Depression Analysis

A B.Tech Project report submitted in fulfilment of the requirements
for the degree of Bachelor of Technology

by

Vandita Goyal (2016UCP1004)
Akshita Agrawal (2016UCP1023)
Sankhya Mitra(2016UCP1145)

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY
MAY 2020

Certificate

We,

Vandita Goyal (2016UCP1004)

Akshita Agrawal (2016UCP1023)

Sankhya Mitra (2016UCP1145)

Declare that this B.Tech Project titled, "Depression Analysis" and the work presented in it are our own. We affirm that:

- This project work was done wholly or mainly while in candidature for a B.Tech. degree in the Department of Computer Science and Engineering at Malaviya National Institute of Technology, Jaipur (MNIT).
- Where any part of our B.Tech Project has been submitted earlier for a degree or for any other qualification in Malaviya National Institute of Technology, Jaipur or in any other institution, this has been clearly stated. Wherever we have consulted the published work of others, this is always clearly credited. Where we have quoted from the work of others, the source is always given. With the exception of all such quotations, this B.Tech Project report is entirely our own work. If there is any uncited plagiarism to be found, it is our (the students') sole responsibility.
- We have acknowledged all of the main sources of help.

Signed:

Date: Dr. Dinesh Kumar Tyagi

Assistant Professor

23.05.2020

Department of Computer Science and Engineering

Malaviya National Institute of Technology, Jaipur

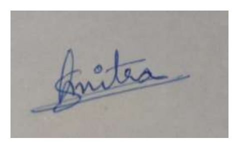
Abstract

The project aims at creating a framework encompassing a psychological model containing depression diagnostic tools capable of increasing accuracy and precision in testing while minimizing human intervention. Depression is a widespread mental illness which, if left untreated, can cause life-long problems for not only the individual suffering from depression but people surrounding him. Awareness of depression and its testing is extremely low owing to social and economic factors. The stigma associated with depression is high and hence prevents testing when it is required. Testing can also be extremely subjective due to the variance in methods used by different medical practitioners. The project aims to research the feasibility of various parameters including but not limited to different components of speech (audio parameters) and visual parameters like gaze, action units etc. in diagnosing depression. The project further aims to create an application that automates the process of depression diagnosis with the use of various tools and machine learning models. This is to ensure that testing can be readily available, objectively executed and highly accurate. It is observed that people are more willing to talk to an online application rather than a human for the fear of being judged. Hence to further minimize human interaction, we experimented with using a Chatbot DINA (Depression Interpretation Navigation Assistant) to collect the data from users. This data can be further used for testing purposes and diversification of the dataset.

Combining the above parameters, we aim to create an integrated machine learning model to test depression to ensure that testing is objective, easy to access and accurate.

Acknowledgements

We would like to express our deepest gratitude and thankfulness to Dr. Dinesh Kumar Tyagi, our project supervisor, for his patient mentorship, enthusiastic encouragement and useful critiques of this work. Constant encouragement and constructive feedback made this possible and helped us to have hands-on experience with various new technologies.



Vandita Goyal
(2016UCP1004)

Akshita Agrawal
(2016UCP1023)

Sankhya Mitra
(2016UCP1145)

Contents

Certificate ... i

Abstract ... ii

Acknowledgements ... iii

List of Figures ... vi

List of Tables ... vii

1. Introduction ... 1

1.1 Situation in India ... 1

1.2 Barriers to effective care ... 1

1.3 Objective ... 2

1.4 Exploration of Possible Solution ... 2

1.4.1 Consultation with an expert ... 2

1.4.2 Field Survey ... 4

1.4.3 Formulation of proposed methodology ... 7

2. Literature Survey ... 8

2.1 Convolution Network ... 8

2.1.1 Convolutional Layer ... 9

2.1.2 Pooling ... 9

2.1.3 ReLU Layer ... 10

2.1.4 Fully Connected Layer ... 10

2.2 Long Short Term Memory ... 11

2.2.1 Recurrent Neural Network ... 11

2.2.2 Long Term Dependency Problem ... 12

2.2.3 Input Gate ... 13

2.2.4 Forget Gate ... 14

2.2.5 Output Gate ... 14

2.3 Open Face ... 14

2.4 Media Recording API ... 14

2.5 Flask ... 15

2.6 Chatter Bot ... 15

2.7 Background Scheduler ... 17

2.7.1 Triggers ... 17

2.7.2 Job stores ... 17

2.7.3 Executors ... 17

2.7.4 Scheduler 18
2.8 Web Speech API 18
2.9 MySQL 18
3._Related Works 19
3.1. An Approach for automatically measuring facial activity in depressed subjects 19
3.2. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews 20
3.3. Detecting Depression from Facial Actions and Vocal Prosody 22
3.4. Automated Audiovisual Depression Analysis 23
3.4.1 Visual Behavior Analysis 23
3.4.2 Acoustic Behavior Analysis 24
3.5. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender 25
3.6. Automatic audiovisual behavior descriptors for psychological disorder analysis 26
4._Proposed Method 27
4.1 Dataset used 27
4.1.1 Features of the dataset: 27
4.2 Preprocessing and Processing of Data 27
4.2.1 Audio 27
4.2.2 Gaze 28
4.2.3 Action units 29
4.3 Model Integration 30
4.4 Application Architecture 31
4.4.1 Process Flow 32
4.4.2 User Interface 33
4.4.3 Database Design 35
4.5 Future methodology 37
5._Experimental Results 38
6._Conclusion 52
References 53
Appendix I 55
Plagiarism Report 61

List of Figures

Figure 1: Convolution Neural Network 8
Figure 2: LSTM chain.. 11
Figure 3: RNN chain.. 12
Figure 4: Input Gate Layer.. 13
Figure 5: Code of "Action Units"_……………………………………………………………29
Figure 6: Flow Diagram for the application ………………………………………………………………… .31

Figure 7: Chatbot flow...…..33
Figure 8: Results displayed to the Counselor…………………………………………………34
Figure 9: Login Page…………………………………………………………………….....34
Figure 10: Registration Page………………………………………………………………....35
Figure 11: Reg_log table…………………………………………………………………...36
Figure 12: dep_ana table…………………………………………………………………...36
Figure 13: LSTM model for gaze data ………………………………………………………..38
Figure 14: Results from LSTM model for gaze data…………………………………………....39
Figure 15: CNN model for audio spectrogram images…………………………………………39
Figure 16: CNN model layers for audio spectrogram images…………………………………..40
Figure 17: Training and accuracy from the CNN model………………………………………..41
Figure 18: LSTM model for AUs……………………………………………………………..50
Figure 19: Results for LSTM model for AUs…………………………………………………50
Figure 20: Data from Student survey 1…………………………………………………………55
Figure 21: Data from Student survey 2…………………………………………………………56
Figure 22: Percentage of students who have experienced at least one symptom of depression.......... 56
Figure 23: Percentage of students who have consulted a mental health professional or counsellor . 57
Figure 24: Percentage of students who believe that there are adequate facilities and awareness to help those suffering from mental health problems... 57
Figure 25: Percentage of students who would feel comfort or ease in using an online application that could screen for early signs of depression.. 58
Figure 26: Data from medical professional survey……………………………………………...58
Figure 27: Percentage of doctors who have experienced at least one symptom of depression 59
Figure 28: Percentage of doctors who believe that there are adequate facilities and awareness to help those suffering from mental health problems 59
Figure 29: Percentage of doctors who have consulted a mental health professional 60
Figure 30: Percentage of students who would feel comfort or ease in using an online application that could screen for early signs of depression.. 60

List of Tables

Table 1. Mapping of AUs to RUs 20

Table 2. AU 01 42

Table 3. AU 02 42

Table 4. AU 04 43

Table 5. AU 05 43

Table 6. AU 06 44

Table 7. AU 09 44

Table 8. AU 10 45

Table 9. AU 12 45

Table 10.AU 14 46

Table 11. AU 15 46

Table 12. AU 17 47

Table 13. AU 20 47

Table 14. AU 23 48

Table 15. AU 25 48

Table 16. AU 28 49

Table 17. AU 45 49

Chapter 1

Introduction

Depression is a widespread illness and mood disorder, affecting more than 264 million people worldwide. Depression differs from normal mood variations or short-term emotional responses to difficulties in daily life. Especially when long-lasting and having moderate or severe intensity, depression can evolve into a severe health condition with serious consequences. It can result in the depressed person suffering greatly and can lead them to function poorly at every level, whether it is at work, at school or at home. Depression can also lead to suicidal attempts and ideation. [1]

1.1 Situation in India

Depression has been growing steadily in the Indian population, growing by 18.4% between 2005 and 2015. This worrying trend has also been reflected in any increased number of suicides, particularly in adolescents. Only 20% of those who have any kind of mental illness have access to a healthcare facility and only 10% receive treatment. [1]

1.2 Barriers to effective care

Some barriers to effective care include a lack of resources and awareness, lack of trained healthcare. providers and a general social stigma associated with mental disorders. Mental healthcare generally carries with it negative social connotations. Yet another barrier to effective care is an inaccurate assessment. In countries all over the world, of all developmental levels, those who are depressed are very often not diagnosed correctly, and many others who do not have the disorder are too often misdiagnosed and prescribed antidepressants.

When left untreated, depression can threaten the safety of the individual and those who live with or depend on them. There are certain critical professions such as doctors, pilots and police where depression can have more serious consequences. In 2015, a German wings flight was crashed by a co-pilot suspected of having depression, killing all on board. These are not just isolated incidents. It is thus essential that the testing for depression becomes more accessible and common place.

Workplaces too do not cover regular mental health care checkups. As a result, it is common to label depression as “corporate burnout”. The stress of corporate environments leaves many individuals vulnerable to depression and other mental illnesses.

Another problem that we identified in the diagnosis of depression is that in the present method, the diagnosis relies completely on the judgement and experience of the psychologist. Psychologists use methods like interviewing the patient to glean information about certain habits such as fatigue, lack of sleep etc. to determine whether the patient has depression.
As a result of these subjective methods, a large number of people are misdiagnosed each year and are prescribed harmful antidepressants. This is also known as a “false positive”. [1]

1.3 Objective

As discussed in the above section, problems causing lack of effective mental health care served as the main motivator for the project. Therefore, our objective is to examine the feasibility of and aim to use different behavioral indicators for depression, consisting of, but not limited to, visual and audio features to design an effective testing model which can be made more accessible than traditional testing methods.

1.4 Exploration of Possible Solution

1.4.1 Consultation with an expert

In the process of researching the applicability of Computer Science in diagnosing psychological conditions such as depression, we interviewed Mr. Ruchir Sodhani, an experienced and practicing counselor and therapist from the Australian College of Applied Psychology currently working in the Department of Neurosciences, Rukmani Birla Hospital, Jaipur and the Department of Deaddiction, Santokba Durlabhji Memorial Hospital Jaipur. According to his expert opinion, the problem of diagnosing depression lends itself well to a technological supported solution. We have summarized the original document provided by Mr. Ruchir below:

Counselling is an extensive process involving complex and adaptive interaction for effective engagement, understanding, insights and ultimately providing support and help. Depression, in particular, is rarely recognized unless it reaches an advanced stage where the everyday life of the individual is impacted. Easier, accessible, and more acceptable access to testing can be significant in discovering depression before it manifests in more severe forms. Diagnosis is based on understanding behaviors, thoughts and feelings of the subject and how these are impacting the daily life of the subject. It is largely based on the experience of the medical practitioner in analyzing verbal and non-verbal cues of the subject. Technology can play an important role here by streamlining the process while providing comfort and privacy to the subject. Working in conjunction with mental health professionals, technology can make it possible to administer tests in private, safe, confidential and potentially self-administered, low-cost settings where people are more comfortable.

A. SOCIAL STIGMA AND FEAR

Mental health is still associated with stigma and a general lack of awareness makes it difficult to dispense the resources required by those dealing with various mental health issues. Mental illness can have a major impact on the person's livelihood which can affect their future prospects. Due to the stigma, there are many who are reluctant to seek the help they require until it is too late.

Technology can certainly play a vital role here. It can help provide quick, more accessible testing and build a preliminary screening model before proceeding to more intensive counselling.

