International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Analysis of Voice: Machine Learning Technique to Detect the Internalizing Disorders

Harshavardhana D[1], Dharshitha S[2], Deepthi H P[3], Deepthi Reddy S N[4], Amita Chaitra Heggade[5]

[1]Assistant Professor, Dept. of CSE, S J C Institute of Technology, Chickballapur, Karnataka, India
[2]Student, Dept. of CSE, S J C Institute of Technology, Chickballapur, Karnataka, India
[3]Student, Dept. of CSE, S J C Institute of Technology, Chickballapur, Karnataka, India
[4]Student, Dept. of CSE, S J C Institute of Technology, Chickballapur, Karnataka, India
[5]Student, Dept. of CSE, S J C Institute of Technology, Chickballapur, Karnataka, India

---

Abstract - *Childhood depression is something that is neglected, where there will be risk in future. Untreated depression can lead to some risk of suicide. They should be given proper treatment. There are many effective tests done to check whether the child is in depression or not. In that one such method is trier-social stress task (TSST). This test is designed where the participant has to prepare a five minute presentation and speak in front of panel of three judges along with a video camera in a room. The panel of three judges tries to disturb the participant and observe the participant presentation. Later they analyze the result and send the results after two or three months where the results will be late then the child would have already gone a long way in depression. This paper presents a new approach of identifying the children depression using three minute speech task. By using machine learning technique we can detect whether the child is in depression or not with accuracy level of 93%. These results help the child overcome his depression and have a bright future in life without suffering from depression even after he grows up. Where he will get confidence to overcome his internalizing disorder. This helps in future enhancement of children for internalizing disorders so that interventions can be deployed when they have highest chance for long term success.*

***Key Words*: Machine Learning, Voice audio detector, Support vector machine, internalizing disorder, Regression Algorithm.**

1. INTRODUCTION

Anxiety and depression are one of the most common disorders that are observed in the people. It can also appear in the children as young as 4 years old. But symptoms are unnoticed until and unless the kid is able to express his or her discomfort. If this is not noticed serious health problems would arise later. Hence effective early assessments are needed. At present clinical diagnostic assessment is conducted to the children. This diagnostic Assessment involves 60 to 90 minutes partially structured interview with clinicians and the parents or guardian of the kid. The primary caregiver should know the complete information about the kid. Improper or poor report could misguide the clinician; this would affect the children at times. In order to solve these problems another approach has been emerged and it is through the voice. This paper explains about the voice analysis. A 3 minutes speech task is conducted to the children between the age 3 to 7 years old. A machine learning technique is used to detect the internalizing disorders in the children.

2. METHODS

2.1 PROCEDURE

This study was done from the Michigan University of Institutional Review Board. Here the child and the caretaker are brought to the university lab and provide them a set of tasks to complete.

The caretakers complete the self and parental report sessions like interview them about their child and to assess the child's diagnosis physically, where the child undergoes some behavioral task in the next room. These tasks are to obtain the features like panic, anxiety and their influence. Moreover the experiment team visits the home of the participants (patient) to note the additional behavior of the child. And for the higher studies they examine the patient's reply to the speech spoken by the examiner. This task is proposed to extract the anxiety report, meanwhile the survey and the diagnostic process are used to evaluate the symptoms.

2.2 CLINICAL MEASURES

The speech task here is the accommodate version of TSST -C Trier Social Stress Task for children which is to get the anxiety level in children 7 and older.

The task which is conducted at the time of home visit is standardised and all the other assistants of the research are trained to carry out the work according to the rules even displaying blunted affect during the period of the task. In this speech task the clients are instructed to record their speech for three minutes and the caretaker will be judging it based

on how interesting the speech is, the participants has to prepare for the speech of three minutes and a buzzer will be there for the interruption which tells about the remaining time left for the task. The buzzer rings at the 90 and 30 sec to inform the remaining time using the standardised script. The caretaker will response to participant question if any. And the whole task is recorded using a video camera, truncated to add three minutes speech task and audio was obtained for further analysis. The clinical interview for 1-2 hours duration was also conducted by the Master's level psychology students along with the caretaker.

