CS4246 AI Planning and Decision Making - Project 2
Planning and Decision Making Automation on Depression

Team 01
Antoine Charles Vincent Garcia - A0159072A
Chan Jun Wei - A0112084U
Chen Tze Cheng - A0112092W
Eric Ewe Yow Choong - A0112204E
Han Liang Wee, Eric - A0065517A
Ho Wei Li - A0094679H

Abstract

Depression is a debilitating mental illness that has excellent prognosis given early detection and treatment. In recent years, youth suicide rates in Singapore have increased at an alarming rate. Suicide is often a byproduct of depression; however, depression can be easily treated. General practitioners are the most accessible source of help but the lack of time, attention, skills and the presence of social stigma prevent an accurate diagnosis of depression. In our project, we propose a general framework to pre-screen patients before they make appointments in general practice for depression. Outputs from the Gaussian Process will be used to determine if the patient requires medical attention and if so, the urgency of the case. Besides, we produced a proof-of-concept prototype illustrating our framework. We also further note that the framework can be extended to other psychiatric disorders.

Introduction

Depression has a severe, and at times long-term, negative impact on an individual's quality of life. Major depression is the 3rd leading cause of disability worldwide with 65 million life years spent living with the disability or lost due to early death [World Health Organization, 2004]. Depression's annual toll on U.S businesses amounts to about $80 billion in medical expenditures, lost productivity and suicide. Among the costs, close to $10 billion accrued in lost workdays each year and more than $33 billion in other costs accrue from decreased productivity due to symptoms that sap energy, affect work habits, cause problems with concentration, memory, and decision-making [Greenberg et al., 2015].

Depression, if left unchecked, increases risk for morbidity, suicide, decreased cognitive and social functioning,

self-neglect, and early death [Fiske, Wetherell, and Gatz, 2009]. Death from suicide is one of the top 10 causes of death, above the mortality rate for chronic liver disease, Alzheimer's, homicide, arteriosclerosis or hypertension [Jiaquan Xu et al., 2016]. Despite the severe consequences, depression is one of the most treatable mental illnesses, but it is also one of the most under-diagnosed globally.

General practitioners (GPs) are often dubbed as the first line in the provision of healthcare as they form the interface between specialized healthcare providers and patients. Compared to specialists, GPs are considered to be more accessible, less stigmatizing and more comprehensive. Thus they are expected to be a depression patient's best chance at getting preliminary help [Rothman and Wagner, 2003]. Indeed, in the US, almost 75% of depression patients who seek help, do so with a general practitioner [Goldman, Nielsen, and Champion, 1999].

However, in general practice, 48.4% of patients suffering from depression go unrecognized [Kessler et al., 2003]. During a visit to the GP, a patient's symptoms(non-physical) may not be observed. Even if some GPs show interest in patients' psychosocial issues, they likely lack the requisite interviewing skills to properly elicit the relevant history while others lack the sensitivity to affective and nonverbal patient cues [Badger et al., 1994]. GPs have to complete a plethora of tasks during each consultation such as assessment, prevention, and management of medical conditions in addition to filing paperwork. This severely limits the time budget that remains for assessing mental health problems such as depression and further raises the barrier of accurate diagnosis and proper treatment [Telford et al., 2002].

Motivation

The most recent study conducted by the Institute of Mental Health has shown that Major Depressive Disorder (MDD) is the most common mental disorder in Singapore. An estimated 1 in 17 people having suffered from MDD at some point in their lives which is almost twice that of the next most common disorder, alcohol abuse [Chong et al., 2012]. In 2015, despite falling overall suicide rates, and a shrinking proportion of youths in the aging population

of Singapore, the number of suicide deaths in youths between 10-19 of age was twice as many as the year before, and it is also the highest in 15 years [Samaritans of Singapore, 2016]. The large increase in suicide rates is a worrying trend, as suicide is closely correlated with depression. In light of all the available data, we extrapolate that the prevalence of MDD has increased over the last five years.

