International Journal of Scientific Research & Engineering Trends
Volume 7, Issue 6, Nov-Dec-2021, ISSN (Online): 2395-566X

# Automated Mental Illness Analysis Using Voice Samples

**Sowhardh Honnappa Gowda**
**Illinois Institute of Technology,**
**Chicago**

**Abstract- One in 24 suffers from critical mental illness like Schizophrenia, Psychosis, Clinical Depression, Anxiety Disorder,Obsessive Compulsion Disorder (OCD), Autism, Bipolar Disorder, Attention Deficit Hyperactivity Disorder (ADHD) etc. found that the average Vector Similarity between adjacent sentences in free speech, along with other variables like Number of words/phrases, pauses, tone, intensity, frequency and other Low- Level Descriptors form the raw audio recording could be used to identify clinically high-risk patients with great accuracy. Audio visual hallucination and thought insertion appear to be the top side effects in case of patients suffering from Schizophrenia [3]. Acoustic studies between healthy and depressed individuals [4] shows us that the top audio features which help identify depression in mental illnesses are Loudness, MFCC5 and MFCC7. One of the studies dealing with "Automated Depression detection using Audio features" [5], suggests that the lacking objective Clinical depression assessment methods is the key reason that several patients can't be treated appropriately on time. This study aims to find an optimal approach to calculate depression scores amongst people suffering from mental illnesses using Artificial Intelligence techniques.**

**Keywords- Depression, Mental Disorder, Audio Analysis, PHQ8, Artificial Intelligence, Deep Learning**

## I. INTRODUCTION

According to the American Psychiatric Association, "Mental Illness" is a health condition which involves changes in behavior or thinking and constantly fluctuating emotions [11]. Nearly 19% of the U.S. adults experience mental illnesses (over 47 million). One in 24 suffers from critical mental illness like Schizophrenia, Psychosis, Clinical Depression, Anxiety Disorder, Obsessive Compulsion Disorder (OCD), Autism, Bipolar Disorder, Attention Deficit Hyperactivity Disorder (ADHD) etc.

Amongst these, Schizophrenia is one of the most severe mental illnesses affecting over 20 million people worldwide. Human beings suffering from Schizophrenia are 2-3 times more likely to die sooner than other individuals [11]. The most common syndrome accounting to all of the mental illness is "Depression", which leads to eventual mortality. Mental illnesses like Schizophrenia, Psychosis takes away an individual's ability to think and feel [2]. Early intervention and detection can help stop the decline in Cognitive functioning of individuals suffering from mental illnesses [2]. The onset of mental illnesses like Schizophrenia, Psychosis has a prodromal phase which includes abnormalities in thought, perception and communication. The problem is that, detecting signs of mental illnesses at an early stage proves to be challenging as the symptoms are very subtle [10].

Digital Phenotyping - Elvesiag et al. observed that "Cosine Similarity" between adjacent words used by patients suffering from Schizophrenia was a lower value when compared to healthy controls [10] Bedi et al. found that the average Vector Similarity between adjacent sentences in free speech, along with other variables like Number of words/phrases, pauses, tone, intensity, frequency and other Low- Level Descriptors form the raw audio recording could be used to identify clinically high-risk patients with great accuracy. Audio visual hallucination and thought insertion appear to be the top side effects in case of patients suffering from Schizophrenia [3]. Acoustic studies between healthy and depressed individuals [4] shows us that the top audio features which help identify depression in mental illnesses are Loudness, MFCC5 and MFCC7. The research also states that depressed voices have a slow, monotonous and disfluent speech pattern [3].

One of the studies dealing with "Automated Depression detection using Audio features" [5], suggests that the lacking objective Clinical depression assessment methods is the key reason that several patients can't be treated appropriately on time. It goes on to add that, Therapy does not solve depression and the mental illness ends up staying in the patient in the form of 1. Excessive Sleep, 2. Fatigue, 3. Loss of Energy. These symptoms are easy to stay unidentified.

