A Decision Tool Based on Mobile Sensing for Social Anxiety Disorder among University of Virginia (UVA) Students

Shichen Li
sl4bq@virginia.edu

***Abstract*—In the recent years, it is high time that we should pay attention to the treatments of individuals who have social anxiety disorders. In this project, we collected physical and physiological data of 6 subjects from Shimmer 3 ECG and GSR sensors and extract features to identify the activity recognition accuracy by selecting the best classification model. Then we chose some treatment methods to provide interventions for these subjects according to their social interaction anxiety scale (SIAS) score changes. Finally we evaluate the accuracy of score changes prediction with the best regression model. The result shows that the Random Forest Classifier can lead to the best performance in terms of accuracy, and Linear Regression model can yield the best prediction in terms of the Root Mean Square Error (RMSE). But there is a lot of future work to deal with, for instance, we need to collect data from more subjects, to test these models in a larger scale.**

I. Introduction

Social anxiety is defined as an intense fear of being judged and scrutinized by others that is beyond what would normally be expected in a situation [1]. Individuals struggling with such mental health problems face a significant amount of obstacles in their daily life.

The recent advancements in m-Health field have demonstrated significant potential for alleviation or even removal of some obstacles. [2] For instance, acute stress is the result of demands and pressures of the recent past as well as those anticipated in the near future. This can derive from instances in anxiety when meeting new people. Chronic stress, can also be ascribed to long-standing pressures and demands experienced as a result of difficulties in interpersonal relationships. If left unmanaged, chronic stress can have detrimental consequences on those experiencing it.

Due to the expensive process of data collection, especially the monetary and time expenses, in our tight schedule, what we can do here is to increase the types and amount of data collected in each session, and it is this kind of way that can dramatically improve the data amount gained to cost ratio. The increment of types in data collection makes it possible for us to do some in-depth data analyses, which is quite important for us to identify which predictors could be the best to predict changes in anxiety. This provides the bedrock for us to provide immediate interventions to individuals who is detected with an increase in social anxiety.

There have been recent studies investigating which features may be indicative of changes in anxiety, with the goal of being able to provide immediate interventions to individuals when an increase in anxiety is detected. Xin Zhao et. al. conducted a study on the effects of working memory (WM) training on various indices potentially related to social anxiety, in terms of electroencephalography (EEG) [3] . It turned out to be that trained students showed decreased social anxiety relative to pre-training assessment, and also better WM performance in transfer task. These students showed changes in event-related potentials (ERPs) to facial stimuli, but these changes did not have great influence on WM training.

Half a year ago, we did a comprehensive analysis through various physiological, audio and motion metrics in order to predict the anxiety from the self reported and measured baseline, but due to some difficulties in the data collection and processing, we were not able to provide an overall view in the project. For example, the survey questions before the pilot study were not fully utilized. This might influence the accuracy of the data processing. Also the ability of activity recognition of all measures here was not very clear, and in this project we need to put energy on this identification.

In this project, we focuses more on the mathematical modeling for the decision system. So we need to identify the variables from the environment, and the decision variables, to define the utility functions. Based on these, our decision making process could be more sound and reasonable.

The rest of paper is organized like this: The related work, followed by our emphasis on what we can do and the relationship to the related work, will be discussed in Section II. Section III will describe the predictive tools we used and the formalized model for the decision analysis. Then we will discuss the results in the Section IV. Finally, we will make a conclusion and clarify the future work directions in the Section V.

II. Related Work and Motivation

According to the literature, [1] a survey about the detection and recognition of emotions as affective states for the sake of identifying the connection between emotional states and physical health, focused mainly on the physiological and physical measures of stress using some sensing technologies. This is a basis for our research here. It pointed out that stress has effects on one's physical condition whether it is chronic or acute, but it also presents itself via biomarkers and physical expression, so it can be measured by means of activity signals. For example, the brain activity can be measured by electroencephalography (EEG), heart activity can be measured by ECG, skin conductance can be explained by GSR and EDA. After examination of these technologies, it was concluded that these modes of affective computing could be very promising and creative, and as it would be increasingly used in commercial fields, it would be beneficial for consumers in the decision making process.