Diagnosis and treatment of depression in Singapore are difficult. Apart from the highlighted difficulties that GPs face, there are also significant delays in specialized treatment, which can extend up to several months due to the presence of an extended referral chain to a psychiatrist, inherent in the public health system. Sociocultural factors such as stigmatization compound the problem. In Asian societies, where a majority of the public, including mental health professionals, think that they should be protected from people with mental disorders or that these people should be locked up. [Chong et al., 2009]. These issues justify need for a better process for diagnosing depression.

Objective

In the past decade, researchers have had successfully correlate emotion with voice production and speech acoustics [Johnstone, 2001]. A corollary to that, a recent research was successful in using voice acoustics as predictors of clinical depression scores, proving that it is an effective indicator of depression severity [Hashim et al., 2016]. In a recent study, Gaussian Process Dot Product trained using the MFCC feature set of acoustic analysis was found to be a good model for the prediction of depression incidence and severity with better accuracies than other state-of-the-art machine learning models [Han et al., 2016].

We propose a framework based on the outputs of the Gaussian Process Dot Product on MFCC model. The framework lowers the barriers to diagnosis and treatment of depression by supplanting the diagnosis process in general practice with the predictions of the model in most cases, requiring the GP's attention only in ambiguous cases. In univocal cases, patients are referred directly to specialists for treatment, thereby reducing the referral chain to a single step. Specialized healthcare providers are also given the predicted Patient Health Questionaire 8 (PHQ-8) score of the patient to prioritize urgent cases better, improving appointment scheduling. On the side of GPs, time and resources saved can be utilized to provide better quality healthcare in all other aspects.

Modelling and Approach

Leveraging on the success of the modeling of depression prediction of Personal Health Questionaire depression scale (PHQ-8) scores in project 1, we extend and apply the work done to solve the aforementioned problem of under-staffing. Our solution is to implement pre-screening and automate the process of the deciding if the person needs an appointment or otherwise. In this manner, we cut down on the number of appointments reducing the workload on the staff. Moreover, our process allows prioritizing of patient's appointments by their PHQ-8 score which represents the severity of depression. Hence, we would need to use two Gaussian Processes (GP): Gaussian Process Classifier (GPC) and the Gaussian Process Regressor (GPR). In this section, we will firstly describe the process that we are proposing, and we will go into detail on each component of the process. We note that the process is a general framework and that different clinics have different operating processes, clinics should tweak the framework to better fit their needs. Hence, we would omit some detail in our model and leave them to the implementer.
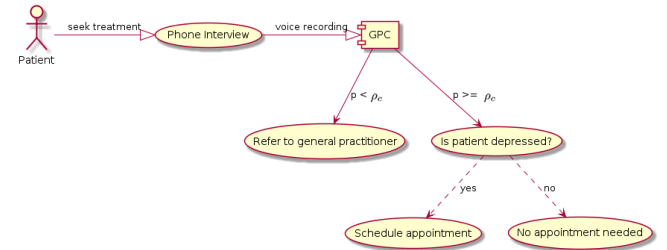
Automation Flow

Figure 1: Automation Flow

Figure 1 represents the flow of a person who is seeking for medical attention from a psychiatric clinic. Prior to medical treatment, the patient would be calling up the clinic to arrange an appointment time and date. Before the clinic fixes the appointment, a short phone interview consisting of several questions is conducted. Refer to the sample questions in the section titled 'Proof of Concept'. The objective of the interview is to record the patient's audio signal as he/she answers the questions. The recordings will serve as an input for GPC, which is a classifier that will output both a probability estimate and a label, determining if the patient is suffering from depression or not. The probability estimate represents the confidence on the predicted label. We can then define a specific pre-defined probability estimate ρ_c, such that we have two two scenarios as listed below. In the illustration below, we make the argument based on a probability estimate p which is based on some audio recording collected.

1. Confident of prediction ($p \geq \rho_c$)
   GPC is confident of its prediction, allowing us to make decision with confidence on the predicted label. Then, we make the decision based on whether the patient is depressed or not depressed as predicted by GPC. A depressed patient will be given an appointment (refer to Figure 2) for treatment, whereas a non-depressed patient would be informed that he/she is not required to go for an appointment.