According to World Health Organization (WHO), 350 million people suffer from mental illness globally which

is an alarming insight. Cohn et al [12] states that "Building an automatic Depression analysis system is possible". This will greatly benefit the Clinical Theory and practice. There has been another research which studies "Depression analysis" in mental illnesses using Visual and Vocal features. It has found out that human cognitive systems can be stimulated by Artificial Intelligence, which can also rate depression on standard assessment scales like Beck's Depression Inventory form vocal and visual expressions (BDI-II), PHQ-8, PHQ-9 etc. This approach uses Facial Action Coding System (FACS) which analyses the face emotions using combination of Facial muscles and ECG (Echocardiogram), EDA (Electro dermal activity) of patients for depression analysis [6].

Another research has developed automated depression analysis systems using "Deep learned features" [7] using Convolutional Neural Networks (CNN) and Deep Convolutional Neural Networks (DCNN). The Deep Convolutional Neural Network is built to learn deep learned features from spectrogram and raw speech, then manually extract texture descriptors, "Median Robust Extended

Local Binary Patterns (MRELBP)" from the spectrogram of input raw audio. There are a wide variety of depression diagnostic methods used by clinicians such as: -

- Diagnostic & Statistical Manual of Mental Disorder (DSM - IV)
- The Quick Inventory of Depressive symptoms – Self report (QIDS)
- Beck's Depression Inventory (BDI)
- The 8-item PHQ-8 [Patient Health Questionnaire]
- The 9-item PHQ-9 [Patient Health Questionnaire]

Automated depression analysis using branches of AI like Machine Learning, Natural Language Processing, and Deep Learning can facilitate automated scoring on standard depression measurement scales. Although we have a number of approaches and research conducted on automated depression analysis for mental health using Machine Learning, Deep Learning and Open Source tools for audio analysis, one of the biggest missing pieces in the puzzle is the utilization of Natural Language Processing approach in combination with Machine learning and Deep learning methodologies like "Latent Semantic Density"/ "Idea Density", "Semantic Coherence" to develop efficient and high performance Depression analysis system [9].

Latent Semantic Density of sentences can be measured by breaking the components of a sentence into meaning vectors using a technique called "Vector Unpacking". This gives us an opportunity to explore the linguistic markers which can help identify depression at an early stage. In patients suffering from mental illnesses, the semantic density/sentence richness/idea density is usually low when compared to normal patients.

The approach is to develop an extremely intelligent and accurate Artificial Intelligence eco system which will be capable enough to process audio and video components of Physical Interview recordings of patients suffering from Schizophrenia, process the data based on an ensemble approach of Machine Learning + Deep Learning + Natural Language Processing methodologies and possess the ability to provide Depression scores on standard Depression scales. The future work of this research would be to assist the users on appropriate actions like Therapy, Books to read, Podcast, Consultation etc. to users based on the extent/ severity of depression. The tool can be utilized for children's mental health analysis at Schools, it can be used at Organizations to monitor employee's mental health on a regular basis, the tool can also be provided as an Open-Source tool to the Veterans, Third World countries and so on. The applications for such a mental health assessment tool seem to be endless.

**Why is this Research worth doing?**

Mental Illnesses such as Schizophrenia, Psychosis etc. evolve over 1-5 years before the first contact with Mental Health Services [8]. The changes in mood can be the only premonitory sign of an imminent mental disorder for years [13]. The Psychopathological phenomena that can be used for early detection include: 1. Characteristic Prodromal Signs & Symptoms 2. Neuropsychological Deficits 3. Characteristics of Illness course.

There are two opportunities of early detection before the condition of Schizophrenia grows worse. They are: -

- Prepsychotic Prodromal Stage: This is the phase from the first sign of Schizophrenia until the first psychotic symptom. It lasts approximately for 4.8 years.
- Psychotic Prephase: This stage is from the first positive symptom until first admission of the patient (Mean duration of 1.3 years)

In about 75% of the cases, Schizophrenia onset occurs with slowly mounting depressive and negative symptoms involving increasing functional impairment and Cognitive dysfunction [14]. An awareness program educating and alerting both the population and health services in Norway helped reduce the undetected Psychosis treatment from 2.5 years to 1.5 years [15]. McGlashan and Johannessen suggest that the "Plasticity of brain" can be preserved by both antipsychotic medication and social stimulation at a sensitive stage [16]. The administration of proper dozes of Olanzapine can halt the process of toxic brain damage. Some of the early signs of Schizophrenia include Restlessness, Depression, Anxiety, Trouble with Thinking, Lack of energy, Lack of self-confidence, poor work performance, social withdrawal etc. leading to toxic brain damage.

