Smart Companion Agent for Mental Well-being through Deep Learning and NLP

by

Rafiur Khan
(16101087)
Abdullah Al Sohel
(19341008)
Farhana Azad Shreyashee
(16101096)
Shamima Hossain
(17101429)
Mahin Fiaz
(16101269)

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc in Computer Science

Department of Computer Science and Engineering
Brac University
January 2021

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Rafiur Khan

Rafiur Khan
(16101087)

F.A. Shreyashee

Farhana Azad Shreyashee
(16101096)

Shamima Hossain

Shamima Hossain
(17101429)

Abdullah Al Sohel

Adullah Al Sohel
(19341008)

Mahin Fiaz

Mahin Fiaz
(16101269)

Approval

The thesis titled "Smart Companion Agent for Mental Well-being through Deep Learning and NLP" submitted by

1. Rafiur Khan (16101087)
2. Abdullah Al Sohel (19341008)
3. Farhana Azad Shreyashee (16101096)
4. Shamima Hossain (17101429)
5. Mahin Fiaz (16101269)

Of Spring, 2021 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on June 6, 2021.

Examining Committee:

Supervisor:
(Member)

Md. Golam Rabiul Alam,PhD
Associate Professor
CSE Department
BRAC University

Co Supervisor:
(Member)

Mr. Annajiat Alim Rasel
Senior Lecturer
CSE Department
BRAC University

Head of Department:
(Chair)

Sadia Hamid Kazi, PhD
Chairperson and Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

Mental disorders are an unfortunate reality among the general population nowadays. Conditions like anxiety; depression may seem trivial on the surface but have serious consequences on an individual's life. These disorders have shown to be detrimental to health and hamper a person's general well being. In severe cases, if mental disorders go unnoticed and untreated they can cause permanent damage to one's personality, drive him/her to social isolation and in worst cases compel the person to commit suicide as a means to end their suffering. Therefore, a need for proper detection and awareness of such diseases in a person emerges. Mental disorders may not show physical symptoms in a person but it is possible to find patterns in people with a potential mental disorder and detect them with the help of modern Machine learning techniques. In addition to this, such methods are completely automated and non-invasive; as a result these systems can also help continuously monitor a person's mental state. We propose a system that can take various physiological signal readings from the human body as a way to predict distress. Upon detecting a user's distress, the system tries to converse with the user trained by a knowledge base of conversations of people suffering from mental disorders and can interact with the user in a conversation-like interface as a companion. For this we used a system consisting of BioBERT models(separately for questions and answers) and a couple of FCNN layers.

Keywords: BioBERT, Transformer, Mental health, Machine learning techniques,Signals.

Motivation

Mental health has been a real challenge for this world nowadays. It is one of the keys to keep yourself happy and healthy. But often, our mental health does not get the same priority as our physical health. It has always been ignored for the last few years. People don't talk about mental well being that much. Moreover, people having anxiety or stress disorder often carry a fear of getting judged while expressing their own problems to someone else. That is one of the main reasons when a person is having any mental or anxiety disorder he prefers to keep it inside himself. A lot of time it has been observed that people don't have the courage to seek professional help or talk about their mental well being. Due to this reason, the rate of suicide has also been increased over last few years. That is why we want to come up with a solution with the help of deep machine learning and natural language processing, where people can talk about their mental well beings with an automated bot. This bot will detect whether a person is under any stress or anxiety through PPG, EMG, EDA signals and will continuously talk with the person to make him feel better.

Acknowledgement

First and foremost, we want to show our gratitude to Allah for helping us complete this work without any kind of major difficulties. Next, we want to thank our parents who supported us utmost in reaching where we are now. Furthermore, we want to thank our honorable Supervisor who helped us by giving continuous feedback and suggestions on improving our work and reaching the ultimate goal.

Table of Contents

List of Tables

List of Figures

List of abbreviations(alphabetically)

1. ANN - Artificial Neural Network
2. ASR - Automatic Speech Recognition
3. AUC - Area Under The Curve
4. BLSTM/Bi-LSTM - Bidirectional Long Short Term Memory
5. BP - Blood Pressure
6. BSN - Body Sensor Network
7. COM - Center of Mass
8. CTC - Connectionist Temporal Classification
9. DBP - Diastolic BP
10. DDQN - Double Deep Q-Learning
11. DNN - Deep Neural Network
12. DQN - Deep Q-Learning
13. DTW - Dynamic Time Warping
14. E2E - End to end
15. EDA - Electrodermal Activity
16. EMG – Electromyography
17. EOS - End Of Sentence
18. FG - Fine Gained
19. GRU - Gated Recurrent Units
20. HMM - Hidden Markov Model

21. k-NN - K-nearest-neighbor
22. LCS - Largest common substring
23. LDA - Linear Discriminant Analysis
24. LPC - Linear Predictive Coding
25. LSTM - Long Short Term Memory
26. MFCC - Mel Frequency Cepstral Coefficient
27. MIST - Montreal Imaging Stress Task
28. MLP - Multilayer Perceptron
29. MT - Translation modules
30. PPG – Photoplethysmogram
31. PTSD - Post-traumatic stress disorder
32. PWTT - Pulse Wave Transfer Time
33. PWV - Pulse Wave Velocity
34. RNN - Recurrent Neural Network
35. ROC - Receiver Operating Characteristic
36. SBP - Systolic BP
37. Seq2Seq - Sequence-to-Sequence
38. ST - Speech to Text
39. TTS - Text-to-Speech
40. WER - Word error rate environment.

Chapter 1

Introduction

1.1 Introduction

Mental health is a crucial aspect of well-being. Despite being so, it is a much-neglected part when it comes to taking care of one's mental health compared to other physical diseases. Diseases such as depression, anxiety and other traumas may seem trivial at glance as many people do not even consider them diseases since they do not show physical symptoms. However such diseases are as detrimental to well-being and people suffering from them can be compared to fighting an invisible and lethal enemy that slowly eats away a person's peace. On the other hand, while seeking professional help, many people face social stigma and discrimination. Many are still not aware that mental disorders exist and that it should be treated like any other diseases. Therefore, in case of mental diseases, it is a challenge to detect, diagnose and seek treatment. Hence, this calls for a way to monitor and help people detect and make them aware if they are suffering from such diseases.

Recently in this era of big data, IoT and advancements in fields of medicine with applications of ML and other Artificial Intelligence areas, it is now possible to learn from people's behaviours online and from daily activities and predict if they have a disease or not. This has shown results with good accuracy when applied in many medical conditions and we believe such an approach could help us design a smart system which could detect and monitor mental disorders as well.

1.2 Problem Statement

In recent years, the world has seen a surge of cases involving mental disorders. Different varieties of such diseases such as depression, anxiety, PTSD are continuously on the rise. Moreover according to [1], such diseases have affected populations of varying age from adolescents to people of old age. Past research has focused mostly on identifying, diagnosing and discussing preventive measures to curb these disorders. In addition to these, other problems for example: proper treatment of mental disorders such as social stigma and discrimination are a hindrance to the patients receiving proper treatment. However, little work has been done to put corrective measures and help people already suffering such disorders with proper care and assistance in our modern world where we have little time to look after ourselves. If these problems persist and all we do is apply traditional treatment methods of going to a psychiatrist and counselors, which may be possible for some of the affected population, but for the vast majority having access to such treatment facilities is a luxury. Finally, according to [2], extensive studies on population of renowned Western countries have shown that if people with mental disorders remain unaware of their condition or are left untreated, their conditions take turn for the worst as time goes and may lead them to cause personal harm and in extreme cases even commit suicide. Therefore, we intend to solve this problem by using modern technology and ideas that would help detect and monitor people's mental state. This system would learn from some physiological signals readings as well as apply analysis and detection techniques. And tie up the findings to identify if users are suffering from aforementioned conditions. In the long run, we envision that such a system would help prevent mental disorders as people would be aware of the gradual changes in their mental state and take preventive measures to curb on set of disease.

1.3 Research Objective

Our primary objective of this research is to design a system that can assess the user's mental health conditions. Our research aims to improve mental health conditions of users especially who might not have access to professional help or are afraid of the social stigma behind mental illness. In addition, it would serve as a companion to the user in his/her time of need. Our model would accurately determine a user's mental state by analyzing and applying advanced machine learning algorithms on their textual data. We will ensure accuracy with the help of several physiological signals which would reinforce the model's prediction. To sum up our research goals are-

1. Creating a system that can predict the user's mental health condition
2. Let the users monitor mental condition through continuous conversation like a human.

1.4 Research Organizing

In our paper there are many chapters. Chapter 1 contains the introduction and research objective of the paper. Chapter 2 contains review of papers and works related to our proposed system. Chapter 3 elaborates the datasets that we used in our system. Chapter 4 describes our overall proposed system. Chapter 5 elaborates the implementation of our model and its results. Chapter 6 concludes our paper with how we can further improve our system and the limitations we faced during our research.

Chapter 2

Literature review

2.1 Literature Review

In this section, we present relevant works for achieving a good conversational agent and also provide background on physiological signals to process emotions. Emotions and mental states have always seemed arbitrary as there does not seem to be concrete laws governing their behaviour. However in the recent decades MIT Researchers have instigated great deal of enthusiasm among researchers who work with technologies involving emotions or affective computing [3]. Machine learning algorithms have shown promising results with problems involving detection and the technology keeps evolving for the better as hardwares becomes more robust and our understanding of application of ML algorithms increases. Our work aims to bridge two different technologies namely a conversation agent and a system which recognizes the emotional state of a person, coupled together provide a wholesome system which helps distressed people in their time of need. We have seen their separate applications in various studies either independently with the usage of physiological signals or by learning from user's conversational data. For instance in a study done by JI-Won Back Kyungyong Chung [4]. Mental illness is influenced by different factors like physical activities, and diseases. There were signs of continuous depression, irregular sleep, and suicide chances. In another research[5]. the researchers conducted a study on posts of a group of reddit users who published about mental health information and then moved to explain suicidal ideation in the future. Several features estimated this shift: heightened self-focus, poor linguistic style matching with the community, reduced social engagement, and expressions of hopelessness, anxiety, impulsiveness, and loneliness. Furthermore, the researchers in [6]. demonstrated that found social media contains useful signals for characterizing the signs of depression in personals, as calculated through decrease in social activity, gained negative effect, heavily clustered ego networks, heightened relational and medicinal concerns, and larger expression of the involvement of religion. Also in [7]. it was discussed that a serious challenge in personal and public health formed from depression. Tens of millions of people suffer from depression every year and only a little portion come for adequate treatment.

2.2 Physiological signals used for psychoanalysis

We plan to incorporate 3 sensor datasets namely EDA, PPG and EMG signals which are relevant to our scope of work. These were previously used in various studies involving human emotions and are widely accepted.The most important signal for our purpose is EDA- electro dermal activity which is extracted from the DEAP dataset.

2.2.1 EDA signals

EDA measures continuous variation in the electrical characteristics of our skin. These changes in the skin can be triggered by external factors or internal factors in the human body which is why this signal is of great interest to us. Various studies [8] [9] [10] show that the nervous system is connected to the sweat glands on our skin. These glands secrete sweat in response to external or internal stimuli which causes changes in the skin's electrical conductivity. According to [11] EDA is one of the earliest signals to be used for quantifying emotional excitement. In [12] the researchers bring forth an extensive system which analyses human emotional arousal with the help of EDA. EDA was also used to measure stress, where an open source tool was proposed which was able to predict stress with an accuracy of 92%. It is also used in the field of virtual reality where the researchers used a kernel based extreme-learning machine [13] to classify different levels of stress in an environment which simulated stressful conditions, achieving 95% accuracy. These are just a few examples of the vast area of research that is ongoing in the field of psychoanalysis and undoubtedly EDA plays a huge role in it. To summarize theories in [14] EDA measures the emotional arousal we experience when we are stressed, fearful, happy and various other emotional states. It is possible to do so because whenever we experience emotional stimulation, sweat glands on our skin secrete sweat causing changes in the skin's electrical conductivity. These electrical changes are indicative of our levels of arousal measured as skin potential, resistance, conductance etc. EDA is measured in microSiemens(mS) .This signal enables us to access the otherwise autonomously controlled nervous system of the body to detect emotional arousal that can be translated to emotional states offering us an unadulterated view of our own psychological processes in a completely non invasive way. Therefore we can use skin conductance to reflect and measure cognitive and sympathetic responses. As a result, emotional states like stress, happiness and sadness can be detected using EDA. Lastly, owing to the fact that measuring EDA is relatively affordable, easy to operate and non-invasive it has been widely used in the field of psychoanalysis and various psychological researches. As studied from [8] [9] [10] EDA or GSR signals consist of two main components that are indicative of our emotional behaviourSCL - skin conductance level SCR - skin conductance response

The skin conductance level is also referred to as the tonic component of the signal. It slowly changes over time. The time scale is usually of tens of seconds to tens of minutes. It is not particularly informative on its own. It is shown that tonic level varies widely from person to person as it depends on their hydration, skin dryness etc. Therefore the vast majority of professionals and researchers have deemed it to be a poor measure of skin's conductance. The other component, skin conductance response is of particular importance to us. It is also known as the phasic component. They are shown as GSR bursts or peaks if a graph of their measure is plotted against time. This phasic component is sensitive to specific stimuli that are emotionally arousing. They occur when EDA amplitudes cross a certain threshold value in a time period within 1-5 seconds after a stimulus is initiated.