2. Not confident of prediction ($p < \rho_c$)
   GPC is not confident of its prediction, we cannot rely on the predicted label. In this case, patient will be referred to a general practitioner for further observations

(Figure 2). We note that he/she would need to be diagosed by the clinician and treatment if appropriate. We would also obtain the ground truth from the diagnosis and use that to train the GPC/GPR on this data-point.
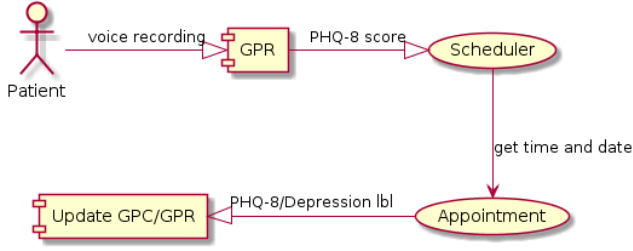

Figure 2: Flow for Scheduling an Appointment

Figure 2 represents the case that he/she needs an appointment. Similarly, the patient's voice is use as an input for the GPR which predicts the PHQ-8 score. PHQ-8 score determines the levels of depression (from 0-24, 0 means no depression and 24 means very depressed) of the patient, allowing the staff to prioritize appointments based on the PHQ-8 score. We would prioritize patients who are more depressed (predicted by the GPR with a higher PHQ-8 score) will have higher priority and vice versa. Through the appointment(s) with the psychiatrist, the ground truth of the depression and PHQ-8 labels will be obtained and can be used to train both GPC and GPR if required (if this data-point was predicted with no confidence i.e. $p < \rho_c$). As more data points are observed and subsequently use for training, both GPC's and GPR 's accuracy would improve incrementally.

Insights to Our Model

Gaussian Processes : In our modeling, we used two GPs for regression and classification tasks respectively. The GP models that we used here are modeled similarly to the modeling done in project 1. GPR predicts the PHQ-8 score, while GPC predicts whether the person is depressed or not. Similar to project 1, we trained GPR and GPC with data obtained from Audio/Visual Emotion Challenge and Workshop (AVEC 2016) [Valstar et al., 2016b], with PHQ-8 and depression labels (provided as part of the dataset) respectively. We applied the same audio signal processing techniques to the audio files as per project 1. We used GP with a dot-product kernel, with the Mel Frequency Cepstral Coefficients (MFCC) feature subset as we have seen the good theoretical and experimental results the feature subset had produced in project 1. More importantly, we make the same two assumptions so that the GP model planning and decision making is suitable for depression prediction:

1. Depression prediction is an event-based which provides a single depression estimate over a time period [Valstar et al., 2016a].

2. Speech signals extracted from people suffering from depression should share some similarities and thus admissible for prediction with the Gaussian Process models [Cummins et al., 2015].

In our formulation GPs must be able to train on new data. Since the GP models receive new data incrementally, we cannot use an offline GP as we have described in project 1. Hence, we would need to tweak the GP to fit this problem, needing a GP that learn dynamically, adapting to new data as it becomes available. Online machine learning is a method that allows the data to be updated when it becomes available. We have noted that there are online variants of GP which will be used in our project.

Phone Interview : We noted that regardless of the questions asked, a depressed person will still exhibit signs of depression in his speech [Kroenke et al., 2008]. Hence, the questions asked are irrelevant. A sample of questions is presented in the section titled 'Proof of Concept'.

Decision making : The greatest advantage of using GP over other machine learning algorithms is that it provides us with a probability estimate representing confidence of the prediction. With that probability, we can decide if the predicted label can be relied on or that it cannot be trusted and needs to learn this data point. In our model, we exploit this property, unique to GPs. After the interview, GPC will predict the depression label with a confidence p on the audio recording. As mentioned in the modeling, we define a particular probability ρ_c. If the GPC predicts with $p \geq \rho_c$, then we can trust the predicted label and continue to decide appropriately based on the label whereas if the GPC predicts with $p < \rho_c$, then we cannot trust the predicted label and refer the patient to a general practitioner to obtain more data. We determine the probability ρ_c experimentally by observing the prediction quality of the labels in the training data, a summary is given in Table 1. We observe that with a $\rho_c = 70\%$, it would predict the depression label with 100% accuracy. Henceforth, in this paper we will define $\rho_c = 70\%$.