Furthermore, recent studies indicated that this ideology would be quite promising in the future. For example, Gong et. al. [7] conducted a 2-week study of 52 undergraduates from University of Virginia (UVA), and investigated some primary research questions that if behavioral patterns occurring around social interactions vary across the social anxiety level with accelerometer, in addition, the communication media or semantic locations. They concluded that there existed a significant difference in the behaviors of these subjects in terms of social anxiety levels as well as locations, for instance, higher SIAS scores meant more movement around the time of phone calls, especially in some unfamiliar areas. Boukhechba et. al. [8] conducted another pilot study in UVA lasting 3 weeks with 20 college students, collecting data from both passive ways like PPG and accelerometer, and active ways like interpretation bias questionaires. By the intra-model analysis of heart rate and accelerometer, it was found that the correlation between them could be used to predict the changes in self-reported negative effect. And the results showed that the changes in correlations had the same pattern as the changes in interpretation bias and negative effect.

In spite of these magnificent achievements in the field of social anxiety detections and predictions, no study to our knowledge has do something effective in choosing the right way to provide the right means to alleviate the symptoms. On top of that, there were few studies using the Bayesian decision analysis model to choose the best option for each individual patients about their symptoms. So now what we need to do is to choose some therapy options to make this possible.

From the studies done in the past few years, it is concluded that cognitive behavioral therapy (CBT) is one of the most effective way to treat the social anxiety disorder, which aims to provide techniques and practice to instruct the patients in terms of their thoughts and behaviors, especially in situations terrifying them. The general CBT usually cover a time span of 12 to 16 weekly sessions lasting for 1 or 1.5 hours per session. [4]

One of the most popular way to reduce the social anxiety, according to Scheurich et. al. [5] is the exposure therapy. In this form of therapy, the psychologists create a safe environment and these individuals with symptoms will be "exposed" to the things they tend to fear and shirk. After several times of similar experiments, they would become more calm and get used to these situations. 6 Participants' social distress and stuttering severity, which are incorporated into the comprehensive level of Social Anxiety Rating ranging from 0 to 10, were recorded in Baseline (BL), Progressive Muscle Relaxation Therapy (PMR) and Exposure (EXP) periods. After that the post experience and the level after 6 months should also be incorporated. The results showed that there were dramatical reductions in social anxiety and the improvements in affective, behavioral and cognitive experiences of stuttering.

Another CBT treatment frequently applied in SAD treatment, Social Skills Training (SST), studied by University of Murcia [6], showed great results in reduction of the numbers of social situations patients fear. 108 adolescents with the age range of 14 to 17 were involved in this study, and the dropout rate of treated adolescents was reduced. The Intervention Program for those people proved to somewhat effective.

Apart from CBT, the medical options are often chosen for the generalized form of social anxiety disorder. [4] The most prevalent medical options include selective serotonin reuptake inhibitors (SSRIs), and serotonin and norepinephrine reuptake inhibitor (SNRI). Generally 50% to 80% of patients will respond after 8 to 12 weeks of treatments.

In this model, we will use the EXP for the treatment of high social anxiety in the anticipatory period, and SST for the experience period, SSRIs for the post-event periods.

III. Approach of Experiment

In this section, we will introduce the whole process of our study, including the survey and activities the need to complete, and the data collection methods, as well as the information related to participants. Then goes the data processing and modeling.

A. The Procedure and Design

There were 6 participants involved in the pilot study, 4 undergraduate students and 2 graduate students, and each subject had received specific course rewards by the corresponding supervisor. This study was done on November 21st, 2019, and it was assumed that they all provided their real feelings, and they had already received necessary training before participating in the survey.

The first step is the collection of survey results: they were given different amount of questions in the 4 HOBBY pilot surveys. Each question, the answers would be scaled in Social Interaction Anxiety Scale (SIAS) scores ranging from 1 to 5. The higher SIAS scores imply higher risk of having fears of general social interactions, and these scales might remain stable unless treatment.

In the first survey, there were three questions about their feelings in the baseline period, then followed the second survey testing their anticipatory feelings about certain activities before the experiments with 13 questions. After this, the third survey with 13 questions aimed to test their real feelings during the experiments, and the final survey recorded their post-experiment conditions by means of 8 questions.

Furthermore, after our further consideration, we decided to choose all questions from each set of the survey questions to implement some machine learning techniques, to determine which machine learning techniques can produce the highest accuracy. They were mainly associated with the calmness or anxiety. When the survey-taker showed a more calm state, the SIAS score would be lower. They were:

Question 1: *I feel Calm or Anxious. (Answered before the test)*

Question 2: *I feel Calm or Anxious. (Answered at the beginning of the test)*

Question 3: *When my feelings were the most intense during the last experience, I felt... Calm vs Anxious*

Question 4: *Looking back on the last experience, I feel... Calm vs Anxious*

During each experiment, each participant should complete three activities, the solo video watching, dyad no-evaluation conversations, and dyad evaluation conversations on specifics (e.g. music). These were experience types in this study. The schedule of experiments is shown in the Appendix part, Figure 4.