B. PARAMETERS FOR THE MODEL

Pathology-based "physical" testing is not the basis for diagnosis. Rather it is based on understanding the symptoms such as behaviors, thoughts, and feelings.

Among the parameters that are significant when determining whether a patient is depressed and the severity of it, facial and audio cues are especially important. In many instances, the patient may answer contrary to how they are actually feeling. This can be for a variety of reasons ranging from unwillingness to admit to their problems to difficulty in expressing. Some subjects display extreme cheerfulness or talkativeness beyond the norm. As a result, facial and nonverbal clues can assist the counsellor or psychologist in effectively determining whether the subject requires further help.

Facial clues may include expressions of discomfort, aversion to eye contact, distant gaze etc. An analysis of different facial points and how they change over time for patients diagnosed with depression can show similarities and provide a reference timeline that may prove helpful in future diagnosis.

The voice of the patient can also contain clues to the state of the subject. The subject may use lower tones, change in pitch, reduced vowel space and other phonetic markers. An analysis of these parameters is worth researching to codify an algorithm that can test for depression while incorporating a human touch to maximize responsiveness of the subject. Thus, a combination of facial and audio clues can be combined in different proportions to develop an effective testing model for depression.

C. SPECIFIC USE CASES

In some cases, the subject may actively seek to avoid being diagnosed with a mental illness, particularly in professions where it may affect their credibility such as healthcare workers, pilots etc. It is essential that these professions be tested regularly since they perform high risk operations which can greatly impact others. Apart from this, many students suffer from depression and do not receive the help they need to be treated. Many schools and universities do not employ a regular counselor, and even with those that do, the population requiring help is much larger than the capacity of the services available. The application can also reach out to sections of the society that normally lack awareness about mental health such as those residing in rural areas, long haul drivers etc. Hence the application, while created to be used by all walks of life, is particularly useful in these cases.

1.4.2 Field Survey

We also surveyed 224 college and school students to find the prevalence of depression in this segment by circulating a questionnaire.

The questionnaire contained the following questions:

1. At any point of time in your life, have you suffered from any of the common symptoms of depression (loneliness, fatigue, changes in appetite, changes in sleep and difficulty in concentrating)?

   The above question was to assess the prevalence of symptoms of depression among college and school students. It was found that 75.7% of the responders had indeed suffered from certain symptoms of depression before.

2. Have you ever consulted a mental health professional or counselor?

   The above question was posed to assess the ease of access of mental health care facilities and the awareness of the responders about the availability of such facilities. It was found that 88.5% of the responders had not consulted a mental health professional before.

3. Do you believe that there are adequate facilities and awareness to help those suffering from mental health problems like depression, anxiety etc.?

   The above question was to assess the prevalent opinion of the responders about whether the facilities such as availability of mental health professionals, adequate medicines and easily available diagnosis are available or not. 72.5% of the responders believed that the facilities are not adequate.

4. Would you feel comfort or ease in using an online application that could screen for early signs of depression?

   The above question was to help formulate the solution application that is being proposed. The aim was to assess the feasibility of the application that is being proposed. 82.1% of the responders felt comfortable using an online application that would screen for early signs of depression.

The responses and the results of the above survey are available in Appendix I.

We also surveyed 32 medical professionals and doctors to understand the stigma associated with mental healthcare in the medical industry using a similar questionnaire. We found that despite having a medical background and hence easier access to psychological care, a majority of doctors had suffered from depression in some form. Surprisingly, most had never consulted a mental health professional for this.

The questionnaire contained the following questions:

1. At any point of time in your life, have you suffered from any of the common symptoms of depression (loneliness, fatigue, changes in appetite, changes in sleep and difficulty in concentrating)?
   The above question was to assess the prevalence of symptoms of depression among college and school students. It was found that 71.9% of the responders had indeed suffered from certain symptoms of depression before.

2. Have you ever consulted a mental health professional or counselor?
   The above question was posed to assess the ease of access of mental health care facilities and the awareness of the responders about the availability of such facilities. It was found that 90.6% of the responders had not consulted a mental health professional before despite having easier access than most.

3. Do you believe that there are adequate facilities and awareness to help those suffering from mental health problems like depression, anxiety etc.?
   The above question was to assess the prevalent opinion of the responders about whether the facilities such as availability of mental health professionals, adequate medicines and easily available diagnosis are available or not. 68.8% of the responders believed that the facilities are not adequate despite being in the medical field.

4. Would you feel comfort or ease in using an online application that could screen for early signs of depression?
   The above question was to help formulate the solution application that is being proposed. The aim was to assess the feasibility of the application that is being proposed. 87.5% of the

responders felt comfortable using an online application that would screen for early signs of depression.

The responses and the results of the above survey are available in Appendix I.

1.4.3 Formulation of proposed methodology

Through these surveys, we can conclude that mental health facilities remain inaccessible to a significant portion of the population. To overcome these challenges, we sought to investigate the possibility of developing a method that combines advanced Machine Learning based techniques to diagnose clinical depression in individuals. We also created a Chatbot DINA (Depression Interpretation Navigation Assistant) to assist in the collection of data from users.

There are several behavioral indicators that are observed with depression, as explained by Mr. Ruchir Sodhani, which can aid in the learning process for the model. Present diagnosis methods and testing criteria for depression are heavily dependent on certain subjective parameters of behavior, like interviews with the subject themselves or reports by family or friends. These may be unreliable and can be colored by personal bias. While clinicians are subject to variance in their diagnosing methods, the accuracy of their diagnosis and their consistency, the automated system in the realm of the parameters provided has the advantage of performance consistency. These tools can be used for verifying certain behavioral indicators of depression that are under research, assist in the timely and fast screening and diagnosis, measure the patient's response to intervention and rehab, and experiment clinical speculations about the underlying mechanisms that can affect an individual’s mental state. [2]

Thus, through this project, we examine the feasibility of and aim to use different behavioral indicators for depression, consisting of, but not limited to, visual and audio features to design an effective testing model which can be made more accessible than traditional testing methods. We also try to minimize human intervention by using a specially trained ChatBot to assist in data collection

Chapter 2

Literature Survey

2.1 Convolution Network

A Convolutional Neural Network, known more widely as a CNN is a class of Deep Learning algorithm that takes as its input an image, assigns specific importance by using some specific learnable weights and biases, to the various aspects or objects in the image and is thus then becomes capable in distinguishing between them. The pre-processing that is needed in a Convolutional Neural Network is far lesser when contrasted to various classification algorithms. In some of the primitive methods, the filters are usually manually engineered, but after adequate training, CNNs possess the ability to learn these characteristics or filters. [3]

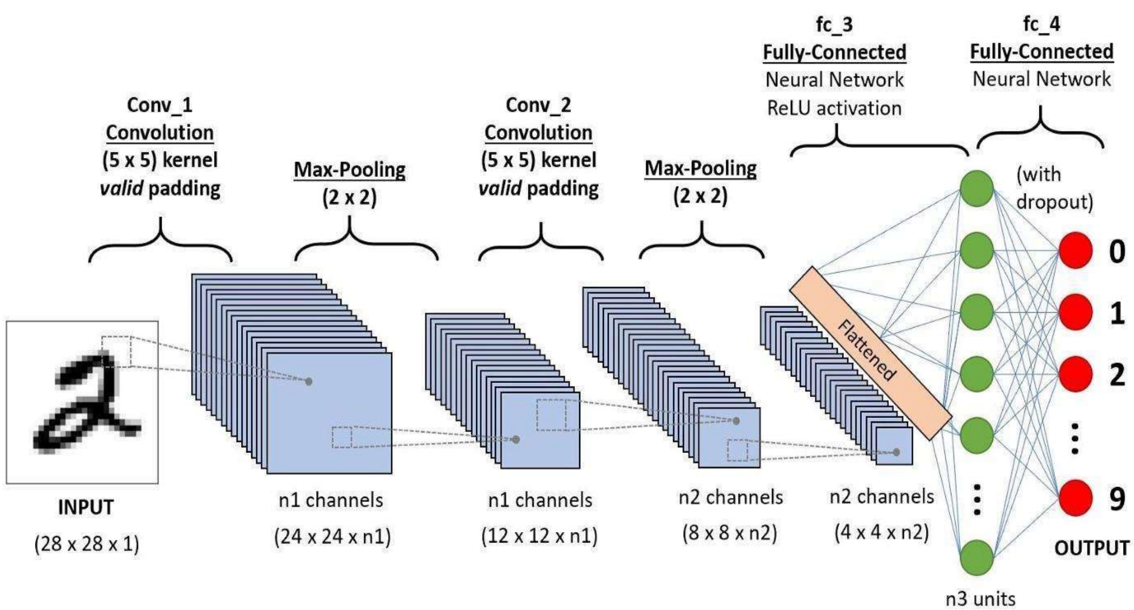


Figure 1: Convolution Neural Network [3]

The above figure depicts the working of a Convolution Neural Network with input being an image (28 height X 28 width X 1 channel). The input is put through a series of layers (conv layer, max pooling layer, fully-connected layer) which are explained below

2.1.1 Convolutional Layer

The initial layer of a Convolutional Neural Network is always a "Convolutional Layer". This layer is also called a "filter" and the region that it is passes across is known as a receptive field. The filter can also be considered as an array of numbers which are known as parameters or weights. When the filter slides or convolves around the inputted image, it also multiplies the values in the filter by the earlier pixel contents of the image. This is also referred to as computing element wise multiplications. These multiplications obtained from the previous step are then all added to result in a single value. After the filter slides over all the locations, an array of numbers is left behind which is known as an activation map or alternatively as a "feature" map.

The whole objective of performing this entire operation is to extract certain the high-level features like edges, from the inputted image. As per convention, the first Convolutional Layer is accountable for capturing some of the low level features like the color, the gradient orientation among others. As and when the layers are added, the architecture gradually adapts to the High-Level features, thus resulting in a network which has a better and more well-rounded understanding of images given in the dataset, in a way that is quite similar to how humans can perceive it. [3]

2.1.2 Pooling

The Pooling layer typically reduces the spatial size of the Convolved Feature. This process is normally done to diminish the computational power demanded for processing the data using dimensionality reduction. Moreover, it is also helpful for deriving some of the dominant features which in nature may be rotational or positional invariant.

There are two main kinds of Pooling: the "Max Pooling" and the "Average Pooling". The Max Pooling echoes the "maximum value" from the part of the image that is covered by the Kernel. Conversely, the Average Pooling gives us the "average" of all the values from the part of the image that is covered by the Kernel. [3]

2.1.3 ReLU Layer

Following each convolution layer, we conventionally use a non-linear layer, which is also known as an Activation Layer. This layer is used with the purpose of introducing non-linearity in a system that so far has only been computing linear operations, such as the element wise multiplications and the summations. Earlier, non-linear functions such as "tanh" and "sigmoid" were applied, however since then, research has suggested that ReLU layers perform much better since the network is now equipped to train much faster (owing to better computational efficiency) without making a notable change in the accuracy. Furthermore, it assists in reducing the “vanishing gradient problem”. In the vanishing gradient problem, the lower layers of the network get trained rather slowly as the gradient declines exponentially over the layers. The ReLU layer then goes on to apply the function "f(x) = max(0, x)" to every value included in the input volume. In simple terms, all the negative activations are changed to 0 by this layer. This layer enhances the non-linear properties of the model and the overall network without having any effect on the receptive fields of the Convolutional Layer. [3]

2.1.4 Fully Connected Layer

The addition of a Fully-Connected layer is a (normally) inexpensive method to learn the non-linear combinations of the high-level features as represented by the output of the convolutional layer. The Fully Connected layer is learning a likely non-linear function in that region. Through this the image is flattened into a column vector. The model, after undergoing many consecutive epochs then becomes capable in differentiating between the dominant and some of the low-level features present in images and is then able to classify them utilizing a technique known as the Softmax Classification. [3]

2.2 Long Short Term Memory

Long Short Term Memory networks are a special type of Recurrent Neural Networks, which have the capability to learn long-term dependencies. LSTMs are specially designed to evade the long-term dependency obstacle. It is their default behavior to remember information for long periods of time. [4]

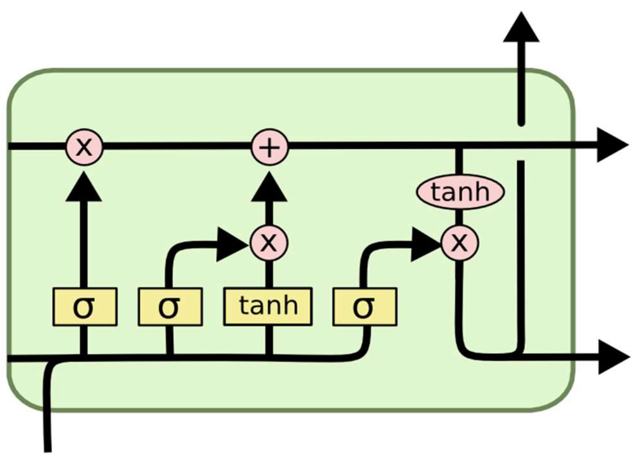


Figure 2: LSTM chain [4]

The above figure depicts the inner workings of a cell in an LSTM chain. This includes input gate, forget gate and output gate which are explained further below.

