2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Classification of Depression and Other Psychiatric Conditions Using Speech Features Extracted from a Thai Psychiatric and Verbal Screening Test

N. Klangpornkun, M. Ruangritchai, A. Munthuli, C. Onsuwan, K. Jaisin, K. Pattanaseri, J. Lortrakul, P. Thanakulakkarachai, T. Anansiripinyo, A. Amornlaksananon, S. Laohawee, and C. Tantibundhit*

***Abstract*— Depression is a common and serious mental illness which negatively affects daily functioning. To prevent the progression of the illness into severe or long-term consequences, early diagnosis is crucial. We developed an automated speech feature analysis application for depression and other psychiatric disorders derived from a developed Thai psychiatric and verbal screening test. The screening test includes Thai's version of Patient Health Questionnaire-9 (PHQ-9) and Hamilton Depression Rating Scale (HAM-D), and 32 additional emotion-induced questions. Case-control study was conducted on speech features from 66 participants. Twenty seven of those had depression (DP), 12 had other psychiatric disorders (OP), and 27 were normal controls (NC). The five-fold cross-validation from 6 settings of 5 classifiers with the combination of PHQ-9 and HAM-D scores, and speech features were examined. Results showed highest performance from the multilayer perceptron (MLP) classifier which yielded 83.33% sensitivity, 91.67% specificity, and 83.33% accuracy, where negative-emotional questions were most effective in classification. The automated speech feature analysis showed promising results for screening patients with depression or other psychiatric disorders. The current application is accessible through smartphone, making it a feasible and intuitive setup for low-resource countries such as Thailand.**

I. INTRODUCTION

Ongoing evolution of technology, social media, economic and social standards, and society has caused a number of teenagers and adults to experience feelings of loneliness, severe sadness and more which eventually lead to depression [1]. It currently affects more than 264 million people worldwide and will only continue to grow [2]. With predictions of depression to be the second leading cause of disability by 2030 [2], a fast, accurate, and suitable screening method is desperately needed.

Depression symptoms may vary from mild to extremely severe and can include symptoms like loss of energy, feeling worthless, changes in sleep schedules and appetite, etc. In more serious cases, symptoms can even lead to acting upon thoughts of death or suicide [3]. Luckily, depression is treatable in some cases. The most common treatments are antidepressant medications and psychological counseling [4]. However, antidepressants and counseling can be expensive in some countries where mental health is not widely discussed. Therefore, early diagnosis for patients is crucial as early treatment decreases the chances that the depression will progress [5].

Current depression diagnosis methods in English-speaking countries usually are based on the gold standard and questionnaires which use rating scales for example, the Patient Health Questionnaire-9 (PHQ-9) [6], a very brief, easy to administer and interpret depression screening method, The Hamilton Depression Rating Scale (HDRS or HAM-D) [7], a clinician-administered depression assessment scale containing 17 questions addressing symptoms of depression, and the Beck Depression Inventory (BDI) [8], an assessment with 21 questions and gives a score from $0-63$ which will determine the severeness of depression. These assessments have been translated to different languages [9] such as French [10], [11], Mandarin [12], and Thai [13], [14].

However, these methods are only paper-based, meaning they are time-consuming as they require specialized doctors to monitor the assessments. Costs to visit the doctor for diagnosis can be expensive as specialized personnel is required for the task. Many young adults may struggle to afford hospital bills [15]. This is especially dangerous as depression is usually common in ages $18-25$ [16]. Fears of being judged by family members, friends, and strangers may also be a contributing factor as to why people may avoid being diagnosed in hospital settings [17].

Several studies have found correlation between speech patterns and depression [18], [19] and noted that that articulation, pitch, speaking rate, and loudness were all factors which differed in patients with depression versus healthy controls. Since then, more researches have been done on developing automated programs and AI to detect depression in patients using speech features. In English-speaking countries, multiple studies have used machine learning [20]–[28] and deep learning techniques [29], [30] using several different classifiers such as SVM and MFCC.