B. Data Collection and Preprocessing

We used two types of Shimmer3 devices, along with the Smartwatch to collect the data we were interested in. The Shimmer3 ECG Unit was configured to measure the electrical signals from the skin, and Shimmer3 GSR+ Unit aimed to measure the Galvanic skin response (GSR). Among the ECG results, we extracted the data streams related to the accelerometer and gyroscope , as well as the heart rate signals, which were calibrated before analysis) and from GSR we extracted the data streams about the skin conductance, EDA and the PPG signals. The example data is shown in Figure 5.

For data preprocessing, it should be noted that in terms of the accelerometer and gyroscope data, in order to make a easier working procedure, we decided to use the Signal Vector Magnitude (SVM) to describe the degree of subjects' movements, which was calculated using equation (1):

$$SVM = (x^2 + y^2 + z^2)^{(1/2)} \tag{1}$$

Then we took windowing approach as standard 2s windows for each of the signals, except for PPG signals. This is because the windowing length of 2 s was not working heart rate calculations, due to the heart beat unavailability. Finally 4s windowing was working after using the timewise unix changing.

C. The Decision Making Model

1) The Decision Maker's Action Sets $\mathbb{A}$: The action sets is divided into three layers. The first layer is the model for the activity recognition, which produces the highest accuracy. In this project, we choose these classification models: Random Forest Classifier (RFC), Gaussian Naive Bayes (GNB), Decision Trees (DT), and k-Nearest Neighbours (kNN).

$\mathbb{A}_1$ = {RFC, GNB, DT, kNN}

Then we need to decide which treatment to take for every subject. In the Section II, we made a complete list of all the options available for us. For the subject i, we need to take one action j from the action set

$\mathbb{A}_2$ = {EXP for video, EXP for no-eval, EXP for eval, SST for video, ... ,SSRIs for eval}

And finally, we need to make a model to predict the score changes for every subject using an appropriate model. In this project, we finally decided to use some regression models: Linear Regression (LR), Least Absolute Shrinkage and Selection Operator (LASSO) and ElasticNet Regression.

$\mathbb{A}_3$ = {LR, LASSO, ElasticNet}

2) The Sample Space $\mathbb{X}$: The sample space here including two parts: one part stands for the subjects' score changes in different activities. In this place, what we need is a system to denote the activities and periods using the subscripts. The first number we use refers to the three activities, 0,1 and 2 respectively for video watching, dyad no-evaluation conversations and dyad evaluation conversations, and use the 0,1,2 and 3 to denote the four periods, baseline, anticipatory, experience and post-event respectively. For example, the score change from baseline period during the video watching activity is denoted as $s_{00,01}$.

The subjects' activities:
$\mathbb{S} = \{s_{00,01}, s_{01,02}, s_{02,03}, \ldots, s_{22,23}\}$

$$a_2^* = \arg\max_{a_2 \in \mathbb{A}_2} \mathbb{S}$$

Another part involves the features extracted from the ECG, GSR signal. For the sake of measuring in a multimodal manner, different from what we have done before, we are going to extract 12 features from accelerometer and gyroscope data, 5 features from skin conductance data and 5 features from PPG data, instead of using totally 12 features from all the data.

The subjects' features:
$\mathbb{F} = \{f_1, f_2, ...f_{20}\}$
and
$\mathbb{X} = \{\mathbb{F}, \mathbb{S}\}$

3) The Parameter Space Θ and Prior Beliefs: All possible values about the parameter space depend on the data stream and the accuracy of classification and prediction is the only factor making great influence on this project.

The reason why we are going to do like this is that according to our work done before, the achieved accuracy was very low for both of the Support Vector Machines (SVM) and Random Forest (RF) classification algorithm. Even if we used a more generalized method, according to the score change in three ways: no change, increase, and decrease, we could only get an accuracy ranging from 75% to 80%. Furthermore, the sizes for PPG data windowing and other data windowing are not always the same.