2.2.1 Recurrent Neural Network

Recurrent Neural Network, popularly known as RNN is a deep learning algorithm which models sequential information. It is extremely popular and is widely used for NLP tasks. Recurrent neural networks work on the idea of using sequential information. In a traditional neural network, it is normally considered that all of the inputs will not be dependent on each other. However, for many tasks this assumption cannot be applied. RNNs are thus designed to make use of the previous inputs as context for their new inputs.

RNNs are networks that contain many loops within them which allows the information to persist. [4]

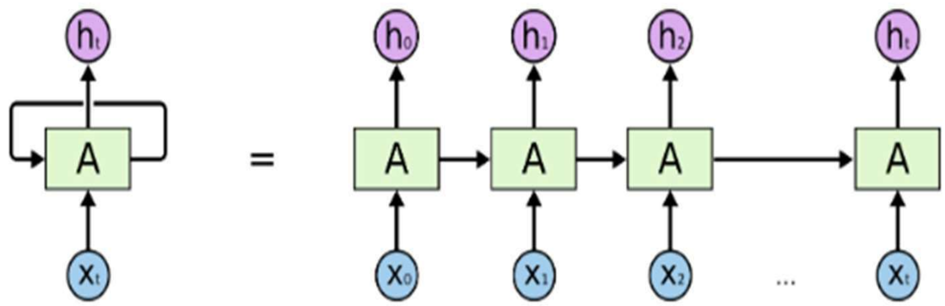


Figure 3: RNN chain [4]

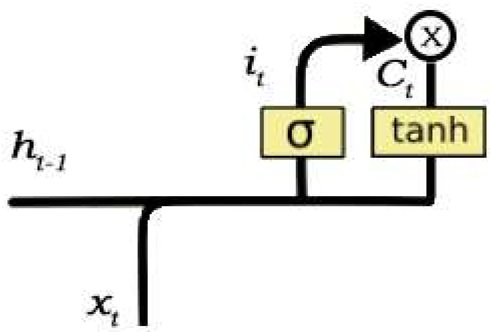
The above figure depicts an unrolled RNN chain. As we can see above, each cell is identical to each other. In the first cell, the input is X0 which gives the output h0. This output is further propagated to the next cell which takes the combination of the previous output and new input X1 as the input to the cell. This process is repeated several times until we get the final output.

2.2.2 Long Term Dependency Problem

In some situations, we only require the information that is recent in order to perform the current task. Consider a particular language model that tries to predict the last word in the sentence "the birds are in the sky". Here, it is relatively easier to predict the subsequent word as sky based on the preceding words. However, the sentence "I live in China, I can speak Chinese well." is much more complex to predict. It is dependent on the previous input also. In such types of sentences, it is totally plausible for the distance between the relevant information and the instance where it is required to grow very large. This problem is known as the Long Term Dependency Problem. RNN is unable to handle this problem. Hence LSTMs were designed. [4]

2.2.3 Input Gate

Input Gate layer is a simple sigmoid gate layer which helps decide what values to update in a cell. This when coupled with a "tanh" layer, which creates a vector of candidate values to be added to the state, is responsible for making the decision about what part of the new information is to be saved in the present cell state. [4]



$$i_t = \sigma\left(W_i\cdot[h_{t-1}, x_t] \ + \ b_i\right)$$
$$\tilde{C}_t = \tanh(W_C\cdot[h_{t-1}, x_t] \ + \ b_C)$$

Figure 4: Input Gate Layer [4]

The above figure depicts the working of the input gate. It takes a combination of previous output and new input and pushes it through a sigmoid gate (i_t). This value is coupled with the value, computed using the same inputs, from tanh helps decide candidate values to be added to the state.

W_i: Weight of input

B_i : bias of input

W_c: Weight for candidate value

B_c: Bias for candidate value

2.2.4 Forget Gate

Forget Gate layer is a simple sigmoid layer which makes the decision about which information to remove from the present cell state. It studies "ht−1" and "xt", and then gives as output a number between "0" and "1" for every number in the cell state "Ct−1". A 1 means "completely keep this" whereas, a 0 denotes "completely get rid of this."[4]

2.2.5 Output Gate

Output Gate layer is a sigmoid layer which along with a "tanh" layer helps decide what to output. [4]

2.3 Open Face

OpenFace is an open source toolkit, developed by Tadas Baltrušaitis, used to analyse the behaviour of the face. It is used to landmark the face, track head positions, track the motion of the muscles of the face, track the direction in which the eyes are looking. It uses a host of models like Convolutional experts constrained local model and Constrained Local Neural Fields to be able to do so.

2.4 Media Recording API

Media Recording API (also known as Media Stream Recording API) is used to record data generated by Media Stream Objects. Often used to record video and audio from source, it provides an event handler to be used when data becomes available. Using the data attribute Blob, the data can be processed and saved.

Steps to use this API:

1. Setup the source for the stream
2. Create the MediaRecorder Object with the source and other options

3. Specify the functionality for event handler to perform when data becomes available
4. When the data streams from the source, the event handler executes the functionality specified. The data provides an attribute called Blob. Using that attribute, the data can be processed, recorded and saved.
5. Call the stop function to stop recording.

2.5 Flask

Flask is a web application framework written in Python. It is quick and easy to get started. It is based on "Jinjja" and "Werkzeug". The micro framework's objective is to simplify the core. It has little dependency on the external libraries and hence is light. It only provides a template engine which helps keep the web application layout uniform and easy-to-maintain. The code for creating an application using Flask while can fit in one page but it doesn't have to. Flask is full of functionalities but it doesn't limit itself to one module for one functionality. It leaves that decision to the creator of the application. It does not have a database abstraction layer but has multiple modules to support different types of the databases.

2.6 Chatter Bot

Chatter Bot is a library written in Python which aids in creation of chatbots. It crafts replies for the user's questions and comments on its own. It uses a host of machine learning algorithms to do as such. The design of this library allows data of any language to be accepted. A new object of the library is generally not able to properly reply to the user, but it learns on the job. To speed up the process, training data can be provided in form of conversation list. These conversations are seen as a back and forth between the user and the chatbot.

Example:

"Hi",

"Hello! How are you?",

"Not so well actually",

"Is everything okay at work?",

"That's what I wanted to talk about",

"Okay. Tell me about it",

"The hours are too long. I'm really tired",

"What do you do in your job?",

"I work as a doctor.",

"That's a tough job.

“How many hours are you working?",

"Sometimes it's 90 hours a week",

"That's a lot. Do you get time for anything else?",

"Not really.",

"What about weekends?",

“I end up working on Saturdays and sometimes Sundays too",

"Do you enjoy your work",

"I do. But it's tiring me out ",

"Perhaps it's time for you to slow down?",

"I don't think I can do that",

"Your job is really important but not as much as your health",

"I don't think people will understand if I tell them that I'm a doctor and I feel depressed",

"You don't need to think about what other people believe. Getting help right now can change your life.”

"I'll think about it",

"Once again, I'd advise you to immediately seek help. Are you eating well?",

"Not really. I don't feel hungry anymore",

"Are you on any kind of special diet?",

"No",

"How many hours do you sleep each night?",

“I lie awake a lot. Though I'm tired, I can't sleep",

"Have you tried talking to a mental health professional about this?",

"No. I want to but I don't get the time",

"I can help you book an online session in which you can speak to the counsellor. Would you like that?",

"Yes. Okay. I think I'm ready",

"will get right to it!",
"Thank you",
"You are welcome!

A chatbot made using ChatterBot chooses its replies by using a Logic Adapter. The Logic Adapter checks the similarity between the data it has and the question or comment given by the user. It does so by calculating the Levensthein distance. Once it has chosen the statement closest to the input, it returns the appropriate reply that it has stored in its database. The database can be structured like SQL or unstructured like MongoDB.

2.7 Background Scheduler

A module of APScheduler which allows the code or function specified to run after a particular period of time. The job is run in the background in a separate thread. New jobs can be added and removed at any time. A database can be used to stored jobs that are to be repeated from time to time or be executed in parts.

It has 4 components as given below:

2.7.1 Triggers

These specify the logic for the scheduler. These determine when and what is supposed to run when a job is run

2.7.2 Job stores

This is used when a job has to be repeated several times or executed in parts. The state of the job is stored in a database

2.7.3 Executors

These are used to actually run and stop the job if any error occurs. This is done by pushing the function object to the thread or pool of threads used for execution

2.7.4 Scheduler

This combines all the elements. It provides the interface to an application to use the above elements to run the job. Adding, removing and changing the job is done through a scheduler.

2.8 Web Speech API

This is used to convert speech into text and text into speech. Speech Recognizer receives the speech through the microphone and converts it into text by comparing the sound against its data. The data it has includes vocabulary and grammar rules. When the matches are found, they are returned. The best result is chosen.

Similarly, Speech Synthesis is used to convert text to speech. It does so using pre-stored sounds mapped against phonetics given. These stored sounds are known as "utterances".

2.9 MySQL

This is used to maintain relational databases. It is an open source software. It is based on Structured Query Language and can be used to perform multitude of functionalities such as creating the database, creating the table schema, inserting, deleting and modifying data, reading and displaying the data.

It can be connected to any framework by using a host of connectors available such as "Flask-MySQL"

Chapter 3

Related Works

3.1. An Approach for automatically measuring facial activity in depressed subjects

Gordan McIntyre et al.

This approach by Gordan McIntyre et al. is to use "Action Units" (AUs) to understand the activity of the muscles of the face in depression. The widely accepted theory is that a depressed person does not show much reaction to negative stimuli. In this study, a study of 116 participants, some with MDD and others with no history of depression was conducted. Participants were made to smile at a comedy clip. They had to review small film clips by giving a score. AUs were extracted from 11 seconds of the video of them smiling. The hypothesis was that people with prior history of MDD were more likely to have controlled smiles than the people with no prior history.

The facial image was then subdivided into regions to track inter-region movement. This was done by dividing the face halfway between topmost and bottom-most points of the face. A second, less dynamic horizontal delineation is used for the intra-RU measurements. The space between features and reference line are measured and normalized. With exception to RU3, only vertical displacement could be derived. In RU3, horizontal displacement for AU20 had to be derived. Inter-region movement of RUs was used to classify prototypical facial expressions and used displacement from midway horizontal lines.

Two sets of Classification models were used. The first one is used to classify on the basis of prototypical expressions. The second one is used to classify the RUs that can be seen. Mapping of AUs to RUs that are used for the second set of classification models is shown below:

Table 1: Mapping AU to RU

AU	RU
1,2,3	1
5,6,7	2
9,10,12,15,17,20,25,26,27	3

Pre-trained AAM model is then used to derive the features which are then used as input into MultiBoost classifiers. These finally mark the regions used to report facial activity. To measure facial activity, 5 values are computed:

1. Total number of RU activity in an interval
2. Total number of prototypical expression in a interval
3. Total number of non- prototypical expression in a interval
4. Positive vs Negative prototypes
5. The delay in the changing of expressions [5]

3.2. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews

Stefan Scherer et al.