## II. LITERATURE REVIEW

Some of the earlier research [1] has made use of Audio Depression Regression Model (DR AudioNet). Based on Convolutional Neural Networks. A long short-term network (LSTM) to identify the prevalence of depression in patients. Multiscale Audio Differential Normalization (MADN) for feature extraction. Depression which is one of the major symptoms of Schizophrenia can be cured with psychotherapy, physiotherapy and early medication. Due to inefficient precision of the current AI methodologies, patients are not treated right at an early stage with appropriate intervention [1].

There are a few challenges with early intervention. They are: -

- It is difficult to distinguish between Depression and Bipolar Disorder. As most of the mental illnesses have similar symptoms, it causes a difficulty to appropriately predict the right disease at an early stage.
- Privacy of patients during preliminary Diagnostic screening is a biggest concern. A lot of patients suffering from mental illnesses shy away from physician interview conducted to assess their mental health.

For this study [1], the speech signal characteristics of 2014AVEC dataset was extracted using the following methodologies. Feature Extraction – MADN method was used for speech feature extraction. Since the extracted conventional speech features include the sample's own personalized speaking features, it results in poor generalization of the trained model.

MADN technique avoids this issue. In order to identify the degree of depression – DR AudioNet was the model used for this purpose. The list if Machine Learning and Deep learning algorithms used in this approach is provided here [17] [18].

The detailed process of depression detection is as follows:

- Perform pre-processing and feature extraction on original data. Extracted features include MFCCs, Zero Crossing Rate, Energy and other Low-Level Descriptors.
- Front end is built for data collection and processing is performed using Convolutional Neural Networks and Deep Convolutional Neural Networks.
- Results are displayed on the screen indicating the Depression score on the standard depression scale.

Challenges of this approach include the following:

- Effectiveness of voice feature selection – Different people have different speech characteristics such as Tone, Loudness, and Energy of voice signals. The most effective speech feature is yet to be found as it might be different on different datasets.
- Accuracy varies due to different data characteristics, experimental algorithms, recording hardware equipment conditions etc. The experimental conditions should be continuously improved for better results.
- Datasets for depression analysis – At present, there is no unified standard dataset as the speech source material for depression recognition. Along with this, there are only a few Open- Source datasets available.
- Processing using AudioNet – The input to AudioNet is voice data. The construction process of speech feature data uses the following approach: 1. Extract MFCCs 2. Zero-Crossing Rate 3. Energy
- Other features from each frame of input audio spectrogram.

Select 60 consecutive frames of speech characteristics in each speech sample to construct a 2D matrix (X-axis = Time, Y-axis = Frequency). The Convolutional layer extracts semantic information. The role of the pooling layers is to reduce the dimension of features. The 2D matrix is converted into 1D data after passing through the Convolutional layer and pooling layer. The LSTM layer is used to extract long-term dependency information. The Fully connected layer is sued to encode changes in speech on the X-axis and give predictions of Depression scores. According to this research [8] suggests early intervention can occur but with significant false positives.

There are 2 approaches:

- Enhancing the risk by assessing characteristic premonitory symptoms at a later stage of prodrome.
- Considering pre-morbid risk factors
- A loss of 30 points or more in the Global Assessment of Functioning (GAF) score in 6 weeks along with a family history of Schizophrenia.

The common instruments for basic assessments of mental illnesses are: -

- Bonn Scale of Assessment for basic symptoms (BSABS)
- Assessment Instruments [19]
- Screening Instruments – PROD Screen, developed by Heinimaa et al [20] in Finland on the basis of structured interview for prodromal symptoms (SIPS).

In the paper "Automated Depression Analysis using CNN" [7], a combination of hand crafted and deep learned features are used to measure Depression from speech recordings. According to World Health Organization (WHO), "Depression" is the 4th most mental disorder as of 2020. The research aims at providing depression scores on Hamilton Scale of measurement (HAMD). The datasets used for the study are AVEC 2013 and AVEC 2014. The research is based on the assumption that voice patterns in speech have a close relation with Emotion and Stress [8].

The analysis of voice patterns can be classified into 3 groups:

- Prosodiac

- Vocal Tract
- Glottal Source.

