University, Benguluru-Karnataka, India

---***---

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

1. INTRODUCTION

Anxiety and depression can emerge in children as young as four years old, but symptoms are often overlooked until children can more clearly express their discomfort given the abstract emotions involved, and communicate their impairment with help-seeking adults. The current gold standard diagnostic assessment in young children is to conduct a 60-90 minute semi-structured interview with a trained clinician and their primary caregiver. Limitations such as waiting lists and insurance burden may slow the assessment process, and poor parental report of internal child emotions may also prevent many children from receiving appropriate referrals and diagnoses. Many signs of anxiety or depression at this early age are unrecognized by well-intentioned but unknowing parents or are dismissed as transient. However, we know that nearly 20% of children experience an internalizing disorder during childhood. This psychopathology impairs a child's functioning and development and predicts serious health problems later in life if left untreated (e.g., substance abuse development of comorbid psychopathology increased risk for suicide Thus, there are high individual and societal burdens associated with internalizing disorders that highlight the need for effective early assessment. New tools that can feasibly and objectively screen children for these internalizing disorders during routine pediatric well-visits would support surrounding adults in understanding the intensity and chronicity of their child's distress, and connect them with interventions early in development, when neuroplasticity and potential for symptom improvement is greatest. Applications are Healthcare, Education, Society and environment.

2. ALGORITHMS SPECIFICATION

Binary classification models (logistic regression – LR, support vector machine with a linear kernel – SL, support vector machine with a gaussian kernel – SG, random forest – RF) relating the 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 the data classified as High Quality.

Logistic Regression:

Logistic Regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.). In other words, the logistic regression model predicts P(Y=1) as a function of X.

Linear Kernel:

Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set. One of the examples where there are a lot of features, is Text Classification, as each alphabet is a new feature. So we mostly use Linear Kernel in Text Classification.

A random forest:

A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement if bootstrap=True (default).

2.1 Modules:

2.1.1 Clinical Measures:

The Speech Task is an adapted version of the 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 the home visit, is standardized, and all research assistants were trained to carry out the task according to protocol including displaying flat affect through the duration of the task. In the Speech Task, participants are instructed to prepare and give a three-minute speech and are told that they will be judged based on how interesting it is. They are given three minutes to prepare, and then begin their three-minute speech. A buzzer is used to interrupt the participant's speech with 90 and 30 seconds remaining in the task. At each interruption, the experimenter informs the participant of the time remaining in the task using a standardized script. The experimenter responds to participant questions as necessary. Each speech was recorded using a standard video camera, truncated to include just the three-minute speech task, and the audio was extracted for further analysis.

2.1.2 Audio Data Processing:

Audio data from the speech task were sampled at 48 kHz and processed via a voice activity detector (VAD) that discriminates instances of speaking from background noise. The 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 the baseline noise. Identified raw speech epochs, which included full sentences, phrases, phonemes, and high energy noise, were smoothed 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 the realities of collecting data from children in the home, many recordings had low signal-to-noise ratios (SNRs) and were corrupted by significant harmonic background noise. Thus, each audio file and its detected speech epochs were screened manually for quality.

2.1.3 Extraction of Audio Features:

To characterize the ability of the proposed approach for identifying children with an internalizing disorder, we first partitioned each three-minute speech task into three phases, the boundaries of which were defined by the buzzer interruptions inherent to the task. To parameterize the audio signal within each phase, we computed the features for each speech epoch: Mel frequency cepstral coefficients (MFCC), computed for each feature within each phase.

3. IMPLEMENTATION

Functionality Requirements

This section describes the functional requirements of the system for those requirements which are expressed in the natural language style.

1. Create an desktop application.
2. Load the voice as input.
3. System should preprocess and extract the deep features from the voice frames.
4. Applying the Logistic regression algorithm System should classify the given voice is depressed or normal.
5. Application should provides high accuracy of classification of child voice.