$p_c(\%)$	Depression	No Depression
$[0, 50)$	-	-
$[50, 60)$	$60\%(3/5)$	$80\%(44/55)$
$[60, 70)$	$100\%(1/1)$	$84.2\%(16/19)$
$[70, 80)$	-	$100\%(4/4)$
$[80, 90)$	-	$100\%(1/1)$
$[90, 100]$	-	-
-	-	$90\%(9/10)$

Table 1: Relationship between probability and accuraccy

Scheduling : For the cases when the person needs to be scheduled for an appointment with the psychiatrist, we can use the GPR to determine the predicted PHQ-8 scores. With the predicted scores, we can prioritize certain higher risk individuals over the rest. We were inspired by triage algorithms used in emergency services [Shah et al., 2015; Oredsson et al., 2011], where the priority of one's treatments are decided by the severity of their ailments [Wikipedia, 2016]. We have seen that this method dispenses limited resources with efficiency [Rosedale et al., 2011] throughout many emergency services worldwide. Similarly, we want to apply that idea to psychiatric clinics, which are also facing

a shortage of resources. Since the depression clinic is under-staffed and struggling to keep up with a number of patients, it is wise to prioritize the individuals who are predicted with higher PHQ-8 scores, which indicates that they are likely to be more depressed. Hence, we optimize the scheduling of appointments, prioritizing people with higher predicted PHQ-8 scores.

Qualitative Advantages

The greatest advantage of using a GP is that it provides us with a probability estimate along with the predicted label. We rely on the probability to determine the reliability of the predicted label. As we are dealing with human beings, we would want to rely on the predicted label only if it is reliable. Additionally, we have read in medical literature regarding the dire consequences of misdiagnosis and/or inappropriate treatment [Nasrallah, 2015; Bowden, 2001; Dunner, 2003] in the area of depression. With the probability, we can be confident of the decisions that we make; that can potentially affect a person. Assume that we have a predicted label that can be trusted, then with the label, i.e. true for depressed and false for not depressed, we can decide if the person needs to come to the clinic for an appointment. Hence, reducing the number of appointments that are made, easing the workload of the staff in the clinic.

In addition to reducing the workload of the staff, the GP models can potentially improve as more patients go through the pre-screening process. It can be the case that the GP is not confident of its prediction, then we should not trust the label that the GP had predicted. Then, a psychiatrist would need to determine with the help of a physical examination if the person in question is depressed or not and administer the appropriate treatment. From the appointment with the psychiatrist, he/she can determine if the patient is depressed or not. With the ground truth, we can now train the GPs with the new data point. We are also careful only to update the GP if the data point is predicted with low confidence.

In our model, we do not require a trained staff to administer the phone interview. GP will determine the depression label of the person objectively, not considering the content of the interview but relying on certain depression indicators in speech [of Health and Services, 2015]. This allows the removal of any bias (e.g. gender, racial) in pre-screening, which can potentially cause misdiagnosis. Additionally, the clinic can save skilled human resources as they can hire anyone or can use an automated system to perform the pre-screening task. We also prioritize the clinic's resources based on a person's depression severity, allowing better allocation of resources. This ensures that the clinic's resources are directed to people who need them most.

Hence, our model introduces an unbiased pre-screening process reaping the following benefits: leading to a reduction in the number of appointments, reduction of human resources required, a pre-screening process whose accuracy improves incrementally over time, and a better allocation of resources.

Proof Of Concept

In order demonstrate our framework, a prototype has been created. Before we are going deeply into the prototype, we will first talk about the success of our GPC. Then we will discuss about a scenario of a busy doctor and how our prototype is going to help him.

