ICIC Express Letters
Volume **17**, Number **3**, March **2023**
 ISSN 1881-803X
pp. **303–314**

DEPRESSION ANALYSIS USING BEHAVIOR TRACKING AND FACIAL EXPRESSION RECOGNITION DURING PHQ-9 ASSESSMENT

Natratanon Kanraweekultana[1], Sajjaporn Waijanya[1,*]
Nuttachot Promrit[1,*], Apisada Korsanan[2] and Sansanee Poolphol[2]

[1]Center of Excellence in AI and NLP
Department of Computing
Faculty of Science
Silpakorn University
Sanam Chandra Palace Campus, Rajamunka-nai Road, Muang District
Nakhon Pathom 73000, Thailand
kanraweekultana_n@su.ac.th; *Corresponding authors: { waijanya_s; promrit_n }@silpakorn.edu

[2]Psychiatry and Drug Addiction Subdivision
Somdet Phra Phutthaloetla Hospital
708, Prasit Pattana Road, Mueang District, Samut Songkhram Province 75000, Thailand
{ Apisada_ak; Sansanee88 }@hotmail.com

Received June 2022; accepted August 2022

Abstract. *The objective of this research is to comparatively study the analysis of depression potentiality from videos recorded during PHQ-9 assessment and clinical psychologists' expertise (psychometric) in major depression disorder diagnosis by implementing experimental research from three subject groups. The subjects were selected by using the purposive sampling method, and they consented according to the Human Research Ethics Committee. In addition, the data were collected via experimentation at the Psychiatry and Drug Addiction Subdivision of the hospital by using a research tool created from the application of facial expression recognition pre-training model, user behavior tracking, and PHQ-9 questionnaire in accordance with clinical manifestations. The relationship between PHQ-9 questions and the subjects' emotions, which occurred while they are taking the questionnaires, are analyzed by using correlation statistics. After that, the results are compared by three clinical psychologists who have been specializing in depression diagnosis for more than seven years. The relationship exhibits the following interesting outcomes: 1) The correlation result showed a significant relationship between PHQ-9, emotion, behavior tracking, and reaction time with real subjects (Experimental group) but not significant with fake subjects (Experimental excluded group); and 2) Most of the emotions derived from the experiment have similar trends. To elaborate, the untreated subjects and the subjects who are undergoing treatment are prone to sadness, whereas the normal subjects tend to be happy, and the result barely has a relationship with PHQ-9; 3) The occurring behaviors can be detected more easily by the system than by naked-eye observation; 4) In terms of reaction time, the system is able to provide a more clearly detailed description of the reaction time of the PHQ-9 experiment than the time tracking observation method.*
Keywords: PHQ-9, Facial emotion recognition, Depression

1. **Introduction.** Major depression disorder (MDD) is a common disorder which occurs worldwide, and more than 264 million people are suffering from it. MDD differs from normal emotional swings which normally happen to ordinary people. The distinguishable difference is that the shortterm reaction of people with MDD toward emotional changes can drastically affect their life. As a result of depression, almost 800,000 people die of

DOI: 10.24507/icicel.17.03.303

suicide. Suicide is the second leading cause of death among people aged 15-29 years old [1]. Depression can be divided into two types: 1) situational depression, which is short-term depression that happens from traumatic incidents, life-changing events, or abnormal self-adjustment by using depression to cope with bad situations (e.g., divorce, underemployment, death, and terrible accidents), and 2) clinical depression, which is more severe than situational depression. This type of depression is considered a disorder according to The Diagnostic and Statistical Manual of Mental Disorder, 5th Edition (DSM-V) or as an abnormality of neurotransmitters [2]. With regard to Thailand's mental health statistics during the COVID-19 pandemic, around 10-12 Thai people commit suicide daily, with four times more men committing suicide than women. The screening and diagnosing processes can be implemented in various ways, such as through a questionnaire, behavior observation, discussion and consultation, and background information analysis. All these can be applied to clinical psychiatrists' considerations with the Patient Health Questionnaire (PHQ-9) in the first couple of weeks, which solely focus on responders' physical and emotional clinical symptoms in conjunction with the international practices for depression screening and assessment questionnaires [3-5]. Nevertheless, both situational and clinical depression are internal states. Hence, some deviations may be generated by cumulative stress that happens from a situation that intervenes with or disturbs people's daily life. This condition causes nondepressed people who are simply stressed but whose scores are at the depression level to needlessly undergo hospitals' depression screening processes. Even worse is that people who actually have depression lose the opportunity to receive treatments because their assessment scores do not actually represent depression. Such problems are derived from the lack of behavior observations during PHQ-9 assessment.