4) The Utility Function or Loss Function: In this project, we use the Root Mean Square Error to calculate the standard deviation of the prediction errors:

$$\begin{aligned} RMSE = (\frac{1}{N}\sum_{s \in S}(&predict\ score\ change \\ &-actual\ score\ change)^2)^{1/2} \end{aligned} \tag{2}$$

TABLE I
THE ACCURACY OF ACTIVITY RECOGNITION WITH INDIVIDUAL ECG

Algo	ECG			
Act	Total	0	1	2
kNN	45.22	66.96	69.12	68.32
NBC	10.83	28.78	34.28	31.36
DT	62.85	87.37	80.54	80.27
RFC	68.43	88.40	84.76	85.09

$$a_3^* = \underset{a_3 \in \mathbb{A}_3}{\arg\max} \ \{\}$$

5) The Payoff Function: Based on what we have discussed above, we want to maximize the accuracy of the classification, i.e. recognition accuracy, and we also want to have a lower RMSE. In this way, we can define the payoff function, the final performance as:

$E(a_1, a_2, a_3|\theta) = acc_{class} \cdot (RMSE)^{-1}$

IV. RESULTS

A. The Classification

For the classification part, since something went wrong during the pilot experiment, the GSR files for subject 3 and 4 had entirely wrong timestamps, and there was also a great deal of data lost. So we choose 18 features from ECG for these two subjects, and for other subjects, we choose 14 subjects from ECG signals and 5 features for GSR signals, 5 features for PPG signals.

In the ECG file, what we are most interested in should be the gyroscope data and accelerometer data. A gyroscope is a device using gravity to identify the orientation, and its design consists of a freely-rotating disk, rotor, who is attached to a spinning axis located in the center of a larger and more stable wheel. Thus the axis ‘ turning will be recorded in magnitude of x, y and z directions when parts of the body rotates. A accelerometer measures the acceleration other than gravity, and it uses microscopic crystals going under stress when vibrations occur, and from that stress a voltage is generated to create a reading on any acceleration.

In this experiment, by referring the literature related to activity recognition[9], we choose mean, median, energy, variance, skewness, kurtosis as features for both accelerometer and gyroscope signals. By using sliding window algorithm, we use the First, we are going to show the result for the subject 1,2,5,6 in Table I and II.

It is obvious that using the RFC proves to be the best way for both ECG and GSR signals. Furthermore, when using kNN, DT and RFC, ECG statistical features seems to have a much better performance than GSR, in terms of both individual activities accuracy and the total accuracy. and it is quite possible that there are more ECG features than GSR.

What we found interesting is that when using NBC, the performance of ECG is quite worse than GSR, contrarily. No matter individual activities accuracy or total accuracy, the performance of the former is dwarfed compared to the latter.

TABLE II
THE ACCURACY OF ACTIVITY RECOGNITION WITH INDIVIDUAL GSR

Algo	GSR			
Act	Total	0	1	2
kNN	34.28	48.45	59.86	53.57
NBC	20.20	36.56	38.23	42.55
DT	51.38	75.92	72.65	57.92
RFC	55.36	76.51	76.32	66.15

TABLE III
THE ACCURACY OF ACTIVITY RECOGNITION WITH INDIVIDUAL PPG

Algo	PPG			
Act	Total	0	1	2
kNN	18.02	33.73	43.01	40.44
NBC	16.16	26.33	37.53	43.89
DT	27.22	45.26	46.84	43.26
RFC	28.30	45.86	47.12	46.08

And similarly, we use the same algorithm for PPG, extracting the mean, median, standard deviation, variance and mode[9]. The results are shown in Table III. Here we are astonishing to discover that the individual activities accuracy and total accuracy of using PPG to identify the activities are both very low. Using all these algorithms cannot yield a total accuracy of over 30%.

Then we discuss the reason why PPG features cannot guarantee a high-level performance. Let us take the RFC as an example, the confusion matrices is shown in Table IV.

So we finally decide to discard the PPG features, and we go on to try to combine the features of ECG and GSR, according to the literature[9], the combination of these features often proves to be more effective. And after our trial, it turns out to be that using the RFC for the combined features of ECG and GSR can still yield the highest accuracy, which is shown in Table V.

Besides this, another striking fact is that by combining these two features, the accuracy values for DT have a dramatic increase, and so do RFC. But for NBC there is only a very

TABLE IV
THE CONFUSION MATRIX FOR PPG ACTIVITY RECOGNITION WITH RFC ALGORITHM

Act	0				1				2			
P	0	1	2	3	0	1	2	3	0	1	2	3
0	2	6	0	16	3	2	8	7	3	5	9	2
1	0	51	8	44	4	27	62	8	2	12	50	4
2	0	28	39	42	2	16	122	6	0	15	115	6
3	0	25	14	63	2	9	68	20	1	6	72	17

TABLE V
THE ACCURACY OF ACTIVITY RECOGNITION WITH COMBINED ECG AND GSR FEATURES

Algo	Combined ECG and GSR			
Act	Total	0	1	2
kNN	45.22	66.96	69.12	68.32
NBC	10.97	28.63	34.38	31.37
DT	85.04	93.24	94.28	84.78
RFC	90.09	95.88	96.32	91.61

TABLE VI
THE ACCURACY OF ACTIVITY RECOGNITION WITH 18 ECG FEATURES FOR SUBJECT 3 AND 4

Algo	Combined ECG and GSR			
Act	Total	0	1	2
kNN	49.74	71.15	75.56	80.69
NBC	14.96	32.41	44.81	37.07
DT	77.87	85.37	91.11	83.40
RFC	87.85	87.75	95.93	92.27

slight increase for these values. The most important irony here is that the accuracy values for kNN is totally the same as the individual ECG.