The success of GPC

As Gaussian Process (Dot-Product) works well in the depression regrassion task, we believe that it will work well in the classification task too, and we tested it. The result is shown in figure 3 and table 2.
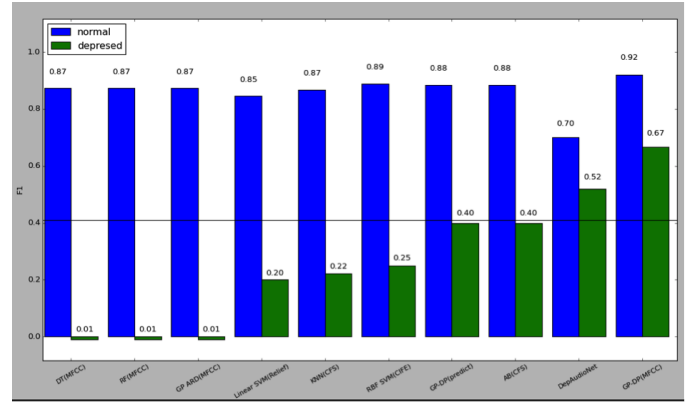

Figure 3: Results of the classification with the GP Dot Product

	F1	Precision	Recall	Accuracy
Gaussian Process (Dot-Product)	0.67(0.92)	0.57(0.96)	0.8(0.88)	0.87
DepAudioNet	0.52(0.7)	0.35(1.0)	1.0(0.54)	-
BaseLine	0.41(0.58)	0.27(0.94)	0.89(0.42)	-

Table 2: Our GPC beats the related study (DepAudioNet) and the baseline.

As shown in graph 3, GPC (with Dot-Product Kernel) yields the best result compared to other existing machine learning methods. In summary, the table 2 contrasts the fact that our GPC beats the related study (DepAudioNet) and the baseline.

A busy doctor

In a clinic, there is only one doctor who is called Antoine and he is very busy. So, he sets a rule: in order to be consulted by him regarding on depression, one has to make a phone call to the clinic and has a phone interview one week before the consultation. However, he is still very busy and would like to minimize the number of person he is consulting. Therefore, an autonomous agent is designed for him to help him to perform a pre-screening and schedule the time for him. In this scenario, the agent is written in Java and the agent is designed to have the workflow shown in figure 4,
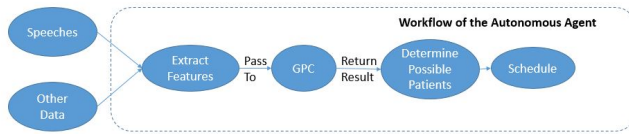
Figure 4: The workflow of the autonomous agent.

Phone Interview : Right now, Antoine has to ask his nurse, Joy to do the phone interview as this task is not supported by the autonomous agent yet. In the future, the phone interview might be able to be conducted by the autonomous agent. Antoine has given his nurse a guideline to follow, and the guideline is the following:

1. The aim of the phone interview is to collect the speeches of the person.

2. The phone interview is preferably to be over 1 minute.

3. The interviewer should use the same application such as Call Recorder to record sound.

4. The interviewer must ask the person some questions during the recording. The interviewer could:

   (a) Ask for the person's personal contact, such as email, or phone number, or address.

   (b) Ask the person's availability to come to the clinic, such as the date and time, or period of time

   (c) Ask the person's individual health screening history: Is this the first time you find a psychologist to talk with? Do you have any medical records in the past?

   (d) Ask the person, is he/she is willing to help out in a questionnaire if the phone interview is less than 1 minute.

5. It is not necessary to ask all the questions above as long as the length of the phone interview is around 1 minute.

6. Avoid saying something like, "Cheer up!", "Lighten up!" to provide help, or giving suggestions like "Take a hot bath. That's what I always do when I upset!" to the patients. Each phone call should obey the principle of "Ask for information and NOT providing unprofessional guidance". [?]