This method is to use speech as a way to predict depression and suicidal tendency. Speech can be noticeably affected due to slight physiological and cognitive changes. Stefan Scherer et al. used reduced frequency in vowel space to classify and characterize participants suffering from depression. A speaker's vowel space which is the difference between the frequency of first and second formant of the vowels /i/, /a/ and /u/, is measured using an automated machine learning algorithm. There were 253 participants in the experiments conducted which showed a significant reduction in vowel space of a depressed participant. 54% participants had reported that they were suffering from depression. To conduct the experiment, a robust fundamental frequency tracker and voice analysis tool (provided in COVAREP 1.0.0 toolbox) to identify areas of interest. The vowel space analysis will be done in these areas. Formants are tracked throughout the voiced component of speech. Next step is to track the said Formants. Vowels are characterized by the first two formants. Vowel space assessment and Articulation rates are used in the method since people suffering from PTSD and depression tend to take longer to answer basic questions.

The results of the experiments show that the participants suffering from depression show a smaller difference between first and second formant than those who are not suffering. None of the observed vowel space differences could be explained by the articulation rate. The vowel space peaked at the initial 5 minutes of the conversation; however, the length of the conversation did not affect the size of the vowel space. Vowel space did not differ based on the demographic differences shown by the participants.

Based on the observations of the experiments conducted, 3 hypotheses were proposed and examined:

1. Effect of Psychological Conditions on Vowel Space

The results showed that the people who scored negatively in PHQ9 questionnaire have larger differences between the frequencies of formants than those who scored positively.

2. Limited data does not affect the measurements

Tracking of formants can be mess. But it could be seen that the fluctuation in the measurements disappeared after the first 5 minutes. The length of the conversation did not affect the size of the vowel space. Hence the robustness of the vowel space measure can be verified.

3. The demography and articulation rate does not affect the measurements

The overall measure of the vocal space is not affected by differences in race, ethnicity, gender or education. The dialect did cause differences in the measurements. [6]

3.3. Detecting Depression from Facial Actions and Vocal Prosody

Jeffrey F. Cohn et al.

This approach by Jeffrey F. Cohn et al. sets out to prove the feasibility of automatic depression analysis. The participants of this trial consisted of 20 men and 37 women who had been diagnosed for Major depressive disorder (MDD). Some were given anti-depressants while some were not before the clinical interviews had taken place. Video was taken using 4 cameras - two focusing on the face and shoulders from the left while the third one was focusing on the complete body. Audio was digitised at 48Mhz and 10 msec clip was taken as the sample per second.
This recording was subjected to 3 experiments -

1. Manual FACS

   The activity of the muscles of the face was recorded for each participant's response to the first 3 questions. The questions were related to feelings of guilt, sadness and desire to commit suicide. Four features were computed for every action unit - mean duration of the activity, ratio of onset phase to offset phase, ratio of onset phase to total duration and the portion of the interview which the AU occurred. For each AU 16 possible combinations were computed. These were input to a Support Vector Machine using leave-one-out cross validation. Using all AUs, the accuracy for classification came out to be 79%. In particular, AU14 gave the best results of 89% accuracy.

2. AAM

   Active Appearance Models are used to decouple the shape and appearance of a face. Hence the AAMs prepared were person specific. These were used to obtain a shape-based vector of the face taken every 10s. The mean, median and standard deviation of the velocity was computed as coefficients to shape vectors. These were then further combined by computing mean, median, maximum value and minimum value. A vector of 12 numbers corresponding to 3 by 4 was outputed which were further concatenated to 10 corresponding shape vectors. This yielded 120 features which were then used as inputs to SVM. The accuracy for such a classification came out to be 79%

3. Vocal prosody

Using a publicly available software, vocal prosody was computed for the answers given to the first 3 questions. The frequency and delay in answering the question were computed. The hypothesis was that the frequency would decrease while the delay would increase. Using Logistic Regression, the accuracy for depressed was 88% and non-depressed was 64%. [7]

3.4. Automated Audiovisual Depression Analysis

This approach was proposed by Jeffrey M, Girard and Jeffrey F. Cohn in 2014 and it analyzes the feasibility of studying depression using automated audiovisual information. It divides these "behavioral indicators" into two main types: acoustic and visual.

3.4.1 Visual Behavior Analysis

Several visual features like the facial expressions of the individual, such as the "head and body movements, eye gaze, posture, and gestures" have been identified as possible indicators for depression. Some types of visual indicators are a smaller average distance between the eyelids and a faster blinking rate (Alghowinem S, Goecke R, Wagner M, Parker G, Breakspear M), a slower head movement (Nonverbal Social Withdrawal in Depression: Evidence from manual and automatic analysis, Girard JM, Cohn JF et al).

Lesser head motion, looking down for longer durations, decrease in smiling, decrease in frowning, and increase in mouth dimpling have also been observed.

3.4.2 Acoustic Behavior Analysis

Most of the approaches for depression detection focus on semantic or lexical analysis on text, spoken or written. This is done by obtaining transcripts of spoken conversations or through written conversations such as social media activity. Some indicators include pitch, loudness, rhythm, pauses, voice quality, the speaking rate and articulation [8]

Several methods have been discussed for supervised learning and for feature extraction. While some use only visual features, the others use audio features as well. This automated behavioral analysis can then be used to help in proving and verify many clinical theories linking depression and behavior. Apart from recognizing those who are at risk, these automated behavioral measures can also help record symptoms over time to serve as a precautionary instrument.

The studies conducted to determine the behaviors intrinsic to the participants suffering from depression also revealed that when critically depressed, participants tend to show "reduced head movement, decreased smiling and frowning and an increase in dimpling". When extremely depressed, participants also tend to show longer pauses before answering questions.

3 methods were proposed for the diagnosis:

1. Comparing the individual behavior to the groups determined by diagnosis or severity of symptoms.
2. Using classification algorithms to divide the participants into two mutually-exclusive groups by utilizing high-dimensional audio-visual features.
3. Using various regression algorithms to estimate the severity of any participants' depressive symptoms using high-dimensional audio-visual features. [9]

3.5. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender

Jonathan Gratch et al.

This method sets out to establish a relation between non-verbal indicators of depression and gender. The following indicators are taken as measures of depression and hence the effect of gender over them is studied:

1. Intensity of expressions of anger, disgust, contempt and joy

   For this, facial muscle movements were used. Intensity of AU4, AU7, AU9 and AU12 were measured since these appear predominantly in the emotions mentioned above. In this, the disparity between the trends of both genders could be seen. It was noted that depressed males showed more intense expression of anger and disgust than non-depressed males while the opposite is true for females. While it could be seen that depressed females show contempt more than non-depressed females, no such conclusion could be derived for males.

2. Emotional Variability

   Emotional variability can be measured by seeing how intense the neutral expression is. It was observed that for both genders, depressed participants displayed higher intensity f neutral expression than non-depressed participants

3. Motor Variability

   For this, head movement was tracked and hence measured. It could be seen that for both genders, depressed participants displayed less head movement than non-depressed participants [10]

3.6. Automatic audiovisual behavior descriptors for psychological disorder analysis

Giota Stratou et al.

This method is to use speech and visual data to identify indicators of depression and suicidal tendencies. The participants of this trial were left alone for some time to fill out certain surveys – PCL-C, PHQ9, STAI-T, BIDR, RME, BFI and PANAS. After filling of the survey, they were made to sit 7ft apart from the interviewer. They were recorded using a high-quality webcam and depth sensor while microphones were attached to the lapels of their clothes. The raw data obtained is analyzed by MultiSense and Covarep. This method seeks to investigate the following:

1. Automatic gaze descriptor

   For this, position of the head and gaze were measured. It could be seen that the distressed participant's gaze was downwards more that the non-distressed participants. But the position of head did not factor in this. The probability of a distressed participant's position of head being downwards was the same as the probability of such in a non-distressed participant.

2. Automatic smile descriptor

   For this, smile intensity and smile duration were measured. It could be seen that the distressed participant's smile intensity and duration both was lower than that the non-distressed participants.

3. Automatic voice assessment

   Two common parameters NAQ and QOQ was measured for this. These are both related to vocal tension. The smaller the value, the higher the tension. [11]

Chapter 4

Proposed Method

4.1 Dataset used

The database that we used for the training is a section of a bigger corpus, the "Distress Analysis Interview Corpus (DAIC)" (Gratch et al., 2014), that comprises of clinical interviews that were created to assist in the detection and the diagnosis of various psychological distress conditions like anxiety, depression, and PTSD. The Data accumulated includes some audio and video recordings and comprehensive answers to a questionnaire. The interviews were taken by an animated virtual interviewer named Ellie, who was being controlled by a human interviewer from a different room. Data was then deciphered and interpreted for a number of verbal and nonverbal features.

4.1.1 Features of the dataset:

1. Includes 189 folders
2. Average size of folder: 500 Mb
3. Folders contain audio recordings, visual input data – facial points in 3D space, gaze points and "action units"

After observing the work that has been done in this field, we take three separate approaches of analyzing audio, gaze and action units separately and then make predictions on the result that is acquired from these methods.[12]

4.2 Preprocessing and Processing of Data

4.2.1 Audio

The audio used is a mono channel of 16 Khz. The audio can be processed in various ways. The most suitable method was to generate Spectrogram images of the audio and train a CNN to learn the various regions of the image.

First, we convert all the audio files which are in wav format to spectrogram images with dimensions of (512,512) pixels.

Then we feed the Audio spectrograms into a Convolutional Neural Network. At first when we tried to train the model with a very basic architecture, the hardware we had was not able to run CNN. Then we switched to Google Colab. Our current model has 2 convolution layers with the first layer having 64 filters and with a kernel size of (3,3) and the second layer has a filter size of 64 with kernel size of (3,3). After each convolution layer a max pooling layer has been added with the size of (2,2) with stride of 1. After the convolution layers a flatten layer has been added to flatten the features that are acquired from the convolution layers. After the flatten layer 2 dense layers have been used with hidden node size of 512 each and with relu activation function. After this layer a dropout layer has been added to reduce overfitting and to randomly deactivate nodes in the final dense layer. The last layer is the output layer with 2 nodes and softmax activation function.

Softmax is used here because we wanted to distribute the result over the target value.

4.2.2 Gaze

The gaze data is provided for each interview in separate files with the name XXX.CLNF_gaze.txt. There are 4 vectors used as gaze data: First two are coordinates of the eyes in world space "x_0, y_0, z_0, x_1, y_1,z_1" and the next two are coordinates in headspace "x_h0, y_h0,z_h0, x_h1, y_h1, z_h1". The coordinates which are in the head space provide information about the position of the eyes relative to the position of the head.

1. Detecting movement using only the coordinates in head space: By taking the difference of the coordinates between frames. Then we find out the number of times the difference is more than 0.001(threshold that we have set for eye movement). Then we pass these data to a basic neural network to make predictions.
2. Classifying using LSTM and all coordinates in world space and headspace as the gaze data we have is in sequential form. Using success as a parameter to see whether to use the coordinates or not. Since the Gaze data for each candidate is not uniform in length, this can lead to inconsistencies. As a result, first we needed to

preprocess the data by padding the data by using 0s. In this data we also have many zero values where the data is not recorded properly and another challenge that we face is that not all the records for all the patients are the same so in order to tackle both the problems, we collected all the data in a single place and padded it with 0's. Then we use a masking layer to mask the 0 values so that the network does not trail on these values. 3 LSTM layers are used sequentially, followed by 2 dense layers, each having a sigmoid activation function.

4.2.3 Action units

Facial Action Units attributes to a set of facial muscle movements which correspond to a display of emotion.

AU	Name
1	Inner brow raise
2	Outer brow raise
4	Brow corrugator
5	Upper lid raise
6	Cheek raise
7	Lower lid tight
9	Nose wrinkle
11	Nasolabial furrow
12	Lip corner pull
15	Lip corner depress
17	Chin raise
20	Lip stretch
23	Lip tighten
24	Lip press
25	Lips part
26	Jaw drop
27	Mouth stretch
44	Eye squint

Figure 5: Code of "Action Units" [13]

The data that is provided gives the activation of an action unit frame-by-frame.