2.3 AUDIO DATA PROCESSING

The audio data speech is taken at 48 kHz and processed through a VAD (Voice activity detector) that separates the background noise from the data. VAD works on the signal energy and has developed to have high responsiveness towards speech. The speech span was noticed when the energy is in a sliding window which was greater than the baseline noise. Identified raw voice span also includes the full statements, phonemes, phrases and also high energy noise. Which are then smoothed using a medium filter with window length 0.21 sec.

It makes sure that the natural pauses in the voice are contained within a single span of speech and those short duration pauses and the noise were removed. As the data is collected in the children's home, many recordings had low SNR's (signal to noise ratio) and Android was corrupted by background noise. Thus every audio file in the detected speech spans where screened manually for quality. Here in this process the two research assistants manually label each voice span detected by the VAD. Differences between the labels are discovered by the third researcher. All the audio files which contain patient speech are classified by one researcher into four categories.

1) High quality which means medium to a very strong form of the speech content and frequency. 2) Low quality which has a very poor form of speech and frequency. 3) Participant deviation with the patient (participant) not speaking during the task being done. 4) Protocol deviation where the task won't run as it is planned. Eg. Tash does not last for 3 minutes, buzzers don't ring. Therefore data are taken from the high and low quality group for further analysis.

2.4 PARTICIPANTS

Data from the high and the low quality categories are collected from the 71 children (63% female) under primary care takers (95% mothers). Participants are enlisted from either an ongoing observation or from the bulletin being posted in the psychiatrist clinics and community to get samples with the symptoms. The eligible children ages of 3 and 8 who speak fluent English and whose parents are 18 years and older than that. Exceptional criterias are suspected and diagnosed developmental disorders which have serious medical conditions for taking the medicine that affect the nervous system. (eg. autism) the sample of those children were aged between 3 and 7 and was 65% white non-latinx and 83% who lived in two parental houses. The Multimode assessments that also include diagnostic interviews are conducted between August 2014 and 2015, so based on the several assessments and agreement coding N= 20 patients are identified as having internalizing disorders according to DSM IV (Diagnostic and Statistical Manual of Mental Disorder)

2.5 MODEL DEVELOPMENT AND ANALYSIS

Binary regression models - LR, Veteran support machine with direct kernel - SL, veterans support machine with gaussian laser - SG, random forest - RF) relating to the signals of the audio signal from each stage to intracranial hemorrhage was diagnosed with clinical K-SADS-PL consistency are trained using a supervised learning approach information classified as high quality. The functioning of the self was developed using a one-leave-one-subject cross (LOSO) confirmation. In this way, data from all participants (N = 42) were categorized as training dataset and converted into z-scores before performing the Davies-Bouldin Index Feature selection. This yields eight features with zero zero and unit variants that better discriminate between diagnostic groups. Therefore, 42 observations were made of these eight factors training predictive binary split models to diagnose the disease internally. Eight similar features were released, converted to scaled-up scores (e.g. mean, difference) from the training set, and used as input to the model to predict the availability of a single test subject. The diagnostic marker is set for each iteration using the amount required for the procedure required by Misdiagnosis in the ROC curve for training information. This ratio is improved to obtain the right limit, the estimate of the number of patients that need to be monitored to select one anonymous. The process was repeated 42 times until the diagnosis of each subject was predicted. The predicted diagnosis was used to calculate standard methods for performance classification including accuracy, sensitivity, and specificity. We have also integrated the area under the effective finder (ROC) curv (AUC) to comment on the general discriminatory ability of invaders. We also examined eight selected factors. It is more common in discriminating between participants and without internal interference to give an indication that lectures provide an index of basic psychopathology. A consent test used error rates (error rate = number of incorrect / complete guesses prediction value = 1 - classification accuracy) very different from what we would see a random chance. To complete this experiment, we rated distribution of possible error rates for each deviation model as a beta distribution measured by a number of negative predictions and number of observations, as is mentioned, and in a random sample 100 is possible error rates from this distribution. Next, we repeat the model with a predefined training process with 100 random

permissions for diagnostic labels, and a computer error rate for each partition. Finally, the Mann-Whitney used for experiments to detect transient fruitful phases the models for classification of error rates are quite different for the prospective opportunity for this data.