For Mandarin-speaking citizens, there have been studies which have attempted to develop new classification systems [31] and have looked into using machine learning to yield the best results [32]. A speech corpus database was also created by Lu *et al.* to allow more research to be conducted on depression screening based on the Mandarin language [33]. This is partially because Mandarin is a tonal language so the voice recordings and screening methods may have different results than non-tonal languages like English. Additionally, culture differences may also affect the way patients answer

*C. Tantibundhit (e-mail: tchartur@engr.tu.ac.th), N. Klangpornkun, and A. Munthuli are with Center of Excellence in Intelligence Informatics, Speech, and Language Technology, and Service Innovation (CILS), Faculty of Engineering and C. Onsuwan, T. Anansiripinyo, A. Amornlaksananon, and S. Laohawee are with CILS and Faculty of Liberal Arts, Thammasat University, Thailand. M. Ruangritchai is with CILS and NIST International School, Thailand. K. Jaisin, K. Pattanaseri, J. Lortrakul, and P. Thanakulakkarachai are with Department of Psychiatry, Faculty of Medicine, Siriraj Hospital, Mahidol University, Thailand.

questions [34], [35]. Therefore, screening methods and results might be unique depending on each language.

Current diagnostic methods for depression in Thailand include แบบทดสอบภาวะซึมเศร้า PHQ-9 [13], the translated version of the PHQ-9 assessment and other translated versions of worldwide assessments which have been used for non-tonal languages such as English. However, these translated versions do not fully captivate the Thai culture and language. So far, there have been no Thai-based papers which have investigated creating a questionnaire or attempting to automate screening methods. Therefore, an inexpensive, non-invasive, accurate, and fast screening method is still needed in Thailand.

Seeing as there is a desperate need for an appropriate screening method for Thailand, the aim of this study is to develop a screening tool consisting of 3 sections (see Table I). Using the speech features in this study, we attempt to extract the critical features using the openSMILE program [36] such as fundamental frequency (F0), loudness, intensity and mel-frequency cepstral coefficients (MFCC), and accurately classify the subjects into 3 groups: depression, other conditions, and normal controls.

In Section 2, the process of developed Thai psychiatric and verbal depression screening test is discussed. In Section 3, the verbal part of the assessment is discussed. Section 4 details the data collection process. Section 5, the experimental setup, includes further details on the speech features extracted from the audio files. Sections 6 and 7 discuss the results yielded from the investigation, the significant, and the future work.

II. DEVELOPMENT OF THAI PSYCHIATRIC AND VERBAL SCREENING TEST

A Thai paper-based screening tool has been designed by our research team to screen for depression among Thai patients. It was developed based on previous studies and screening tests [6], [7], [37], [38] and was evaluated by a team of psychiatrists from Siriraj hospital, Thailand. It is categorized into 3 sections, including subject information, pre-existing assessments, and a verbal assessment, which is outlined in Table I. This paper-based screening tool will then be used to develop a mobile application so that the assessment is easily accessible.

III. VERBAL ASSESSMENT

The verbal assessment sessions were recorded and used to screen for speech features. The 5 subsections include verbal instructions, reviews of symptoms, descriptions of emotional experiences, story telling tasks, and imagination tasks.

A. Verbal instructions

The verbal instructions include 3 tasks which ask the subject to follow straight-forward instructions such as, reading a short paragraph, counting from numbers 1 to 20, and repeating certain phrases as fast as possible. These tasks aim to find significant changes in the patient's tone of voice and are used to extract speech features regarding loudness and speech rhythm.

TABLE I: Depression screening test sections

Subject information	1. Age 2. Gender 3. Relationship status 4. Religion 5. Occupation 6. Education level 7. Residence 8. Native language 9. Health conditions
Existing assessment	10. Thai translated version of Hamilton Rating Scale for Depression (HDRS or HAM-D) 11. Thai translated version of the PHQ-9
Verbal assessment	12. Verbal instructions (Items 1-3) 13. Reviews of symptoms (Items 4-9) 14. Descriptions of emotional experiences (Items 10-15) 15. Story telling tasks (Items 16-26) 16. Imagination tasks (Items 27-32)

B. Reviews of symptoms

The reviews of symptoms include 6 tasks based on clinical symptoms of depression, personal feelings and current living conditions. For example, questions include describing emotions for the past 2 weeks or describing any noticeable change in emotion recently. These types of questions were designed to find issues which may be a factor causing depression or signs of depression based on current feelings and mood.

C. Descriptions of emotional experiences

The descriptions of emotional experiences include 6 tasks divided into positive, neutral, and negative emotional experiences. Tasks include describing an experience when you felt happy (positive experience) or describing what the subject would do if you had a year of spare time (neutral experience).