Hand crafted features have proven to provide better accuracy. But it requires a lot of time, effort, domain knowledge etc.
From the input audio samples, various tools are available to extract Low Level Descriptors from raw input audio such as: Open SMILE, COVAREP, SPTK, KALDI, YAAFE, Open EAR etc. Deep learning methods have several models such as Single Layer Learning models, Probabilistic models, Auto- Encoders, Convolutional Neural Networks (CNN) etc. CNN has been widely used to perform State of the Art performance. Convolutional Neural Networks matched the State of the art for dataset with macroscopic and microscopic images.

The Dep learned features from Spectrogram has not yet been explored. In this approach, the author suggests using CNN in learning spectrogram patterns from speech. It can be a Regression/ Classification problem. Various depression recognition approaches have been suggested in "Depression Recognition sub challenges". Regression methods have been developed on AVEC 2013 and AVEC 2014 datasets and Classification methods developed on AVEC 2016 and AVEC 2017. For AVEC 2013, researchers have used Open EAR tool.

In AVEC 2013 Depression challenge, combination of Eigen value Spectra and co-ordination features to analyze the relationship between vocal behavior and depression scale [21]. With the Co-ordination and phoneme rate-based features, the researchers designed a Gaussian Staircase Regression system to predict Depression score on BDA-II. Principle Component Analysis (PCA) was also used for dimensionality reduction. The model performance metrics were RMSE and MAE which stood at 7.42 and 5.75 respectively.

Moore et al. explored prosodics, vocal tract and parameters extracted directly from glottal waveform to discriminate the depressed speech from a normal one. Extracted about 200 prosodics, vocal tract and glottal waveform measures from the depression database [22].
Nicholas et al. provided a comprehensive and exhaustive conclusion about depression. They reviewed the important characteristics of paralinguistic speech affected by depression and suicide [23].

A number of audio features such as
- Estimated Articulatory trajectories during speech production
- Acoustic characteristics
- Acoustic-phonetic characteristics
- Prosodic features were explored.

Different models were run to predict the Back's depression rating scale (Models – Support Vector Regression (SVR), Gaussian Backend, Decision Trees etc.) [24]

Williamson et al fused different feature domain to obtain a better performance. The researchers combined Gaussian Staircase Regression with extreme machine learning classifiers (ECM) and obtained a test RMSE of 8.12. There were a couple approaches used for Feature extraction,
- Hand crafted Feature extraction
- Deep Learning based feature extraction [25].

**1. Hand crafted Feature Extraction:**
In this approach, 2 kinds of descriptors were adopted. (1. Median Robust Extended Local Binary Patterns (MRELBP) and 2. Audio features extracted from Open SMILE toolkit). There were about 2268 baseline audio features of AVEC 2013 and AVEC 2014 with up to 40 Low Level Descriptive features like Spectral, Cepstral, Prosodic, Voice Quality information etc. Processing the audio files included overlapping fixed length segments shifting forward at a rate of 1 second, size of window was set to 20 seconds, which can capture slow- changing long range characteristics.

**2. Deep Learning based feature extraction:**
These features were extracted from 2 models: 1. The first deep network extracts deep learned audio features from frame level raw waveforms 2. The other deep network directly learns feature representations from spectrogram images. Audio features and Texture features were extracted from this network.

**2.1 Deep learned Audio Features –** Frame level raw waveform was fed into the first CNN Convolutional layer to learn a filter bank representation which was equivalent to filter kernels in a time- frequency representation. In this methos, if the raw waveform is filtered by the first stride of Convolutional layer, the output feature map will have the same features as the spectrogram

**2.2 Deep learned Texture Features –** CNN has a few limitations such as 1. It cannot process high resolution images 2. It requires a lot of samples for training. In order to extract vocal patterns from CNN, 6 second and 20 second audio segments were considered.

## III. DATA AUGMENTATION

Audio features were extracted from Frequency domain of spectrogram
The spectrogram image sequences were horizontally flipped/rotated (-15 Degree, -10 Degree, -5 Degree, 5 Degree, 10 Degree, 15 Degree).