For each of the SCR peaks detected, we can further extract a number of features such as amplitude, rise-time, latency. These features are indicative of a person's emotional state quantified by valence and arousal. For our purposes, we will be putting this information into use and apply various machine learning models to find out which features best reflect a person's valence and arousal and thus build a classification model.

A limitation EDA signals is that it can only identify emotional arousal but does not give us any additional information on what kind of emotional arousal for example excited or distressed an individual is in. So for future iterations we plan to use a more sophisticated classifier to improve accuracy and incorporate two other datasets namely EMG and PPG as they all together would reinforce the classification produced by EDA.

2.2.2 PPG Signals

To summarize theories in [15], Photoplethysmography (PPG) is usually used nowadays for continuous sign monitoring. It measures the blood volume changes non-invasively at the skin surface. An oximeter is used in PPG that shows the skin and measures changes in light absorption. In different words, it is used for measuring the blood pressure. In order to measure the blood, the instruments have to be either cuff-based or invasive. It also cannot be used continuously because of their characteristics .PPG based continuous rate monitoring is often utilized in many domains. As our research is about assessing the mental states, we are able to estimate mental distress from PPG. Mental stress is harmful to cardiovascular health. Excessive stress can trigger cardiac death. A spike in pace, pressure, and viscus contraction are all typical vascular responses to mental stress. Because of its negative effects on health and productivity, such an aberrant response to mental stress has become a social liability.

In this paper [16], the authors demonstrated the practicality of getting accurate heart rate variability (HRV) by analyzing the photoplethysmography (PPG) signals that are recorded by a reflectance-mode pulse oximeter sensor that can be attached to the forehead. According to them, the HRV data can be really useful for combat medics to determine the health condition of a soldier and it is even possible to predict the possibility of mortality using it. They used a reflectance-mode pulse oximeter which was attached to the head and analyzed the data using computer software (LabVIEW). This demonstrated that it is possible to extract the HRV signals using the method along with breathing rate, HR, and SpO2 and these reports can be used by the combat medics to optimize their care and can potentially change a life and death situation for a wounded soldier on the battlefield.

According to [17], HRV can be collected from PPG as an alternative to electrocardiography (ECG). In order to use the HRV, ECG signal is essential which is used to detect the RR intervals. However, ECG signals are not always available easily for the personal health application. It also requires more instruments for the setup. On the other hand, PPG is easier to use as it requires a single optical sensor whereas ECG requires multiple electrodes. The authors used pulse rate variability (PRV) features from the PPG signal as proxy of HRV indexes and compared each other. For that purpose, they used three techniques: 1. Spearman's rank correlation (SRC), 2. Normalized root mean square error (NRMSE) and 3. Wilcoxon's rank sum test (WRST). The result showed that PRV can be used as an alternative to HRV as the correlation was 82% higher in both time-domain and frequency-domain features. So, it confirms that PRV from PPG is a representative of ECG-based HRV and PPG signal can be used for the HRV analysis.

This paper [18] discussed the measurement of HRV from PPG signal as compared to ECG signal. The authors showed the high correlation between PPG signal and ECG signal. PPG signals are derived from the PP intervals whereas ECG signals use RR intervals. HRV features are categorized in three domains: 1. The time domain, 2. frequency domain and 3. Poincare´ plot (non-linear). They computed both PP interval method for PPG signal and RR intervals method for ECG signal. The differences and errors were much insignificant in those two methods. Total ten volunteers had been experimented in the process and features from both PPG and ECG showed in table 1. The actual error between those two signals' intervals were much lower.

Parameters	**ECG**	**PPG**	**Actual Error**
Time domain			
Mean NN(ms)	800.0 ± 94.7	800.0 ± 94.7	0.1±0.1
SDNN(ms)	52.0 ± 15.8	52.4 ±15.4	0.9±0.5
RMSSD(ms)	49.0 ± 25.0	50.3 ± 23.1	2.4±1.4
SDSD(ms)	49.1 ± 94.7	25.0 ± 50.4	23.1±0.1
NN50(Count)	45.1± 24.5	46.2 ±24.1	3.9 ±2.4
Frequency domain			
LF nu	44.72± 13.17	44.39 ±12.43	1.93± 2.02
HF nu	55.28 ± 13.17	55.61± 12.43	1.93± 2.02
LF/HF ratio 0.92 ±0.54	0.90± 0.51	0.04± 0.04	
Poincare domain			
SD1(ms)	34.7 ±17.7	35.6± 16.4	1.7 ±1.0
SD2(ms)	64.4 ±16.3	64.6± 16.1	1.1± 0.8
SD ratio	0.51 ±0.15	0.53± 0.13	0.03± 0.02

Table 2.1: The comparison of three features parameters derived from ECG and PPG signals [18]

To sum up, it is clear that HRV features can also be reliably extracted from PPG signal with PP interval method. It can be collected from the finger-tips using an oximeter and computes the HRV.
According to [19], a python package, "Neurokit", can be used to conduct interval related analysis. So we can compute the HRV indices and get MeanNN, SDNN, and RMSSD intervals. This package also includes a dataset that can be used for testing. It does not come with the package to avoid its weight. After analysing from the dataset, the package finds the peaks from the PPG signals of individuals. Then it can extract HRV indices such as time-domain, frequency-domain and non-linear(Poincare´ plot).
So these intervals are essential to determine mental state from PPG

signals. Those intervals can be analyzed and used with other signals. That's how it can determine an individual's mental condition.

2.2.3 EMG signals

EMG signal is a combination of actions of muscle fibers. This signal measures electrical activity of muscle both at rest and stress. We can extract information about muscle fibers by analyzing this signal. EMG signal accelerates activity of different motors. According to [20], if there is any problem with muscle or any neuromuscular junction getting affected, then a huge impact on the signals can be seen. In [20] this paper the authors reported an experiment in which they monitored EMG signals from the upper trapezius muscle under three distinct stressful conditions. The stress tests comprised mathematical work , a logical puzzle activity, and a memory challenge. The results demonstrate that the amplitudes of the EMG signals are much larger during stress compared to rest conditions. Furthermore, mean and median frequencies were significantly lower during stress than during rest. EMG characteristics were shown to be linked to subjectively stated stress levels.According to their research [20] amplitude of EMG signals are quite higher when a person is under any stress condition whereas amplitude of EMG Signal is lower when somebody is resting or at rest position. From the recorded EMG signals and findings of desired features value it has been observed that, in rest conditions,average Mean frequency , median value, mean value , RMS value is quite lower than the stress condition. Using this property of EMG signals we can easily differentiate when a person is under stress [20]

	R1	R2	R3	R4	R5	S1	S2	S3	Mean R	Mean S	S-R
RMS (%RVC)	11.3	14.6	11.6	13.7	13.4	15.7	15.1	15.9	12.9	15.5	+2.6
Static (%RVC)	8.3	11.6	8.1	9.5	9.6	8.1	9.5	8.4	9.4	8.7	-0.8
Median (%RVC)	10.0	13.7	10.8	12.0	12.7	15.4	14.8	14.8	11.8	15.0	+3.1
Peak (%RVC)	16.3	18.8	15.4	18.9	17.8	23.9	19.9	22.7	17.4	22.2	+4.7
Gaps/min	8.8	5.0	4.8	7.8	7.7	6.6	4.2	6.4	6.8	5.7	-1.1
Relative time with gaps (%)	44.7	32.3	47.5	41.2	42.7	28.5	20.7	33.1	41.7	27.4	-14.3
MNF (Hz)	147.5	141.4	140.6	149.9	136.3	152.5	140.5	147.2	147.7	139.1	-8.6
MDF (Hz)	112.7	107.7	106.1	116.1	102.8	119.8	106.0	112.4	113.7	105.0	-8.8
Trends (%RVC/s) $*10^{-3}$	-4.1	-29.6	36.1	-17.7	-2.7	-73.6	-12.1	-58.3	-36.7	7.1	+43.8

Figure 2.1: Features value of EMG Signals during both rest conditions (R1-R5) and stress condition (S1-S3) [20]

2.3 Advanced NLP

2.3.1 Seq2Seq models

One of the basic criteria that our model should fulfill is that it should be able to converse with the user. In this section, we are trying to explore how such a feat can be achieved. The underlying technology used to make conversation between humans and machines possible is a Machine Learning architecture called Sequence-to-Sequence model, which is uniquely suited to this task. Having RNNs at the core of them which can act as artificial persistent memories like that of the human brain. In this section we compare some Seq2Seq models, explore their applications to audio and text in context to our problem. Then we explore two possible approaches to solving our problem.

2.3.2 Comparing seq2seq models with different RL methods:

In paper [21] the authors have compared many variations of seq2seq models using different RL methods and evaluated them by scores from ROUGE, BLEU, CIDEr, METEOR, SPICE and WER. The purpose of the paper was to summarize some of the most significant works that tried to combine two separate strategies, Reinforcement learning and seq2seq models, and include an open-source library for the abstractive text summary problem that demonstrates how to train a seq2seq model using various RL strategies. They tried to perform some tasks like- Machine Translation, Text Summarization Headline Generation, Question Generation, Question Answering, Dialogue Generation, Semantic Parsing, Image Captioning, Video Captioning, Computer Vision, Speech Recognition, Speech Synthesis etc. The paper explained the procedure in 5 parts- Algorithm 1 Training a simple seq2seq model, Algorithm 2 REINFORCE algorithm, Algorithm 3 Batch Actor-Critic Algorithm, Algorithm 4 Deep Q-Learning, Algorithm 5 Double Deep Q-Learning. The authors later discussed how they developed an open source library that implements the following features:Adding temporal attention and intra-decoder attention that was proposed in [22] Adding scheduled sampling along with its differentiable relaxation proposed in [23] and E2EBackProb [24] for solving exposure bias problem Adding adaptive training of REINFORCE algorithm by minimizing the mixed objective loss Providing self-critical training by adding as the baseline the greedy reward. Providing Actor-Critic training choices for model training using Value Network, DQN, DDQN, and Dueling Network asynchronous training. Providing scheduled sampling options for Q-Function training in DQN, DDQN, and Dueling Net

From the comparison of results shown in the paper we tried to implement a chatbot for our paper with a focus on the seq2seq model.

2.3.3 Using a single seq2seq model to translate audio of one language to text of another language)

In paper [25] the authors wanted to prove that the seq2seq model alone was powerful enough to convert speech from one language

and translate it to some other language. They used three jointly trained neural networks- A recurrent encoder that transforms an input sequence into a sequence of hidden activation's, frames. A decoder network that, by next stage prediction, emits a sequence of output tokens: emitting one output token per stage.
One of its features is that the design of the model is essentially the same as that of an ASR neural system based on attention. Within the same computational footprint as speech recognition, direct speech-to - text

2.3.4 Using a seq2seq model to generate relative conversations

This research [26] is for conversational agent development that generates conversations using recurrent neural networks and their memory unit. The authors trained a single layer LSTM during this experiment using stochastic gradient descent with gradient clipping. Their suggested model was represented in very few measures. Next, it is to determine what knowledge is going to be thrown away from the state of the cell. A sigmoid layer called the forgotten gate layer shapes this decision. Second, the sigmoid layer selects which values to modify in the 'input gate layer.' The tanh layer then generates a vector that may be used to add new candidate values to the state. In the following stage, these two are combined to display state changes. The performance will be based primarily on the state of the cell, but will be an iterated model. They started with a sigmoid layer, which dictated which sections of the cell state they would produce. Then they use tanh to multiply the cell state by the sigmoid gate output, ensuring that just the items they wish to output are output. The vocabulary consists of 20 K words, including special tokens indicating turn and actor, which are the most common. They used a sampled softmax function as their loss function. Their proposed model had the attention mechanism with the seq2seq model. They used Cornell movie-dialog dataset. For our model we followed similar architecture suggested by this paper.

