Biosc.Biotech.Res.Comm. Special Issue Vol 13 No 14 (2020) Pp-305-309

BBRC
Bioscience Biotechnology Research Communications

Mental Illness Detection Through Audio Signal Processing

Pravin Karmore[1*], Swapnili Karmore[2] and Satyajit Uparkar[3]
[1],[3]*Assistant Professor, Department of Computer Application, Shri Ramdeobaba College of Engineering and Management, Nagpur, India*
[2]*Associate Professor, Department of Computer Science, G H Raisoni Institute of Engineering and Technology, Nagpur, India*

ABSTRACT

Mental illness is considered as great problem and sometimes it is incurable. It creates severe psychological issues as well affects the physical condition of a person badly. Causes may vary from person to person and different situations in their life. The physical and mental health in a human body goes hand in hand but if any one of them is disturbed the other automatically is devastated. The human body and mind is balanced between physical and mental world, when the harmony is disturbed it causes decease. Mental health is furthermore important in our life because the cause of mental disorder is usually undetectable. Most of the youth is mentally unstable nowadays because of the current lifestyle. Speech is an important factor for mental health. It is the physical expression of the mental state of mind. In this paper, we have discussed and developed a novel neural network model that can examine the audio signals from interview sessions to discover voice patterns that could indicate stress level. The user-generated data helps to distinguish between different disturbed groups and abnormal symptoms which can manifest in people with various mental illnesses in different ways. In particular this would automatically predict the stress level scale and differentiate disturbed mental condition from other mental disorders using the patient's psychiatric illness history and dynamic descriptions extracted from the user inputs. The proposed framework is an extension of the pre-existing frameworks, replacing the handcrafted feature extraction with the Deep feature extraction technique.

KEY WORDS: AUDIO SIGNAL PROCESSING, DEPRESSION DETECTION, DEEP LEARNING, MENTAL ILLNESS.

INTRODUCTION

Nervousness is commonly severe medical disorder. The difference between manic-depressive psychosis and major mentally disturbed episodes is the regular incidence of obsession within the latter, a state of mind with lack of confidence, discontent sleep, purposeful action, impulsivity, and enlarged activity [Cacheda F, 2019]. Each disease is a genetic disorder, and maybe well acknowledged as a hormonal imbalance to the atmosphere distressing the inner genetic circumstances and probably causing mood swings. Anxiety is related with non-continuous genetic rhythm caused by environmental annoyance like seasonal revision in hours of daylight, change of social rhythms thanks to as an example shift work or line of longitude wandering; moreover joined to lifestyles related with everyday rhythms unpredictable with the normal day to day cycle [M. Deshpande, 2017]. The looks of mental illness indications relate moreover to disturbed physical health and issues associated with it.

Medical aspect effects, community factors, and life measures, besides alcohol and matter abuse, and such factors may probably cause indication of mental illness. The world lifespan generality of disturbed mind condition based anxiety is roughly fifteen percent, however, the prevalence of episodes with an extremity level not

ARTICLE INFORMATION

*Corresponding Author: karmorepy@rknec.edu
Received 19th Oct 2020 Accepted after revision 29th Dec 2020
Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRCBA

Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal

NAAS Journal Score 2020 (4.31)

Online Contents Available at: http//www.bbrc.in/
Doi: http://dx.doi.org/10.21786/bbrc/13.14/71

meeting the wants for a depressive identification are much more. Actigraph record action of motor movement are thought of as an unbiased technique for observant depression, though this subject is far from thoroughly studied at intervals in medical specialty research. Machine learning models have been developed that can recognize words and speech intonation that could indicate depression based anxiety.

Though speech functions have been shown to be very useful in predicting anxiety and stress, following research will examine more robust classification models to brace clinical depression diagnosis [M. Deshpande, 2019]. Language analysis could also be combined with motion analysis as a technology - a supported diagnostic tool to identify children at risk of anxiety without suspecting that something is wrong. The technology would also be combined with motion analysis to help diagnose sadness in children more accurately, and would support diagnostic tools that help identify them as vulnerable. The risk of anxiety and depression before they are suspected of doing anything wrong [Islam, 2018].

Literature Survey: Researchers have developed a brand new neural network for deep learning that may establish speech patterns that indicate mental illness. In an essay to be bestowed at the Inter speech conference, the Massachusetts Institute of Technology researchers describe their methodology of discovering speech patterns that indicate anxiety, stress and sadness. This is often the primary time that such a large-scale, powerful, profound learning rule has been shared. We have a tendency to introduced gender-based vowel level analysis to push language recognition supported disturbed mindset [Cacheda Fidel, 2019]. In the first part of the experiment, Researchers have tried to validate the validity of a deep learning model for classification using EEG images. In the second approach, a metal severity model was accustomed predict mental disturbed state supported thresholds.