For subject 3 and 4, we extract 9 features for both accelerometer and gyroscope data, and they are: mean, median, maximum, energy, variance, skewness, kurtosis, entropy, and mean absolute deviation (MAD). The accuracy result is shown in Table VI.

It is evident that with more features, the performances with DT or RFC classifiers are significantly improved, compared to the Table I. But with kNN or NBC, the total performances do not have great improvement. Moreover, compared to Table IV, the best total performance is still yielded from RFC, though it is a bit lower.

Now we finally decide to choose the $a_1^* = RFC$, and $\theta_{1\ for\ 1,2,5,6}^* = 90.09\%$. And for other two subjects, $\theta_{1\ for\ 3,4}^* = 87.85\%$.

B. The Choice of Treatment

Participant	Activity	Baseline	Anticipatory	Experience	Post	B-A	A-E	E-P
1	0	2.667	2.923	2.077	1.714	0.256	-0.846	-0.363
1	1	1.667	2.846	2.923	2.143	1.179	0.077	-0.78
1	2	1	3.077	2.923	2.286	2.077	-0.154	-0.637
2	0	3	3	1.923	2	0	-1.077	0.077
2	1	2	2.846	2.769	2.857	0.846	-0.077	0.088
2	2	2.333	3.154	2	3	0.821	-1.154	1
3	0	3.667	2.692	2.231	2.286	-0.975	-0.461	0.055
3	1	2.667	3	2.308	1.571	0.333	-0.692	-0.737
3	2	2.333	2.538	2.846	2.714	0.205	0.308	-0.132
4	0	2	3.077	3.077	3.143	1.077	0	0.066
4	1	3	3.308	3.538	3	0.308	0.23	-0.538
4	2	1.333	2.692	3	2.413	1.359	0.308	-0.587
5	0	3	3.308	2.769	2.571	0.308	-0.539	-0.198
5	1	3.333	2.615	1.846	2.571	-0.718	-0.769	0.725
5	2	3.333	3.154	2	2	-0.179	-1.154	0
6	0	3	1.615	1.077	1	-1.385	-0.538	-0.077
6	1	2.667	2	1.923	1	-0.667	-0.077	-0.923
6	2	1	1.923	1.923	1	0.923	0	-0.923

Fig. 1. The Score Changes for Each Subject

In this project, we finally decide to use the average score of these subjects for each period during each activity. And there will be 3 score changes in 3 periods, for example, the score change during video activity from baseline to anticipatory period (B-A), will be denoted as $s_{00,01}$. So our final decision for a_2^* will be:

a_1^* = EXP for eval
a_2^* = SSRIs for eval
a_3^* = EXP for no-eval
a_4^* = EXP for eval
a_5^* = SSRIs for no-eval
a_6^* = EXP for eval

C. The Prediction Tools

This time we are going to implement some linear models to train the data, to best predict the score changes with the least RMSE. In the feature extraction done before, there are 33 features for the subjects 1,2,5 and 6, and 20 features for the subjects 3 and 4. To eliminate the unimportant factors influencing our prediction, we are going to leverage on the Principal Components Analysis (PCA) to reduce the dimensions.