Putting results from high quality data into context, we performed a few additional analyzes. First, we trained classification models for all labeled data as high quality and we used it to guess the subject's findings from the data labeled Low quality. Outgoing predictors have been used for writing accuracy, sensitivity, specificity, and AUC in comparison results from high quality data. We checked again the use of CBCL as an internal assessment tool interference (according to the clinical K-SADS-PL consistency) in this sample using a pre-established clinic cutoffs (T score ≥ 70) for manual use and more Conservative cutoff (T score ≥ 55) was suggested for improvement screen efficiency.

2.6 SOUND FEATURES

To demonstrate the power of the proposed method of identification of children with internal disorders, we first divided each of the three-minute work of speaking into three sections, in which their boundaries were defined by the buzzer disturbance work-related. Inserting an audio signal inside each paragraph, lists the following features for each presentation peak time: length of expression, zero crossing rate (ZCR) of audio signal, frequency cepstral coefficients (MFCC), prominent frequency, mean frequency, influence mirror Centroid (PSC), spectral flatness, skew and kurtosis of Power spectral density (PSD), ZCR for the z-ratio of PSD (ZCR zPSD) for all talk times, first, second, and third formulas, and signal strength percentages above 200, 500, 700, 1000, and 2000 Hz. We have also released which means, on average, standard deviation, maximum, and minimum ZCR zPSD from sliding windows on time and frequency domains inside each head dialogue (ZCR zPSDsw). Descriptive statistics (i.e., mean, standard deviation, average, maximum, minimum) was subject to each element in each category. With us count the number of speaking days completed by patient and physician who provide a total of 164 individual features in three stages. Signal performance and feature extraction were performed in MATLAB (Mathworks, Natick, MA, USA). Many of these features have been suggested before in diagnostic and depressive diagnostic literature on adults.

3. RESULTS

The result shows whether the child is in depression or not by using four machine learning Binary classification models. They are logistic regression, Random Forest, SVM Linear Kernel, SNM Gaussian Kernel. We need to find the accuracy to evaluate a particular model which is an essential process of creating a machine learning model. So that we will know how well the model is performing. The MSE (Mean Squared error), MAE (Mean Absolute error), RMSE (Root Mean Squared error), R-SQUARED (Coefficient of Determination) metrics are mainly used to evaluate the prediction error rates and model performance.

Table 1 reports the performance measure in logistic regression binary classification model in terms of MSE, MAE, RMSE, R-SQUARED and accuracy.

TABLE -1: Performance Measure

Binary classificati-on model/ Metric	Logistic Regress-ion	Rando-m Forest	SVM Linear Kernel	SVM Gaussia-n Kernel
MSE	0.1666	0.0555	0.1666	0.3333
MAE	0.1666	0.0555	0.1666	0.3333
R-SQUARED	0.2987	0.7662	0.2987	0.4025
RMSE	0.4082	0.2357	0.4082	0.5773
ACCURACY	0.8333	0.9444	0.8333	0.6666

(MSE) Mean squared error, (MAE) Mean absolute error, (R-squared) when higher the value the better the results will be, (RMSE) Root mean square error), and Accuracy is measured using logistic regression, Random Forest, Support vector machines with Linear and Gaussian kernels models trained on data.