Procedure Feature Extraction(audio files):

Input: audio files

Output: Extract the MFCC core features

Begin

Step1: Read the dataset files

Step 2: for each file

Step 2.1 :Extract the core features from the file using librosa library

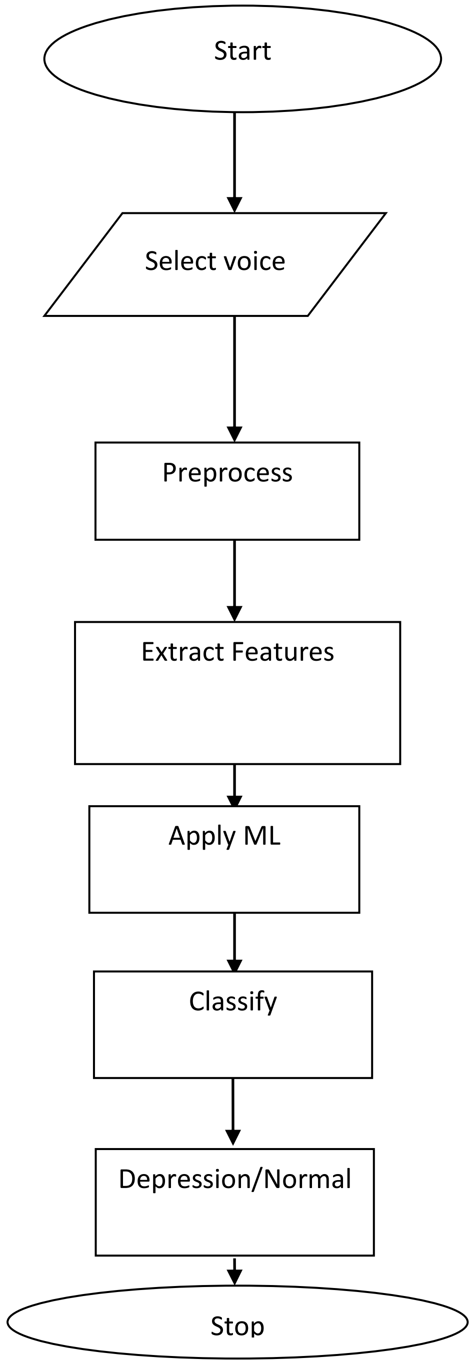
End for

Step 3: write the extracted features in corresponding csv file.

End

We are collecting the child audio dataset Using the VAD we are performing the preprocessing, then we are extracting the features from the audio, the applying the machine learning algorithm to classify whether they are depressed or normal.

Fig 3 Flow chart



5. DATAFLOW DIAGRAM

The DFD is also referred to as bubble chart. It is an easy graphical formalism that can be used to represent a system in terms of input files to the system, varied process distributed on this data, and therefore the output data is generated by this system.

Level 0: Describes the overall process of this project. we are passing voice as a input the system will classify the voice using logistic regression to determine the person is in depression or not.

Level 1**:** Describes the first stage process of this project. we are passing voice as a input the system will detect the frames in voice then extract the features

Level 2: Describes the final stage process of this project. we are passing extracted features from level 1and trained data as a input the system will classify the given voice is normal or depressed using logistic regression algorithm.