D. Story telling tasks

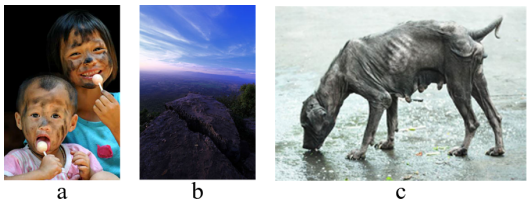
The story telling includes 5 tasks which assess subject's creativity and memory. For example, telling an imaginary story with specific prompts such as being in a loud environment. The subject has 10 seconds to think of the story and 2 minutes to tell the story. The prompts are said out loud through an automated voice and a timer is coded into the application.

E. Imagination tasks

The imagination tasks include 2 tasks. The first is asking the subject to describe their feeling towards happy and sad (feelings of loss, guilt, and worthlessness) scenarios such as having dinner with someone you love or attending a funeral of someone you love. The second task is describing their feelings towards pictures which represent sad, neutral, and happy emotions. Figure 1 shows the kinds of pictures included in the task.

IV. DATA COLLECTION

This section details how we collected participant data from a computerized version of the developed screening test. Importantly, the verbal assessment is used in analysis of speech feature data.



a b c

Fig. 1: Samples of pictures used in imagination tasks a) "smiling kids" representing happiness, b) "sunset" representing neutral emotions, and c) "stray dog" representing sadness [38].

a
b
c
d
e
f
g
h

Fig. 2: Screenshot of the data collection for verbal assessment: a) Participant code, b) Dimmed box indicates the number of tests. Solid color box indicates the number of tests performed, c) Timer, d) Instruction: "Please answer the following questions with as much detail as possible.", e) Question: "How are you doing lately?", f) Trigger command that is displayed every 20 seconds: "Please give me more details.", g) Control buttons: Return, Start, Stop, and Next buttons in order from left to right, and h) The left part shows a video with a face mask. The right section shows descriptions of the control buttons.

A. *Computerized version of the developed screening test*

This section will be detailing the verbal assessment sessions where participants come from 3 groups: those with depression (DP), other psychiatric conditions (OP), and normal controls (NC). The include audio and video recordings of 32 questions, consisting of 5 subsections: verbal instructions, reviews of symptoms, descriptions of emotional experiences, story telling tasks, and imagination tasks. The total time of collecting data per subject is approximately 40 minutes to 1.5 hours.

The data for each participant was collected in video and audio formats in order to find consistency in speech, part of speech, emotions, and detect facial landmarks. The data were collected using mobile phone via applications developed by the research team on Android and iOS systems in resolution 1920×1080 pixels or approximately 2 mega pixels at 25 frames per second. The mobile phone was held vertically in order to be able to film subjects from head to shoulder. Through the application, commands, pictures or videos would appear in the center of the screen. The subject's self-view was shown in the lower left corner of the screen as shown in Figure 2. Their face was masked using the Expo FaceDetector so that the subject cannot see their face while answering questions.

B. *Sampling method*

For this study, data were collected through a mobile application developed by our research team. The data were collected at the psychiatric outpatient clinic, Siriraj hospital. All of the participants were older than 18 years old and spoke central Thai. Sixty six participants were classified into 3 groups: 27 DPs, 12 OPs, and 27 NCs. DPs were patients diagnosed with major depressive disorder, bipolar disorder with depressive episodes, or persistent depressive disorder by the diagnostic and statistical manuals of mental disorders, forth (DSM-IV) or fifth (DSM-5) edition in their latest visit. OPs were patients diagnosed with other diagnosis by the DSM-IV or DSM-5 apart from major depressive disorders, bipolar disorder, or persistent depressive disorder in their latest visit. Finally, NCs reported themselves without psychiatric diagnosis. The PHQ9 scores (less than 9) and HAM-D scores (less than 8) were used to confirm that NCs were deprived of depression.

TABLE II: Gender, HAM-D score, PHQ9 score and age of participants

Group	Female (%)	Male (%)	HAM-D ±SD	PHQ9 ±SD	Ages ±SD
DP	18 (66.67)	9 (33.33)	13.96 ±7.12	13.70 ±6.30	37.63 ±16.16
OP	9 (75.00)	3 (25.00)	10.25 ±5.59	12.25 ±6.73	41.08 ±13.37
NC	20 (74.07)	7 (25.93)	2.11 ±2.26	3.00 ±2.95	35.96 ±13.58

V. EXPERIMENTAL SETUP

Voice recordings from the 32 questions and speech recognition features are used in the analysis to train a model to classify the participants according to 3 groups: depression patients, other psychiatric patients, and normal controls. This audio data were then taken to normalize the energy of the sound according to the peak normalization. The characteristics of the audio from the participants were studied using the open-source program: The Munich open-Source Media Interpretation by Large feature-space Extraction (openSMILE) [36] in conjunction with the Weka program [39]. The features extracted from the audio signals using emobase resulting in 988 features per audio file which can be applied to the extractable characteristics to the science of signal processing and machine learning.