Researchers remove the silence from the recordings using algorithms for voice activation in the MATLAB voice box toolkit and tested it with monitored learning methods. The methodology had split into a spread of audio and connected tasks corresponding to speech recognition, speech translation to speech translation and speech recognition to speech recognition [Havigerová Jana M, 2019]. Researchers analyzed a number of different models that are developed in recent years for the automated detection of anxiety and stress level which leads mental illness. First, analysis of binary supply regression shows that speech operates contribute considerably to predicting mental illness. These results show that speech function is able to predict anxiety and shows that additional refined models for clinical identification are often developed on this basis. The researchers have develop comprehensive voice biomarkers for anxiety more accurate diagnosis of mental illness, will verify the suitable treatment choices for folks at risk. However, current ways of detective work depression are human -intense, and their results rely upon the expertise of the doctor. However, this methodology of detecting depression is human intensive and its outcome depends on the expertise of doctors [Lin Chenhao, 2020].

However, the employment of voice operates offers the simplest way to automates the detection of disturbed mind states and increase screening capacity, as voice samples and questionnaires are often crammed in. Language is clinically simple to capture and its combination with anxiety has been extensively analyzed and considered, though the particular prognostic result of speech has not been studied. However, voice data can predict depression and totally different emotions and mood, which implies that depression detected by speech function, is reliable and has potential in clinical situations. The voice is believed to possess been utilized in the past to forestall and treat disturbed mind condition [Kumar Ravi, 2020].

Although AI may ultimately play a task in treating diagnosed disturbed mind conditions, most AI research regarding mentally disturbed state is targeted on mistreatment machine learning to assist with initial identification and current monitoring. However, detective work for disturbed mind condition on-line and on social media are often a significant challenge, as there are varied hurdles to overcome, from knowledge assortment to learning the parameters of a model from scant or advanced data. Therefore, in our initial phase, we have a tendency to used applied math ways to research whether or not speech data will considerably predict disturbed mind condition. We have a tendency to train our own CNN - LSTM neural network on the very fact that current ways of mood analysis don't seem to be ready to directly infer disturbed mind condition, and thus trained it mistreatment knowledge from voice samples and questionnaires [Verma Bhanu, 2020].

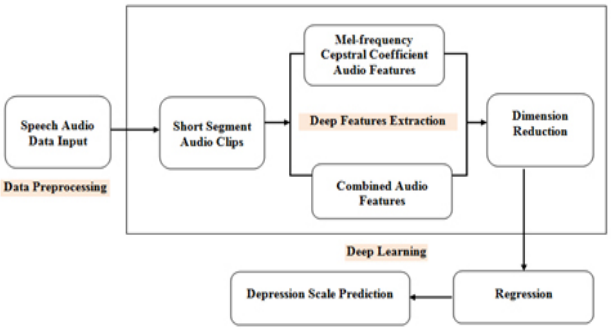
Design Methodology:

A. Dataset representing stress & anxiety: For this experiment, we have used a dataset prepared by the University of Stanford which contains more than three hundred audio clips with each audio clip containing a recorded interactive session by patient-computer of duration ranging from minimum fifty six seconds to maximum sixteen minutes [Malviya Aastik, 2020]. The dataset contains audio segments of patients ranging from age of sixteen sixty four with mean age calculated to be thirty three point five years, with a deviation of fifteen point three years. The anxiety-based BDI-II scale ranges from zero to sixty three where each range has its significance. The range of zero to ten is considered to be normal with no anxiety, range of eleven to sixteen is considered to be mild mood disturbance or stress, range of seventeen to twenty is considered as borderline clinical anxiety, range of twenty one to thirty is considered as moderate anxiety, range of thirty one to forty is severe anxiety, and over forty is extreme nervousness. The highest score recorded from the dataset for anxiety is forty seven which indicates that the dataset includes patients coming under each category [Zheng Wenbo,

2020]. The dataset undergoes audio signal processing technique using deep learning algorithms to calculate the anxiety BDI-II score if a patient.

B. System Overview: The automated audio signal processing based anxiety detection model takes visible input as audio signals from the dataset used.

Figure 1: Block Architectural diagram for Audio Signal Processing using Deep Feature extraction Process



Initially, data pre-processing techniques were carried out on the dataset as the dataset was not as per the model requirements. The audio data was damped into small audio segments and deep audio feature extraction and dimensionality reduction techniques were eventually enforced to all the audio segments for converting the input parameters into feature vectors and reducing their dimensions. Using the victimization regression technique and reducing spatiality throughout the feature vector capturing the dynamic patterns for calculating the anxiety scale assessment. From the fragmented audio segments, Mel frequency cepstral constant (MFCC) and combined audio features are extracted using deep neural networks. In the deep characteristic process, the temporal facts for each pattern are broken down into short audio segments which may be pre-processed with the help of using scaling and subtracting the given suggested segment. These segments are further pushed into deep networks for feature extraction using neural networks. Initially, the deep characteristics are extracted from the audio segments and then ranked and normalized according to the FDH set of rules into a pattern of 0s and 1s converting the output into a single row vector.