Logistic regression is a technique of predicting the probability of one or more independent variables, which has outcome such as 0 or 1. Mean Square Error (MSE) is calculated based on the sum of squared distances between our target variable and predicted values.

$$MSE=\frac{\sum_{i=0}^{n}(Yi - Yi^{P})^{2}}{n}$$

Where a graph is plotted, let the true value be 100, and the predicted values range between -10,000 to 10,000. The MSE loss (Y-axis) reaches its minimum value at prediction (X-axis) =100. The range is 0 to ∞. MSE is the mean $(1/n\sum_{i=0}^{n})$ of the squares of the errors $(Yi - Yi^{P})^{2}$.

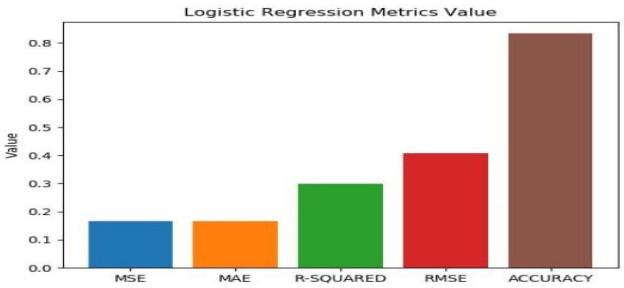

Chart-1: The graph plotted to rate the errors for logistic regression model to detect the child is in depression or not.

Random Forest is an algorithm that merges multiple decision trees into one forest. Its goal is to depend on single learning model, but there will be many decision models which will increase the accuracy. Mean absolute error it measures the errors between paired surveying the observations which are expressing the same occurrences of an event.

$$MAE=1/n\sum_{i=0}^{n}|Xi - X|$$

Where n is the number of errors, |Xi-X| is the absolute errors. Where it can be explained like this also E=Xmeasured-Xtrue.

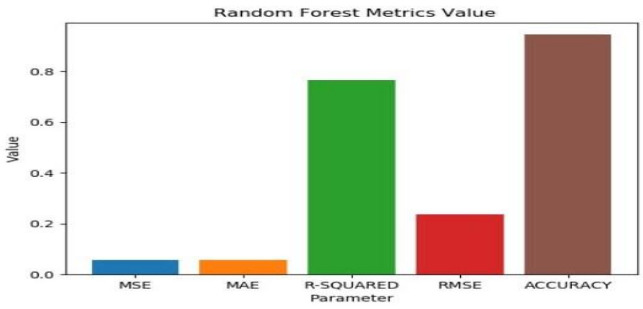

Chart-2: The graph plotted to rate the errors for Random Forest model to detect the child is in depression or not.

SVM Linear Kernel is a linear support vector machine which can be separated using a single line. Which is used when there is large number of features in a particular data set. R-Squared will indicate the percentage of variance and the independent variables. R-Squared will measure the strength of the relationship between our model and the dependent variable.

$$R^2=\frac{Variance\ explained\ by\ the\ model}{Total\ variance}$$

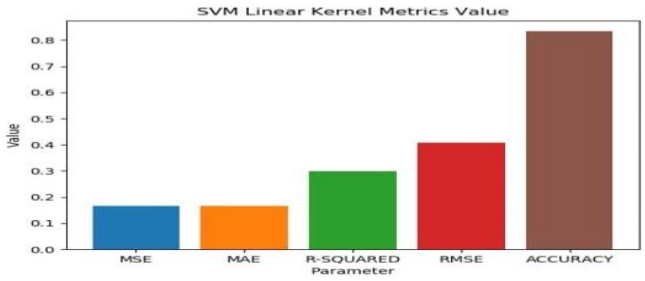

Chart-3: The graph plotted to rate the errors for SVM Linear Kernel model to detect the child is in depression or not.

SVM Gaussian kernel is a Gaussian support vector machine whose value depends on the distance from the original point or from some other point. By using distance of these original spaces we calculate the values of this distance between two points. RMSE is called as root mean squared error which measures the average immensity of the error.

$$RMSE=\sqrt{\sum_{i=0}^{n}\frac{(Yi^{\wedge}-Yi)^2}{n}}$$

Where Yi^ is the predicted values and Yi is the observed values and n is the number of observations made.