The data augmentation process is necessary in audio analysis due to the limited number of available audio datasets for the analysis of mental illnesses. The Data

Augmentation approach increases the input data (Original data + Flipped Images + Rotated images with 6 angles) which results in an 42X increase in the training sample size. Depression Score Calculation is considered as a Regression problem. Therefore, the Euclidean Loss was used as the loss function.
Research was conducted on using AI to mark depression on Beck's Depression Inventory (BDI-II) from vocal and visual expressions. AVEC2014 was the input depression dataset.

This research focused on the following features:

- Facial Features obtained from Deep Learning techniques
- Spectral Low-Level Descriptors and MFCCs from raw audio
- Feature Dynamic History Histogram which was used to capture temporal movement on the feature space.

Some of the pre-trained solutions like Alex Net, VCG Network (trained on millions of images) were used in the study for Facial emotion analysis. Partial Least Squares (PLS) and Linear Regression (LR) algorithms were used to model the mapping between dynamic features and depression scale [6].

Visual Feature Extraction approaches: Hand crafted Image feature extraction. This involves 3 different textual features: Local Binary Patterns (LBP), Edge Orientation Histogram (EOH), and Local Phase Quantization (LPQ). VCG Face pre-trained deep models like VCG-S, VCG-M, VCG-F (Slow, medium, fast), VCG-D and VCG-E were tested out on ImageNet data for object classification task.
In the research paper, "Automated Depression Analysis using Audio Features" a combination of "f0" and switching pauses were used to predict depression scores. Cannizzaro found that a change in the severity of depression covaried with vocal prosody.

As per the study, depression is heavily associated with Neuroticism, Introversion and Conscientiousness. Also, Interpersonal Gap span and speaking rate are closely associated to changes in the depression severity. Data Processing – The Audio is cut into 15 second samples using Python and SOX to capture 2-3 sentences per recording. The features extracted from raw audio files are Cepstral, Glottal and Spectral. Some of the other inspiring voice features identified in the research are pitch, loudness, speaking rate, rhythm, voice quality, articulation, sentence length, intonation, fundamental frequency and MFCCs.

The audio pre- processing includes Silence Removal, Speaker Diarization and Audio Thumbnailing. In the research, Feature extraction includes extracting short term and mid-term audio features (MFCC, Chroma Vectors, Zero Crossing Rate etc.) and serve them as inputs to the classification algorithm.

One of the major challenges of this approach is that these audio features are lower-level representations of audio, but the complex speech features of depression might go undetected. Each speech stimulus is represented as Spectrogram. Spectrogram maintains a high level of detail including noise which can present challenges to neural network learning [5].

As per the research conducted in "Acoustic differences between healthy and depressed people", which examines the acoustic difference between healthy and depressed individuals, focus is placed on the significance of acoustic features like Statistical significance and magnitude of effect size.

As per the paper, there were 3 acoustic features which stood out: Loudness, MFCC5 and MFCC7. The study states, depressed people have less vocal tract compared to healthy controls. This symptom is termed "Psychomotor Retardation". Also, findings state that the Mean of MFCC5, MFCC7, and Loudness were high for healthy people when compared to folks suffering from mental illnesses like Schizophrenia. The decrease in MFCCs in depressed people is an outcome derived from the reduction of neural responses in "Inferior Frontal Gyrus" resulting in less speech motor [4].

## IV. RESEARCH METHODOLOGY

Below is the research methodology which I would like to implement during the course of my PhD.

The first step which is the Data collection process for my research would be collecting the following depression and mental illnesses recordings.

- AVEC 2013, AVEC 2014, AVEC 2016 depression datasets – These are depression dataset provided as part of Audio/Visual emotion and depression recognition challenge.
- DAIC-WOZ database – This is a corpus of Distress Analysis Interview Corpus (DAIC). It contains clinical interviews designed for mental illnesses like Anxiety, Depression, and Post-traumatic stress disorder. The dataset contains about 189 sessions ranging between 7 – 33 minutes.
- Depression dataset (Open-Source Datasets)

## V. DATA IMPUTATION & DATA PRE-PROCESSING TASKS

Due to the limited availability of datasets, Data Imputation comes in very handy for this research. Following are the steps which can be implemented to increase the training data.

- Split audio into 10 second segments containing 2-3 sentences.