C. Audio Signal Processing: The dataset contains overall two thousand two hundred seventy eight features which are further used to investigate the most dominant feature and calculate the BDI-II anxiety level. The descriptors for audio features are provided by the University of Stanford which is further used to sort limited labelled data and calculate Mel-frequency cepstral coefficients (MFCC). The features extracted are then compared to the performance of the baseline results. The validation dataset is then tested with each feature vector where the top performing descriptors and paired with each other are kept and used. The extracted short-term and mid-term features are fed as input to support vector machine and Random forest algorithm. Short term feature matrices of fifty seconds audio segment extracts thirty four short-term features using PyAudioAnalysis.

Figure 2: Flow chart for Audio Signal Processing using Deep Feature extraction process

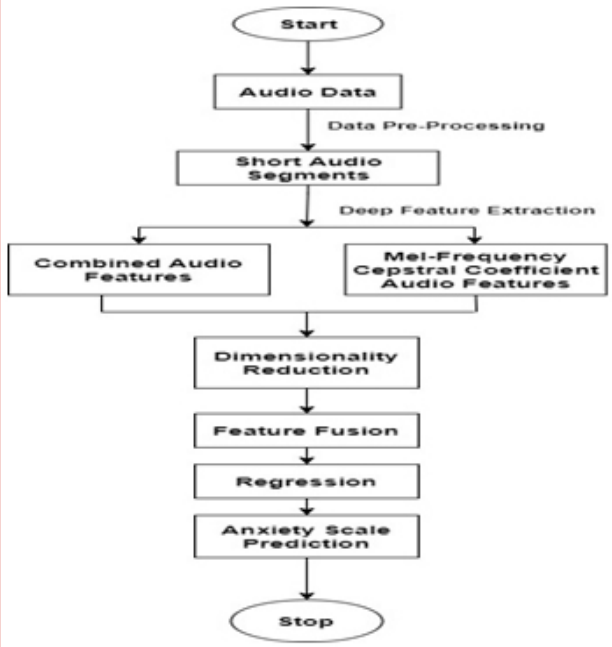
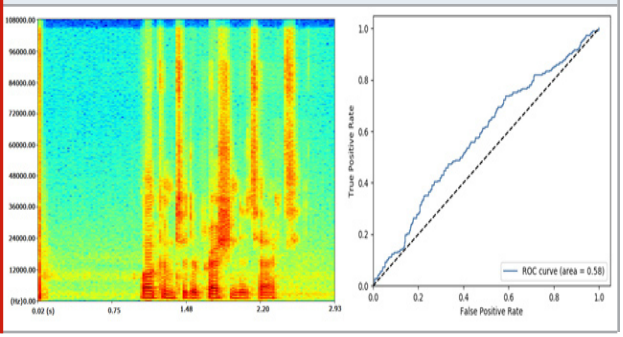


Figure 3: Spectrogram created for extracting deep features from audio segments and ROC curve plot for ResNet Architecture



RESULTS AND DISCUSSION

We performed all our experiments and tested our model using Google Collaboratory platform using NVIDIA Tesla K80 graphics card. For training the neural networks, we used PyTorch and Matlab for evaluating and loading the pre-trained network architectures by varying the training parameter values. We have used parameters such as accuracy value, precision value, F1-Score value and recall value for rating and analyzing the performance of our analysis model. We have divided the dataset for training, testing and validation. From three hundred audios, we have used two fifty audios for training the model, forty for testing and ten for validation.

The MatConvNet architecture is used for extracting deep features. This tool has been opted for the experiments because it permits full management over deep networks with access to data across any layer together with simple visualization. The experimental setup included stages like data preparation, data pre-processing, convolutional neural network training and fine tuning of procedures to obtain accurate results for calculating the anxiety of a patient. We have evaluated five convolutional neural network ResNet architectures like ResNet-18, 34, 50,101,152 in the deep feature extraction process and checked the impact over classification results by generating high resolution based spectrograms. After experimenting on larger input spectrograms we found that they significantly do not improve the computation results. We confirmed that generating a group of larger input segments wouldn't considerably improve the results. We tend to additionally perform check Time Augmentation (TTA).