2.3.5 Using Listen, Speller and Attention to enhance decoder and language model with seq2seq model

In paper [27], an attention-based seq2seq speech recognition system was analyzed by the authors that transcribe recordings directly into characters. Two shortcomings were observed: overconfidence in its predictions and a propensity when language models are used to generate incomplete transcriptions. They suggest realistic solutions to both problems that achieve competitive speaker-independent word error rates on the Wall Street Journal dataset. Their method of speech recognition builds on the recently proposed network of Listen, Attend and Spell [28]. It is an attention-based model of seq2seq that can transcribe an audio recording directly into a sequence of characters delimited by room. It uses an encoder-decoder architecture, like other seq2seq neural networks: an acoustic-modeling listener module, a speller module emitting characters and an attention module acting as an intermediary between the speller and the listener. The listener is a Bi-LSTM multilayer network that converts a sequence of N frames of acoustic features into a sequence of hidden activations that may be shorter. The speller measures the likelihood of a series of characters based on the listener's activation. The speller uses the attention function to find a collection of appropriate activations of the listener to emit a character and summarize them in a context. Using a single LSTM layer, they implemented the recurrent phase. Using a SoftMax function, the output character distribution is computed. Using beam quests, the best results are received. By separating its effects on model accuracy and efficacy in beam search, they investigated the influence of model trust. Three techniques designed to avoid incomplete transcripts were contrasted. The first strategy does not change the beam search criteria, but prevents the EOS token from being released unless its probability is within a specified range of the most likely token. Alternatively, to allow long transcripts, the necessity for beam search can be extended. Finally, they suggest using a coverage definition that counts the number of frames that have received more than the threshold cumulative publicity. On the Wall Street Journal dataset, all of their studies were performed. Using 4 bidirectional LSTM layers, their base configuration introduced the Listener. A single LSTM layer was the Speller. 3 convolution filters were used by the attention MLP. Via excellent results, they have shown that the sequence-to - sequence approach to speech recognition can be competitive with other non-HMM methods, such as CTC, with successful regularization and careful decoding. For our model we tried to implement the attention mechanism as it showed in this paper to boost seq2seq models performance.

2.3.6 Chatbot using BioBERT model

To summarize from [29], Bidirectional Encoder Representations from Transformers (BERT) is a language representation model. It is transformer-based ML technique that is used in Natural Language Processing (NLP) pre-training. It can be used to analyze the context of words from user's search queries. Instead of using directional model which reads the text input from left to right or right to left, Bidirectional from Transformer model reads the entire text by categorizing words (or sub-words) with its encoder. It uses the attention mechanism which is used to learn contextual relations between the words from a sentence with its encoder and decoder. BERT only needs encoder part as its aim is to create a language model through NLP pre-training tasks. The authors addressed two approaches to fine-tuning pre-trained language representations: feature-based and feature-based fine-tuning. They argue that these procedures, particularly fine-tuning procedures, limit the power of pre-trained representations. So, by introducing the BERT model, they hope to improve fine-tuning-based approaches. Pre-training and fine-tuning are two steps in their BERT architecture. BERT is trained on unlabeled data during pre-training by completing various pre-training tasks. The pre-trained parameters that use labeled data from downstream activities are used in fine-tuning mode. For their BERT model, the authors ran four experiments: GLUE, SQuAD v1.1, SQuAD v2.0, and SWAG. The General Language Understanding Evaluation (GLUE) is a set of natural language comprehension exercises. For the data utilized in GLUE tasks, they employed a batch size of 32 and three epochs. The Stanford Question Answering Dataset, or SQuAD v1.1, contains around 100k crowd-sourced question/answer pairings. Its goal is to anticipate the answers from the questions and excerpts in Wikipedia. SQuAD v2.0 is an upgraded version of v1.1 that assumes the possibility if the paragraph lacks a quick answer. They fine-tuned batch size 32 with 3 epochs and batch size 48 with 2 epochs in SQuAD v1.1 and v2.0, respectively. The GLUE, SQuAD v1.1, and SQuAD v2.0 scores were 80.5 percent, 93.2 percent, and 83.1 percent, respectively, after performing eleven NLP tasks. Overall, these results win deep unidirectional representations of low-resource tasks, and their goals are to generalize by creating deep bidirectional structures to tackle other NLP tasks.

The authors of this research [29] developed a model called BioBERT that may be utilized for text mining in the biomedical sector. It is a pre-trained biomedical language representation model that aids in retrieving relevant information about medical objects using natural language processing (NLP). Because of the shift in word distribution from general domain corpora to biomedical corpora, it still does not produce effective results. The authors' goal is to introduce BioBERT, which is based on BERT, another pre-trained language model. BERT outperforms the current best model, which is also utilized in biomedical text mining. BioBERT, on the other hand, can outperform both of those models. BERT and the most up-to-date model tackle three text mining tasks: 1. biomedical named entity recognition (NER), 2. biomedical relation extraction (RE), and 3. biomedical question answering (QA). By improving on the results of those tasks, the authors show that BioBERT can outperform BERT and state-of-the-art models. To do so, they suggest that the current state-of-the-art paradigm, which relies on general domain corpora, has limitations for NLP approaches. As a result, estimating performance on datasets containing biomedical literature is problematic. As a result, it needs to be trained on biomedical corpora, and it was pre-trained on eight Nvidia GPUs for 23 days. They enhanced their performance on the following biomedical text mining tasks: NER (F1 score 0.62 percent), RE (F1 score 2.80 percent), and QA (F1 score 2.80 percent) (MRR 12.24 percent). The BioBERT was pre-trained and then fine-tuned for the authors' research, like the BERT model. Overall, by adapting BERT for biomedical corpora, BioBERT can achieve more solid results on various NLP tasks for biomedical purposes

This article [30] described a medical chatbot that used the BERT and GPT2 models to provide an alternative to immersive doctor-patient interaction. The author demonstrated the value of a chatbot in the medical field, where a patient can utilize it to get faster answers to questions in an emergency while the doctor is absent. When a patient asks a question, the chatbot is taught to respond with appropriate facts using the question-answer pair method. If the patient is unsatisfied with the responses, the chatbot will be able to recommend some previously answered questions and answers pairings. The functionality of this chatbot is made up of two primary components and numerous sub-components. To begin, the chatbot will use a seq2seq task with supervised train using prior question-answer pairs to construct an answer. Second, if the patient is not happy, the chatbot will suggest an alternate answer using semantic search in an unsupervised train. This approach will employ neural network characteristics to construct supervised learning binary classification question-answer embedding. A medical question-answer dataset [4] with 25k question-answer pairings was employed to do this. This chatbot's overall architecture is made up of seven steps. The first section requires adding a question from the patient, after which the query will be forwarded to the following section, which involves the BioBERT model. BioBERT employs semantic search to extract embedding for previously answered question-answer pairs. The fourth stage employs FAISS to look for answers by utilizing the embeddings extractor network. The replies will be "cosinely" comparable to the main question and will return the same question-answer pair in sorting order as Q1-A1, Q2-A2, Q3-A3. With question-answer pairings like this Q3A3Q2A2Q1A1Q, the main question will be included. Following this, the question-answer pair with the main question will be concatenated and submitted to the GPT2 model, which will produce an output with the desired response for the main question, such as Q3A3Q2A2Q1A1QA', where A' is the generated answer. When fine-tuning the BioBERT model, the authors choose to ensure that no tag is included in the main question in the embedding section rather than generating random negative samples. As a result, the chatbot will not classify similar question-answer pairs as negative examples. To summarize, this chatbot presents not only the produced response, but also the ranking previously replied question-answer pairs that are comparable to the patient's primary inquiry

Chapter 3

Datasets

This section describes the signal datasets used in this research . We also describe CBT dataset that is used to train the chabot to produce context relevant responses

3.1 DEAP dataset

In order to build a valid and reliable model, we used a well-known and comprehensive multidimensional dataset called DEAP [31] for studying human affective states. DEAP dataset is widely used to study human affective states. This dataset was recorded with evidence from 32 participants who viewed 40 one-minute music video extracts. While watching 40 one-minute videos, few physiological signals were recorded from 32 participants. The recorded physiological signals are the horizontal electrooculogram (hEOG), vertical electrooculogram (vEOG), zygomaticus major electromyogram (zEMG), trapezius major electromyogram (tEMG), GSR, respiration belt results, plethysmogram, and body temperature. Among them we have used 3 signal's recordings from this dataset and they are EMG, PPG and GSR. EMG signal is divided into two parts - zEMG and tEMG.The subjects rated the levels as continuous values of arousal, valence, liking, dominance, and familiarity. The rating was done on a scale of 1-9.The participants who scored themselves less then 5 are assigned to a value of 0 and the participants who scored themselves greater than 5 or equal to 5 are assigned to a value of 1.

3.1.1 Data pre-processing

The data is preprocessed by applying downsampling, EOG removal, filtering, segmenting etc for EMG, PPG and GSR recordings. Data was first downsampled to 128Hz before being divided into 60 second trials with a 3 second pre-trial baseline removed. We have also mean imputed missing values and applied standard scaling to ensure reliable results for the classification task.

3.2 Feature extraction

3.2.1 GSR signal

In order to effectively classify emotional state a number of features from the signals need to be separated. These features have a correlation with the label class of our interest namely the valence, arousal of each of the participants that they self-assessed. For each of the SCR peaks detected in the GSR signals, we extracted time-domain features as follows: rise-time, latency, amplitude, half-amplitude, half-amplitude index,half-amplitude index Pre, EDA at apex, decay-time, SCR-width. The figure below shows the decomposed EDA signal for the first participant and shows his SCR peaks using neurokit2 python package [19]

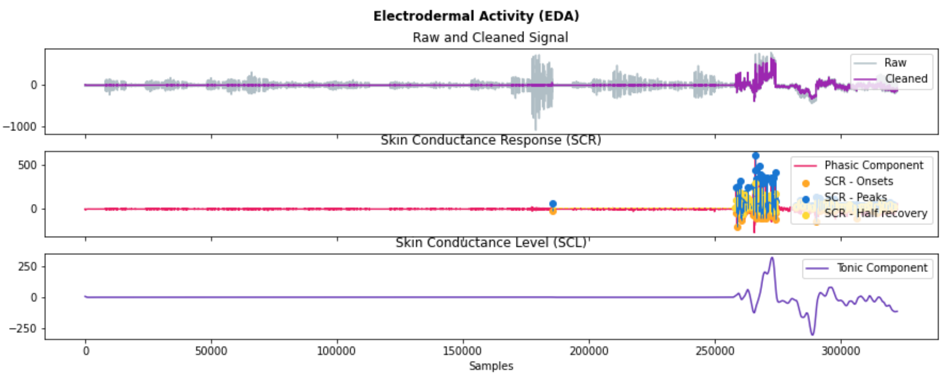


Figure 3.1: Decomposed GSR signal for the first participant

3.2.2 PPG signal

We have collected various features from the mentioned dataset for the PPG signal. We have found total 47 features by processing the signal from the dataset. The features are known as Heart rate variables (HRV). These variables are categorized into three indices. These indices are 1. Time domain 2. Frequency domain and 3. Non-linear domain (Poincare´ plot).

Time domain HRV metrics: RMSSD, MeanNN, SDNN, SDSD, CVNN, CVSD, MedianNN, MadNN, MCVNN, IQRNN, pNN50, pNN20, TINN, HTI

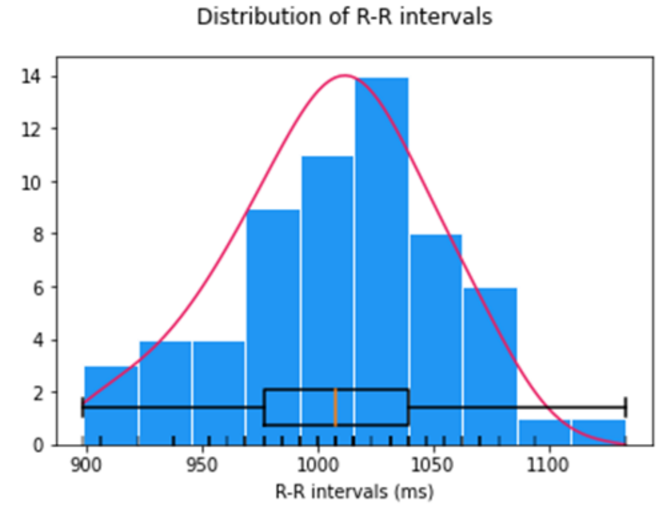


Figure 3.2: Time domain HRV metrics

These are frequency-domain indices of Heart Rate Variability (HRV) metrics: HF, VHF, HFn, LnHF

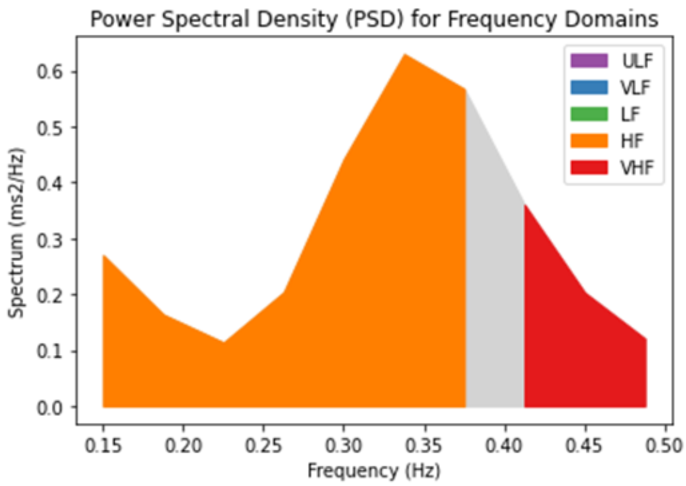


Figure 3.3: frequency domain HRV metrics

The remaining indices are nonlinear (Poincar´e plot) indices of Heart Rate Variability (HRV): SD1, SD2, SD1SD2, CSI, CVI, PIP, IALS, PSS, PAS, GI, SI, AI, PI, C1d, C1a, SD1d, SD1a, C2d, C2a, SD2d, SD2a, Cd, Ca, SDNNd, SDNNa, ApEn, SampEn