7. End with, thank you Sir/Ms for your time. We will contact you within 3 working days.

8. Extracting the speeches of the person from recording manually and pass it to the agent.

The autonomous agent : Suppose Joy collected many persons' speeches and relevant data such as the number of hours they needed to consult the doctor and their names. She ran the autonomous agent to process them at night to produce a schedule for Antoine. The agent first extracts the audio features from the speeches. MFCC is the audio feature that is extracted by the agent. They are passed to the GPC

to classify if the person is depressed. As the GPC is implemented using Python, the agent would execute the GPC by executing these lines of codes:

```java
private ArrayList<DiagnosedPatient>
    classifyDepression(ArrayList<DiagnosedPatient> patients) {
  Process process;
  try {
    String command = "python onlinegp.py " +
        AudioFeaturesGenerator.EMOTION_DCAPSWOZ_ALL;
    //System.out.println(command);
    process = Runtime.getRuntime().exec(command);
    InputStream is = process.getInputStream();
    InputStreamReader isr = new InputStreamReader(is);
    BufferedReader br = new BufferedReader(isr);
    String line;

    int i = 0;
    while ((line = br.readLine()) != null) {
        String[] tokens = line.split(" ");
        int depressedNum = Integer.parseInt(tokens[0]);
        boolean isDepressed = depressedNum == 1;
        patients.get(i).setIsDepressed(isDepressed);
        patients.get(i).setGpConfidence(Double.parseDouble(tokens[depressedNum
            + 1]));
        i++;
    }
    br.close();
  } catch (IOException e) {
    // TODO Auto-generated catch block
    e.printStackTrace();
  }
  return patients;
}
```

Then, GPC will reply its predition on the person with the confidence level. The agent will trust the GPC if the person is predicted as not depressed with the confidence level $\geq$ 70, and these persons will be suggested to be rejected and will not be scheduled a consultation as the agent assumes Antoine is very busy and he would like to consult as little person as he could.

These are a few sample data that is obtained from the GP:

1. Daniel: GP is 91.9163306736% confidence to say that the patient is normal and required 3.0h.

2. Miguel: GP is 54.36852983429999% confidence to say that the patient is depressed and required 1.0h.

3. Maria: GP is 59.406768555% confidence to say that the patient is normal and required 2.5h.

According to these data, Daniel is not going to be scheduled a consultation, while the others will. The scheduling is just done using greedy algorithm as the agent is designed to be used by Antoine only.

After the scheduling, a schedule of Antoine for the next week is generated and it is looked like this:

```
Antoine:
Miguel : 14-11-2016 (Mon) 08:00 14-11-2016 (Mon) 09:00
Davi : 14-11-2016 (Mon) 09:00 14-11-2016 (Mon) 10:30
Ethan : 14-11-2016 (Mon) 13:00 14-11-2016 (Mon) 15:00
Maria : 15-11-2016 (Tue) 08:00 15-11-2016 (Tue) 10:30
Lucia : 14-11-2016 (Mon) 10:30 14-11-2016 (Mon) 12:00
Elena : 14-11-2016 (Mon) 15:00 14-11-2016 (Mon) 16:30
Free: 14-11-2016 (Mon) 12:00 14-11-2016 (Mon) 12:00
Free: 14-11-2016 (Mon) 16:30 14-11-2016 (Mon) 17:00
Free: 15-11-2016 (Tue) 10:30 15-11-2016 (Tue) 12:00
Free: 15-11-2016 (Tue) 13:00 15-11-2016 (Tue) 17:00
Free: 16-11-2016 (Wed) 08:00 16-11-2016 (Wed) 12:00
Free: 16-11-2016 (Wed) 13:00 16-11-2016 (Wed) 17:00
Free: 17-11-2016 (Thu) 08:00 17-11-2016 (Thu) 12:00
Free: 17-11-2016 (Thu) 13:00 17-11-2016 (Thu) 17:00
Free: 18-11-2016 (Fri) 08:00 18-11-2016 (Fri) 12:00
Free: 18-11-2016 (Fri) 13:00 18-11-2016 (Fri) 17:00
```

Note that, in the future if Antoine has a few new partners, a new scheduling algorithm is needed to ensure that everyone's work will be distributed clearly. It is not hard to observe that it is a Constraint Satisfaction Problem (CSP), and thus we suggest genetic algorithm for this case, as it is one of the famous and useful algorithm for scheduling. PHQ-8 score is not used in the scheduling algorithm right now due to the high error-rate. However, we think that it will be good if we take the severity of depression into account in the future. Online GP is not implemented yet but we also think that it will be good if we implement it to improve the accuracy or error rate of the model.