1. Using Classifiers : we had computed 3 features (ratio of number of frame of activation vs total number of frames, mean duration of activation, portion of the interview for activation) for each Action Unit combined for all the interviews and had tried to classify on the basis of those using different classifiers such as Random Forest classifier, Naive Bayes classifier, SVM classifier and Logistic Regression. However, the results were not very promising. [14]
2. Using AUs as sequential data and feed into LSTM. [14]
3. Combine the AU 1,2,4 - movement of eyebrows, AU9,10 - nose wrinkle and AU12,14,15,17,20,23,25,26,28 for full lip movement, AU45 for blinking and feed it into the LSTM. [15]

4.3 Model Integration

Further, we integrated the 3 models by use of Weighted Average. Weighted average is the multiplication of the predicted values with their respective weights and summation. The class is chosen on the basis of which predicted value is more. We tested various combinations of the weighted average in order to decide the final weights that will be used to integrate the results from the 3 individual models i.e., audio model, gaze model and the action units model. The weights that we chose were according to the accuracy of individual models. We gave the highest weight to Audio model since it had the highest accuracy. Some of the weights that we used were:

- 50% Audio, 30% Action Units, 20% Gaze
- 40% Audio, 30% Action Units, 30% Gaze

We tested this against the dataset to obtain the model which gives the highest accuracy among all the test combinations.

4.4 Application Architecture

We also created a ChatBot called DINA (Depression Interpretation Navigation Assistant) using Chatterbot module to interact with the user for collection of data. This ChatBot interacts with the user much like a human. The user speaks out the reply and this is then converted to text. For the server we used Flask. To collect the audio and video data, we used a Media Recording API to record the stream and save it. For the extraction of the features, we utilized OpenFace. It is an open source tool that is used for feature extraction and face land marking. Through this, we collected data pertaining to the action units and gaze. We then used HTML, CSS and JS to design a simple User Interface.

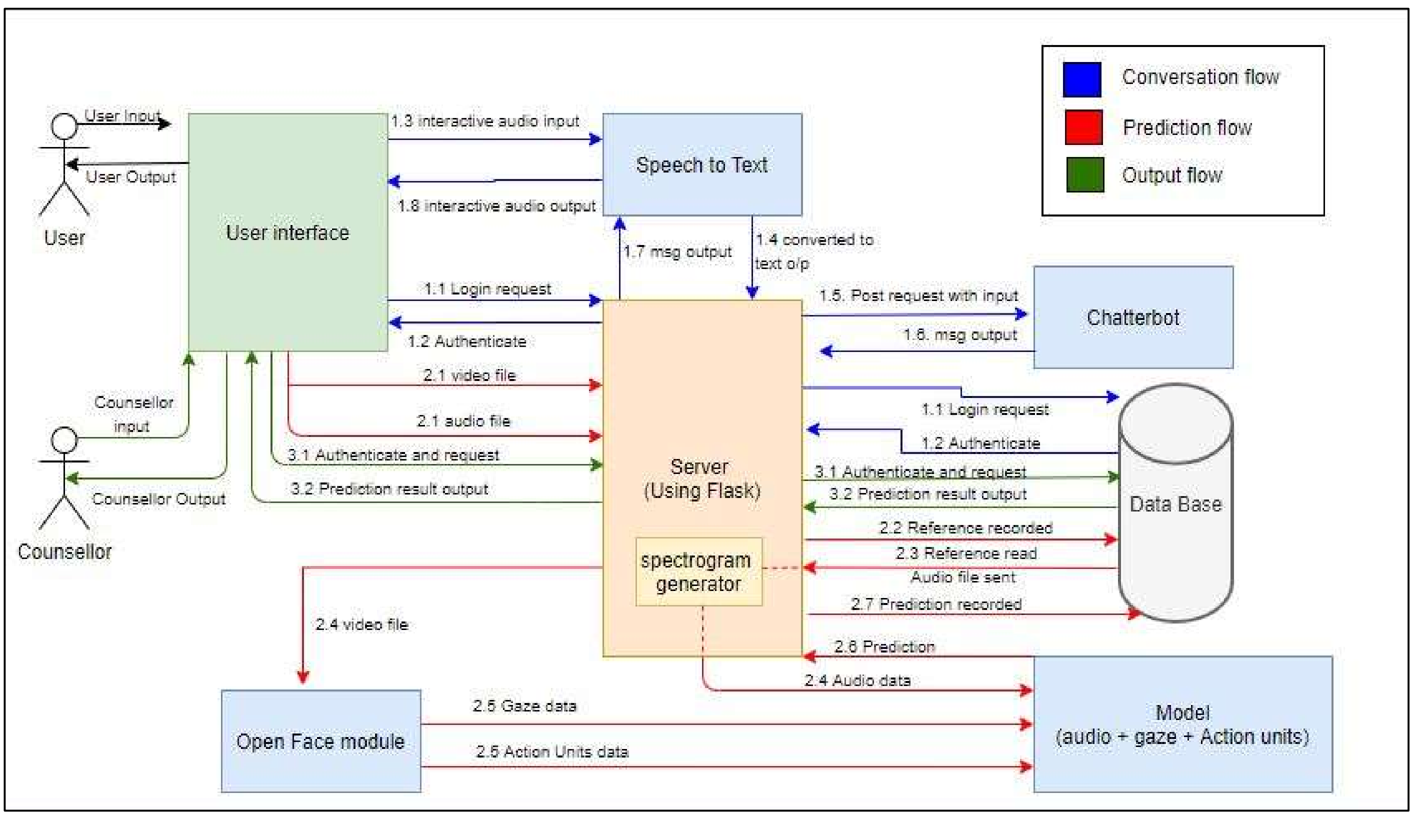


Figure 6: Flow Diagram for the application

The above figure depicts the flow of data and action of our application. The conversation flow (shown in blue) shows us how the user can interact with the Chat Bot. The predication flow (shown in red) shows us how the audio and video data is processed and stored. The output flow (shown in green) only runs when the counsellor logs in. This shows us how the results are displayed to the counsellor.

4.4.1 Process Flow

1. User logs in
2. User chats with the chatbot. The speech input is converted into text using WebSpeech API
3. The text input is then given to the chatbot which responds. The response is converted back into speech.
4. When the user is done, he can stop recording. He will be automatically logged out.
5. Meanwhile, the video and the audio are being recorded and is stored in the database.
6. A scheduled job runs every hour to process the audio and video file
7. It takes the audio file and converts it into a spectrogram
8. The video file is processed by OpenFace software to provide Gaze and "Action units" data
9. These data files are then sent to the model for prediction.
10. The predictions are stored in the database
11. The counsellor can log in and view the results through the User Interface.

4.4.2 User Interface

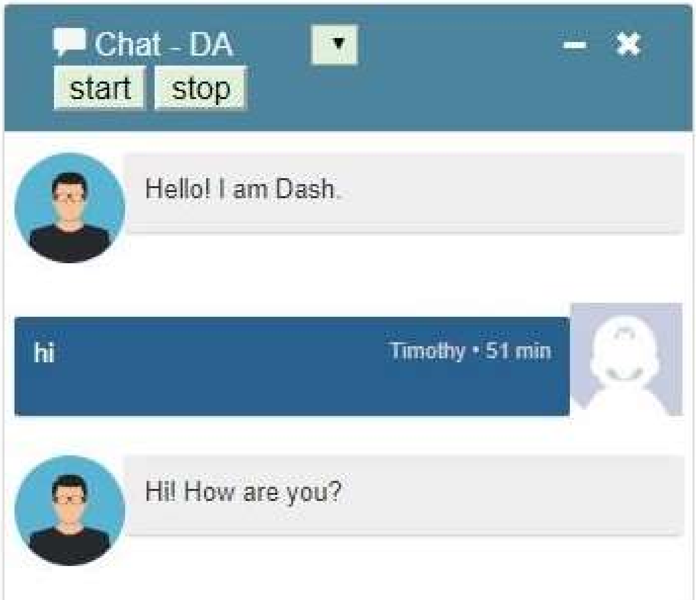


Figure 7: Chatbot flow

User is speaking into the microphone and the chat bot is replying. User text is in blue while the chatbot's replies are in grey username get records

Log Out

date	registration id	prediction
2020-05-18	2016ucp1004	Probability of not being depressed higher
2020-05-18	2016uec1013	Probability of not being depressed higher
2020-05-19	2016ucp1004	Probability of not being depressed higher
2020-05-19	2016ucp1023	Probability of not being depressed higher
2020-05-19	2016uec1013	Probability of not being depressed higher
2020-05-20	2016ucp1004	Probability of not being depressed higher

Figure 8: Results displayed to the Counselor

The counselor can log in and view the results of the users who have used the application in the past 3 days. The Counselor can also access the records for a specific user.

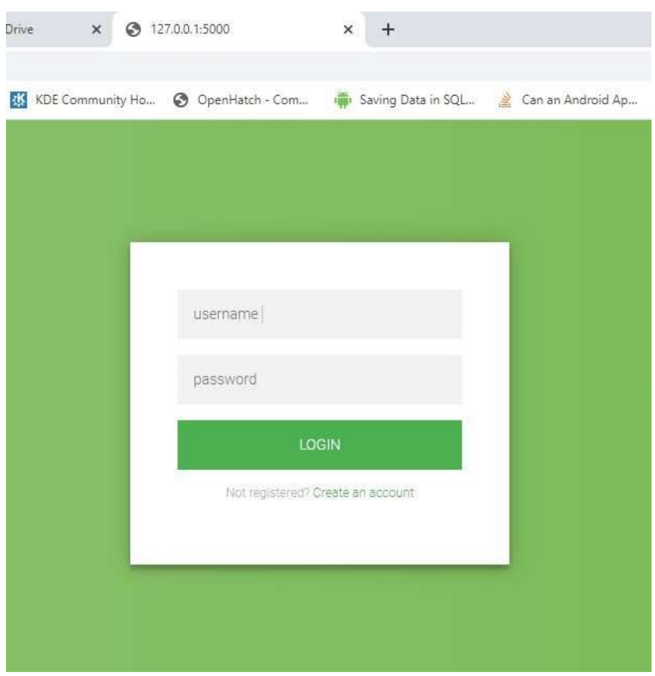


Figure 9: Login Page

The login page allows the user to authenticate by putting in the username and password. The password is converted into an encrypted word using SHA256 algorithm and compared with the stored algorithm.

The user is only allowed to log in once per day to maintain unique records of the audio and video files. The user is redirected to the login page after he has finished registration. This is done so that he can immediately start using the application.

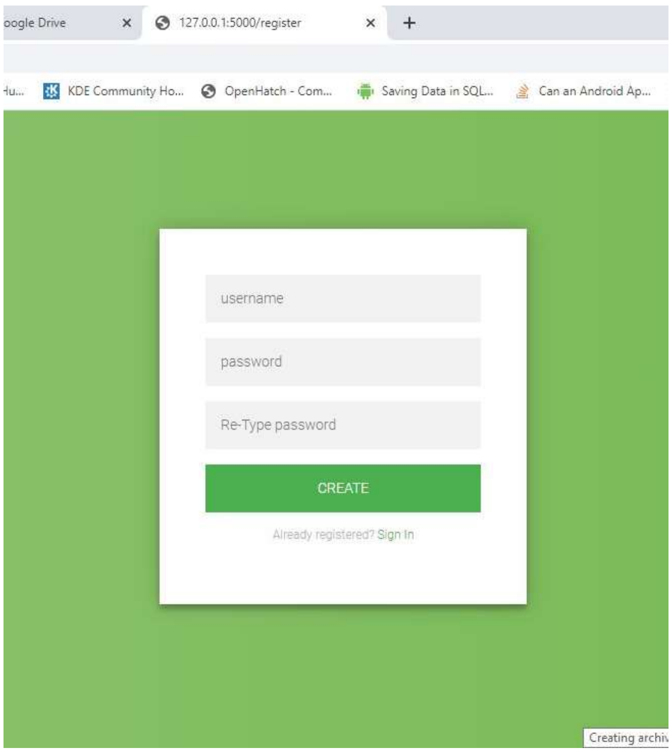


Figure 10: Registration Page

4.4.3 Database Design

To make our application multi-user, we had to store the conversations per user per session. This would enable the counselor to view all the conversations of a particular user and track their progress.