Once the features were obtained, they were standardized so that each dimension had the same width, allowing the data to be split to create 5-fold cross validation, divided into 90% training set data, containing voice feature data of the 25 DPs, 10 OPs, and 25 NCs, while testing set data, 10% consisted of voice feature data of 2 DPs, 2 OPs, and 2 NCs. Then, the data were used to train machine learning in Python with package scikit-learn using multiclass classifier for every multiclass classifier. The scikit-learn package contained five schemes: OneVsRest (OR), OneVsOne (OO), OutputCode (OC), MultiOutput (MO), and ClassifierChain (CC) using classifier LinearSVC.

VI. EXPERIMENTAL RESULTS

This study started with the hypothesis that the voice of each group had some significantly different speech features.

The overall audio were put together, and the features were extracted resulting in 988 features when running 5-fold cross validation and measured with ROCs. These were used to measure and classify the model's performance in dichotomous outcome (positive/negative test results). The performance can be compared with the AUC value. If the AUC is large (maximum is 1), the model has a good classification of the two groups. The test results from the validation data set gave the mean AUC results as shown in Table III. It is found that the results were not able to be classified well as they should. The average AUC of all classifiers is 0.6880±0.1740, 0.5600±0.1307, and 0.7371±0.1583 for DP, OP, and NC, respectively. The average AUC modeled with 988 features is 0.6617±0.1707.

Validation results of models with the 988 features of continuous speech show low average AUC at 0.6617. Therefore, 5 more features were added including the HAM-D score, PHQ-9 score, age, gender, and education level. The average AUC of all classifiers is 0.7531±0.1349, 0.5580±0.1463, and 0.8017±0.1453 for DP, OP, and NC, respectively. The average AUC modeled the 988+5 features is 0.7043±0.1759. However, when selecting a feature with AttributeSelection in Weka, it was found that only the features extracted from the HAM-D and PHQ-9 scores play an important role in classification (Table III).

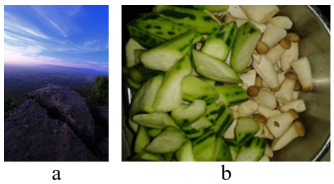
Although the validation results of models with 988+5 features of the continuous speech increased, the results are still not satisfactory. The modified hypothesis was that the voices in response to each question had different characteristics which made them possible to classify. The speech features of 32 questions were separately extracted, totaling $(988 \times 32) + 5 = 31,616 + 5$ features. The average AUC results are shown in Table III. The average AUC of all classifiers is 0.6931±0.1634, 0.6080±0.1754, and 0.8286±0.1307 for DP, OP, and NC, respectively. The average AUC of model with 31,616+5 features is 0.7099±0.1804.

The results are averaged lower due to the excessive number of features. These features were selected with AttributeSelection in order to filter for only the essential features: 77 features for the first fold, 70 features for the second fold, 96 features for the third fold, 71 features for the fourth fold, and 69 features for the fifth fold which included 311 unique features. The average AUC result is shown in Table III. The average AUC of all classifiers is 0.7594±0.1284, 0.5600±0.2483, and 0.9246±0.1051 for DP, OP, and NC, respectively. The average AUC modeled with feature selection of 31,616+5 features is 0.7480±0.2268. All questions were selected for at least 1-fold of the model except for the last question: "Please describe your feelings and thoughts about the following picture (sunset).".

Modeling from feature selection shows that the 31,616+5 features give better results. Therefore, the model was experimented with union 311 features and yielded more satisfactory results (Table III). The average AUC of all classifiers is 0.9251±0.0663, 0.8920±0.1170, and 0.9709±0.0563 for DP, OP, and NC, respectively. The average AUC modeled with union 311 features is 0.9293±0.0892.