- 2 second overlaps between audio snippets during Speech transcription to make sure, words are not missed out.
- Sound Diarization using Google's UIS-RNN for the initial proof of concept. Implementing a post processing approach for efficient Sound Diarization of audio recordings.
- Audio Pre-Processing techniques like background noise removal, silence removal, audio format conversions, resampling- dimensionality reduction etc.

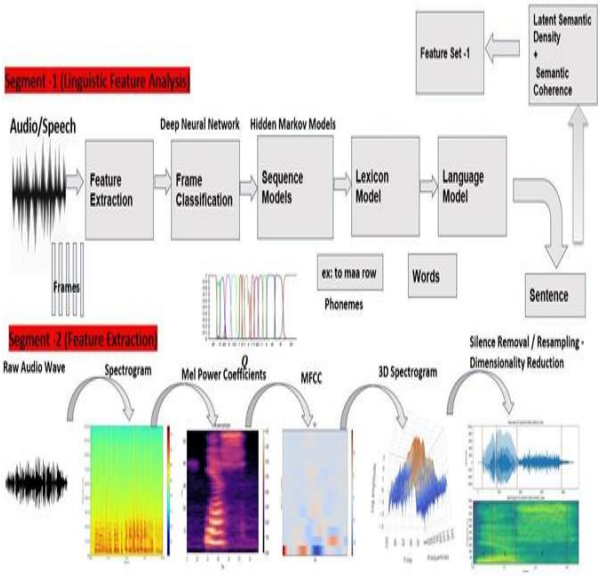


Fig 1. Component -1 (Audio Feature Extraction + NLP Tasks).

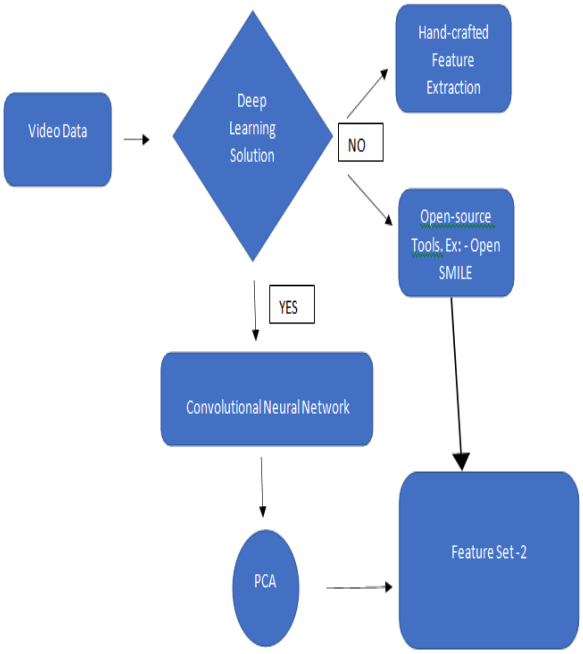


Fig 2. Component -2 (Face Emotion Analysis from Videos).

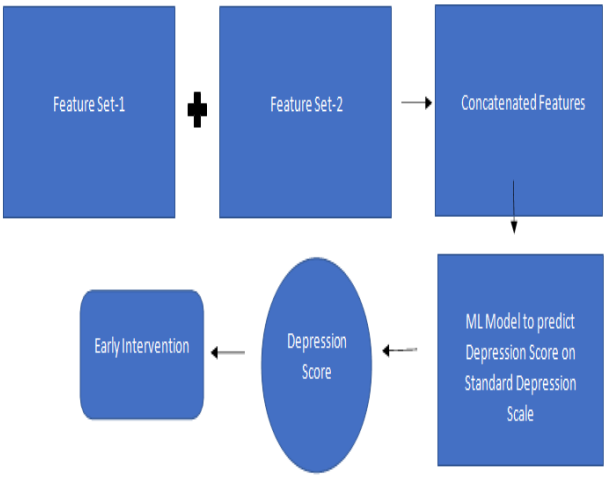


Fig 3. Component -3 (Feature Concatenation and Depression Score Calculation).

## VI. METHODOLOGY

The proposed methodology includes 3 components:

- Feature Extraction & Linguistic Feature Analysis from Audio
- Face Emotion Analysis from Video
- Concatenation of Features and Machine Learning algorithm to predict Depression scores.

As evident from the above architecture, the approach is an ensemble of Machine Learning, Deep Learning and Natural Language Processing, which is an unexplored solution and has the potential to give out results of extreme accuracy which can be practically put into use in real world scenarios.