Conclusion

Our work has succesfully shown that GP is a good model for this problem and can predict PHQ-8 better than state-of-the-art machine learning models. In addition to being on par or better at prediction, GP can inherently provide an estimate of prediction uncertainty. This allows the user to gauge the model's confidence of the prediction, and to make more informed decisions based on both the prediction and its uncertainty. We can also intelligently supplement more data to our training set based on the prediction uncertaintly. Therefore, after considering both results and GP's advantages, we conclude the GP Dot Product trained using MFCC feature set is a good model for depression prediction.

Further Work

For this experiment, we only used machine learning algorithms with their default parameters. An aspect that deserves further exploration is to perform automatic hyper-parameter optimization across all the machine learning algorithms to fine-tune each model's performance. In particular, we can try Hyperopt-sklearn [Komer, Bergstra, and Eliasmith, 2014] or GP based hyper-parameter tuner. We opine that with hyper-parameter tuning, we can predict PHQ-8 scores better and can have a more objective comparison of the different learning algorithms.

Contributions

- **Antoine Charles Vincent Garcia**: Scripting the program, setting up machine learning libraries, running tests and generation of the utility function.

- **Chan Jun Wei**: Scripting the program, setting up machine learning libraries, running tests and generation of the utility function.

- **Chen Tze Cheng**: Scripting the program, setting up machine learning libraries, running tests and generation of the utility function.

- **Eric Ewe Yow Choong**: Formatting the report as well as research and writing up of the technical approach section.

- **Han Liang Wee, Eric**: Retrieving data, testing as well as research and writing up of the technical approach section.

- **Ho Wei Li**: Research, vetting of the report and writing up of the motivation and introduction of the experiment.

References

Badger, L.; deGruy, F.; Hartman, J.; Plant, M.; Leeper, J.; Ficken, R.; Maxwell, A.; Rand, E.; Anderson, R.; and Templeton, B. 1994. Psychosocial interest, medical interviews, and the recognition of depression. *Archives of Family Medicine* 3(11):974.

Bowden, C. L. 2001. Strategies to reduce misdiagnosis of bipolar depression. *Psychiatric Services*.

Chong, S.; dS Verma; Vaingankar, J.; Chan, Y.; Wong, L.; and Heng, B. 2009. Perception of the public towards the mentally ill in a developed asian country. *Social Psychiatry and Psychiatric Epidemiology* 42(9):734–739.

Chong, S. A.; Abdin, E.; Vaingankar, J. A.; Heng, D.; Sherbourne, C.; Yap, M.; Lim, Y. W.; Wong, H. B.; Ghosh-Dastidar, B.; Kwok, K. W.; and Subramaniam, M. 2012. A population-based survey of mental disorders in singapore. *ANNALS Academy of Medicine Singapore* 41:49–66.

Cummins, N.; Scherer, S.; Krajewski, J.; Schnieder, S.; Epps, J.; and Quatieri, T. F. 2015. A review of depression and suicide risk assessment using speech analysis. *Speech Communication* 71:10–49.

Dunner, D. L. 2003. Clinical consequences of under-recognized bipolar spectrum disorder. *Bipolar disorders* 5(6):456–463.

Fiske, A.; Wetherell, J. L.; and Gatz, M. 2009. Depression in older adults. *Annual Review of Clinical Psychology* 5:363–389.

Goldman, L. S.; Nielsen, N. H.; and Champion, H. C. 1999. Awareness, diagnosis and treatment of depression. *Journal of General Internal Medicine* 14(9):569–580.