TABLE III: Average AUC of 5 classifiers modelled with 988 features, 988+5 features, 31,616+5 features, feature selection of 31,616+5 features, union 311 features, and critical 16 features

Model	Scheme	Average AUC±SD of conditions				
		DP	OP	NC	All	Avg
988 features	OR	.6971 ±.1791	.6000 ±.1696	.7486 ±.1973	**.6819 ±.1805**	.6617±.1707
	OO	.7086 ±.1659	.4600 ±.1245	.7429 ±.1702	.6371 ±.1939	
	OC	.6400 ±.2228	.5600 ±.1084	.7200 ±.1118	.6400 ±.1603	
	MO	.7029 ±.1858	.5800 ±.0274	.7314 ±.1839	.6714 ±.1561	
	CC	.6914 ±.1845	.6000 ±.1696	.7429 ±.1884	.6781 ±.1784	
988+5 features	OR	.7657 ±.1531	.6200 ±.1823	.8229 ±.1544	.7362 ±.1756	.7043±.1759
	OO	.7486 ±.1531	.4400 ±.1084	.8000 ±.1565	.6629 ±.2101	
	OC	.7086 ±.1292	.5600 ±.0962	.7886 ±.1423	.6857 ±.1510	
	MO	.7657 ±.1361	.5500 ±.1173	.7743 ±.1788	.6967 ±.1729	
	CC	.7771 ±.1531	.6200 ±.1823	.8229 ±.1544	**.7400 ±.1763**	
31,616+5 features	OR	.7143 ±.1750	.6100 ±.2162	.8629 ±.1168	.7290 ±.1937	.7099±.1804
	OO	.7429 ±.1726	.6000 ±.2208	.8457 ±.1364	**.7295 ±.1966**	
	OC	.6600 ±.1500	.6500 ±.1458	.7771 ±.1194	.6957 ±.1419	
	MO	.6343 ±.1916	.5700 ±.1304	.7943 ±.1853	.6662 ±.1862	
	CC	.7143 ±.1750	.6100 ±.2162	.8629 ±.1168	.7290 ±.1937	
feature selection of 31,616+5 features	OR	.7943 ±.1150	.5300 ±.3012	.9429 ±.1125	.7557 ±.2541	.7480±.2268
	OO	.8057 ±.1268	.6800 ±.1956	.9429 ±.1125	**.8095 ±.1774**	
	OC	.7114 ±.1788	.5500 ±.2550	.9029 ±.1076	.7214 ±.2309	
	MO	.6914 ±.1043	.5100 ±.2485	.8914 ±.1163	.6976 ±.2250	
	CC	.7943 ±.1114	.5300 ±.3012	.9429 ±.1125	.7557 ±.2536	
Union 311 features	OR	.9429 ±.0571	.8800 ±.1351	.9771 ±.0511	.9333 ±.0929	.9293±.0892
	OO	.9371 ±.0586	.9100 ±.1084	.9829 ±.0383	**.9433 ±.0757**	
	OC	.9057 ±.0611	.9200 ±.1255	.9514 ±.0753	.9257 ±.0871	
	MO	.8971 ±.1012	.8700 ±.1255	.9657 ±.0767	.9110 ±.1041	
	CC	.9429 ±.0571	.8800 ±.1351	.9771 ±.0511	.9333 ±.0929	
Critical 16 features	OR	.9200 ±.0867	.6300 ±.2775	.9943 ±.0128	.8481 ±.2251	.8396±.2093
	OO	.9314 ±.0592	.7700 ±.1857	1.0000 ±.0000	**.9005 ±.1443**	
	OC	.8286 ±.1367	.6000 ±.2424	.9714 ±.0391	.8000 ±.2182	
	MO	.8429 ±.1558	.5900 ±.2608	.9657 ±.0767	.7995 ±.2329	
	CC	.9257 ±.0772	.6300 ±.2775	.9943 ±.0128	.8500 ±.2248	



a b

Fig. 3: a) "sunset" picture in item 29, b) "precut vegetables" picture in item 32.

Then we attempted to create a model with less features by considering the 16 critical features, which was determined by finding features with greater than or equal to 3 folds. The features in the critical features model consist of the HAM-D score and PHQ9 score, and 14 speech features: mfcc_sma[7]_iqr1-3 (mfcc: Mel Frequency Cepstral Co-efficients, sma: simple moving average, and iqr1-3: inter-quartile range: q3-q1) of item 1, pcm_loudness_sma_de_kurtosis (pcm: pulse-code modulation and de: delta) of item 2, lspFreq_sma[2]_iqr1-2 (lspFreq: line spectral pairs frequency) of item 7, mfcc_sma[9]_minPos (minPos: absolute position of the minimum value) of item 8, lspFreq_sma[6]_linregc2 (linregc2: offset (t) of a linear approximation of the contour) of item 10, mfcc_sma[7]_skewness of item 12, F0_sma_de_minPos (F0: fundamental frequency) of item 20, lspFreq_sma[6]_amean (amean: arithmetic mean of the contour) and pcm_zcr_sma_minPos (pcm_zcr: Zero-crossing rate of time signal in frame-based) of item 21, mfcc_sma[6]_range of item 23, F0env_sma_maxPos (F0env: envelope of fundamental frequency and maxPos: absolute position of the maximum value) of item 25, mfcc_sma[8]_linregc2 and mfcc_sma_de[6]_quartile2 (quartile2: first quartile (50% percentile)) of item 29, and F0_sma_kurtosis of item 31.