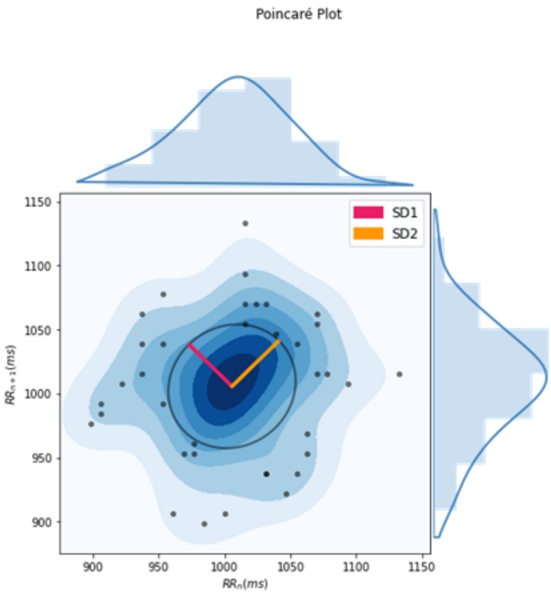


Figure 3.4: non-linear HRV metrics

Equation of Poincaré plot:

$$\begin{aligned} \mathrm{SD1}^2 &= \gamma_{RR}(0) - \gamma_{RR}(m) \\ \text{or,}\mathrm{SD1} &= \mathrm{F}(\gamma_{RR}(0), \gamma_{RR}(m)) \end{aligned} \quad (1)$$

AND

$$\begin{aligned} \mathrm{SD2}^2 &= \gamma_{RR}(0) + \gamma_{RR}(m) - 2(\bar{RR})^2 \\ \text{or,}\mathrm{SD2} &= \mathrm{F}(\gamma_{RR}(0), \gamma_{RR}(m)) \end{aligned}$$

Where SD1 is the dispersion perpendicular to the line of identity (minor axis) and SD2 is the dispersion along the line of identity (major axis). RR (m) is the autocorrelation function for lag-m RR interval. [32]

3.2.3 EMG signal

For the EMG signal, we extracted several features from the DEAP dataset. Here, EMG signals are divided into two groups: zEMG and tEMG. We found a total of 13 features by analyzing each signal from the dataset.The collected features are -wave length,zero crossing,willison amplitude,Number of times the slope of the EMG signal changes sign (SSC),Simple Square Integral (SSI),Mean Absolute Value (MAV),MAV1,MAV2,Mean Power using PSD,Mean Frequency of the PSD,Power spectrum ratio of the EMG power spectrum,RMS of EMG,Integrated EMG.

3.2.4 Cognitive behavioral therapy (CBT) dataset

In our paper we needed CBT datasets to train our bot to have context-relevant communication with our users. However, due to the confidentiality issue and the HIPAA (Health Insurance Portability and Accountability Act) law we could not get hold of such a dataset. Instead, we will be using a combination of YouTube videos' captions where they role play some interaction with patients as well as some questions and answers dataset from prominent medical websites such as eHealth forum, iCliniq, QuestionDoctors, WebMD where real doctors have provided public answers to the questions asked by patients to train our chatbot. In order to train our model, we separated the questions and answers. The total dataset is about 300,000 questions and answers. Our current method shows the potential of training a chatbot in this context. It also gives the prospect of promising results when trained with a more proper CBT dataset .

Our current method shows the potential of training a chatbot in this context. It also gives the prospect of promising results when trained with a more proper CBT dataset .

Chapter 4

Proposed Method

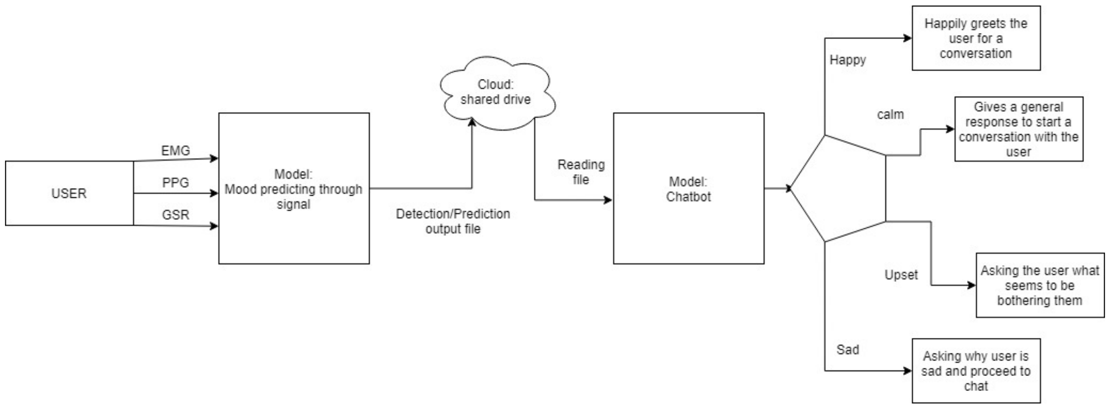


Figure 4.1: Overall System Architecture

For our system the signals feed the mood predicting model that tries to predict the current mood of the user receiving the signals. The prediction is sent to a shared cloud folder in google drive. The chatbot model reads the file and if it finds that the user is in a sad state then prompts a response asking why the user is sad and proceeds to chat with the user. However if the user is happy then the chatbot does not generate any responses but there still remains an option to chat with the bot.

4.1 Overall Proposed Model

At the very first step, we took readings from 3 signals -GSR, EMG and PPG . These three signals data are chosen from the DEAP dataset. Secondly, features were extracted from these 3 signals individually. After extraction, these features were fed into the XGBOOST model for accurate prediction. The model was trained by tuning hyperparameters consisting of objective function called binary logistic, maximum tree depth of 100, learning rate of 0.1, alpha value of 10 and

100 estimators. Furthermore retrieved features are passed into the trained model for prediction and it applies binary classification based on the Arousal and Valence label combination and consequently produce happy, sad, calm or upset states. These four mental states are sent to a file in the cloud. Finally, based on that reading, if our chatbot receives any negative emotions from the cloud file such as sad or upset our chatbot will initiate conversation with the user accordingly. On the other hand if our chatbot receives any positive emotions from the cloud file such as happy or calm , it responds and converses with the user accordingly.

The chatbot system architecture of our system consists of separate fine-tuned BERT models for both the questions and the answers. These models turn questions and answers into embeddings. The BERT model for questions and answers are different but they share the same BERT weights among themselves. The embedding of questions and answers are fed to a feed forward neural network layer but there are two: one for questions and one for answers. The FFNN produces a similarity embedding. These similarity embeddings are used for similarity check. Usually the embedding similarity training involves negative samples for example word2vec uses NCE (Noise contrastive Estimation) loss. In this case since the embeddings as well as weights change on each step of the training procedure there is a use of a dot product matrix. Basically, what the matrix represents is the dot multiplication of similarity embeddings of both the questions and the answers. A softmax operation is then run on the matrix across its rows so that the questions and answers dot product become softmaxed. For the loss function the system uses a cross entropy loss with inputs from the softmaxed Questions and answers matrix and the ground truth matrix. The ground truth matrix is just the combination of 0 and 1s where the correct questions and their respective answers are mapped. For example, if there were 5 questions and 5 answers then for q1 row only on the a1 column there would be a 1 and the rest of the q1 rows will have 0s. We wanted our chatbot to have meaningful conversation with our user to help them feel accompanied and not alone. Now in order to have as human-like conversation as possible we chose the BERT model. The pre-trained model is used to fine tune on various biomed-

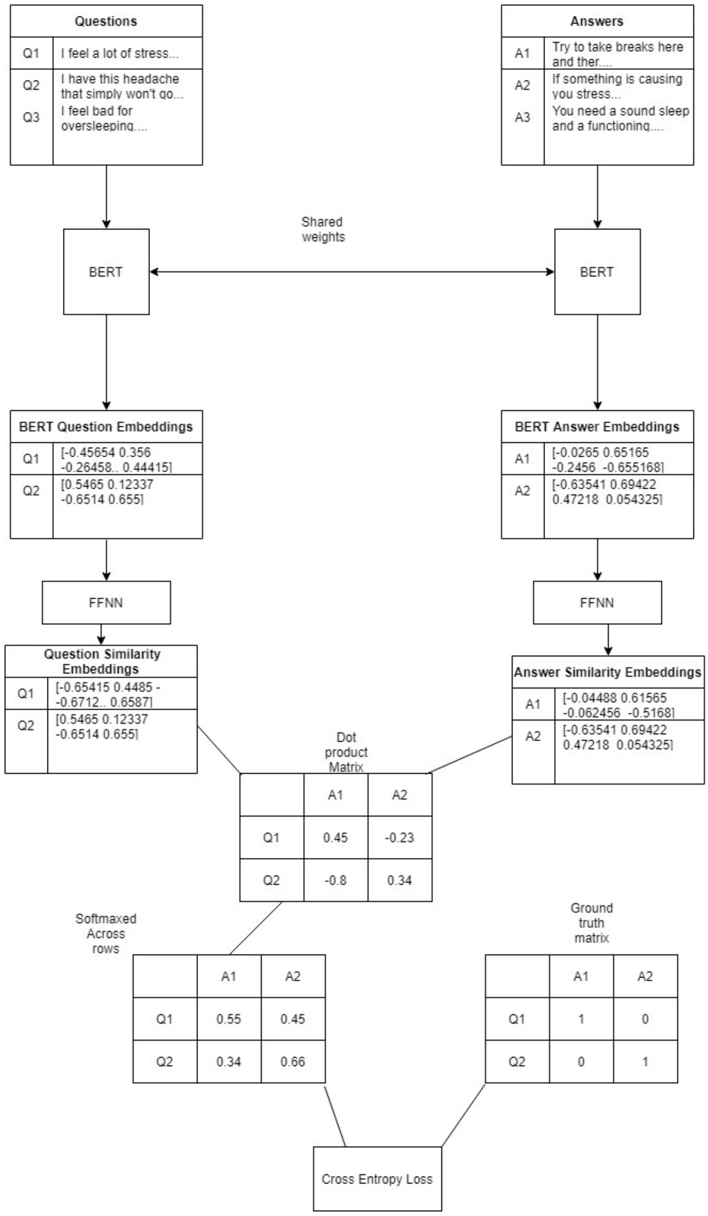


Figure 4.2: Chatbot Architecture

ical tasks like NER, question and answers etc. BERT is short for bidirectional Encoder Representations from transformers. It is a research paper by Google AI language that showed that a model trained bidirectionally can have a deeper sense of the language with context. It is the first ever bidirectional unsupervised technique for language representation. BERT applies bidirectional training of transformers, an attention model to language modelling. The model unlike directional models that read a text sequentially from left to right or right to left, reads the complete sequence of words at once. This enables the model to learn a word's context depending on its surroundings (both left and right of that word). There is a problem with prediction goals while training language models. Usually models are trained to predict the next word in a sequence. For example: The boys are playing . In these scenarios it becomes a challenge for the model to learn context properly. To overcome this BERT uses two methods.
A. Masked Language Modeling, or MLM. Prior to training the word sequences into BERT, 15% of the words in every sequence are replaced with a [Mask] token. The model then tries to predict the original words replaced by the mask tokens based on the context of other words surrounding it. The prediction requires –

1. Addition of a classification layer above the encoder's output.
2. Multiplying the embedding matrix with the output vectors, transforming them into dimensions of the vocabulary.
3. Using softmax to calculate the probability of each word in the vocabulary.