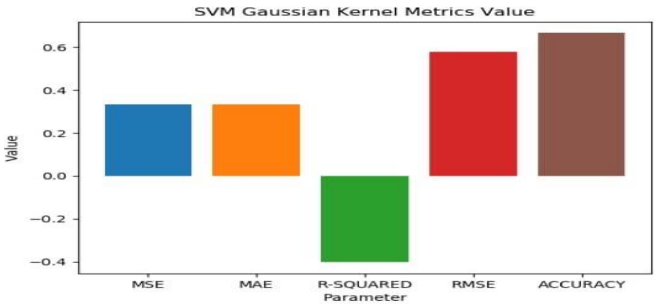

Chart-4: The graph plotted to rate the errors for SVM Gaussian Kernel model to detect the child is in depression or not.

Finally accuracy is checked that which Binary classification gives the most accurate value. In this paper accurate prediction of child depression is given by Random Forest. Hence that it will be proved that the child is in depression or not using the child's audio which is spoken for 3 minutes and further the child will be given proper treatment if the prediction is positive from any Psychiatrist where the child will be cured completely before any self-destructive happens in his/her future.

4. DISCUSSION

For identifying children suffering from internalizing disorders, we make use of audio data from three-minutes speech tasks and machine learning algorithms. The audio dataset from the child is collected using voice Audio Detector

(VAD). The dataset is then subjected to pre-processing and features from the audio are extracted. By applying machine learning algorithms, the participant is classified as depressed or normal.

There is a significant need for an objective method for screening young children with internalizing disorders. We propose the use of data from a sensor during a 90-second induction task and machine learning to fulfill this need. Herein, we take an initial step toward this goal by training classifiers for detecting early indications of internalizing diagnoses using data sampled from each of three phases of a mood induction task and discussing the implications of these results. We further examine the specific features identified as being especially indicative of an internalizing diagnosis and discuss the behaviors described by these features in the context of internalizing disorders. The proposed approach is the first step toward creating an objective method for screening children for internalizing diagnoses rapidly and at low cost.

Audio data from speech tasks are tested at 48 khz and processed via VAD that separates an audio stream into time intervals that contain speech activity and time intervals where speech is absent i.e., VAD detects the presence or absence of human speech. In reality, when data is collected from children at home, most of the recordings may contain low signal-to-noise ratios (SNRs) and can possibly be interrupted by background noise. When the energy within a sliding window is above the baseline noise, raw speech epochs are identified which includes full sentences, phrases, phonemes and high energy noise. Median filters with window length of 0.21 seconds are used to remove noise from audio signals. Each of the three-minutes speech tasks is partitioned into three phases, boundaries of which are determined using buzzer interruptions. To convert the audio signal into parameters within each phase, we compute speech epoch duration, zero crossing rate (ZCR) of the audio signal, Mel frequency cepstral coefficients (MFCC), dominant frequency, mean frequency, perceptual spectral centroid (PSC) and many other features within each phase. Logistic Regression model identifies children with internalizing disorders with accuracy of 83.33%. Support Vector Machine (SVM) model with linear kernel has the accuracy of 83.33% and that of SVM model with Gaussian kernel and Random Forest Model are 66.66% and 94.44% respectively which concludes that Random Forest is the most accurate binary classification model.

TABLE-2: Accuracy Measurement Parameters

Label	Feature	Description
F1	MSE (Mean Squared Error)	Values closer to zero are better
F2	MAE (Mean Absolute Error)	Tells us how big of an error we can expect on average
F3	R-squared parameter	Higher the value, better the accuracy
F4	RMSE (Root Mean Square Error)	Smaller the RMSE value, better the accuracy

However, this study comes with limitations. Future research must be done and it may require adapting this technique of identifying children with internalizing disorders in small screen compatible devices like mobile phones.

5. CONCLUSIONS AND FUTURE ENHANCEMENT

A detailed analysis of audio helps us to identify the internalizing disorders. The results that machine learning technique of speech analysis is able to detect the internalizing disorders in the children.