To do this, we created the following tables:

1. **User Registration table:**
   This contains specific user information like username and password. This would help authenticate the user when he tries to use the system.

Field	Type	Null	Key	Default	Extra
reg_id	varchar(20)	NO	PRI	NULL	
pwd	varchar(100)	NO		NULL	

Figure 11: Reg_log table

The above table is used for login and registration. It contains the records of users who are registered with the application. The reg_id is a primary key.

2. **User Conversation table:**
   Everytime the user has a conversation with the bot, the records are stored for the conversation. The records contain the links to video and audio files with their respective MD5 values. The MD5 values are computed to maintain the integrity of the files. The files are stored on the server file system. The scheduled job runs on an hourly basis to process the unprocessed files after checking the integrity of the files. The predictions are then stored in the table. These conversations and predictions can be accessed by the counselor.

Field	Type	Null	Key	Default	Extra
date	date	NO	PRI	NULL	
reg_id	varchar(20)	NO	PRI	NULL	
link_audio	varchar(500)	NO		NULL	
link_video	varchar(500)	NO		NULL	
md5_audio	varchar(50)	NO		NULL	
md5_video	varchar(50)	NO		NULL	
executed	varchar(2)	NO		N	
prediction	varchar(45)	NO		none	

Figure 12: dep_ana table

The dep_ana table is used to store the records for the user logging in the system, the video files and audio files generated and the prediction results. The date and reg_id attributes are the primary key while the reg_id is also the foreign key referencing the reg_log table. Any deletion from the reg_log table will cause cascading deletion from dep_ana table.

4.5 Future methodology

1. Diversifying our training data through further data collection
2. Making our models more robust by using the said data
3. Using previous records of a user for prediction over a period of time and exploring relevant weightages for the same
4. Implementation of security features to ensure user privacy

On obtaining more diverse data, we can train the mode for specific ethnic traits and acoustic uniqueness of various accents and languages.

Chapter 5

Experimental Results

```
[1]  mod=Sequential()
     mod.add(Masking(mask_value=0, input_shape=(None, 13)))
     mod.add(LSTM(6,return_sequences=True ))
     mod.add(LSTM(3,return_sequences=True ))
     mod.add(LSTM(3,return_sequences=False))
     # mod.add(Flatten())
     mod.add(Dense(3,activation='sigmoid'))
     mod.add(Dense(2,activation='sigmoid'))
     mod.compile(optimizer='sgd', loss='categorical_crossentropy', metrics = ['categorical_accuracy'])
```

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
masking_3 (Masking)          (None, None, 14)          0
_________________________________________________________________
lstm_10 (LSTM)               (None, None, 6)           504
_________________________________________________________________
lstm_11 (LSTM)               (None, None, 3)           120
_________________________________________________________________
lstm_12 (LSTM)               (None, 3)                 84
_________________________________________________________________
dense_6 (Dense)              (None, 3)                 12
_________________________________________________________________
dense_7 (Dense)              (None, 2)                 8
=================================================================
Total params: 728
Trainable params: 728
Non-trainable params: 0
_________________________________________________________________
```

Figure 13: LSTM model for gaze data

Here we used 3 LSTM layers sequentially, followed by 2 dense layers to train the gaze data that we had extracted through Open Face.

```
[39] mod.fit(X_train,y_train, validation_split=0.1 ,epochs=2, batch_size=128)
```

```
Train on 124 samples, validate on 14 samples
Epoch 1/2
124/124 [==============================] - 387s 3s/sample - loss: 0.5906 - categorical_accuracy: 0.7258 - val_loss: 0.7146
Epoch 2/2
124/124 [==============================] - 381s 3s/sample - loss: 0.5905 - categorical_accuracy: 0.7258 - val_loss: 0.7147
<tensorflow.python.keras.callbacks.History at 0x7f451b59f9b0>
```

```
[41] # mod.predict(X_test)
     mod.evaluate(X_test,y_test)
```

```
16/16 [==============================] - 82s 5s/sample - loss: 0.5694 - categorical_accuracy: 0.7500
[0.5694002509117126, 0.75]
```

Figure 14: Results from LSTM model for gaze data

As we can see here, the categorical accuracy of implementing the gaze LSTM model is 75%.

For audio, we used a Convolutional Neural Network. As input, we had given images of size 512x512 pixels.

```
[ ] model = Sequential()
```

```
[ ] model.add(Conv2D(32, kernel_size=3,strides=(1,1) , activation='relu', input_shape=(512,512,3)))
    model.add(MaxPooling2D(pool_size=(2, 2), strides=None, padding='valid', data_format=None))
    model.add(Conv2D(64,  kernel_size=3,strides=(1,1), activation='relu'))
    model.add(MaxPooling2D(pool_size=(2, 2)))
    model.add(Flatten())
    model.add(Dense(512, activation='relu'))
    model.add(Dense(512, activation='relu'))

    model.add(Dropout(0.5))
    model.add(Dense(num_classes, activation='softmax'))
    model.compile(loss=tensorflow.keras.losses.categorical_crossentropy,optimizer=tensorflow.keras.optimizers.Adadelta(),metrics=['accuracy'])
```

Figure 15: CNN model for audio spectrogram images

```
Model: "sequential_6"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_36 (Conv2D)           (None, 510, 510, 32)      896
_________________________________________________________________
conv2d_37 (Conv2D)           (None, 508, 508, 32)      9248
_________________________________________________________________
max_pooling2d_18 (MaxPooling (None, 253, 253, 32)      0
_________________________________________________________________
conv2d_38 (Conv2D)           (None, 249, 249, 64)      51264
_________________________________________________________________
conv2d_39 (Conv2D)           (None, 245, 245, 64)      102464
_________________________________________________________________
max_pooling2d_19 (MaxPooling (None, 122, 122, 64)      0
_________________________________________________________________
conv2d_40 (Conv2D)           (None, 118, 118, 128)     204928
_________________________________________________________________
conv2d_41 (Conv2D)           (None, 114, 114, 128)     409728
_________________________________________________________________
max_pooling2d_20 (MaxPooling (None, 56, 56, 128)       0
_________________________________________________________________
flatten_6 (Flatten)          (None, 401408)            0
_________________________________________________________________
dense_18 (Dense)             (None, 512)               205521408
_________________________________________________________________
dense_19 (Dense)             (None, 512)               262656
_________________________________________________________________
dropout_6 (Dropout)          (None, 512)               0
_________________________________________________________________
dense_20 (Dense)             (None, 2)                 1026
=================================================================
Total params: 206,563,618
Trainable params: 206,563,618
Non-trainable params: 0
```

Figure 16: CNN model layers for audio spectrogram images

After training over the CNN for audio, we observed the following results:

```
model.fit(X_train,y_train,validation_split=0.1 ,epochs=10)

Train on 124 samples, validate on 14 samples
Epoch 1/10
124/124 [==============================] - 52s 422ms/sample - loss: 3.0058 - acc: 0.5484 - val_loss: 2.0587 - val_acc: 0.8571
Epoch 2/10
124/124 [==============================] - 53s 427ms/sample - loss: 2.9587 - acc: 0.5968 - val_loss: 1.0326 - val_acc: 0.8571
Epoch 3/10
124/124 [==============================] - 53s 425ms/sample - loss: 2.7470 - acc: 0.5565 - val_loss: 0.5424 - val_acc: 0.8571
Epoch 4/10
124/124 [==============================] - 52s 420ms/sample - loss: 2.6864 - acc: 0.5565 - val_loss: 0.9706 - val_acc: 0.4286
Epoch 5/10
124/124 [==============================] - 52s 420ms/sample - loss: 2.2019 - acc: 0.5806 - val_loss: 1.4337 - val_acc: 0.1429
Epoch 6/10
124/124 [==============================] - 52s 415ms/sample - loss: 1.8938 - acc: 0.5323 - val_loss: 0.8677 - val_acc: 0.8571
Epoch 7/10
124/124 [==============================] - 52s 416ms/sample - loss: 1.7266 - acc: 0.5726 - val_loss: 1.0915 - val_acc: 0.8571
Epoch 8/10
124/124 [==============================] - 52s 417ms/sample - loss: 1.7627 - acc: 0.6210 - val_loss: 0.4824 - val_acc: 0.8571
Epoch 9/10
124/124 [==============================] - 53s 424ms/sample - loss: 1.2707 - acc: 0.6452 - val_loss: 1.0198 - val_acc: 0.3571
Epoch 10/10
124/124 [==============================] - 52s 423ms/sample - loss: 1.7866 - acc: 0.5968 - val_loss: 0.6081 - val_acc: 0.7857
<tensorflow.python.keras.callbacks.History at 0x7f75d48faa58>

model.evaluate(X_test,y_test)

16/16 [==============================] - 1s 48ms/sample - loss: 0.5512 - acc: 0.8750
[0.551202118396759, 0.875]
```

Figure 17: Training and accuracy from the CNN model

From the above image, we can observe that the accuracy of the CNN model for audio 87.5%

Next, we applied 2 different approaches for the Action Units, the results of both of which can be observed below.

Approach 1 for Action units:

This method used multiple classifiers to classify into depressed and non-depressed using each AU. The classifiers used are “Naïve Bayes”, “Logistic Regression”, “Support Vector Classifier” and “Random Forest Classifier”

Table 2: AU01

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.80	0.55
Logistic Regression	0.86	0.50
Support Vector Classifier	0.90	0.40
Random Forest Classifier	0.72	0.36

Table 3: AU02

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.80	0.55
Logistic Regression	0.86	0.50
Support Vector Classifier	0.80	0.30
Random Forest Classifier	0.72	0.36

Table 4: AU04

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.77	0.40
Logistic Regression	0.79	0.25
Support Vector Classifier	0.79	0.25
Random Forest Classifier	0.72	0.36

Table 5: AU05

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.75	0.30
Logistic Regression	0.79	0.25
Support Vector Classifier	0.87	0.25
Random Forest Classifier	0.72	0.30

Table 6: AU06

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.79	0.25
Logistic Regression	0.86	0.50
Support Vector Classifier	0.74	0.22
Random Forest Classifier	0.74	0.30

Table 7: AU09

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.77	0.40
Logistic Regression	0.60	0.15
Support Vector Classifier	0.79	0.25
Random Forest Classifier	0.50	0.18

Table 8: AU10

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.58	0.17
Logistic Regression	0.79	0.25
Support Vector Classifier	0.79	0.25
Random Forest Classifier	0.69	0.20

Table 9: AU12

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.67	0.40
Logistic Regression	0.71	0.40
Support Vector Classifier	0.60	0.25
Random Forest Classifier	0.72	0.36

Table 10: AU14

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.86	0.20
Logistic Regression	0.54	0.40
Support Vector Classifier	0.81	0.25
Random Forest Classifier	0.61	0.34

Table 11: AU15

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.71	0.40
Logistic Regression	0.62	0.25
Support Vector Classifier	0.69	0.50
Random Forest Classifier	0.82	0.46

Table 12: AU17

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.57	0.20
Logistic Regression	0.79	0.36
Support Vector Classifier	0.66	0.45
Random Forest Classifier	0.83	0.17

Table 13: AU20

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.59	0.30
Logistic Regression	0.79	0.25
Support Vector Classifier	0.80	0.15
Random Forest Classifier	0.83	0.36

Table 14: AU23

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.84	0.40
Logistic Regression	0.75	0.35
Support Vector Classifier	0.89	0.45
Random Forest Classifier	0.82	0.32

Table 15: AU25

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.77	0.50
Logistic Regression	0.69	0.25
Support Vector Classifier	0.75	0.35
Random Forest Classifier	0.87	0.46

Table 16: AU28

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.84	0.30
Logistic Regression	0.79	0.35
Support Vector Classifier	0.80	0.56
Random Forest Classifier	0.82	0.41

Table 17: AU45

Classifier	Accuracy (f1 score) for nondepressed	Accuracy (f1 score) for depressed
Naive bayes	0.88	0.45
Logistic Regression	0.77	0.25
Support Vector Classifier	0.87	0.05
Random Forest Classifier	0.60	0.26

Approach 2 for Action Units

Here we trained an LSTM model over the "Action Units" data that we had extracted using OpenFace.