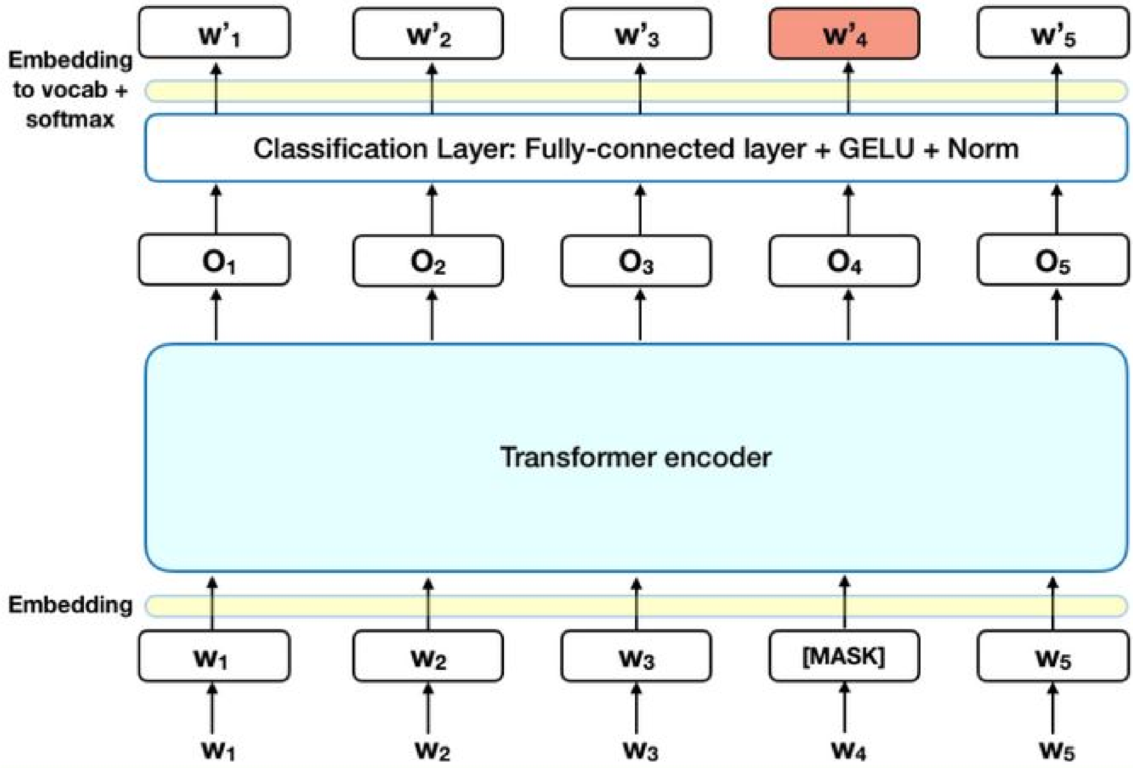


Figure 4.3: the Transformer encoder receives inputs in a sequence of tokens in which the 4th word is masked [33]

In this figure the Transformer encoder receives inputs in a sequence of tokens in which the 4th word is masked. These tokens are then passed to the transformer encoder. The BERT loss function only takes the prediction of masked words into account and not the non-masked words. Hence, the model converges slower than directional models.

B. Next Sentence Prediction, or NSP. BERT is trained by feeding it pairs of texts and learning to anticipate the second sentence in the pair. During training, 50% of the inputs are a pair, with the second sentence being the next sentence in the original text and the remaining 50% being a random sentence from the corpus. The assumption is that the first sentence will be split from the random sentence. To establish this-

1. A [CLS] token is added at the start of the input sequence and a [SEP] token is added at the end of each sentence.
2. Each token during the training period has a sentence embedding that indicates whether the token belongs to Sentence A or

Sentence B. These Sentence embeddings are similar to token embeddings in concept, but with a vocabulary of 2.

3. Each token is given a positional embedding to denote its place in the sequence. The Transformer paper explains the concept and implementation of positional embedding.

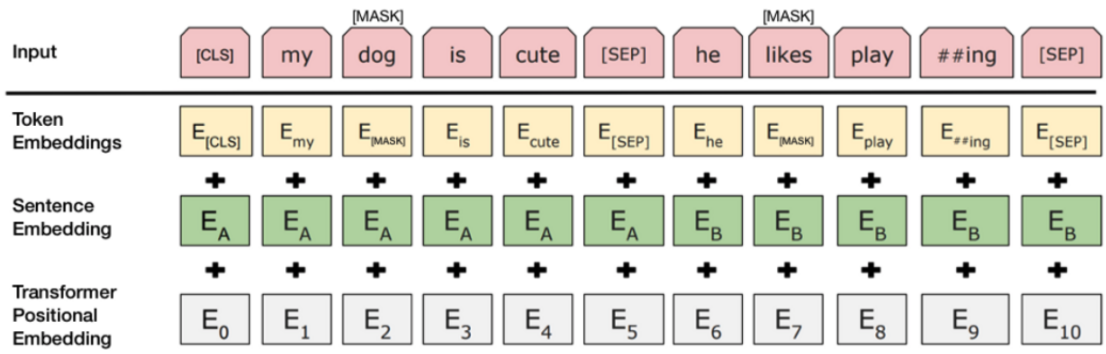


Figure 4.4: The input "my dog is cute he likes playing" we see that MLM as "dog" is replaced with a [Mask] token and there is [CLS] and [SEP] tokens both at start and end of the sentence [29]

In the above figure we see that BERT taking a pair of sentences. Since BERT can not accept strings the words are turned into tokens. The Sentence embeddings added to the token embeddings carry the information of the sentences the words belong to. Transformer position embedding adds value of the position of the words in the sentence to keep the order. Now, in order to see whether the second sentence is indeed connected to the first the following steps are-

1. The entire input goes through the Transformer model
2. Using a simple classification layer, the output of the [CLS] token is transformed into a 2×1 shaped vector.
3. Calculating the probability of IsNextSequence with softmax.

MLM and NSP are both learned simultaneously throughout Bert's training to reduce the combined loss function of the techniques.[33] One of the techniques to make models learn about language with context is called attention. Let us see this through an example:

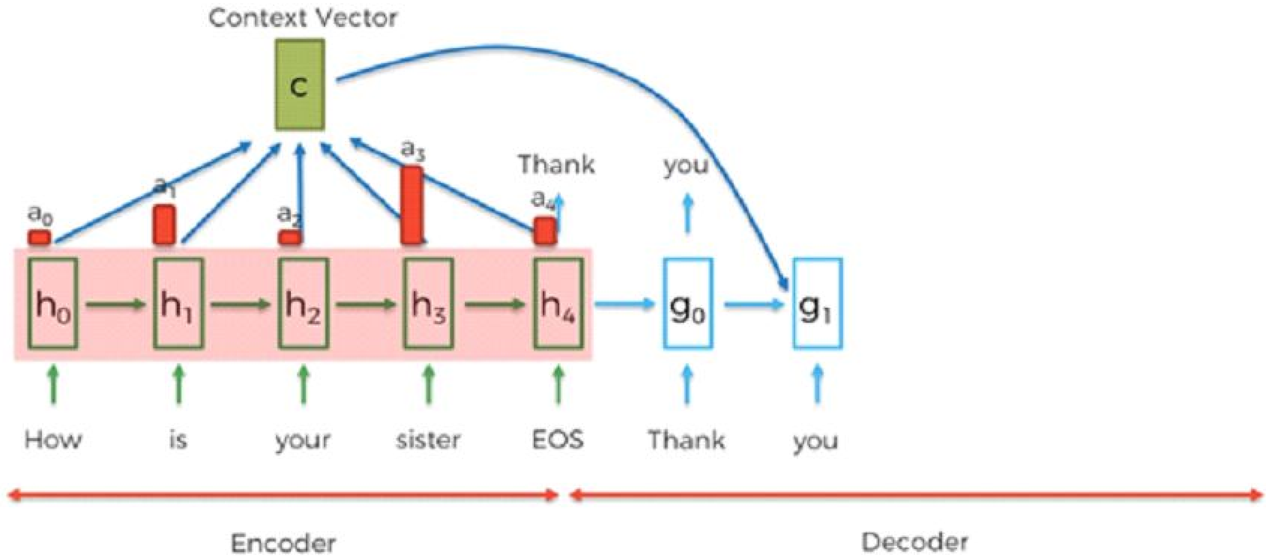


Figure 4.5: Context Vector providing additional input

Here each "h" and "g" represents a whole layer of neurons not just an individual neuron and these directed arrows are vectors that are giving directions of where the flow is. For this example as you can see there is a question from the user that "How is your sister" here EOS represents the end of sentence so we can assume it was a question mark (?). So the input is "How is your sister?" Say the answer would be "Thank you she is fine". As shown in the picture the encoder layer takes the input from the user and runs it through its RNN which then passes its weights to the decoder to give an output for the asked question. However the decoder also has a RNN. Through training the decoder RNN learns to give importance weights (ranging from 0 to 1) to the input words. These weights through the softmax function are kept at a certain range so that they all add up to the value 1. With the importance weights we multiply them with the h vectors for example the context vector stores h0a0, h1a1, h2a2, h3a3 and h4a4 in this case. So the context vector is a weighted sum of the encoder layers of the RNN. So with the context vector it is fed to the decoder's layer as an additional input. So for example the decoder RNN's g1 gets an input from g0, the output of g0 and an input from the context vector C. So with the given inputs the decoder finally outputs "she".

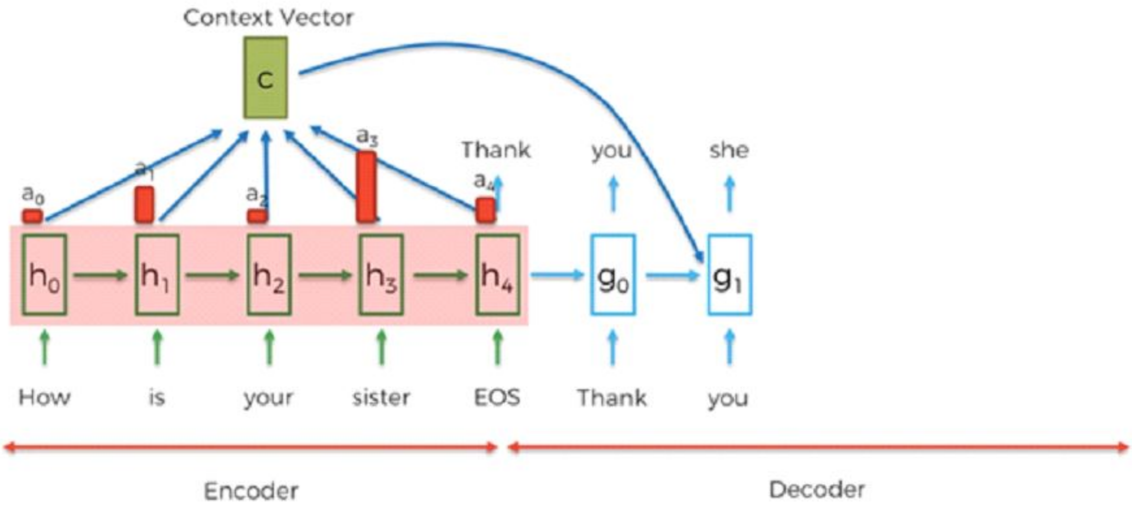


Figure 4.6: Context Vector providing additional input

Now the decoder was supposed to give output based on the input from the encoder layers and the outputs of the previous decoder layers then why was the additional context vector needed? Well in the case of the language English there is a gender difference in pronouns and the model does not understand nouns or pronouns but through training it learns these things. Basically the model just learns for this example when a he or she is needed it needs to go back to the original sentence and from the input there generates an output she. Hence "sister" is shown to have higher importance. Also the word "is" is given emphasis which actually helps to get the information of the number for example if it were "are" that would make the model output "they". As you can see below:

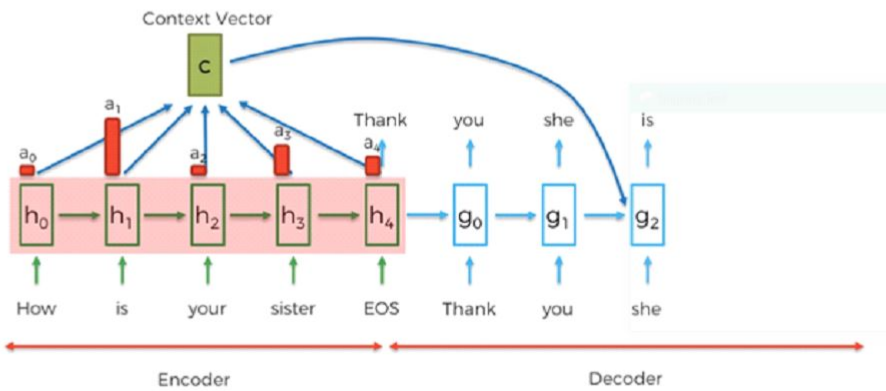


Figure 4.7: Output quot;shequot; from the additional input of context vector

While predicting the verb "is" in the output the weights were changed.

Now we assume that the decoder RNN has access to change and adjust weights of the inputs to be fed into the context vector at every output time and that is illustrated here. Here when the verb was expected as output, the importance of "is" from the input is of higher value and so on.One of the techniques to make models learn about language with context is called attention. Transformer is a deep learning model that uses an attention mechanism to weight the importance of different parts of the input. Multi-Head attention and feed forward layers make up the Transformer model. Because we can't use strings directly, the inputs and outputs are first embedded in an n-dimensional space. Transformers' positional embeddings are critical for keeping track of the word sequences in the input. Because a sequence is dependent on the order of its elements, we must provide a relative position to each word/part in our sequence. These coordinates are added to the embedded representation of each word. [34]
Transformer uses self attention mechanism. The self attention mechanism consists of metrics query, keys and values.

