

PREDICTION OF BECK DEPRESSION INVENTORY (BDI-II) SCORE USING ACOUSTIC MEASUREMENTS IN A SAMPLE OF IIUM ENGINEERING STUDENTS

Muhamad Fikri Zanil¹, Nik Nur Wahidah Nik Hashim² and Huda Azam³

¹⁻³ Department of Mechatronics Engineering, International Islamic University Malaysia

Abstract. Psychiatrist currently relies on questionnaires and interviews for psychological assessment. These conservative methods often miss true positives and might lead to death, especially in cases where a patient might be experiencing suicidal predisposition but was only diagnosed as major depressive disorder (MDD). With modern technology, an assessment tool might aid psychiatrist with a more accurate diagnosis and thus hope to reduce casualty. This project will explore on the relationship between speech features of spoken audio signal (reading) in Bahasa Malaysia with the Beck Depression Inventory scores. The speech features used in this project were Power Spectral Density (PSD), Mel-frequency Cepstral Coefficients (MFCC), Transition Parameter, formant and pitch. According to analysis, the optimum combination of speech features to predict BDI-II scores include PSD, MFCC and Transition Parameters. The linear regression approach with sequential forward/backward method was used to predict the BDI-II scores using reading speech. The result showed 0.4096 mean absolute error (MAE) for female reading speech. For male, the BDI-II scores successfully predicted 100% less than 1 scores difference with MAE of 0.098437. A prediction system called Depression Severity Evaluator (DSE) was developed. The DSE managed to predict one out of five subjects. Although the prediction rate was low, the system precisely predict the score within the maximum difference of 4.93 for each person. This demonstrates that the scores are not random numbers.

1.Introduction

Based on The National Health and Morbidity Survey on 2015, 29 percent of Malaysians had depression which is 17 percent higher than the statistic in 2011[1]. The percent increment shown by this statistic demonstrated that this issue has become very alarming and shows that this project have very high potential to help this problem. Depression does not only has effect on the individual, it also adds problems to the development of country as each individual has a value to the country.

However, we still having difficulties in solving this problem since it is very challenging for clinician to identify the depression severity of a patient. Researchers have struggled to find a way to identify the severity of depression since long ago, as a result, Aaron T. Beck has published an instrument to measure the severity of depression named Beck Depression Inventory. It is actually a self-report, which contain 21 questionnaires and each one have four options which indicate the score from 0 (not present) to 3 (severe). Each of the item or question have its own category, including sadness, pessimism, past failure, loss of pleasure, suicide thought, crying, agitation, loss of interest, indecisiveness, worthlessness, loss of energy, change in sleep pattern, irritability, change in appetite, concentration



difficulty, fatigue and loss interest in sex respectively [2]. The severity of depression can be determined by the rating of overall BDI-II scores as in the Table 1.

Table 1. Depression indication according to scores range [2]

BDI-II Scores	Depression Indication
10 and lower	Normal
11 to 16	Sad
17 to 20	Mild
21 to 30	Moderate
31 to 40	Severe
40 and more	Need Serious treatment

Nowadays, the advancement of technologies has changed our daily lives. Many things have been invented and at the same time encourage researchers to look at new approaches in identifying the severity of the depression. Some use imaging device like camera to do image processing and some use audio device to do audio processing and some even do both. This new approach has the goal to accurately measure the severity of depression instantaneously.

As for this project, we focus on audio processing using the reading speech of the subject. It is good to review others works related to this project. Based on other research by others [3] [4] [5], most of them extract Mel-Frequency Cepstral Coefficients (MFCC) speech features, but the approach to predict the scores were different. They try to predict using Gaussian Mixture Model [3], Gaussian Staircase Regression [4] and Multiple Linear Regression Model [5]. The best method to apply is using Linear Regression as it gave the minimum mean absolute error.

2. Data

Data is an important requirement to complete this project as this data are needed to train the prediction system. The subject need to fill the BDI-II Inventory and then the need to be interviewed and read a passage which will be recorded using Tascam DR-05 with sample rate at 44.1 kHz. All the recorded session is done in a closed room. After the whole data collection is done, the subject will receive RM10.00 as honorarium.

The subjects were male and female with the age range 18-29 which are engineering students at International Islamic University Malaysia (IIUM). All of the subjects do not have any respiratory problem and not under the influence of alcohol. The recording session was done in a closed room, but not fully soundproof to make sure the robustness of the system.