```
Model: "sequential_3"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
masking_3 (Masking)          (None, None, 14)          0
_________________________________________________________________
lstm_11 (LSTM)               (None, None, 20)          2800
_________________________________________________________________
dropout_3 (Dropout)          (None, None, 20)          0
_________________________________________________________________
dense_6 (Dense)              (None, None, 10)          210
_________________________________________________________________
dense_7 (Dense)              (None, None, 2)           22
=================================================================
Total params: 3,032
Trainable params: 3,032
Non-trainable params: 0
_________________________________________________________________
```

Figure 18: LSTM model for AUs

```
Train on 124 samples, validate on 14 samples
Epoch 1/2
124/124 [==============================] - 2519s 20s/sample - loss: 0.6585 - acc: 0.7500 - val_loss: 0.6954 - val_acc: 0.5714
Epoch 2/2
124/124 [==============================] - 2669s 22s/sample - loss: 0.6368 - acc: 0.7581 - val_loss: 0.6976 - val_acc: 0.5714
<tensorflow.python.keras.callbacks.History at 0x7f6cbcfaca20>
```

Figure 19: Results for LSTM model for AUs

For the integration of models, we applied 2 approaches for the average weightage and their individual accuracies can be observed below:

Approach 1 for integration of models:

40% Audio, 30% Action Units, 30% Gaze

```
63.63636363636363 is the final accuracy
```

Approach 2 for integration of models:

50% Audio, 30% Action Units, 20% Gaze