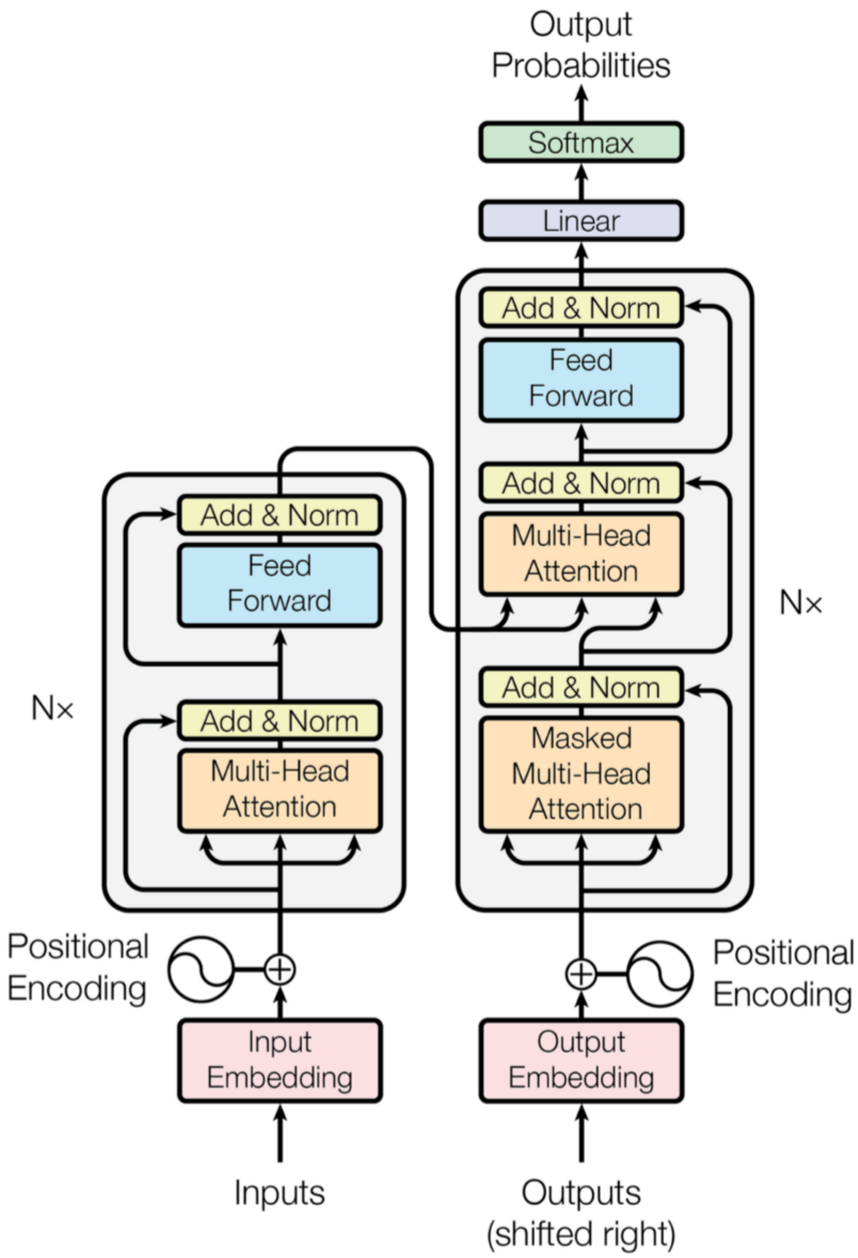


Figure 1: The Transformer - model architecture.

Figure 4.8: The transformer Model[35]

In the figure above, the left block the encoder, and the right block is the decoder. Both Encoder and Decoder are made up of modules that may be layered on top of one another several times, as shown in the diagram by Nx. We can see that the modules are largely made up of Multi-Head Attention and Feed Forward layers. Because we can't utilize strings directly, the inputs and outputs (target phrases) are first embedded in an n-dimensional space.

The positional encoding of the various words is a minor but crucial aspect of the model. We need to give every word/part in our sequence a relative position since a sequence depends on the order of its pieces, and we don't have any recurrent networks that can recall how sequences are fed into a model. These coordinates are added to each word's embedded representation (n-dimensional vector).

$$\text{Attention(Q,K,V)= softmax(QK}^T/(\sqrt{d_k}))V \qquad (2)$$

In the equation above Q represents query metrics, vector representation of a word in a sentence. K represents keys metrics, vector representations of all the words in a sentence. V represents all the values of vector representations of all the words in the sequence. Here, Q and K metrics are multiplied and normalized by dk, the length of the embeddings. They are passed through the softmax function which is then multiplied by the Value metrics. The value matrix V, differs from the sequence represented by the query matrix Q in the attention module since it takes into consideration the encoder and decoder sequences.

To put it another words, the values in V are multiplied and summed with some attention-weights a, where our weights are specified as follows:

$$\text{a= softmax(QK}^T/(\sqrt{d_k})) \qquad (3)$$

In the above equation, the weights,a is determined through a softmax function that takes the query matrix and key matrix and normalizes them by the length of the embeddings. Basically, the weights determine how each word in the sentence is impacted by the rest of the words in the sentence. After that, the weights are applied to all of the words in the sequence.[36] The Feed Forward Neural Network (FCNN) model is a deep learning model where the connected noted do not form a cycle betweeen them. Feed forward network are very simple as the information is only processed in one direction. A FFNN may have multiple layers however, the information always moves in one direction and never backwards.

4.2 Background knowledge on Emotion classification

We will be using machine learning models to find out the relationship between various features of the individual signals and emotional state. As demonstrated in Table 4.3, we will be classifying emotions based on the Arousal and Valence combinations. An output of 0 will be considered as low and 1 as high. We will be using Russell's circumplex model [37] as a source of truth for classifying the different combinations of Arousal And Valence of a particular user. We encoded the Arousal And valence to 0 and 1 to turn this into a binary classification problem. If the rating was higher than or equal to 5, the label was "high," and if it was less than 5, the label was "low." As a result, four labels were created: high arousal low valence (HALV), low arousal high valence (LAHV), high arousal high valence (HAHV), and low arousal low valence (HALV) (LALV). . The input and output data were kept in separate folders before being divided into 40 files for each participant. There were a total of 1,280 participant videos. Furthermore,based on Russell's circumplex model [37] as a source of truth for classifying the different combinations of Arousal And Valence of a particular user, four mental states have been defined accordingly.

No	Emotion Level	Mental Status
1	When Arousal is Low and Valence is High	Calm
2	When Arousal is High and Valence is Low	Upset
3	When both Arousal and Valence is High	Happy
4	When both Arousal and Valence is Low	Sad

Table 4.1: Categorized emotion level and their status

4.3 Emotion Classification Method

4.3.1 Feature selection and emotion classification

We have comparatively studied effects on the accuracy of valence and arousal using four different machine learning models namely: KNN, SVM, ADABoost and Xgboost. The features from all three signals were merged into a single csv file as the label classes valence and Arousal are the same for three signals. Then in order to select features we calculated correlation of the features and removed one of the highly correlated features i.e features which showed correlation greater than 0.6.This was done to prevent overfitting and increase accuracy of our model. The newly selected features for each of the signals are listed below:

EMG	wavelength zero crossing willison amplitude changing signal slope frequency SSI wave length zero crossing willison amplitude changing signal slope frequency SSI MAV2 Power spectrum ratio
PPG	HRV RMSSD HRV MeanNN HRV MadNN HRV HTI HRV HFn HRV CSI HRV PIP HRV GI HRV PI HRV C1a HRV C2d HRV ApEn HRV SampEn
GSR	peakEnd peakStart decayTime

Table 4.2: Selected features for each signals

4.4 Description of Models used

4.4.1 KNN

KNN is a supervised algorithm. It is non parametric and instance based which makes it a simple but robust ML algorithm that can be used for both classification and regression. Due to these characteristics, this algorithm does not have any strong assumptions about the data and hence it is widely used in many research. Furthermore training Machine learning models with KNN is very fast. The algorithm works by computing the distance between every points provided fro training, then selects the ones having least distance and takes into account their corresponding labels. It creates a cluster or class of points which are close to each other. Finally it outputs the label which occur most frequently in that cluster. Different variants of distance can be calculated for KNN for example: Manhattan distance

$$\text{Midst} = \mid x2 - x1 \mid + \mid y2 - y1 \mid \qquad (4)$$

Where x and y are co-ordinates of the data points The distance function used and the hyper parameter K affects the model's performance and the best value for K varies from dataset to dataset.

4.4.2 SVM

Another classifier that we used was Support Vector Machines (SVM). Similar to KNN, this algorithm can also be used for both classification and regression. Under the hood SVMs are just linear separators that divide the data points based on some given criteria. SVMs essentially work by finding a hyperplane or separator that distinguishes between the given data points. The goal of an SVM classifier is to maximize the distance between classes. Hyperplanes in SVM are considered as a kind of decision boundary which divides data into different classes. The dimensions of the hyperplane is proportional to the number of features. The more features in the dataset, the more complex it gets. The data points close to the hyperplane are called support vectors.They are used to maximize the margins. The loss function for SVM with regularization is

$$\min_w \lambda \mid\mid w \mid\mid^2 + \textstyle\sum_{i=1}^{n}(1 - y_i(x_i, w))_+ \qquad (5)$$

Where x and y are coordinates of the data points and w is the weight assigned to them
From the loss function, we find the gradients by taking partial derivatives with respect to weights, and can update the weights which are assigned to data points during classification. The hyperparameters Kernel function and C affect the performance of SVMs. Polynomial K functions can give higher accuracies for data with higher dimensions but may also overfit for simpler data points.

$$\mathrm{k(x,x')} = (1 + x')^k \qquad (6)$$

Another kind of K function is Radial Basis where K is directly proportional to the value of gamma.

$$\mathrm{k(x,x')} = e(-y \mid\mid x - x' \mid\mid^2); \gamma = hyperameter \qquad (7)$$

Consequently higher C gives a more robust model than lower values of C.

4.4.3 ADABoost

Furthermore, we have also used two boosting algorithms, first of which is ADABoost. ADABoost or Adaptive Boost is a superviswed machine learning algorithm. In boosting algothims, models are trainerd sequentially and in each iteration, the model tries to improve performance in regions where performance was poor in previous iteration. Boosting algothms train a series of weak learners and adjust their accuracies iteratively. This is achieved by weighing the errors in an iteration. We assign more weight to points which were misclassified and in the next iteration those particular point will be emphasised more. ADABoost algorithm works as follows: In each iteration weak classifiers h(t), is taken which minimizes the error ϵ_t

$$\epsilon_t = \sum_{i=1}^{m} w_t(i)[y_i \neq h(x_i)] \qquad (8)$$

It then computes weight of the classifier that was chosen by

$$\alpha_t = \frac{1}{2} ln \frac{(1-\epsilon_t)}{\epsilon_t} \qquad (9)$$

And then updates weights of training samples by

$$h(x) = \text{sign}(\alpha^1 h^1(x) + \alpha^2 h^2(x) + ... + \alpha^T h^T(x)) \quad (10)$$

And goes back and repeats error minimization steps.

4.4.4 XGBoost

Another variant of a boosting algorithm that we used is called XG-Boost also known as Extreeme gradient Boosting. It uses gradient boosting under the hood. Gradient boosting differs from plain boosting by the way it updates weak learners in each iteration. It updates weights by Gradient Descent

$$w = w - \eta \nabla w \quad (11)$$

$$\nabla w = (\delta L / \delta w); where\ \ L\ \ is\ Loss$$

Where w is the weight and η is the learning rate.

4.5 Chatbot Model Specification

The whole architecture of our system is composed of fine-tuned bio-BERT models for both questions and replies. These models are used to turn questions and replies into embeddings. When a word is "embedded", it means that the input to the model is a dense N-dimensional vector that represents the word instead of a sparse one-hot encoded vector with the dimension of the words' index in the entire vocabulary set to 1. Although the bio-BERT model for questions and responses differs, they all have the same BERT weights. Question and answer embeddings are input into a fully convolutional neural network layer, however there are two of them: one for questions and one for replies. The transformers of the BioBERT uses the following attention mechanism:

$$\text{Attention(Q,K,V)} = \text{softmax}(QK^T / (\sqrt{d_k}))V \quad (12)$$

In the equation above Q represents query metrics, vector representation of a word in a sentence. K represents keys metrics, vector representations of all the words in a sentence. V represents all the values of vector representations of all the words in the sequence. Here, Q and K metrics are multiplied and normalized by dk, the length of the embeddings. They are passed through the softmax function which is then multiplied by the Value metrics. The value matrix V, differs from the sequence represented by the query matrix Q in the attention module since it takes into consideration the encoder and decoder sequences.

$$\text{a= softmax}(\text{QK}^T/(\sqrt{d_k})) \qquad (13)$$

In the above equation, the weights,a is determined through a softmax function that takes the query matrix and key matrix and normalizes them by the length of the embeddings. Basically, the weights determine how each word in the sentence is impacted by the rest of the words in the sentence. After that, the weights are applied to all of the words in the sequence. The FFNN produces a similarity embedding. These similarity embeddings are used for similarity check. For our loss function we used the sequence loss function. This loss function is just a weighted softmax cross entropy loss function. [38]

$$\text{D(S,L)=-}\textstyle\sum_i L_i.\log(S_i) \qquad (14)$$

Here, S represents the output of the softmax function and L represents the Ground truth matrix where only the correct answers are labeled 1 and the rest of the cells are 0. For each entry of the output vector, softmax function takes the log of the initial entry which is usually less than 1. So the value becomes a very negative number. Then the negative value is multiplied by the Ground Truth Matrix. We sum up all the losses and multiply the sum by -1 to get a positive value. The log only applies to the softmax output since in the ground truth matrix there are multiple 0 values and we can not take log of 0.