During the recording session the subject reads a passage. The passage should be a rainbow passage, but since there is no Bahasa Malaysia passage that is equivalent to the rainbow passage, we decided to go with any controlled passage.

3. Speech Features

Several types of speech were collected to be tested to identify the best combination of speech features to be implemented in DSE system. The speech features include 3 Power Spectral Densities (PSD), 13 MFCC, 6 Transition Matrix (TM), 2 formant, maximum and minimum pitch which in total of 26 features. This system, however only applies PSD, MFCC and TM since the best combination of speech features only contain these 3 types of speech features.

3.1. Power spectral Density (PSD).

Three features are collected from PSD. PSD is actually a power intensity of a signal in frequency or for a sound wave it is loudness. Most energy in a speech of an adult contains in a range 0 to 2000Hz. To minimize the number of features, the range of 0Hz to 2000Hz is divided into 4 segments with equal width. Thus, the first band would be 0Hz to 500Hz and the second band from 500Hz to 1000Hz and the third band would be 1000Hz to 1500Hz. Only the first three bands are collected as the fourth band can be predicted by the first three bands.

3.2. Transition Matrix.

This feature is a set of probability of state transition as shown in figure 1. Each block represents 0.04 second segment and is categorized into number 1, 2 and 3 which represent voiced, unvoiced and silence respectively. The unvoiced speech segments are basically a noise-like signal which has high frequency components. Whereas the voiced signal has low frequency components. Butterworth Filter was used to categorize this signal. However, silence speech sound wave also has low frequency components like voice speech sound wave. In order to categorize between them, a threshold is needed. [6]

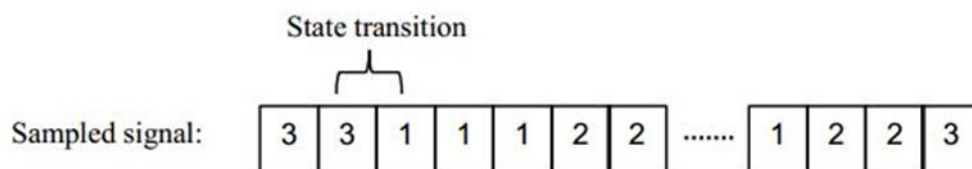


Figure 1. Graphical representation of state transition [5]

3.3. Mel-Frequency Cepstral Coefficients (MFCC)

Mel-Frequency Cepstral Coefficients are quite popular speech features and commonly used for speech analysis. MFCC represents the logarithmic perception of the human auditory system. MFCC also can represent the spectrum of an audio signal in the Mel scale. It acts like human hearing system.

4. Methodologies

The system development was divided into two parts which are identifying the best speech features for part 1 to achieve the first two objectives and develop Depression Severity Evaluator (DSE) to achieve the third objective.

In part 1 the flowchart is shown in Figure 2. This process is very important in order to continue to part 2. The goal of this process is to find the combination of speech features that give the lowest mean absolute error in predicting the BDI-II score using the data that have collected. Too many features may result in high error, so the Sequential Forward Selection (SFS) method was used with linear regression to find the predicted scores. Each feature was selected one-by-one and evaluated using linear regression until all features are evaluated. Feature that gives the least mean sum of absolute error was selected and add it to the empty set one at a time.

Jackknife method also used in this process to train the system. The implementation of the jackknife method in this process is on the basis of leave-one-out. The process involves discarding out one student's score data from the data set and develops a training data set with the remaining N-1 students. The excluded data is then tested. This process is repeated by excluding the next students from the overall set of data until all students have been chosen to be excluded as testing data.

The combination of speech features that give the least mean absolute error were selected as the best and used to find the model coefficient, a using the formula of linear regression $b = D \times a$ where b is the column vector of the actual value of BDI-II scores, D is the input vectors (combination of speech features). Note that the model coefficient, a , is in matrix form.