Fourteen percent of verbal instructions, 16% of reviews of symptoms (clinical questions), 20% of descriptions of emotional experiences questions, 43% of story telling tasks and 7% of imagination tasks were selected to be critical features. In addition, item 29 was selected as a critical feature as it had up to 7 folds (16%) and item 29 was selected because it had up to 6 folds (14%). Item 21 statement ("You leave home, go to work alone, and go to bed with feeling alone.") and Item 29 statement ("Please describe your feelings and thoughts about the following pictures (pre-cut vegetables).") were also selected. The average AUC of all classifiers is 0.8897±0.1098, 0.6440±0.2386, and 0.9851±0.0386 for DP, OP, and NC, respectively. The average AUC modeled with critical 16 features is 0.8396±0.2093.

The experiment was then run using the model with 311 union features and 16 critical features with 5 meta-estimators: LinearDiscriminant, LinearSVC, LogisticRegression, MLPClassifier, and RandomForestClassifier. It is found that MLPClassifier using the OneVsOne scheme provided the best average of AUC at 0.9548 for 311 union features and at 0.9824 for 16 critical features. The second fold average AUC is the highest at 0.9993 for union 311 features and at 0.9830 for critical 16 features. There are folds with AUC of 1 for all conditions including: 2^{nd} and 5^{th} fold for 311 union features, and 2^{nd}, 3^{rd}, and 5^{th} fold for 16 critical features, allowing random selection to possibly be 5^{th} fold. The result of the 5^{th} fold of both yielded 83.33% sensitivity, 91.67% specificity, and 83.33% accuracy.

TABLE IV: Confusion matrix of the 5th fold of 311 union features and 16 critical features

		Predict		
		Normal	Depression	Others
True	Normal	2	0	0
	Depression	0	2	0
	Others	0	1	1

VII. DISCUSSIONS

Speech features can be used to classify three groups: depression condition, other condition, and normal control, by analyzing the features of each of the 32 questions combined with the HAM-D and PHQ9 scores. The union 311 selected features included all questions except for those describing feelings towards pictures of the sky and a cliff. Additionally, 16 features which were selected had more than 3 folds or were critical features. This affected the results which show 83.33% sensitivity, 91.67% specificity, and 83.33% accuracy. There was one patient who was in OP group but wrongly classified as having depression. When exploring in details, this patient was diagnosed with bipolar disorder, manic episode but the HAM-D score of current visit was 10. Therefore, this patient had polarity switching and currently was experiencing depressive episode.

When analyzing the critical features, the description of negative emotional experiences is the most effective and weighs 43% in the classification. The depression patients are more neurologically responsive to negative stimuli [40], [41]. When considering the feature types, each feature has a weight in the classification as follows: F0_sma_de, F0_sma, F0env_sma, mfcc_sma_de[1-12], pcm_loudness_sma_de, and pcm_zr_sma has 7% each feature, lspFreq_sma[0-7] weighs 22%. and mfcc_sma[1-12] weighs 36%. Item 21 statement ("You leave home, go to work alone, and go to bed with feeling alone.") and item 29 which asks to describe their feelings after seeing the image of vegetables could be considered as highlighted items in this study.

The critical features were analyzed using principal component analysis (PCA). Any feature that the first principal component provides at least 80% of the total variance was chosen. There are 8 features including the HAM-D score, mfcc_sma[6]_range of item 23, F0_sma_kurtosis of item 31, F0_sma_de_minPos. of item 20, mfcc_sma[7]_iqr1-3 of item 1, pcm_zcr_sma_minPos of item 21, mfcc_sma[8]_linregc2 of item 29, and pcm_loudness_sma_de_kurtosis of item 2, respectively. The p-value of 0.7213 showed that the feature factors did not make a significant difference while class division did when the p-value was at 0. Additionally, both factors also made a significant difference at $p = 0$. The class division showed a significant difference between the normal patients and patients with other conditions.