```
72.72727272727273 is the final accuracy
```

Chapter 6

Conclusion

Through our research, we have found numerous audio and visual cues that can be used as inputs to develop a model for depression detection. The advantage of using these cues is that testing can be made more accessible, easy and frequent. It can also be used in high risk, high impact professions, across different age groups and economic backgrounds. As the model relies on a set of parameters which have been chosen after careful research, the assessment has the benefit of consistency that cannot be guaranteed when human judgement is applied for diagnosis. As a result, our model can pave the way more objective testing. We intend to explore a combination of models, some of which we have already explained and implemented. For current purposes, we have used a database which uses a carefully developed AI powered voice assistant for data collection. We also created our own framework consisting of a Chabot DINA embedded into a larger application for data collection so as to diversify the demographics of the dataset as well as improve accuracy. At present, our CNN has an accuracy of 87.5% and our LSTM shows an accuracy of 75%. After integration of models we observe an accuracy of 72%

References

[1] Kshitija Iyer, Zaved Khan Depression – A Review Research Journal of Recent Sciences Vol 1(4), 79-87 2012.

[2] J. M. C. J. F. Girard, Automated audiovisual depression analysis 2014 Current Opinion in Psychology 4:75-79

[3] Saad Albawi, Tariq A. Mohammed, Saad Al-Zawi Understanding convolution neural network 2010 IEEE Conferencex

[4] Sepp Hochreiter, J. Schmidhuber Long Short Term Memory Neural Computation 9(8) 1735-1780, 1997

[5] Gordan McIntyre, Roland Goecke, Melissa JG, Michael Breakspeaker, Matthew Hyett. An Approach for automatically measuring facial activity in depressed subjects Conference: Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference

[6] Stefan Scherer, Gale Lucas, Jonathan Gratch, Albert Rizzo, Louise-Phillipe M. Self-reported symptoms of depression and PTSD are associated with reduced vowel space in screening interviews IEEE Transactions on Affective Computing 2015

[7] Jeffrey F. Cohn, Tomas SK, Iain M, Fernando De La Torre et al. Detecting Depression from Facial Actions and Vocal Prosody Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference

[8] Yang Y, Fairbairn CE, Cohn JF. Detecting depression severity from vocal prosody. IEEE Transactions on Affective Computing. 2013;4:142–150

[9] Jeffrey M Girard, Jeffrey F Cohn Automated Audiovisual Depression Analysis 2014 Current Opinion in Psychology 4:75-79

[10] Jonathan Gratch, Louise- Phillipe M, Giota Stratou, Stefan Scherer. Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender 2014 Journal on Multimodal User Interfaces 9

[11] Giota Stratou, Gale Lucas, Stefan Scherer, Marwa Mahmoud, Jill Boberg, Jonathan Gratch Automatic audiovisual behavior descriptors for psychological disorder analysis 2013 Image and Vision Computing 32

[12] Jonathan Gratch et. Al DAICWOZ Database 2014

[13] Paul Eckman, Wallace V. Friesen Facial signs of expression of emotion 1978

[14] Jeffrey F. Cohn, Tomas SK, Iain M, Fernando De La Torre et al. Detecting Depression from Facial Actions and Vocal Prosody Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference

[15] Gordan McIntyre, Roland Goecke, Melissa JG, Michael Breakspeaker, Matthew Hyett. An Approach for automatically measuring facial activity in depressed subjects Conference: Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference

Appendix I

Timestamp	Where are you currently studying?	Age	Q1	Q2	Q3	Q4
5/18/2020 12:07	College/university	22	Yes	No	No	Yes
5/18/2020 12:12	College/university	22	Yes	No	No	No
5/18/2020 12:12	College/university	21	No	No	Yes	Yes
5/18/2020 12:16	College/university	22	No	No	No	Yes
5/18/2020 12:19	College/university	22	Yes	No	No	Yes
5/18/2020 12:24	College/university	21	Yes	No	No	Yes
5/18/2020 12:24	College/university	22	Yes	No	No	Yes
5/18/2020 12:25	College/university	22	Yes	No	No	Yes
5/18/2020 12:29	College/university	21	No	No	Yes	Yes
5/18/2020 12:45	College/university	22	Yes	No	Yes	Yes
5/18/2020 12:47	College/university	23	Yes	No	No	Yes
5/18/2020 12:49	School	17	Yes	No	No	Yes
5/18/2020 12:51	College/university	21	Yes	Yes	Yes	Yes
5/18/2020 12:51	College/university	22	Yes	No	No	Yes
5/18/2020 12:52	College/university	21	Yes	No	No	Yes
5/18/2020 12:53	College/university	22	Yes	No	No	Yes
5/18/2020 12:53	College/university	21	Yes	No	Yes	No

Figure 20: Data from Student survey 1

5/18/2020 12:54	College/university	21	Yes	No	Yes	Yes
5/18/2020 12:55	College/university	23	Yes	No	Yes	Yes
5/18/2020 12:57	College/university	21	Yes	No	No	Yes
5/18/2020 12:59	College/university	21	No	No	No	Yes
5/18/2020 13:00	College/university	22	Yes	No	No	Yes
5/18/2020 13:00	College/university	21	Yes	No	No	Yes
5/18/2020 13:01	College/university	20	Yes	No	No	Yes
5/18/2020 13:01	College/university	22	No	No	No	Yes
5/18/2020 13:02	College/university	23	Yes	No	Yes	No
5/18/2020 13:03	College/university	22	Yes	No	No	Yes
5/18/2020 13:04	College/university	22	No	No	Yes	Yes
5/18/2020 13:05	College/university	22	No	No	No	No
5/18/2020 13:05	College/university	22	No	No	Yes	Yes
5/18/2020 13:06	College/university	21	Yes	No	No	Yes
5/18/2020 13:07	College/university	21	No	No	No	Yes
5/18/2020 13:07	College/university	22	Yes	No	No	Yes
5/18/2020 13:08	College/university	21	Yes	No	Yes	Yes
5/18/2020 13:09	College/university	21	Yes	No	No	Yes
5/18/2020 13:10	College/university	21	No	No	No	Yes

Figure 21: Data from Student survey 2

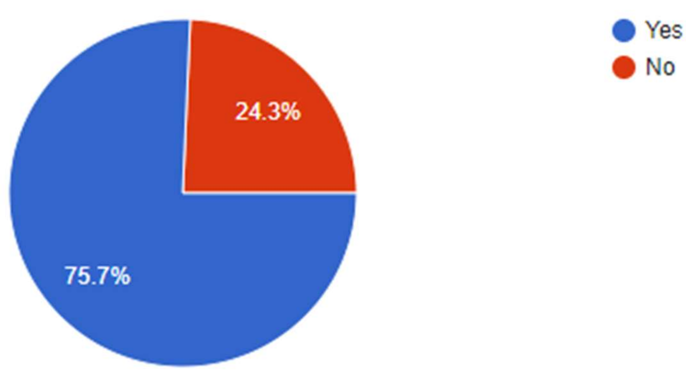


Figure 22: Percentage of students who have experienced at least one symptom of depression (loneliness, fatigue, changes in appetite, changes in sleep or difficulty in concentrating)

In the above graph, responders who have answered yes are students who have experienced at least one symptom of depression while responders who have answered no are students who have not experienced at least one symptom of depression

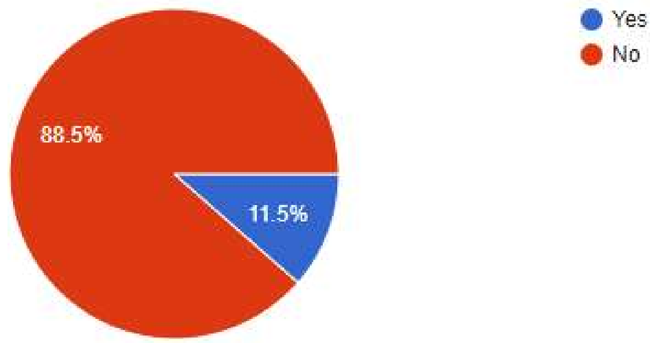


Figure 23: Percentage of students who have consulted a mental health professional or counsellor

In the above graph, responders who have answered yes are students who have consulted a mental health professional or counsellor while responders who have answered no are students who have not consulted a mental health professional or counsellor

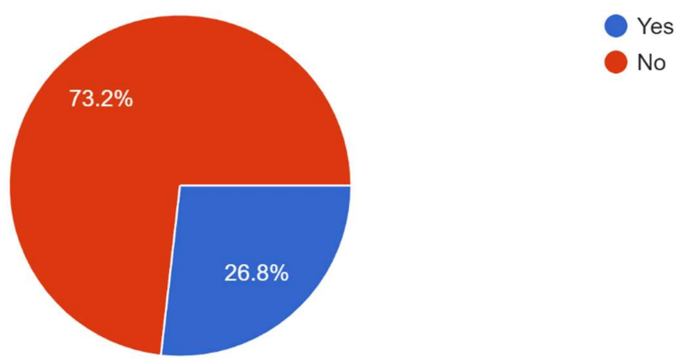


Figure 24: Percentage of students who believe that there are adequate facilities and awareness to help those suffering from mental health problems

In the above graph, responders who have answered yes are students who believe that there are adequate facilities and awareness to help those suffering from mental health problems while responders who have answered no are students who do not believe that there are adequate facilities and awareness to help those suffering from mental health problems

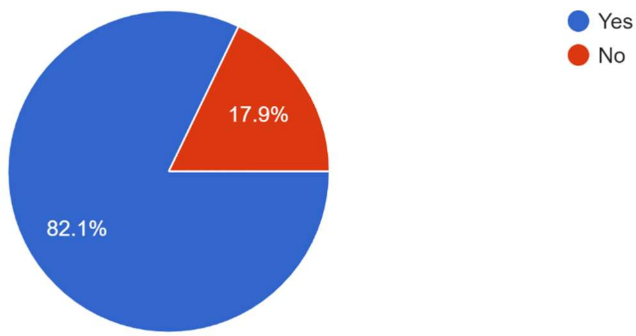


Figure 25: Percentage of students who would feel comfort or ease in using an online application that could screen for early signs of depression

In the above graph, responders who have answered yes are students who would feel comfort or ease in using an online application that could screen for early signs of depression while responders who have answered no are students who would not feel comfort or ease in using an online application that could screen for early signs of depression

Timestamp	Q1	Q2	Q3	Q4
5/18/2020 12:23	No	No	Yes	Yes
5/18/2020 12:33	Yes	No	No	Yes
5/18/2020 13:00	Yes	No	No	Yes
5/18/2020 14:52	Yes	No	No	Yes
5/18/2020 14:52	Yes	No	No	Yes
5/18/2020 14:52	No	No	No	Yes
5/18/2020 14:52	No	No	No	Yes
5/18/2020 14:53	Yes	Yes	No	Yes
5/18/2020 14:53	Yes	No	No	Yes
5/18/2020 14:54	No	No	No	Yes
5/18/2020 14:54	Yes	No	No	Yes
5/18/2020 14:56	Yes	No	Yes	Yes
5/18/2020 14:57	No	No	No	Yes
5/18/2020 14:59	Yes	No	Yes	Yes
5/18/2020 15:05	No	No	No	Yes
5/18/2020 15:11	Yes	No	No	Yes
5/18/2020 16:58	No	No	No	Yes
5/18/2020 17:11	No	No	Yes	Yes
5/18/2020 17:47	Yes	No	Yes	Yes
5/18/2020 17:51	Yes	No	No	Yes
5/18/2020 18:33	Yes	No	Yes	Yes
5/18/2020 19:39	Yes	No	No	Yes
5/18/2020 20:24	Yes	No	Yes	Yes
5/19/2020 9:43	Yes	No	No	No
5/19/2020 11:31	Yes	No	No	Yes
5/19/2020 11:33	Yes	No	No	Yes
5/19/2020 11:35	Yes	No	No	Yes
5/19/2020 11:40	Yes	Yes	Yes	No
5/19/2020 13:42	Yes	Yes	No	Yes
5/19/2020 14:08	Yes	No	Yes	No
5/19/2020 14:54	No	No	No	Yes
5/20/2020 7:42	Yes	No	Yes	No

Figure 26: Data from Medical Professional survey

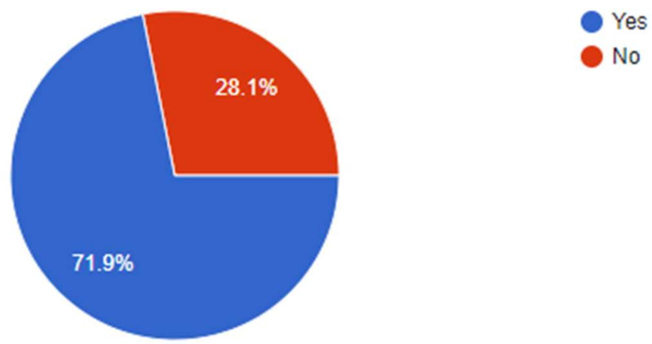


Figure 27: Percentage of doctors who have experienced at least one symptom of depression (loneliness, fatigue, changes in appetite, changes in sleep or difficulty in concentrating)

In the above graph, responders who have answered yes are medical professionals who have experienced at least one symptom of depression while responders who have answered no are medical professionals who have not experienced at least one symptom of depression

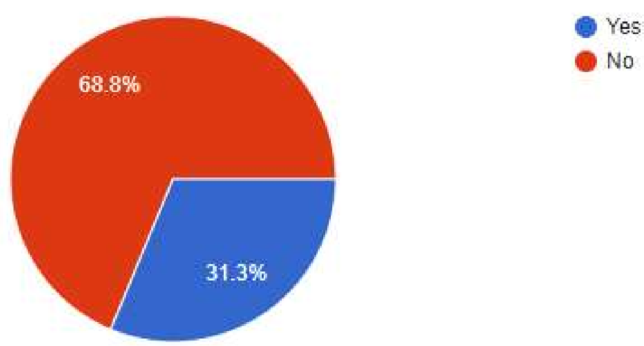


Figure 28: Percentage of doctors who believe that there are adequate facilities and awareness to help those suffering from mental health problems

In the above graph, responders who have answered yes are medical professionals who believe that there are adequate facilities and awareness to help those suffering from mental health problems while responders who have answered no are medical professionals who do not believe that there are adequate facilities and awareness to help those suffering from mental health problems

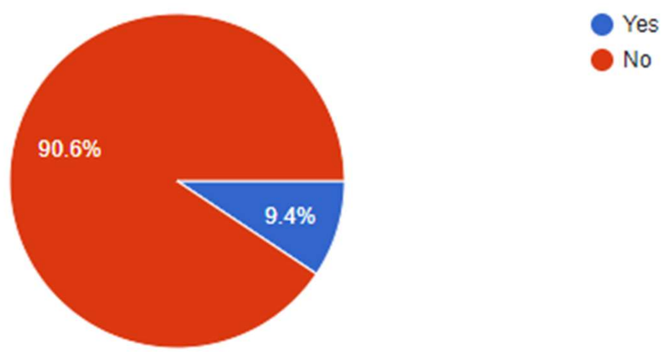


Figure 29: Percentage of doctors who have consulted a mental health professional

In the above graph, responders who have answered yes are medical professionals who have consulted a mental health professional or counsellor while responders who have answered no are medical professionals who have not consulted a mental health professional or counsellor

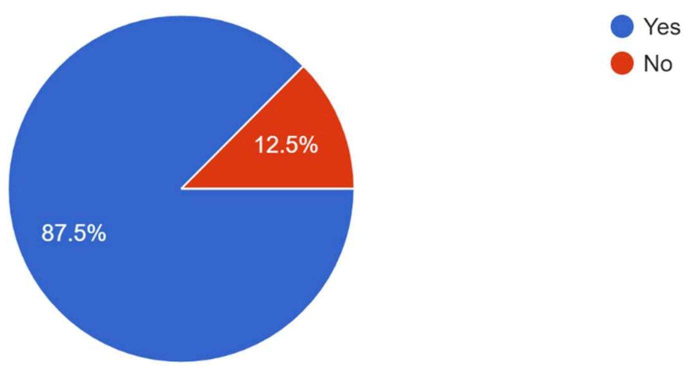


Figure 30: Percentage of doctors who would feel comfort or ease in using an online application that could screen for early signs of depression

In the above graph, responders who have answered yes are medical professionals who would feel comfort or ease in using an online application that could screen for early signs of depression while responders who have answered no are medical professionals who would not feel comfort or ease in using an online application that could screen for early signs of depression

Vandita assignment

ORIGINALITY REPORT

14%	5%	7%	12%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Malaviya National Institute of Technology Student Paper	2%
2	Submitted to CSU, San Jose State University Student Paper	2%
3	Submitted to University of Nottingham Student Paper	1%
4	onlinelibrary.wiley.com Internet Source	1%
5	mafiadoc.com Internet Source	<1%
6	Gordon McIntyre, Roland Gocke, Matthew Hyett, Melissa Green, Michael Breakspear. "An approach for automatically measuring facial activity in depressed subjects", 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, 2009 Publication	<1%

7	www.iainm.com Internet Source	<1%
8	Submitted to DeVry, Inc. Student Paper	<1%
9	towardsdatascience.com Internet Source	<1%
10	Submitted to California State University, Sacramento Student Paper	<1%
11	Submitted to University of Adelaide Student Paper	<1%
12	"Suicide in Adolescence", Springer Nature, 1986 Publication	<1%
13	Submitted to National Institute of Technology, Kurukshetra Student Paper	<1%
14	Submitted to Coventry University Student Paper	<1%
15	arxiv.org Internet Source	<1%
16	Submitted to Murdoch University Student Paper	<1%
17	Submitted to University of Edinburgh Student Paper	<1%

18	link.springer.com Internet Source	<1%
19	Submitted to University of Wolverhampton Student Paper	<1%
20	Submitted to University of Bath Student Paper	<1%
21	aime-static.textpert.ai Internet Source	<1%
22	"Natural Language Processing and Information Systems", Springer Science and Business Media LLC, 2017 Publication	<1%
23	Fabien Ringeval, Maja Pantic, Björn Schuller, Michel Valstar et al. "AVEC 2017", Proceedings of the 7th Annual Workshop on Audio/Visual Emotion Challenge - AVEC '17, 2017 Publication	<1%
24	Submitted to KDU College Sdn Bhd Student Paper	<1%
25	Submitted to Glasgow Caledonian University Student Paper	<1%
26	digitalcommons.wpi.edu Internet Source	<1%
27	docplayer.net Internet Source	<1%

28	Submitted to University of Surrey Student Paper	<1%
29	Submitted to Mississippi State Board for Community & Junior Colleges Student Paper	<1%
30	Submitted to University of Newcastle upon Tyne Student Paper	<1%
31	Submitted to Gulf College Oman Student Paper	<1%
32	Submitted to University of Wales Institute, Cardiff Student Paper	<1%
33	Submitted to Temasek Junior College Student Paper	<1%
34	www.pitt.edu Internet Source	<1%
35	Submitted to Laureate Higher Education Group Student Paper	<1%
36	Submitted to Delhi Technological University Student Paper	<1%
37	Submitted to City University of Hong Kong Student Paper	<1%
38	Submitted to University of Sydney Student Paper	<1%

39	Submitted to University of Lincoln Student Paper	<1%
40	Jeffrey M Girard, Jeffrey F Cohn. "Automated audiovisual depression analysis", Current Opinion in Psychology, 2015 Publication	<1%
41	www.medicalnewstoday.com Internet Source	<1%
42	Submitted to University of Nicosia Student Paper	<1%
43	Submitted to Imperial College of Science, Technology and Medicine Student Paper	<1%
44	Submitted to Loughborough University Student Paper	<1%
45	Submitted to Radboud Universiteit Nijmegen Student Paper	<1%
46	medvr.ict.usc.edu Internet Source	<1%
47	Submitted to University Of Tasmania Student Paper	<1%
48	Submitted to Texas A & M University, Kingville Student Paper	<1%
49	Submitted to University of Warwick	

Student Paper

<1%

50	"Communications, Signal Processing, and Systems", Springer Science and Business Media LLC, 2020 Publication	<1%
51	www.mdpi.com Internet Source	<1%
52	Submitted to Central Queensland University Student Paper	<1%
53	Submitted to The University of Manchester Student Paper	<1%
54	University of Tennessee, Knoxville Publication	<1%
55	Communications in Computer and Information Science, 2015. Publication	<1%
56	"Intelligent Computing and Optimization", Springer Science and Business Media LLC, 2020 Publication	<1%
57	Meredith G Harris, Sandra Diminic, Philip M Burgess, Georgia Carstensen, Gavin Stewart, Jane Pirkis, Harvey A Whiteford. "Understanding service demand for mental	<1%

health among Australians aged 16 to 64 years according to their possible need for treatment", Australian & New Zealand Journal of Psychiatry, 2014
Publication

58	Submitted to RMIT University Student Paper	<1%
59	Qingxiang Wang, Huanxin Yang, Yanhong Yu. "Facial Expression Video Analysis for Depression Detection in Chinese Patients", Journal of Visual Communication and Image Representation, 2018 Publication	<1%
60	Submitted to Aberystwyth University Student Paper	<1%
61	Submitted to Liverpool Hope Student Paper	<1%
62	Stefan Scherer, Louis-Philippe Morency, Jonathan Gratch, John Pestian. "Reduced vowel space is a robust indicator of psychological distress: A cross-corpus analysis", 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015 Publication	<1%
63	"Neural Information Processing", Springer Science and Business Media LLC, 2017 Publication	<1%

64	Silvia Basaia, Federica Agosta, Luca Wagner, Elisa Canu, Giuseppe Magnani, Roberto Santangelo, Massimo Filippi. "Automated classification of Alzheimer's disease and mild cognitive impairment using a single MRI and deep neural networks", NeuroImage: Clinical, 2018 Publication	<1%
65	Submitted to University of Computer Studies Student Paper	<1%
66	Victoria H. Coleman, Michele M. Carter, Maria A. Morgan, Jay Schulkin. "United States obstetrician-gynecologists' accuracy in the simulation of diagnosing anxiety disorders and depression during pregnancy", Journal of Psychosomatic Obstetrics & Gynecology, 2009 Publication	<1%
67	Anastasia Pampouchidou, Panagiotis G. Simos, Kostas Marias, Fabrice Meriaudeau, Fan Yang, Matthew Pediaditis, Manolis Tsiknakis. "Automatic Assessment of Depression Based on Visual Cues: A Systematic Review", IEEE Transactions on Affective Computing, 2019 Publication	<1%