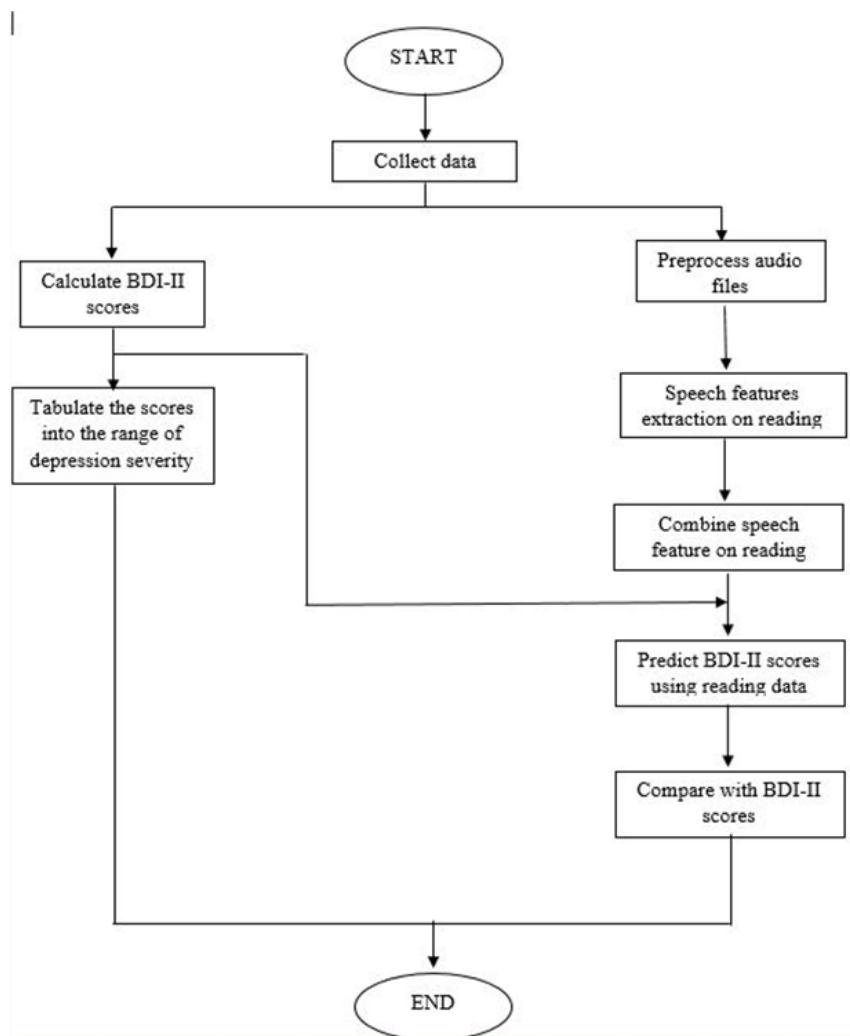


Figure 2. Flowchart of identifying best combination of Speech Features for BDI prediction.

Once we have the model coefficient, **a**, from the best combination of speech features, we can continue to part 2 to build the Graphical User Interface (GUI) of the DSE. The simple block diagram of DSE is shown in figure 3.