The linguistic marker, "Semantic Density" can be obtained using the mathematical method of "Vector Unpacking", which decomposes the meaning of sentences into ideas. Similarly, "Latent Semantic content" can be obtained by contrasting it with contents of conversations generated online. Low semantic density and talks about Voice, sounds, thought insertions are signs of Schizophrenia. With the advancement of technology lead by discoveries in Machine Learning, Natural Language Processing and amplifications of linguistic and behavioral features can be used to construct a "Digital Phenotype", which is a characterization of an individual's knowledge representation and thought process. This supports the theory that features of Natural Language can be mined to predict the onset of mental illnesses like Schizophrenia, Psychosis etc.

One of the critical features for" Digital Phenotype" is Poverty of content/ Low semantic density for candidates with mental illnesses. Auditory hallucination is a core symptom of Schizophrenia, Psychosis. Full hallucination appears in later stage of the illness. In order to account for this, "Topic Modelling can be performed on the extracted

transcription from audio" to identify words relating to voice/sounds and creating a rating scale (0 - 100) for such words which will later be accounted as a feature in prediction of Depression score. In order to derive meanings from sentences, "Nouns, Verbs, Adjectives and Adverbs" can be expressed as Word embeddings.

The word embeddings map the words of a language into vector space of reduced dimensionality and can be created by Skipgram of Word2Vec. So, essentially words that occur in similar context end up having similar embeddings [9].

The word embeddings associated with the words in a sentence when summed produce a resultant sentence vector. Meaning vectors are identified through learning of weights. Patients suffering from mental illnesses are prone to use unrelated semantic in the sentences. This reduces the number of Meaning vectors required to create the sentence. As the meaning vectors are lesser than content words, it results in a reduction in Semantic Density indicating the presence of mental illness.

**Semantic Density = (Number of Meaning Vectors/ Number of Total words).**

The 3 components shown above is the overall approach to deal with Audio and Video interviews of patients suffering from mental illnesses like Schizophrenia to arrive at the depression score on a standard depression scale.

## VII. CONCLUSION

While I worked with Cognizant in the US, I had an opportunity to work for Otsuka Pharmaceuticals, New Jersey. Project: The problem statement at Otsuka Pharmaceuticals was detection and measurement of depression in patients suffering from mental illnesses (Schizophrenia, Psychosis etc.) and scoring the voice samples on the standard Depression scale (PHQ-8).

Methodology: I had obtained Distress Analysis Interview Corpus from the University of Southern California, used it as a training set for the ML Model. I built an ensemble model trained on Data points from Audio Signal + Data Points from Video. The Output was a Depression Score on PHQ 8 scale (0 - 24).

## REFERENCES

[1] A Novel Smart Depression Recognition Method Using Human Computer Interaction System, Lijun Xu, 1 Jianjun Hou,1 and Jun Gao2 1 Institute of Art and Design, Nanjing Institute of Technology, Nanjing, Jiangsu 211167, China 2 Siemens Ltd., China Jiangsu Branch Co., Ltd., Nanjing, Jiangsu 211100, China.

[2] Adult mental health disorders and their age at onset, P. B. Jones, The British Journal of Psychiatry (2013) 202, s5–s10. doi: 10.1192/bjp.bp.112.119164.

[3] Accounting for the phenomenology and varieties of auditory verbal hallucination within a predictive processing framework, Sam Wilkinson, Department of Philosophy, Durham University, 50 Old Elvet, Durham DH1 3HN, UK.

[4] Acoustic differences between healthy and depressed people: a cross-situation study, Jingying Wang1, Lei Zhang2, Tianli Liu3, Wei Pan1, Bin Hu4 and Tingshao Zhu, Wang et al. BMC Psychiatry (2019) 19:300 https://doi.org/10.1186/s12888-019-2300-7.

[5] Automated Depression Detection using Audio Features, Suraj G. Shinde1, Atul C. Tambe2, Avakash Vishwakarma3, Sonali N. Mhatre4, International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 05 | May 2020.

[6] Artificial Intelligent System for Automatic Depression Level Analysis through Visual and Vocal Expressions, Asim Jan, Hongying Meng, Yona Falinie A. Gaus, and Fan Zhang Automated depression analysis using convolutional neural networks from speech.