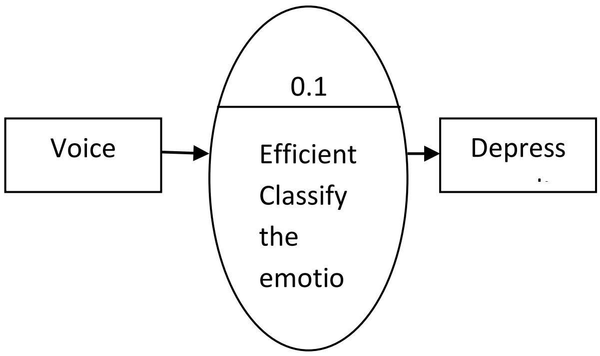


Fig 5.1 level 0

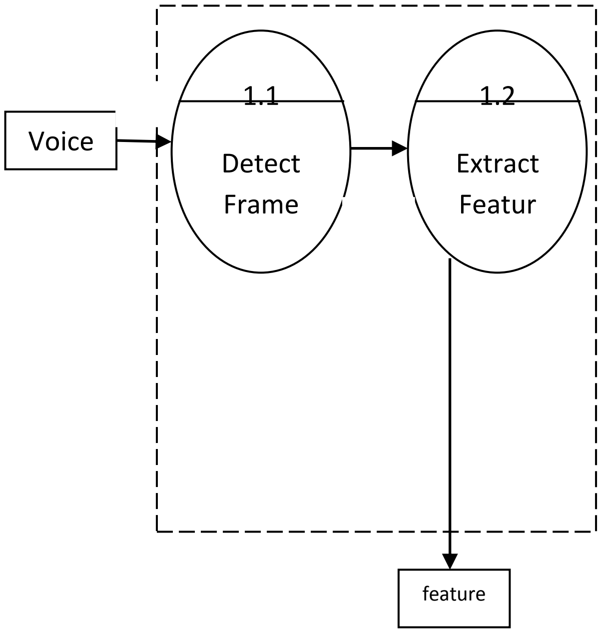


Fig 5.2 level 1

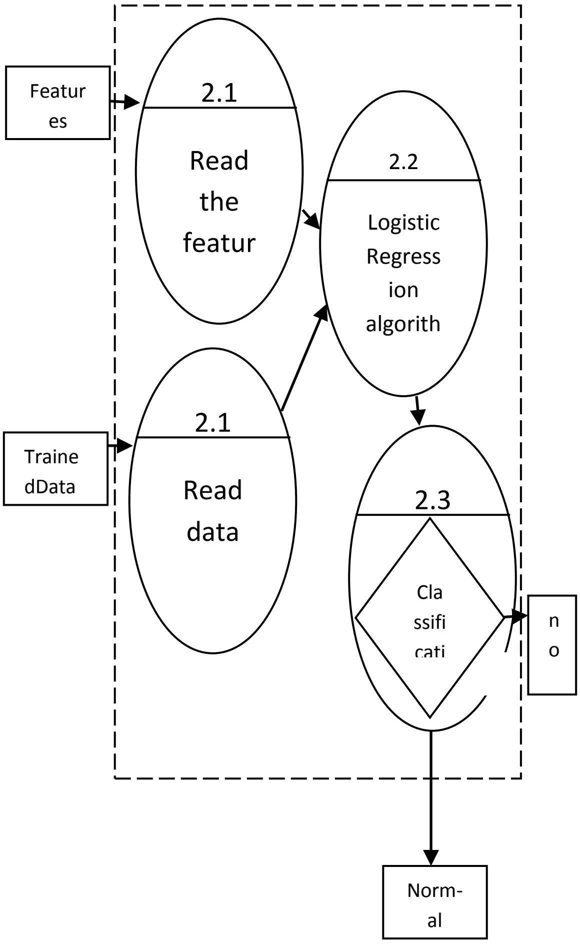


Fig 5.3 level 2

6. CONCLUSION

A desktop application using python to detect depression in young children from their speaking pattern.

We are going to compare various machine learning algorithms and evelovate the performance of those algorithm. The performance result will be shown in graph.

7. REFERENCES

[1] T. E. Chansky and P. C. Kendall, "Social expectancies and self-perceptions in anxiety-disordered children," J. Anxiety Disord., vol. 11, no. 4, pp. 347–363, 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, vol. 47, no. 3–4, pp. 313–337, Apr. 2006.

[3] S. J. Bufferd, L. R. Dougherty, G. A. Carlson, and D. N. Klein, "Parent-Reported Mental Health in Preschoolers: Findings Using a Diagnostic Interview," Compr. Psychiatry, vol. 52, no. 4, pp. 359–369, 2011.

[4] J. L. Luby, A. C. Belden, J. Pautsch, X. Si, and E. Spitznagel, "The clinical significance of preschool depression: Impairment in functioning and clinical markers of the disorder," J. Affect. Disord., vol. 112, no. 1–3, pp. 111–119, Jan. 2009.

[5] N. R. Towe-Goodman, L. Franz, W. Copeland, A. Angold, and H. Egger, "Perceived family impact of preschool anxiety disorders," J. Am. Acad. Child Adolesc. Psychiatry, vol. 53, no. 4, pp. 437–446, Apr. 2014.

[6] A. C. Belden, M. S. Gaffrey, and J. L. Luby, "Relational Aggression in Children With Preschool-Onset Psychiatric Disorders," J. Am. Acad. Child Adolesc. Psychiatry, vol. 51, no. 9, pp. 889–901, Sep. 2012.

[7] S. J. Bufferd, L. R. Dougherty, G. A. Carlson, S. Rose, and D. N. Klein, "Psychiatric disorders in preschoolers: continuity from ages 3 to 6," Am. J. Psychiatry, vol. 169, no. 11, pp. 1157–1164, Nov. 2012.

[8] D. J. Whalen, C. M. Sylvester, and J. L. Luby, "Depression and Anxiety in Preschoolers: A Review of the Past 7 Years," Child Adolesc. Psychiatr. Clin. N. Am., vol. 26, no. 3, pp. 503–522, Jul. 2017.