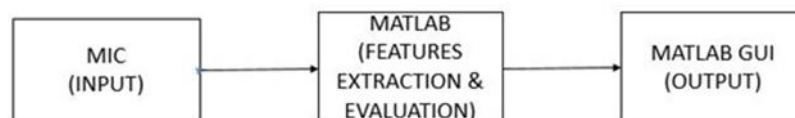
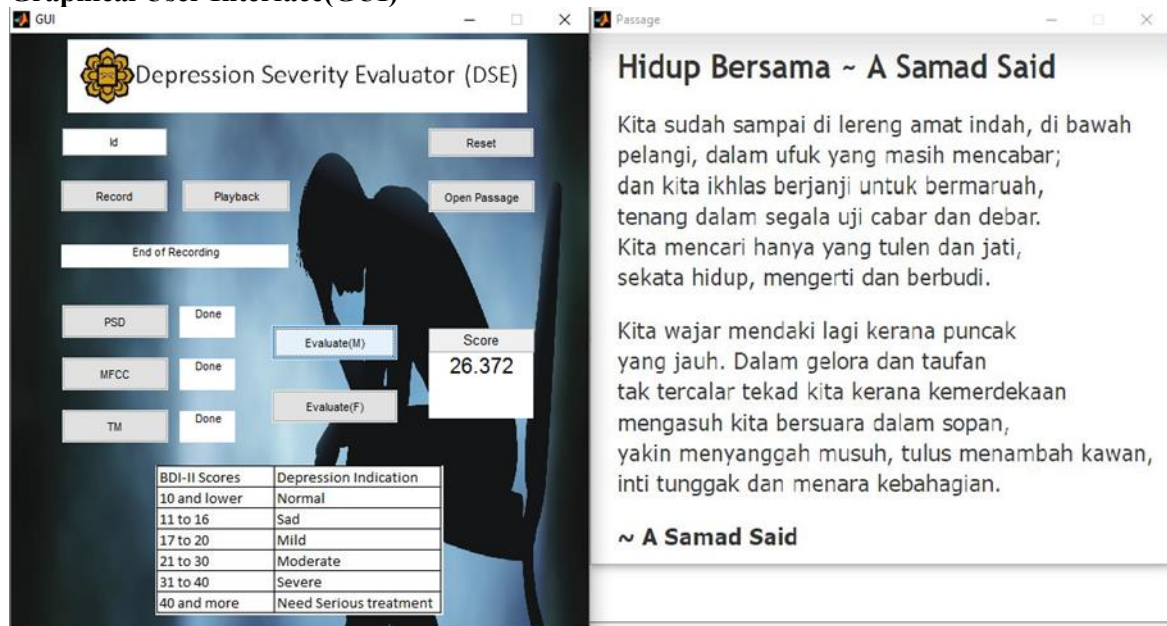
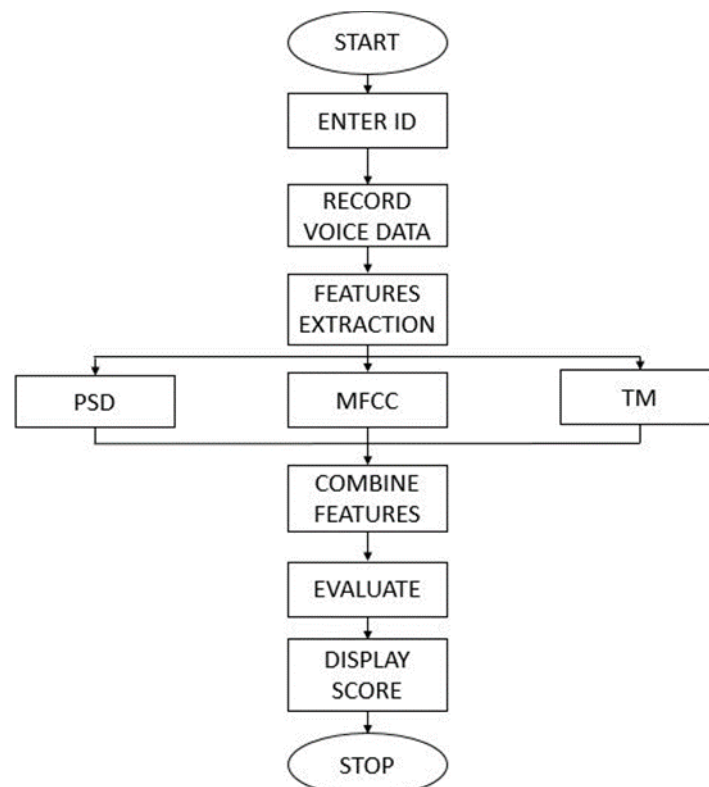


Figure 3. Simple Block Diagram of DSE

Graphical User Interface(GUI)**Figure 4.** Depression Severity Evaluator (DSE) GUI**Figure 5.** Depression Severity Evaluator (DSE) operational flowchart.

The process of this system almost the same with the process in part 1. Only few changes were done on the feature extraction. The major changes in this system compare to part 1 is on

evaluating process. Where the model coefficient, **a**, in part 1 were used to multiply with the best combination of speech features which shown in table 2 for male and female.

Table 2. Best speech features combinations

Male		Female	
Number of features	List of feature	Number of features	List of feature
6	mfcc7, mfcc1, tm9, mfcc9, mfcc11, tm1	6	mfcc2, mfcc11, mfcc8, mfcc7, mfcc10, psd3

5.Result and discussion

For part 1, the result was in term of mean absolute error (MAE). As the process depended on gender separately same goes on the result. For male reading, it can be seen that (SFS) method give the most minimum MAE which is 0.098437 with 6 features selected. The minimum MAE for (SBS) method, however could not outperform the SFS method with MEA of 1.95651 with 8 features left.