[7] Lang Hea, ∗ , Cui Caob, a NPU-VUB joint AVSP Research Lab, School of Computer Science, Northwestern Polytechnical University (NPU),.

[8] Xi'an, China b Moscow Institute of Arts, Weinan Normal University, Weinan, China, , Journal of Biomedical Informatics, Volume 83, July 2018, Pages 103-111

[9] Early detection of schizophrenia: current evidence and future perspectives, HEINZ HÄFNER, KURT MAURER Schizophrenia Research Unit, Central Institute of Mental Health, J5, D-68159 Mannheim, Germany.

[10] A machine learning approach to predicting psychosis using semantic density and latent content analysis, Neguine Rezaii1,2, Elaine Walker3 and Phillip Wolff3, npj Schizophrenia (2019) 5:9; https://doi.org/10.1038/s41537-019-0077-9.

[11] Jones P, Rodgers B, Murray R, Marmot M. Child development risk factors for adult schizophrenia in the British 1946 birth cohort. Lancet 1994; 344: 1398–402. https://www.who.int/news-room/fact-sheets/detail/schizophrenia

[12] Aan, H. Meng, Y. F. B. A. Gaus and F. Zhang, "Artificial Intelligent System for Automatic Depression Level Analysis Through Visual and Vocal Expressions, "in IEEE Transactions on Cognitive and Developmental Systems, vol. 10, no. 3, pp. 668-680, Sept. 2018.

[13] Kraepelin E. Psychiatrie. Ein Lehrbuch für Studierende und Aerzte, 5th ed. Leipzig: Barth, 1986.

[14] Häfner H, Maurer K, Trendler G et al. Schizophrenia and depression: challenging the paradigm of two separate diseases - A controlled study of

schizophrenia, depression and healthy controls. Schizophr Res 2005; 77:11-24

[15] Johannessen JO, Larsen TK, Horneland M et al. The TIPS Project. A systematized program to reduce duration of untreated psychosis in first episode psychosis. In: Miller T, Mednick SA, McGlashan TH et al (eds). Early intervention in psychotic disorders. Dordrecht: Kluwer, 2001:151-66.

[16] McGlashan TH, Johannessen JO. Early detection and intervention with schizophrenia: rationale. Schizophr Bull 1996; 22:201-22.

[17] H. Jiang, B. Hu, Z. Liu et al., "Investigation of different speech types and emotions for detecting depression using different classifiers," Speech Communication, vol. 90, pp. 39–46, 2017.

[18] S. Gao, V. D. Calhoun, and J. Sui, "Machine learning in major depression: from classification to treatment outcome prediction," CNS neuroscience & therapeutics, vol. 24, no. 11, pp. 1037–1052, 2018.

[19] Olsen KA, Rosenbaum B. Prospective investigations of the prodromal state of schizophrenia: assessment instruments. Acta Psychiatr Scand 2006; 113:273-82.

[20] Heinimaa M, Salokangas RKR, Ristkari T et al. PROD-screen - a screen for prodromal symptoms of psychosis. Int J Methods Psychiatr Res 2003; 12:92-104.

[21] J.R. Williamson, T.F. Quatieri, B.S. Helfer, R. Horwitz, B. Yu, D.D. Mehta, Vocal biomarkers of depression based on motor incoordination, Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, ACM, 2013, pp. 41–48.

[22] Moore, Elliot, M.A. Clements, J.W. Peifer, L. Weisser, Critical analysis of the impact of glottal features in the classification of clinical depression in speech, IEEE Trans. Bio-Med. Eng. 55 (1) (2008) 96–107

[23] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, T.F. Quatieri, A review of depression and suicide risk assessment using speech analysis, Speech Commun. 71 (2015) 10–49.

[24] V. Mitra, E. Shriberg, M. McLaren, A. Kathol, C. Richey, D. Vergyri, M. Graciarena, The sri avec-2014 evaluation system, Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, ACM, 2014, pp. 93–101

[25] J.R. Williamson, T.F. Quatieri, B.S. Helfer, G. Ciccarelli, D.D. Mehta, Vocal and facial biomarkers of depression based on motor incoordination and timing, Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge, ACM, 2014, pp. 65–72.