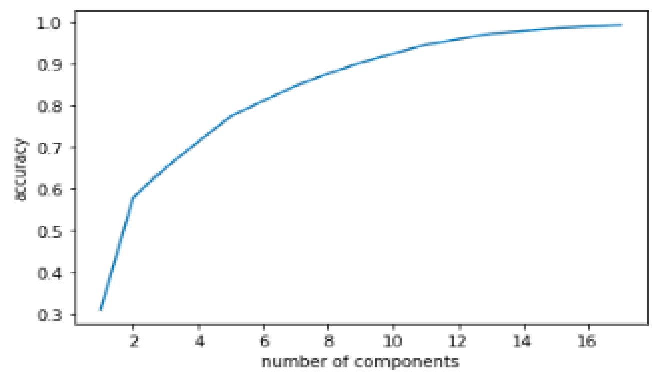

Fig. 2. The PCA Analysis result for Subjects 1,2,5 and 6

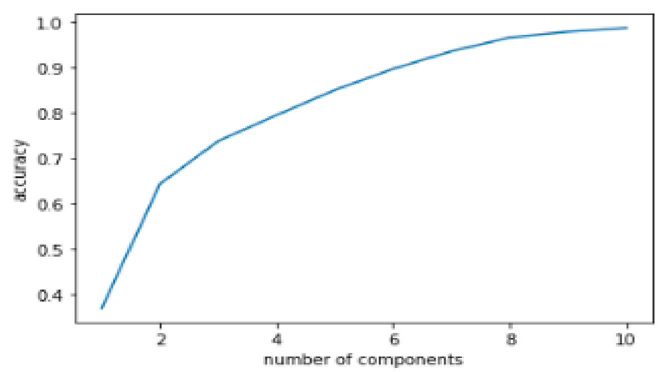

Fig. 3. The PCA Analysis result for Subjects 3 and 4

From Figure 2 and Figure 3, we found that using only 16 features for subjects 1,2,5 and 6, and using only 10 features for subjects 3 and 4 can ensure the accuracy of 99% virtually.

Based on this outcome, we will use the dataset after PCA processing, to predict the score changes using models in $\mathbb{A}_3$ After our implementation, we finally get the results in the Tables VII.

The conclusion points out that it seems no matter which model we choose, the differences between each kind of models are all less than 0.00002. Besides this, we can get a higher performance in terms of the low RMSE by using 10 ECG

TABLE VII
THE RMSE OF PREDICTION ABOUT SCORE CHANGE FOR SUBJECTS 1,2,5 AND 6

Subjects	LR	LASSO	ElasticNet
1,2,5,6	0.672265	0.672277	0.672271
3,4	0.475057	0.475072	0.475065

features to predict score changes of subject 3 and 4, compared to using 16 features to predict score changes of all other subjects.

So $a_3^* = LR$, and $RMSE_{1,2,5,6} = 0.672265$, $RMSE_{3,4} = 0.475057$.

V. CONCLUSION AND FUTURE WORK

From what we have discussed above, first of all, for the activity recognition part, the RFC provides the highest accuracy for both two groups of subjects, and for the subjects 1,2,5 and 6, it is best to combine the features of ECG and GSR, which can get an accuracy of 90.09%. PPG is not a good choice for this recognition. In addition to this, for subjects 3 and 4, using 18 features only extracted from ECG can also ensure the accuracy of 87.85%. Secondly, most of subjects had some social anxiety problems before the no-evaluation or evaluation periods, and some of them were facing the difficulty of high anxiety in post-event periods. Finally, the combination of PCA and Linear Regression models for the prediction of score change can yield the least RMSE.

But the limitation here cannot be neglected. In the first place, 6 students is far away from enough, and next time we are eager to gather more data, at least 30 people should be involved in this experiment. Moreover, as far as we are concerned, when it comes to the data analysis part, we still have a lot of work to expand. For example, the decision of treatment can still depend on more factors, like the fluctuations of each period. On top of that, we need to consider more factors, like the demographics of students, like age, military service, gender and so forth. Besides these, we can try more models on these activity recognition and prediction of score changes, like neural networks, plural voting, stacking, boosted SVM, etc., [10] along with more features, like frequency domain features, and time-frequency domain features [11].

REFERENCES

[1] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders*, 5th ed. Washington, DC: American Psychiatric Association, 2013.

[2] Shalom Greene, Himanshu Thapliyal, and Allison Caban-Holt, *A Survey of Affective Computing for Stress Detection*, IEEE Consumer Electronics Magazine, October 2016, pp. 44-56.

[3] Xin Zhao, Chen Dang, Joseph H.R. Maes, *Effects of working memory training on EEG, cognitive performance, and self-report indices potentially relevant for social anxiety*, Biological Psychology 150 (2020) 107840.

[4] Harvard Mental Health Letter, Treating Social Anxiety Disorder, March 2010.

[5] Jennifer A. Scheurich, Deborah C. Beidel, Martine Vanrycheghem, *Exposure Therapy for Social Anxiety Disorder in People Who Stutter: An Exploratory Multiple Baseline Design*, Journal of Fluency Disorders 59 (2019), pp.21-32. DOI: https://doi.org/10.1016/j.jfludis.2018.12.001.