This research presents a tool for analyzing the depression potentiality of respondents' behaviors during PHQ-9 assessment from videos recorded with the goal of enhancing depression screening for government, private, or educational organizations. Furthermore, this tool will help increase the efficiency of depression screening even if the users have neither the skill nor the knowledge related to depression or psychology. It also supports psychologists and medical personnel as an alternative application for more effective treatments and therapies.

The purpose of this research is to evaluate the results from emotional and behavioral analysis of depressed people by analyzing the relationship between the respondents' facial expression and PHQ-9 results, then comparing the result with specialists' opinions (psychometric) to establish a link with the implicit relationship between the system, experience, and expertise in depression screening.

2. **Related Work.** Mental health issues are derived from environmental or biological factors that can generate adverse effects on individual, family, or working relationships. They also cause negative social interaction [6], depression, narcissistic, or even abnormal behavioral symptoms such as eating disorders. Depression symptoms relate to changes in the function and sensitivity of cortisol in the depression state [7], which can happen due to several factors [1]. Depression can be categorized into two types: 1) situational depression and 2) clinical depression. The primary medical or psychological depression screening process implements various questionnaires, including PHQ-9 or the Hamilton Depression Scale (HAMD-17). All the questionnaires were tested and considered credible and accurate for screening depressed patients in psychiatric hospitals. [8] found that PHQ-9 is a self-managed version of the PRIME-MD diagnostic tool for common mental disorders. PHQ-9 is a depression module that rates all nine DSM-IV criteria "0" (not at all) to "3" (almost every day). In addition, PHQ-9 is reliable, accurate, and adjustable for people with MDD [9]. It is also a convenient and effective tool to screen and evaluate the severity of MDD. Presently, PHQ-9 has been developed into an electronic questionnaire or on the Internet. A study was conducted on a computerized version of PHQ-9

without violating psychometric properties in terms of the medical field for depression self-assessment. In reference to the Faculty of Medicine Ramathibodi Hospital [10], the clinical symptoms of MDD patients may vary as follows: 1) emotional symptoms, 2) neurovegetative symptoms, 3) psychomotor symptoms, and 4) cognition symptoms.

Primary diagnosis implemented with depression assessment questionnaires is probably insufficient. Psychiatrists or psychologists' roles, including history taking and medical observations, are required to accurately diagnose whether patients suffer from MDD or other mental disorders that consist of similar clinical symptoms. Medical skills, expertise, and experiences are needed in these processes, especially for patients whose symptoms are ambiguous [11]. Thus, the idea of applying artificial intelligence as a tool to enhancing depression screening and diagnosis of clinical psychologists or psychiatrists has been proposed. A study on the application of emotional detection with psychological theories pointed out that facial emotion recognition (FER) relates to patients' internal symptoms and social issues by analyzing four emotions detected from facial expressions, namely, fear, sadness, anger, and happiness [12]. In addition, research on the recognition of nonverbal emotional expression in depressed people illustrates that emotions are shown via facial expression. Also, one research [13] presented a diagnosis that was derived by comparing depressed patients and nondepressed people to analyze nonverbal facial expression. The depressed subjects remarkably expressed negative responses. Moreover, a study developed and tested medical processes that can detect depression potentiality and support psychiatrists through quantitative meta-data analysis of the results. This approach can be applied in future research on automatic depression assessment via visual and auditory cues. This research also predicts the depression level from face images by analyzing emotions via an FER2013 model [14,15], namely, the person-specific active appearance model, based on facial emotion feature discovery and depression classified by the changes and movements of the eyes, eyebrows, and corners of the mouth by using support vector machine. Research has shown that these features are effective for automatic identification of depression patients. Nonetheless, facial expression analysis still has room for improvement; the ambiguity of the results needs to be clarified because sadness or happiness may be linked to situational depression only and not to clinical depression [16]. Hence, research on the application between PHQ-9 and facial expression recognition is conducted to better cover all behaviors of the clinical symptoms (emotional, neurovegetative, psychomotor, and cognition symptoms) due to the increase in MDD patients. The patients reveal facial expressions related to their emotions while they are taking PHQ-9 [17]. A literature review is conducted in the scope of developing and enhancing a web application that can potentially analyze depression by combining the application with PHQ-9 and FER. This research's literature review is summarized in Table 1.