The proposed approach can be enhanced by exploiting on different platforms. In future we can use the deep learning algorithm to increase the efficiency of the prediction result.

REFERENCES

[1] H. L. Egger and A. Angold, "Common emotional and behavioral disorders in preschool children: presentation and epidemiology," J Child Psychol Psychiatry, vol. 47, no. 3–4, pp. 313– 337, Apr. 2006.

[2] World Federation of Mental Health. DEPRESSION: A Global Public Health Concern. Available at: http://www.wfmh.org/2012DOCS/WMHDay%202012%20SMALL%20FILE%20FINAL.pdf. Accessed 01/16, 2013.

[3] American Psychiatric Association. Diagnostic and statistical manual of internalizing disorders (DSM-IV TR). 4th, text. rev. ed. Washington, DC: Author; 2000.

[4] Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression and decrements in health: results from the World Health Surveys. Lancet 2007 Sep 8;370(9590):851-858.

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

[6] Brauner CB, Stephens CB. Estimating the early childhood serious emotional/behavioral disorders: challenges and recommendations. Public Health Rep 2006 MayJun;121(3):303-310.

[7] Merry SN, Hetrick SE, Cox GR, Brudevold-Iversen T, Bir JJ, McDowell H. Psychological and educational interventions for preventing depression in children. Cochrane Database Syst Rev 2011 Dec 7;(12):CD003380. doi(12):CD003380.

[8] Glied S, Neufeld A. Service system finance: implications for children with depression. Biol Psychiatry 2001 Jun 15;49(12):1128-1135.

[9] M. Tandon, E. Card eli, and J. Luby, "Internalizing Disorders in Early Childhood: A Review of Depressive and Anxiety Disorders," Child Psychiatric Clinics of North America, vol. 18, no. 3, pp. 593–610, Jul. 2009.

[10] A. C. Belden, J. Pautsch, X. Si, and E. Spitznagel, "The clinical significance of preschool depression: Impairment in functioning," Journal of Affective Disorders, vol. 112, no. 1–3, pp. 111–119, Jan. 2009.

[11] J. L. Luby, "Preschool Depression: The Importance of Identification of Depression Early in Development," Current Directions in Psychological Science, vol. 19, no. 2, pp. 91–95, May 2010 .

[12] J. Garber and K. M. Kaminski, "Laboratory and performance-based measures of depression in children and adolescents," Journal of clinical child psychology, vol. 29, no. 4, pp. 509–525, 2000.

[13] 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.

[14] D. J. Kolko and A. E. Kazdin, "Emotional/behavioral problems in clinical and nonclinical children: Correspondence among child, parent and teacher reports," Journal of Child Psychology and Psychiatry, vol. 34, no. 6, pp. 991–1006, 1993.

[15] Luby JL. Preschool Depression: The Importance of Identification of Depression Early in Development. Current Directions in Psychological Science, 201[16] Adrian M, Zeman J, Veits G. Methodological implications of the affect revolution: a 35-year review of emotion regulation assessment in children. J Exp Child Psychol, 2011.

[16] McGinnis RS, McGinnis EW, Hruschak J, Lopez-Duran N, Fitzgerald K, Rosenblum K, et al. Rapid Anxiety and Depression Diagnosis in Young Children Enabled by Wearable Sensors and Machine Learning. 2018 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Honolulu, HI; 2018.

[17] https://github.com/ranju12345/Depression-Anxiety-Facebook-page-Comments-Text

[18] Gaffrey MS, Luby JL. Kiddie Schedule for Affective Disorders and Schizophrenia- Early Childhood Version (K-SADS-EC). St Louis, MO: Washington: University School of Medicine; 2012.

[19] Lopez-Duran NL, Hajal NJ, Olson SL, Felt BT, Vazquez DM. Individual differences in cortisol responses to fear and frustration during middle childhood. J Exp Child Psychol. 2009;103: 285–295. 10.1016/j.jecp.2009.03.008.