The values of 8 features from the normal patients were significantly different from those with other conditions by at least one feature except for the feature F0_sma_kurtosis

of item 31. The HAM-D score of depressed patients was significantly different from the HAM-D score of the normal patients. The feature F0_sma_de_minPos of item 20 for depressed patients was also significantly different from F0_sma_de_minPos of item 20 for normal patients. In addition, the feature mfcc_sma[6]_range of item 23 and mfcc_sma[8]_linregc2 of item 29 for patients with other conditions was significantly different from those of normal patients.

At first, the linear support vector classifier was used to determine the optimal scheme. Then, the different classifiers with similar scheme were being compared with the results of the model with union 311 features and critical 16 features. We plan to conduct more detailed comparisons for all schemes and benchmark classifiers.

Video-derived parts and part of speech (PoS) analysis can and will be used for classification which may result in a decrease in a number of test items. Our goal is that the data from the speech feature, PoS, and facial landmark can be combined and used to classify subjects without any questions or with a few stimuli and that it takes a little time for the subject to complete the screening test.

REFERENCES

[1] A. Petty, "Why Is Everyone I Know Depressed?" 2017. [Online]. Available: https://greatist.com/live/why-is-everyone-i-know-depressed (accessed Feb. 1, 2021).

[2] World Health Organization, "Depression," 2019. [Online]. Available: https://www.who.int/health-topics/depression (accessed Feb. 1, 2021).

[3] F. Torres, "What is depression?" 2017. [Online]. Available: https://www.psychiatry.org/patients-families/depression/what-is-depression (accessed Feb. 1, 2021).

[4] B. Koskie, "Depression: Facts, statistics, and you," 2018. [Online]. Available: https://www.healthline.com/health/depression/facts-statistics-infographic (accessed Feb. 1, 2021).

[5] A. Halfin, "Depression: the benefits of early and appropriate treatment," *Am. J. Manag. Care*, vol. 13, no. 4, p. S92, 2007.

[6] B. Gelaye *et al.*, "Assessing validity of a depression screening instrument in the absence of a gold standard," *Ann. Epidemiol.*, vol. 24, no. 7, pp. 527–531, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1047279714001501

[7] UFHealth, "Measurement resources," 2011. [Online]. Available: https://dcf.psychiatry.ufl.edu/resources/measurement-resources (accessed Feb. 1, 2021).

[8] A. T. Beck *et al.*, "Beck depression inventory–ii," *Psychol. Assess.*, 1996.

[9] Multicultural Mental Health, "Screening for common mental disorders," 2013. [Online]. Available: https://multiculturalmentalhealth.ca/clinical-tools/screening-for-common-mental-disorders (accessed Feb. 1, 2021).

[10] R. L. Spitzer *et al.*, "Questionnaire sur la santé du patient – 9 (phq-9)," 2019. [Online]. Available: https://multiculturalmentalhealth.ca/wp-content/uploads/2019/07/PHQ-9-French-for-France-Pfizer.pdf (accessed Feb. 1, 2021).

[11] B. M. Byrne *et al.*, "The beck depression inventory (french version): Testing for gender-invariant factorial structure for nonclinical adolescents," *J. Adolesc. Res.*, vol. 9, no. 2, pp. 166–179, 1994.

[12] R. L. Spitzer *et al.*, "Chinese Patient Health Questionnaire-9," 2019. [Online]. Available: https://multiculturalmentalhealth.ca/wp-content/uploads/2019/07/PHQ9_Traditional-Chinese-for-Hong-Kong.pdf (accessed Feb. 1, 2021).

[13] R. L. Spitzer *et al.*, " แบบสอบถามสุขภาพผู้ป่วย- 9 (PHQ-9)," 2019. [Online]. Available: https://multiculturalmentalhealth.ca/wp-content/uploads/2019/07/PHQ9_Thai-for-Thailand.pdf (accessed Feb. 1, 2021).

[14] สถาบันพัฒนาสุขภาวะเขตเมือง, " แบบคัดกรองโรคซึมเศร้า." [Online]. Available: http://hpc13.anamai.moph.go.th/ewt_news.php?nid=1449 (accessed Feb. 1, 2021).

[15] J. C. Mundt *et al.*, "Voice acoustic measures of depression severity and treatment response collected via interactive voice response (ivr) technology," *J Neurolinguistics*, vol. 20, no. 1, pp. 50–64, 2007.

[16] National Institute of Mental Health, "Major depression," 2019. [Online]. Available: https://www.nimh.nih.gov/health/statistics/major-depression.shtml (accessed Feb. 1, 2021).

[17] M. Olfson, "Primary care patients who refuse specialized mental health services," *Arch. Intern. Med.*, vol. 151, no. 1, pp. 129–132, 1991.