Table 1. Table for Performance of Deep net Feature on Development set and Test set on dataset

Partition	Methods	Segment type	RMSE Score	MAE Score
Training		Waveform	9.2589	7.6549
		Spectrogram	9.0124	7.8523
Testing	Deep Learning	Waveform	10.2365	8.3698
		Spectrogram	9.4589	8.1278

Table 2. Table for the Classification Report

	Precision	Recall	Support	F1-score
True	0.98	0.80	5	0.89
False	0.91	0.97	10	0.95
Average	0.94	0.93	15	0.93

The Test time argumentation created predictions supporting the initial spectrogram and waveform from our dataset with four partitions of it helped in improving the model accuracy and precision. The calculated mean prognosis from all the segments is eighty seven point six percent. The planned technique created a promising classification accuracy of around eighty percent for a ResNet-34 model and eighty four point two percent for ResNet-50 model trains on spectrograms of 224X224 pixels. During the training stage waveform, RMSE and MAE values were 9.2589 and 7.6549 respectively which were low as compared to other models stating that the precision and accuracy are low. The average precision value for the model was found to be ninety-four percent, accuracy as ninety three percent, F1-score as ninety three percent and support as fifteen. According to the performance metrics ResNet-50 has performed more accurately as it was successfully able to categorize two thirty three voice samples of patients correctly into the scale and failed to categorize.

CONCLUSION

Healthy Mind is furthermore important in our life. Mental imbalance causes severe problems which was difficult to diagnose as well as difficult to cure. After performing deep literature survey, it is found that speech and voice are great Medias to measure its level. In this paper, we have developed an Artificial Intelligent based model for automated anxiety and stress scale prediction based on vocal expressions in recorded video clips using audio feature extraction to get precise results with deep features extracted like combined audio features and MFCCs as this methodology is much beneficial and accurate. The audio clip is extracted from the video dataset and converted in small audio segments. The extracted features later undergo dimensionality reduction and using regression the BDI-II scale is calculated for anxiety. The overall result for the anxiety prediction model on the testing partition show more accuracy and precision than the baseline, and performs much better than other pre-existing models. The precision was found to be ninety-one percent, recall as ninety seven percent, F1-score as ninety three percent and support as fifteen as a classification model.

REFERENCES

Cacheda F, Fernandez D, Novoa FJ, Carneiro V, "Early Detection of Depression: Social Network Analysis and Random Forest Techniques", J Med Internet Res 2019; 21(6):e12554.

C. E. Granger and A. Hadid, "Depression Detection Based on Deep Distribution Learning", IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp4544-4548, doi: 10.1109/ICIP.2019.8803467.

Cacheda Fidel et al. "Early Detection of Depression: Social Network Analysis and Random Forest Techniques.", Journal of medical Internet research vol. 21, 6 e12554.

10 Jun. 2019, doi:10.2196/12554.
Havigerová Jana M., Haviger Jirí, Kucera Dalibor, Hoffmannová Petra, "Text-Based Detection of the Risk of Depression", Frontiers in Psychology Volume 10, 2019, page 513.
Islam, Md Rafiqul & Kabir, Ashad & Ahmed, Ashir & Kamal, Abu & Wang, Hua & Ulhaq, Anwaar, "Depression detection from social network data using machine learning techniques", 2018 Health Information Science and Systems. 6. 8. 10.1007/s13755-018-0046-0.
Kumar, Ravi & Nagar, Santosh & Shrivastava, Anurag, "A Review on Depression Detection Among Adolescent by Face", 2020, Smart Moves Journal IJOScience, 6. 4. 10.24113/ ijoscience.v6i1.257.
Lin, Chenhao & Hu, Pengwei & Su, Hui & Li, Shaochun & Mei, Jing & Zhou, Jie & Leung, Henry, "SenseMood: Depression Detection on Social Media", 2020, pp407-411, 10.1145/3372278.3391932.
M. Deshpande and V. Rao, "Depression detection using emotion artificial intelligence", International Conference on Intelligent Sustainable Systems (ICISS), Palladam, 2017, pp858-862, doi: 10.1109/ISS1.2017.8389299.
Malviya, Aastik & Meharkure, Rahul & Narsinghani, Rohan & Sheth, Viraj, "Depression Detection through Speech Analysis: A Survey", 2020, International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp712-716, 10.32628/CSEIT1952190.
Verma, Bhanu & Gupta, Sonam & Goel, Lipika, "A Neural Network Based Hybrid Model for Depression Detection in Twitter", 2020, 10.1007/978-981-15-6634-9_16.
Zheng, Wenbo & Yan, Lan & Gou, Chao, "Graph Attention Model Embedded With Multi-Modal Knowledge for Depression Detection", 2020, 1-6. 10.1109/ICME46284.2020.9102872.