[6] Pablo Jose Olivares-Olivares, Pedro Francisco Ortiz-Gonzalez, Jose Olivares, *Role of social skills training in adolescents with social anxiety disorder*, International Journal of Clinical and Health Psychology (2019) 19, 41-48

[7] Mehdi Boukhechba, Jiaqi Gong, Kamran Kowsari, Mawulolo K. Ameko, Karl Fua, Philip I. Chow, Yu Huang, Bethany A. Teachman, Laura E. Barnes, *Physiological Changes over the Course of Cognitive Bias Modification for Social Anxiety*, 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), 4-7 March 2018,Las Vegas, Nevada, USA, pp. 422-425.

[8] Mehdi Boukhechba, Alexander R. Daros, Karl Fua, Philip I. Chow, Bethany A. Teachman, Laura E. Barnes, DemonicSalmon: Monitoring mental health and social interactions of college students using smartphones. Smart Health 9-10 (2018), pp. 192-203,ISSN 2352-6483, https://doi.org/10.1016/j.smhl.2018.07.005.

[9] Priyanka Das, Anwesha Khasnobish, Prof. D.N. Tibarewala, Emotion Recognition employing ECG and GSR Signals as Markers of ANS, 2016 Conference on Advances in Signal Processing (CASP), Cummins College of Engineering for Women, Pune. Jun 9-11, 2016.

[10] Nishkam Ravi and Nikhil Dandekar and Preetham Mysore and Michael L. Littman , Activity Recognition from Accelerometer Data, Department of Computer Science, Rutgers University.

[11] C. Berke Erdas, Isil Atasoy, Koray Acici, Hasan Ogul, ntegrating features for accelerometer-based activity recognition, Procedia Computer Science 98 (2016) 522 – 527.

VI. Appendix

					Baseline perdiod		Anticipatory anxiety		Experience		Post-event	
Group #	Date	Time per experience (min)	Experience_type	Participant_id(s)	Start_time	End_time	Start_time	End_time	Start_time	End_time	Start_time	End_time
Pilot	11/21/2019	2 (antic), 2 (video), 2 (post)	Solo Video	1	9:35:27 AM	9:36:41 AM	9:38:06 AM	9:42	9:42:36 AM	9:46:48 AM	9:47:25 AM	9:51:13 AM
		2 (antic), 2 (convo), 2 (post)	Dyad No-eval	1	9:51:45 AM	9:53:21 AM	9:54:36 AM	9:58:15 AM	10:00:01 AM	10:05:39 AM	10:06:19 AM	10:10:46 AM
		2 (antic), 2 (convo), 2 (post)	Dyad eval	1	10:17:05 AM	10:18:28 AM	10:21:56 AM	10:24:35 AM	10:27:15 AM	10:33:18 AM	10:33:51 AM	10:37:58 AM
		2 (antic), 4 (convo), 2 (post)	Group eval	1								
		2 (antic), 4 (convo), 2 (post)	Group No-eval	1								
		2 (antic), 2 (video), 2 (post)	Solo Video	2	9:31 AM	9:31	9:33	9:36:57 AM	9:38	9:44:45 AM	9:45:39 AM	9:49:06 AM
		2 (antic), 2 (convo), 2 (post)	Dyad No-eval	2	9:50:23 AM	9:51:16 AM	9:54:40 AM	9:58:47 AM	10:00:01 AM	10:06:56 AM	10:07:36 AM	10:10:10 AM
		2 (antic), 2 (convo), 2 (post)	Dyad eval	2	10:17:17 AM	10:18:16 AM	10:20:02 AM	10:24:20 AM	10:27:15 AM	10:33:29 AM	10:34:03 AM	10:37:58 AM
		2 (antic), 4 (convo), 2 (post)	Group eval	2								
		2 (antic), 2 (convo), 2 (post)	Group No-eval	2								
		2 (antic), 2 (video), 2 (post)	Solo Video	3	9:23:04 AM	9:24:13 AM	9:25:09 AM	9:27:25 AM	9:29:41 AM	9:35:19 AM	9:36:11 AM	9:40:52 AM
		2 (antic), 2 (convo), 2 (post)	Dyad No-eval	3	9:41:18 AM	9:42:07 AM	9:44:08 AM	9:48:08 AM	9:49:04 AM	9:54:49 AM	9:56:51 AM	10:00:58 AM
		2 (antic), 2 (convo), 2 (post)	Dyad eval	3	10:03:01 AM	10:03:54 AM	10:05:11 AM	10:09:13 AM		10:15:28 AM	10:17:56 AM	10:21:52 AM
		2 (antic), 4 (convo), 2 (post)	Group eval	3								
		2 (antic), 4 (convo), 2 (post)	Group No-eval	3								
		2 (antic), 2 (video), 2 (post)	Solo Video	4	Need to impute	Need to impute	Need to impute	9:27:30 AM	9:30:21 AM	9:32:15 AM	9:32:42 AM	9:34:17 AM
		2 (antic), 2 (convo), 2 (post)	Dyad No-eval	4	9:38:02 AM	9:38:20 AM	9:40:06 AM	9:41:05 AM	9:49:04 AM	9:54:49 AM	9:57:06 AM	9:58:00 AM
		2 (antic), 2 (convo), 2 (post)	Dyad eval	4	10:03:43 AM	10:04	10:06:35 AM	10:09:11 AM	10:11 AM	10:15	10:16:40 AM	10:19:36 AM
		2 (antic), 4 (convo), 2 (post)	Group eval	4								
		2 (antic), 4 (convo), 2 (post)	Group No-eval	4								
		2 (antic), 2 (video), 2 (post)	Solo Video	5	9:45:20 AM	9:45:49 AM	9:53:42 AM	9:59:04 AM	10:00:06 AM	10:03:19 AM	10:07:09 AM	10:11:08 AM
		2 (antic), 2 (convo), 2 (post)	Dyad No-eval	5	10:19:22 AM	10:19:45 AM	10:25:39 AM	10:29:44 AM	10:31:09 AM	10:37:46 AM	10:38:52 AM	10:43:37 AM
		2 (antic), 2 (convo), 2 (post)	Dyad eval	5	10:45:18 AM	10:45:40 AM	10:48:04 AM	10:50:49 AM	10:51:12 AM	10:57:05 AM	10:57:34 AM	11:00:41 AM
		2 (antic), 4 (convo), 2 (post)	Group eval	5								
		2 (antic), 4 (convo), 2 (post)	Group No-eval	5								
		2 (antic), 2 (video), 2 (post)	Solo Video	6	9:41:46 AM	9:43:11 AM	9:51:57 AM	9:56:03 AM	9:57:11 AM	10:02:13 AM	10:03:04 AM	10:07:39 AM
		2 (antic), 2 (convo), 2 (post)	Dyad No-eval	6	10:19:22 AM	10:19:45 AM	10:25:39 AM	10:29:44 AM	10:31:09 AM	10:37:46 AM	10:38:52 AM	10:43:37 AM
		2 (antic), 2 (convo), 2 (post)	Dyad eval	6	10:45:18 AM	10:45:40 AM	10:48:04 AM	10:50:49 AM	10:51:12 AM	10:57:05 AM	10:57:34 AM	11:00:41 AM
		2 (antic), 4 (convo), 2 (post)	Group eval	6								
		2 (antic), 4 (convo), 2 (post)	Group No-eval	6								