Greenberg, P. E.; Fournier, A.-A.; Sisitsky, T.; Pike, C. T.; and Kessler, R. C. 2015. The economic burden of adults with major depressive disorder in the united states. *The Journal of Clinical Psychiatry* 76(2):155–162.

Han, E. L.; Chan, J.; Ho, W. W.; Chen, T.; Ewe, E. Y.; and Garcia, A. C. 2016. Depression predictioin with gaussian process. CS4246 AI Planning and Decision Making Project 1.

Hashim, N. W.; Wilkes, M.; Salomon, R.; Meggs, J.; and France, D. J. 2016. Evaluation of voice acoustics as predictors of clinical depression scores. *Journal of Voice* 0(0).

Jiaquan Xu, M.; Sheery L. Murphy, B.; Kenneth D. Kochanek, M.; and Brigham A. Bastian, B. 2016. Deaths: Final data for 2013. *National Vital Statistics Reports* Vol 64, No. 2.

Johnstone, T. 2001. *The effect of emotion on voice production and speech acoustics*. Ph.D. Dissertation, University of Western Australia.

Kessler, R.; Berglund, P.; Demler, O.; and et al. 2003. The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (ncs-r). *JAMA* 289(23):3095–3105.

Komer, B.; Bergstra, J.; and Eliasmith, C. 2014. Hyperopt-sklearn: Automatic hyper-parameter configuration for scikit-learn.

Kroenke, K.; Strine, T.; Spitzer, R. L.; Williams, J. B.; Berry, J. T.; and Mokdad, A. 2008. The phq-8 as a measure of current depression in the general popoulation. *Journal of Affective Disorders* 114(1-3):163–73.

Nasrallah, H. A. 2015. Consequences of misdiagnosis: Inaccurate treatment and poor patient outcomes in bipolar disorder. *The Journal of clinical psychiatry* 76(11):e1328–e1328.

of Health, U. D., and Services, H. 2015. Depression, what you need to know.

Oredsson, S.; Jonsson, H.; Rognes, J.; Lind, L.; Göransson, K. E.; Ehrenberg, A.; Asplund, K.; Castrén, M.; and Farrohknia, N. 2011. A systematic review of triage-related interventions to improve patient flow in emergency departments. *Scandinavian journal of trauma, resuscitation and emergency medicine* 19(1):1.

Rosedale, K.; Smith, Z.; Davies, H.; and Wood, D. 2011. The effectiveness of the south african triage score (sats) in a rural emergency department. *SAMJ: South African Medical Journal* 101(8):537–540.

Rothman, A., and Wagner, E. 2003. Chronic illness management: what is the role of primary care? *Annals of Internal Medicine* 138(3):256–261.

Samaritans of Singapore. 2016. National suicide statistics.

Shah, S.; Patel, A.; Rumoro, D. P.; Hohmann, S.; and Fullam, F. 2015. Managing patient expectations at emergency department triage. *Patient Experience Journal* 2(2):31–44.

Telford, R.; Hutchinson, A.; Jones, R.; Rix, S.; and Howe, A. 2002. Obstacles to effective treatment of depression: a general practice perspective. *Family Practice* 19(1):45–52.

Valstar, M.; Gratch, J.; Ringeval, F.; Torres, M. T.; Scherer, S.; and Cowie, R. 2016a. AVEC 2016 – Depression , Mood , and Emotion Recognition Workshop and Challenge.

Valstar, M. F.; Gratch, J.; Schuller, B. W.; Ringeval, F.; Lalanne, D.; Torres, M.; Scherer, S.; Stratou, G.; Cowie, R.; and Pantic, M. 2016b. AVEC 2016 - depression, mood, and emotion recognition workshop and challenge. *CoRR* abs/1605.01600.

Wikipedia. 2016. Triage — Wikipedia, the free encyclopedia. `http://en.wikipedia.org/w/index.php?title=Triage&oldid=748740826`. [Online; accessed 10-November-2016].

World Health Organization. 2004. The global burden of disease: 2004 update. Technical report, World Health Organization.