[18] J. K. Darby *et al.*, "Speech and voice parameters of depression: A pilot study," *J Commun Disord*, vol. 17, no. 2, pp. 75–85, 1984.

[19] S. M. Lamers *et al.*, "Applying prosodic speech features in mental health care: An exploratory study in a life-review intervention for depression," in *Proc. NAACL*, 2014, pp. 61–68.

[20] S. Alghowinem, "From joyous to clinically depressed: Mood detection using multimodal analysis of a person's appearance and speech," in *Proc. ACII*. IEEE, 2013, pp. 648–654.

[21] M. Asgari *et al.*, "Inferring clinical depression from speech and spoken utterances," in *Proc. MLSP*. IEEE, 2014, pp. 1–5.

[22] N. Cummins *et al.*, "An investigation of depressed speech detection: Features and normalization," in *Proc. INTERSPEECH*, 2011.

[23] H. Kaya *et al.*, "Cca based feature selection with application to continuous depression recognition from acoustic speech features," in *Proc. ICASSP*. IEEE, 2014, pp. 3729–3733.

[24] S. Mantri *et al.*, "Clinical depression analysis using speech features," in *Proc. ICETET*. IEEE, 2013, pp. 111–112.

[25] L. A. Low *et al.*, "Detection of clinical depression in adolescents' speech during family interactions," *IEEE. Trans. Biomed. Eng.*, vol. 58, no. 3, pp. 574–586, 2010.

[26] S. Scherer *et al.*, "Audiovisual behavior descriptors for depression assessment," in *Proc. ICMI*, 2013, pp. 135–140.

[27] E. W. McGinnis *et al.*, "Giving voice to vulnerable children: machine learning analysis of speech detects anxiety and depression in early childhood," *IEEE J Biomed Health Inform*, vol. 23, no. 6, pp. 2294–2301, 2019.

[28] B. Yalamanchili *et al.*, "Real-time acoustic based depression detection using machine learning techniques," in *Proc. ic-ETITE*. IEEE, 2020, pp. 1–6.

[29] L. He *et al.*, "Automated depression analysis using convolutional neural networks from speech," *J Biomed Inform*, vol. 83, pp. 103–111, 2018.

[30] T. Rutowski *et al.*, "Optimizing speech-input length for speaker-independent depression classification." in *Proc. INTERSPEECH*, 2019, pp. 3023–3027.

[31] H. Jiang *et al.*, "Detecting depression using an ensemble logistic regression model based on multiple speech features," *Comput Math Methods Med*, vol. 2018, 2018.

[32] K.-Y. Huang *et al.*, "Speech emotion recognition using deep neural network considering verbal and nonverbal speech sounds," in *Proc. ICASSP*. IEEE, 2019, pp. 5866–5870.

[33] X. Lu *et al.*, "Design of chinese depression spoken speech corpus based on psychological self-related processing," in *Proc. ICOT*. IEEE, 2018, pp. 1–5.

[34] Y. Zheng *et al.*, "Applicability of the chinese beck depression inventory," *Compr Psychiatry*, vol. 29, no. 5, pp. 484–489, 1988.

[35] M. Pogosyan, "How culture affects depression," 2017. [Online]. Available: https://www.psychologytoday.com/us/blog/between-cultures/201712/how-culture-affects-depression (accessed Feb. 1, 2021).

[36] F. Eyben *et al.*, "Opensmile: the Munich versatile and fast open-source audio feature extractor," in *Proc. ACM-MM*, 2010, pp. 1459–1462.

[37] C. Ngamprom *et al.*, "Development of the affective norms for thai words (thai-anw) bank system," *Research Methodology and Cognitive Science*, vol. 15, no. 2, pp. 162–178, 2018.

[38] T. Sripornngam *et al.*, "Development of the thai affective picture bank system," *Research Methodology and Cognitive Science*, vol. 13, no. 2, pp. 57–70, 2016.

[39] G. Holmes *et al.*, "Weka: A machine learning workbench," in *Proc. ANZIIS*. IEEE, 1994, pp. 357–361.

[40] Q. Dai *et al.*, "Deficient interference inhibition for negative stimuli in depression: An event-related potential study," *Clin Neurophysiol*, vol. 122, no. 1, pp. 52–61, 2011.

[41] J. P. Hamilton *et al.*, "Neural substrates of increased memory sensitivity for negative stimuli in major depression," *Biol. Psychiatry*, vol. 63, no. 12, pp. 1155–1162, 2008.