Fig. 4. The Schedule for Experiment

	ECG1_Timestamp_FormattedUnix_CAL	ECG1_Accel_LN_X_CAL	ECG1_Accel_LN_Y_CAL	ECG1_Accel_LN_Z_CAL
0	yyyy/mm/dd hh:mm:ss.000	m/(s^2)	m/(s^2)	m/(s^2)
1	09:35.3	22.84782609	11.27173913	-17.51086957
2	09:35.3	21.2826087	10.10869565	-14.48913043
3	09:35.3	21	9.923913043	-13.94565217
4	09:35.3	20.97826087	9.913043478	-13.92391304
5	09:35.4	20.9673913	9.902173913	-13.90217391
6	09:35.4	20.94565217	9.891304348	-13.86956522
7	09:35.4	20.93478261	9.880434783	-13.84782609
8	09:35.4	20.92391304	9.869565217	-13.82608696
9	09:35.4	20.89130435	9.858695652	-13.76086957
10	09:35.4	20.85869565	9.836956522	-13.70652174
11	09:35.4	20.84782609	9.815217391	-13.69565217
12	09:35.4	20.81521739	9.804347826	-13.65217391
13	09:35.4	20.80434783	9.793478261	-13.63043478
14	09:35.4	20.79347826	9.782608696	-13.58695652
15	09:35.4	20.77173913	9.77173913	-13.57608696
16	09:35.4	20.75	9.75	-13.52173913
17	09:35.4	20.7173913	9.739130435	-13.4673913
18	09:35.4	20.69565217	9.717391304	-13.44565217
19	09:35.4	20.68478261	9.717391304	-13.41304348
20	09:35.4	20.67391304	9.706521739	-13.40217391

Fig. 5. The Example Data: Accelerometer Data Stream in ECG