For female reading, again, the Sequential Forward Selection method gives the lowest mean absolute error with also 6 speech features. The mean absolute error is quite low which at 0.4096. Whereas, the MEA for SBS method is 4.261898 with 7 features left.

Thus, the best combination of speech features for male reading are mfcc7, mfcc1, tm9, mfcc9, mfcc11, tm1 whereas for female reading mfcc2, mfcc11, mfcc8, mfcc7, mfcc10, psd3. Quite interesting that the remaining best features are MFCC, PSD and TM. MFCC and PSD are a spectrum based measure where spectrum based, in a sense that it involves the energy information, whereas for Transition Matrix, more like speech pattern.

As for part 2, the DSE GUI was tested by using the train data to validate the system as few changes have been made. The result was pretty good, we can say that the system can predict the score accurately within the trained data. Then, the system was tested on a new subject where the subject need to fill BDI-II survey as an actual score and read a passage two times to get scores using DSE as predicted scores. All the result were evaluated and tabulated in table 3. The percentage error were calculated by using actual score and the average of test score 1 and 2. Whereas, the test score difference shows the different between test score 1 and 2. Despite the unsatisfactory on percentage error, the test score difference shows the potential of DSE.

Table 3. DSE analysis result

Subject	Test1	Test2	Test3	Test4	Test5
Actual Score	17	13	4	13	34
Test Score 1	4.29	4.25	25.36	27.63	40
Test Score 2	5.36	2.87	28.03	26.96	35.07
% Error	71.62	72.62	567.38	107.35	37.53
Test Score Difference	1.07	1.38	2.67	1.35	4.93

6.Conclusion

As a conclusion, the relevant speech features from Bahasa Malaysia's speech were obtained. The combination of speech features for BDI-II score prediction also managed to be evaluated and the prediction system Depression Severity Evaluator has developed. Even though the DSE could not accurately measure the severity of depression, the score is not just a random numbers which have

potential to be improved. Thus, it highlights the potential of this project in identifying the severity of depression.

7. References

- [1] Ministry Of Health Malaysia 2015 *National Health and Morbidity Survey of 2015: Non-Communicable Diseases, Risk Factors & Other Health Problem* 2 pp187-184
- [2] Simin H and Zahra G 2011 *A survey Beck test in university students & its relationship between some related risk factor*.
- [3] Lopez-Otero P, Dacia-Fernandez L and Garcia-Mateo C 2013 *A Study Of Acoustic Features for Depression Detection*
- [4] Williamson J R, Quatieri T F, Helfer B S, Ciccarelli G and Mehta D D 2013 *Vocal Biomarkers of Depression Based on Motor Incoordination and Timing*
- [5] Nik Hashim NNW 2014 *Analysis Of Speech Features As Potential Indicators For Depression And High Risk Suicide And Possible Predictors For The Hamilton Depression Rating (HAMD) And Beck Depression Inventory Scale (BDI-II)* 113-41
- [6] Ozdas A 2001 *Analysis of Paralinguistic properties of speech for near-term suicidal risk assessment* 51-48

Acknowledgments

This work is part of the ongoing project on the development of BDI-II score prediction system funded by the Department of Mechatronics Engineering, International Islamic University Malaysia.

Corrigendum: Prediction of Beck Depression Inventory (BDI-II) Score Using Acoustic Measurements in a Sample of Iium Engineering Students

IOP Conf. Ser.: Mater. Sci. Eng. 260 (2017) 012022

Muhamad Fikri Zanil¹, Nik Nur Wahidah Nik Hashim² and Huda Azam³

¹ Department of Mechatronics Engineering, International Islamic University Malaysia

² Department of Mechatronics Engineering, International Islamic University Malaysia

³ Department of Mechatronics Engineering, International Islamic University Malaysia

Page 7:

In the Acknowledgements section, the following text appears:

“ This work is part of the ongoing project on the development of BDI-II score prediction system funded by the Department of Mechatronics Engineering, International Islamic University Malaysia.”

This should read:

“This work is part of the ongoing project on the development of BDI-II score prediction system funded by the Department of Mechatronics Engineering, International Islamic University Malaysia. This research is supported by the Research Acculturation Grant Scheme (RAGS 14-040-0103) by the Ministry of Higher Education Malaysia.”