Chapter 5

Implementation and Results

5.1 Emotion classification model and results

The selected features were fed to the 4 models and their accuracies were compared. After initial preprocessing during the feature extraction phase we also had to process the dataset by inputting some of the missing values with the mean of that particular column. Also, we checked and removed anomalous values e.g outliers so that our model does not include them during training. For KNN classifier our hyperparameters according to sckit-learnn library were n-neighbors = 7, leaf-size = 3 and p = 1 for classifying both Arousal and Valence. For SVM classifier, our hyperparameters according to scikit-learn library were, kernerl = ‘rbf’, random-state = 42 for both arousal and valence For ADABoost classifier, our hyperparameters according to scikit-learn library were, n-estimators=100, learning-rate=0.1 and random-state = 4 For XGBoost classifier, our hyperparameters according to scikit-learn library were, objective=binary-logistic, max-depth = 4, alpha = 10, learning-rate = 0.1, n-estimators = 100 In order to train the model we used KFold cross validation on the dataset, with number of folds = 10 without shuffling. The performance was evaluated based on the Accuracy and F1 score

Results evaluation and emotion classification model selection

	Arousal		**Valence**	
Classifier	avg Accuracy %	avg F1 score %	avg Accuracy %	avg F1 score %
KNN	55.5	63.6	52.3	59.5
SVM	55.8	68.2	54.6	63.0
ADAboost	51.7	62.9	55.9	62.7
XGboost	53.9	62.4	53.9	62.4

Table 5.1: Comparison of different models used for emotion classification

Through the usage of SVM, we achieved a better performance in our classification task and hence it will be used for prediction. The confusion matrix for this model is as below: For valence

	0	**1**
0	39	75
1	35	101

Table 5.2: Confusion matrix for Valence For Arousal

For Arousal

	0	**1**
0	21	84
1	19	126

Table 5.3: confusion matrix for Arousal

5.2 Chatbot implementation

For our system chatbot we implemented 2 BioBERTs one for questions and one for answers. BioBERT is a form of BERT trained on biomedical data with pre trained weight from a generally trained BERT model. Hence, there is already a transformer model within the BERT that is being used here. For our model we used the following parameters while training it-

attention probs dropout prob: We have set the value of the dropout ratio for the attention probabilities to 0.1 or10%.

hidden activation: There are many activation functions such as 'relu'(Rectified Linear Unit), 'gelu'(Gaussian Error Linear Unit). We used 'gelu' for our model.

$$GELU(x) = xP(X \leq \mathrm{x}) = \mathrm{x}\phi(x) = x.\tfrac{1}{2}[1+\mathbf{erf}(x/\sqrt{2})] \;\; if X \; N(0,1) \quad (15)$$

Here, x is the input of the neuron multiplied by m,Bernoulli((x)), where (x) = P(X x), if X N (0, 1) is the cumulative distribution function of the standard normal distribution. Since using Batch Normalization, neuron inputs tend to follow a normal distribution, gelu is a good activation function in this case. Erf represents error function.

hidden dropout prob: We have set the value of the dropout probability for all fully connected layers in the embeddings, encoder and pooler to be 0.1 or 10%.
initializer range: We have set the standard of the truncated normal initializer for initializing all weight matrices range to be 0.02.
intermediate size: We have set the size of the feed-forward layer in the Transformer encoder to 3072.
max position embeddings: max position embedding refers to the maximum number of tokens the model will take as input. Since the model takes input as tokens this parameter value determines the maximum number of tokens the model will take for encoding. Since we are using a model to perform question and answer tasks, we set the value to 512.
num attention heads: We have used 12 attention heads per attention layer in the Transformer encoder.

num hidden layers: We have12 hidden layers in the Transformer model.
type vocab size: Our vocabulary size of token type ids in the model is 2.
vocab-size: The total number of our vocab size is 28996 which indicates the sum of all the input ids in the model.
Here is a few conversations between chatbot and the user on different moods :

A conversation when user is Sad

```
Hey you seem to be sad. What seems to be the matter?
User: I've had anxiety since as long as I can remember and I have this tendency to constantly fluctuate between under confidence and over confidence. I
Search the similarities in answer or question: answers
How many results to return: 5
1it [00:00,  4.97it/s]

Result  1
if it's interfering with your life you should have it check out. there might be an organic cause that could be easily treated. your doctor can run labs

Result  2
either it is a psychological or medical disorder- both are important to your well being and this needs to be evaluated by a pcp.

Result  3
Anxiety means your brain overreacts to negative emotions. You expect the worst, even when there's nothing to fear. Some people with anxiety spend so mu

Result  4
is needed. anxiety may be the issue or part of the issue. there may also be something medical happening like thyroid imbalance or other hormones. pleas

Result  5
everyone's tolerance to changes in lifestyle including schedule is different. it does seem you are having some anxiety. talk to your doctor or your psy
```

Figure 5.1: A Conversation with the chatbot ,When user is sad

---

User: I've had anxiety since as long as I can remember and I have this tendency to constantly fluctuate between under confidence and over confidence. Is this a sign of anything that needs medical attention like therapy sessions?
Bot:

1. if it's interfering with your life you should have it check out. there might be an organic cause that could be easily treated. your doctor can run labs on you. must rule out thyroid disease at the very least.

2. either it is a psychological or medical disorder- both are important to your well-being and this needs to be evaluated by a pcp.

3. Anxiety means your brain overreacts to negative emotions. You expect the worst, even when there's nothing to fear. Some people with anxiety spend so much energy worrying about family, health, money, or work that they find it hard to carry out everyday tasks.You might also have: Muscle tension Fatigue Trouble sleeping Irritability

4. is needed. anxiety may be the issue or part of the issue. there may also be something medical happening like thyroid imbalance or other hormones. please see your doc. he/she can test diagnose etc. if anxiety is a major factor a mental health professional can help w/stress management. peace and good health.

5. everyone's tolerance to changes in lifestyle including schedule is different. it does seem you are having some anxiety. talk to your

doctor or your psychologist or psychiatrist. sometimes a willing set of ears is all you need sometimes more.

A conversation when user is Upset

```
You seem upset. Is there something I could help you with?
User: I don't really know . I have been feeling down lately and I am lacking sleep a lot. What should I do?
Search the similarities in answer or question: answer
How many results to return: 5
1it [00:00,  4.96it/s]

Result  1
at this point you may have a sleep disorder especially if the sleep issues are causing you distress or are impacting function. you may have developed
 
Result  2
your problem is chronic. it is clear you've had some evaluation and it appears you have had a sleep study. you should really see a sleep specialist wh

Result  3
How’s your diet? Healthy diet will lead to more energy. Also, you need to get off the cycle of taking a nap if possible. The use of your phone in bed

Result  4
to determine the reason for your sleepiness. it is possible that you have a hormone imbalance such as hypothyroid. although there are many other reaso

Result  5
daytime fatigue sleepiness snoring at night. frequent naps and waking up not refreshed
```

Figure 5.2: A Conversation with the chatbot ,When user is Upset.

User: I don’t really know. I have been feeling down lately and I am lacking sleep a lot. What should I do? Bot:

1. at this point you may have a sleep disorder especially if the sleep issues are causing you distress or are impacting function. you may have developed delayed sleep phase syndrome which is a type of circadian rhythm disorder. the period of time you sleep is later (per clock time). as a result, your body becomes used to getting tired and falling asleep later at night. because of this if you try to get up early you will probably feel tired. this means that you are likely to get up later in the day. possible causes include caffeine use in the evening or exposure to light. if you want to normalize your sleep times the first thing you will have to do is to cut out naps and your will need to get up earlier in the morning. this is likely to make you tire earlier in the evening. here are some basic sleep hygiene tips. 1. maintain a regular bedtime and awakening time schedule including weekends. get up about the same time every day regardless of what time you fell asleep. 2. establish a regular relaxing bedtime routine. relaxing rituals prior to bedtime many include a warm bath or shower aroma therapy reading or listening to soothing music. 3. sleep in a room that is dark quite comfortable and cool; sleep on comfortable mattress and pillows. 4. use your bedroom only for sleep and

sex. have work materials computers and TVs in another room. 5. finish eating at least 2-3 hours prior to your regular bedtime. 6. avoid caffeine within 6 hours; alcohol smoking within 2 hours of bedtime. 7. exercise regularly; finish a few hours before bedtime. 8 go to bed only when sleepy. lay in bed only for sleeping not for work or watching tv. 9. designate another time to write down problems possible solutions in the late afternoon or early evening not close to bedtime. 10. after 10-15 minutes of not being able to get to sleep go to another room to read or watch tv until sleepy.

2. your problem is chronic. it is clear you've had some evaluation and it appears you have had a sleep study. you should really see a sleep specialist who can work through what might be the underlying cause(s) and a trial of treatments. could try a bedside journal if your mind races. avoid alcohol and tobacco before bed. try massage or relaxation therapy if available. try consistent bed time.

3. How's your diet? Healthy diet will lead to more energy. Also, you need to get off the cycle of taking a nap if possible. The use of your phone in bed is a huge factor for sleep issues. Al though it'll be hard trying not to use your phone when trying to fall asleep. I suggest trying these. If you need help falling asleep to make sure you get enough sleep, melatonin 5 mg - 10 mg is a start. However, I recommend this after you try lifestyle modifications.

4. to determine the reason for your sleepiness. it is possible that you have a hormone imbalance such as hypothyroid. although there are many other reasons such as sleep apnea or poor sleep hygiene. often times a sedentary lifestyle creates low energy. you may need to force yourself into more activity and healthy eating. a health coach can help. if you are obese or gaining weight you must act now.

5. daytime fatigue sleepiness snoring at night. frequent naps and waking up not refreshed

A conversation when user is Upset

```
Hello there! What can I do for you?
User: I was wondering if a proper sleep can calm someone's mind. I have been calm today unlike yesterday and the day before. Is it because I had a gooc
Search the similarities in answer or question: answer
How many results to return: 5
1it [00:00,  5.48it/s]

Result  1
Do you feel fully rested when you wake up at 4am? Not everyone needs 8 hours of sleep per night.

Result  2
Humans naturally have a cycle that makes them want to nap in the afternoon. Western culture just prevents us from doing that. A 1 hour nap in the after

Result  3
No, you have an inverted sleep schedule as do many physicians and healthcare providers. The hours of which you choose to sleep do not pose an effect or

If you are getting an adequate amount of sleep, and a good quality of it at that then there is no indication that a different set of sleeping hours wi

Result  4
There's no research suggesting that there's any alternative to enough sleep besides enough sleep apportioned differently throughout the day. You can ma

Result  5
As long as you are resting a sufficient amount of time there is nothing wrong with not sleeping during the night time. In some parts of Alaska they exp

Furthermore us healthcare guys are known to work the hours you listed above and nap when we can.
```

Figure 5.3: A Conversation with the chatbot ,When user is Calm

---

User: I was wondering if a proper sleep can calm someone's mind. I have been calm today unlike yesterday and the day before. Is it because I had a good sleep last night?
Bot:

1. Do you feel fully rested when you wake up at 4am? Not everyone needs 8 hours of sleep per night.

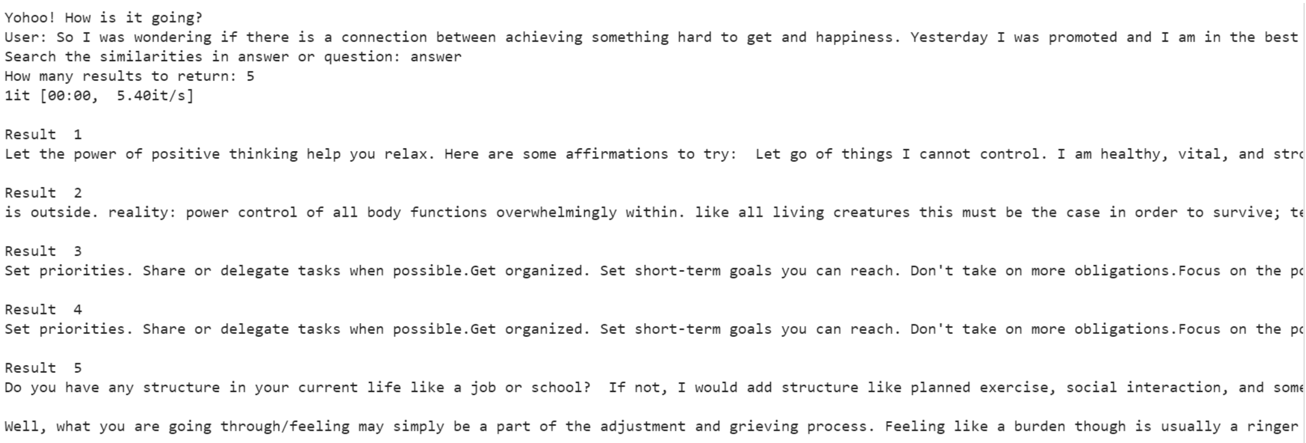
2. Humans naturally have a cycle that makes them want to nap in the afternoon. Western culture just prevents us from doing that. A 1-hour nap in the afternoon plus a night of sleep is actually probably more "natural" than just sleeping at night. As far as naps adding up to 8 hours...that is not as good. There are different stages to sleep, and you need to be sleeping for a certain amount of time before your brain gets into those stages. REM is certainly important, for example, and you don't go into REM sleep right away.

3. No, you have an inverted sleep schedule as do many physicians and healthcare providers. The hours of which you choose to sleep do not pose an effect on your health; *however*, it is the amount and quality of your sleep that dictates whether or not it is sufficient. If you are getting an adequate amount of sleep, and a good quality of it at that then there is no indication that a different set of sleeping hours will cause any problems.

4. There's no research suggesting that there's any alternative to enough sleep besides enough sleep apportioned differently throughout the day. You can maybe get by with multiple chunks, but you need somewhere around 7 hours, give or take for individual variation.

5. As long as you are resting a sufficient amount of time there is nothing wrong with not sleeping during the night time. In some parts of Alaska, they experience 18+ hours of night time during certain times of the year. Furthermore, us healthcare guys are known to work the hours you listed above and nap when we can. Moral of the story is if you get all the hours you need it doesn't matter what time of day it is.

A conversation when user is Happy



```
Yoohoo! How is it going?
User: So I was wondering if there is a connection between achieving something hard to get and happiness. Yesterday I was promoted and I am in the best
Search the similarities in answer or question: answer
How many results to return: 5
1it [00:00,  5.40it/s]

Result  1
Let the power of positive thinking help you relax. Here are some affirmations to try:  Let go of things I cannot control. I am healthy, vital, and stro

Result  2
is outside. reality: power control of all body functions overwhelmingly within. like all living creatures this must be the case in order to survive; te

Result  3
Set priorities. Share or delegate tasks when possible.Get organized. Set short-term goals you can reach. Don't take on more obligations.Focus on the po

Result  4
Set priorities. Share or delegate tasks when possible.Get organized. Set short-term goals you can reach. Don't take on more obligations.Focus on the po

Result  5
Do you have any structure in your current life like a job or school?  If not, I would add structure like planned exercise, social interaction, and some

Well, what you are going through/feeling may simply be a part of the adjustment and grieving process. Feeling like a burden though is usually a ringer
```

Figure 5.4: A Conversation with the chatbot,when user is happy

---

User: So, I was wondering if there is a connection between achieving something hard to get and happiness. Yesterday I was promoted and I am in the best mood ever since.

Bot:

1. Let the power of positive thinking help you relax. Here are some affirmations to try: Let go of things I cannot control. I am healthy, vital, and strong. There is nothing in the world I cannot handle. All my needs are met. I am completely and utterly safe. Every day, in every way, I am getting stronger.

2. is outside. reality: power control of all body functions overwhelmingly within. like all living creatures this must be the case in order to survive; termed homeostasis. control of heart like rest of body from brain many many interactive control cycles per second. thus, mental issues are dominant though some outside influences can interact. no precise way to know/answer; what research is about.

3. Set priorities. Share or delegate tasks when possible. Get organized. Set short-term goals you can reach. Don't take on more obligations. Focus on the positive. Laugh! Humor is a good release. Listen to music that relaxes or revives you. Talk to a counselor or a friend. Remember, things don't have to be perfect. "Good enough" may be fine. Take time for yoga, meditation, or some deep breaths. Get regular exercise. Take some time, even 5 to 10 minutes, for yourself daily.

4. Do you have any structure in your current life like a job or school? If not, I would add structure like planned exercise, social interaction, and some meaningful goal for the future. You're constantly distracted by the thing you're trying to deal with, but what is/are those things? Are they important in the bigger picture or just distractions? It sounds like you have some anxiety, and sometimes inactivity can contribute to that. Are you doing things to make sure you are progressing forward towards a meaningful life? Or just waiting for stuff to happen? Maybe explore new hobbies or find a life coach? I recently lost a big piece of the meaning in my life and I am trying to live for myself.

5. Well, what you are going through/feeling may simply be a part of the adjustment and grieving process. Feeling like a burden though is usually a ringer for depression. You may benefit from a few sessions of counseling to help you sort things out in your head so you can start feeling like yourself again.

In future iterations, we would train our model with CBT dataset, making it an expert for detecting mental disorders.
Overall work flow of system

Chapter 6

Conclusion and Future Work

6.1 Conclusion

Our main purpose of this study is giving users mental support through our chatbot which is pre-trained on large biomedical corpora. In our research, we have tried to demonstrate a system that takes physiological signals as input, detects a user's mental state and converses with them. We have used various machine learning models to accurately detect their emotions and trained our chatbot with context relevant data. Last but not the least, mental well being should not be something to be hidden off or ignored. We should take care of our mental health properly. There are still a lot of people who do not even know how to deal with it. That is why we are trying to come up with an automated chatbot based system where our system can continuously communicate with the person if he is under any stress or anxiety.

6.2 Limitations

We faced few problems collecting CBT dataset as this is a very controversial dataset and is not easily available on the internet .Our CBT dataset is not large enough to run a continuous conversation with the user according to his mental state. Moreover, we realize the number of data fed into a machine learning model directly affects its performance and that more signal data is required to build a robust emotion classifier.

6.3 Future Work

In the future we would like to use deep learning models to classify emotions as recently they have gained popularity in building models with very high accuracy. We have a plan to make our chatbot more automated by taking continuous psychological signal inputs from the user and continuously converse with the user according to the user's mental condition . Additionally we would also like to train our chatbot on a large cbt dataset so that it can perform more accurately and give better results to mental health related issues.

Bibliography

[1] F. W. H. U. Jacobi, *Size and burden of mental disorders in europe-a critical review and appraisal of 27 studies, european neuropsychopharmacology, pp. 357-376,* 2005.

[2] E. R.V de Graaf R.Hiripi, *,r.c.,kohn,r.offord, d.r. wittchen,h.u.bijl.the prevalence of treated and untreated mental disorders in five countries,health affairs*, 2003.

[3] R. Picard, *Affective computing. MIT press.*, (2000).

[4] J.-W. BAEK and K. CHUNG, "Pcontext deep neural network model for predicting depression risk using multiple regression," *Predicting Depression via Social Media,*, vol. vol.8, January 2020.

[5] C. G. Kiciman E Dredze M, "Kumar m dr choudhury m,discovering shifts to suicidal ideation from mental health content in social media," *in Human Factors in Computing Systems*, 2016.

[6] S. C. M. Dr Choudhury and E. Horvitz, "Predicting postpartum changes in behavior and mood via social media," *Predicting Depression via Social Media,*, 2013.

[7] S. C. Michael Gamon and E. H. M. D. Choudhury, "In international aaai conference on weblogs and social media.," *Predicting Depression via Social Media,*, 2016.

[8] . D. Geršak G., "Electrodermal activity patient simulator. plos one," *and Y. Bengio D. Bahdanau,Acoustics, Speech and Signal Processing (ICASSP)*, 15(2), e0228949, (2020).

[9] K. F. D.R. Bach and R. Dolan, "Analytic measures for quantification of arousal from spontaneous skin conductance fluctuations," *Int. J. Psychophysiol*, vol. 76(1), pp. 52–55, Apr. 2010.

[10] M. T. Ko?odziej, "Electrodermal activity measurements for detection of emotional arousal. bulletin of the polish academy of sciences, technical sciences.," *Ko?odziej, Marcin Tarnowski, Pawe? Majkowski, Andrzej Rak, R.*, vol. 67. Pp. 52–55, (2019). DOI: 813-826.10.24425/bpasts.2019.130190.

[11] B. W. E. Activity, "London: Springer london;" *Int. J. Psychophysiol*, vol. second, 2012.

[12] N. Aqajari S. A. H., "Gsr analysis for stress: Development and validation of an open source tool for noisy naturalistic gsr data.," *E. K., Mehrabadi, M. A., Labbaf, S., Rahmani, A. M., Dutt, N.*, 2020. DOI: arXivpreprintarXiv: 2005.01834.

[13] D. Cho, J. Ham, J. Oh, J. Park, S. Kim, N.-K. Lee, and B. Lee, "Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine," *Sensors*, vol. 17, no. 10, p. 2435, 2017.

[14] A. Öhman, F. Esteves, A. Flykt, and J. J. Soares, "Gateways to consciousness: Emotion, attention, and electrodermal activity," in *Progress in electrodermal research*, Springer, 1993, pp. 137–157.

[15] J. Allen, "Photoplethysmography and its application in clinical physiological measurement," 2007.

[16] W.johnston and Y. Mendelson, "Extracting heart rate variability from a wearable reflectance pulse oximeter," 2005.

[17] R. C. N. Pinheiro, "Can ppg be used for hrv analysis," *J. Henriques, J. Muehlsteff, I. Quintal, L. Gonçalves and P. Carvalho*, 2016.

[18] A.J.J.S.K.K.D.a.S.A.N.SELVARAJ, "Assessment of heart rate variability derived from finger-tip photoplethysmography as compared to electrocardiography," 2008.

[19] D.M.P.J.L.B.L.P.S.A.Chen, "Neurokit2: A python toolbox for neurophysiological signal processing," 2020.

[20] J. G. Wijsman, "Trapezius muscle emg as predictor of mental stress. acm transactions on embedded computing systems (tecs)," *, Bernard Penders, Julien Hermens, Hermie*, 2010. DOI: 12.155-163.10.1145/2485984.2485987.

[21] T. S. Yaser Keneshloo, "Deep reinforcement learning for sequence-to-sequence models," *Naren Ramakrishnan, and Chandan K. Reddy,IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–21, August 2019.

[22] C. Xiong and R. P. R. Socher, "A deep reinforced model for," *arXiv preprint arXiv:1705.04304*, May 2017.

[23] C. Dyer and T. B.-K. K. Goyal, "Differentiable scheduled sampling for credit assignment," *ACL*, vol. vol 02, pp. 366–371, 2017.

[24] M. A. S. Chopra and W. Z. M. Ranzato, "Sequence level training with recurrent neural networks," *arXiv preprint arXiv:1511.06732*, 2015.

[25] J. C. Ron J. Weiss, "Navdeep jaitly, yonghui wu, and zhifeng chen ,title:sequence-to-sequence models can directly translate foreign speech," August 2017.

[26] V. Badhe, "And manasi u. kulkarni ,a neural conversational model for automatic generation of conversations," *International Research Journal of Engineering and Technology*, vol. vol. 05, no. no. 06, June 2018.

[27] J. Chorowski and N. Jaitly, "Towards better decoding and language model integration in sequence to sequence models," August 2017.

[28] Q. N. Jaitly, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing ,O. Vinyals W. Chan*, March 2016.

[29] J. Devlin, "Bert:pre-training of deep bidirectional transformers for language," *FCN-Fully Convolution Neural Network*, vol. en, arXiv:181004805(cs). DOI: https://arxiv.org/abs/1810.04805.

[30] S. Jammalamadaka, "Medical chatbot using bert and gpt2," February 2020.

[31] *(PDF) DEAP: A Database for Emotion Analysis Using Physiological Signals.*, DOI: AvailableURLhttps://www.researchgate.net/publication/232651717_DEAP_A_Database_for_Emotion_Analysis_Using_Physiological_Signals[accessedMay022021.

[32] C. K. Karmakar, "Complex correlation measure: A novel descriptor for poincaré plot," *A. H. K. J. G. M. P. Chandan K*, p. 17, August 3,2009.

[33] Raghudeep, "Bert explained:state of the art landuage model for nlp," vol. en, 2021-02-05T20:23:18:143Z. DOI: https://medium.com/@raghudeep/biobert-insights-b4c66fde8fa7#:~:text=BioBERT\%20is\%20a\%20contextualized\%20language,only\%20understood\%20by\%20domain\%20experts..

[34] Sanders, "Seq2seq-chatbot," (2019). DOI: https://github.com/AbrahamSanders/seq2seq-chatbot.

[35] S.-H. Tsang, "Fcn-fully convolution network(semantic segmentation)," vol. en, 2019-03-20T 15:54:42.325Z. DOI: https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1.

[36] Maxime, "Medium," vol. en, 2020-03-05T0:42:48:604Z. DOI: https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04.

[37] M.-R. A. García-Martínez B., "A review on nonlinear methods using electroencephalographic recordings for emotion recognition. ieee trans. affect. comput.," *Alcaraz R., Fernández-Caballero A.*, (2019). DOI: 10.1109/TAFFC.2018.2890636..

[38] M. A.Nielsen, "Neural networks and deep learning," *Determination Press*, 2015.