Therefore, our research motivation is driven by the following development methods with regard to clinical symptoms: 1) facial emotion analysis represents emotion symptoms; 2) user behavior tracking represents psychomotor and cognition symptoms; and 3) PHQ-9 assessment represents neurovegetative and psychomotor symptoms. These methods were used in the development of this research tool to measure depression behaviors in all four major depressive disorder symptoms and the experience of clinical psychologists [18,19]. This led to the research hypothesis that the depression instruments developed with web application.

3. **Methodology.** This study applies an experimental research method that focuses on exploring the truth and potential of the results in accordance with theories, knowledge, and technologies. All elements are applied to studying the concept used to invent the prototype tool for analyzing depression potentiality during PHQ-9 assessment via website application to enhance depression screening. The methodology framework of this research is shown in Figure 1.

TABLE 1. Summary of literature review

Research	Year	Major depressive disorder (MDD)			
		Emotion	Neurovegetative	Psychomotor	Cognition
WHO (Medical Only) [1]	2021	✓	✓	✓	✓
Timothy (Medical Only) [2]	2018	✓	✓	✓	✓
Costantini et al. [3]	2020	✓	–	–	✓
Schuler et al. [4]	2018	–	✓	✓	–
Scoppetta et al. [5]	2021	–	✓	✓	–
Rivera-Bonet et al. [7]	2021	✓	–	–	–
Sun et al. [8]	2020	–	✓	✓	–
Dede et al. [12]	2021	✓	–	✓	–
Hümmer et al. [13]	2021	✓	–	–	✓
Meher et al. [14]	2020	✓	–	–	–
Wang et al. [15]	2018	✓	–	–	✓
Bone et al. [16]	2019	✓	–	–	✓
Van Vleet et al. [17]	2019	✓	✓	–	✓
Our research motivation	**Depression**	✓	✓	✓	✓
	Technique	**Facial emotions**	**PHQ-9**	**User behavior tracking/ PHQ-9**	**Facial emotions/ user behavior tracking**

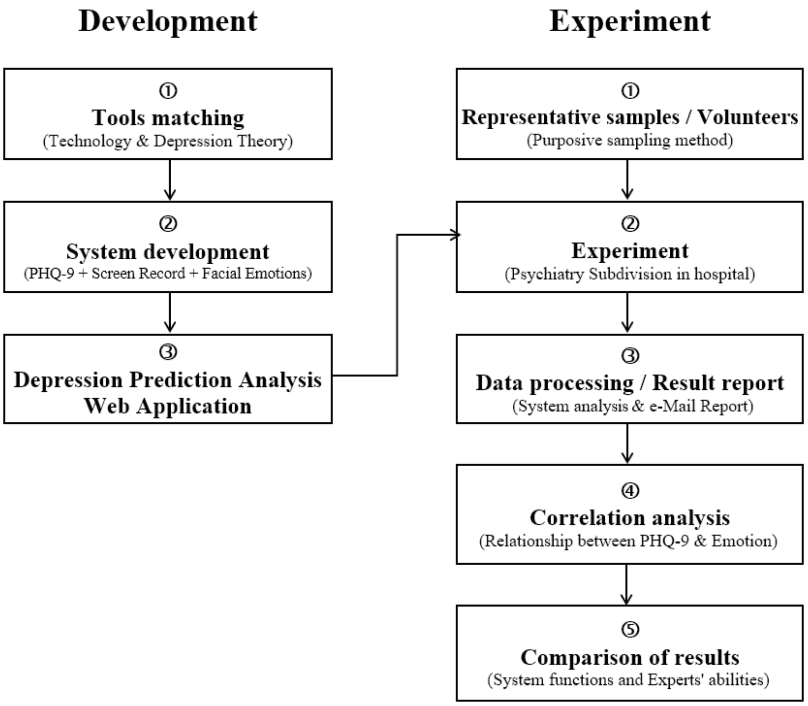

FIGURE 1. Methodology framework

3.1. **User behavior tracking system.** The tools developed and designed for analyzing depression potentiality from behavior observation during PHQ-9 assessment and the experiment of clinical symptoms match the four clinical symptoms: 1) emotional symptoms represent facial emotions; 2) neurovegetative symptoms represent PHQ-9; 3) psychomotor symptoms represent user behavior tracking/PHQ-9; and 4) cognition symptoms represent facial emotions/user behavior tracking.

An in-depth interview is examined by using the walkthrough technique to systematically review all techniques according to the designed model. Then, the results are analyzed by using the most suitable tools derived from the tool matching process between the technological tools and the four clinical symptoms. Afterward, the system is developed in reference to UX/UI design, which emphasizes users' emotions and feelings (UX) as well as responds to product activation or other systems and technical information (UI). The other essential point to consider for UI is color psychology, as illustrated in the development framework and algorithm shown below.

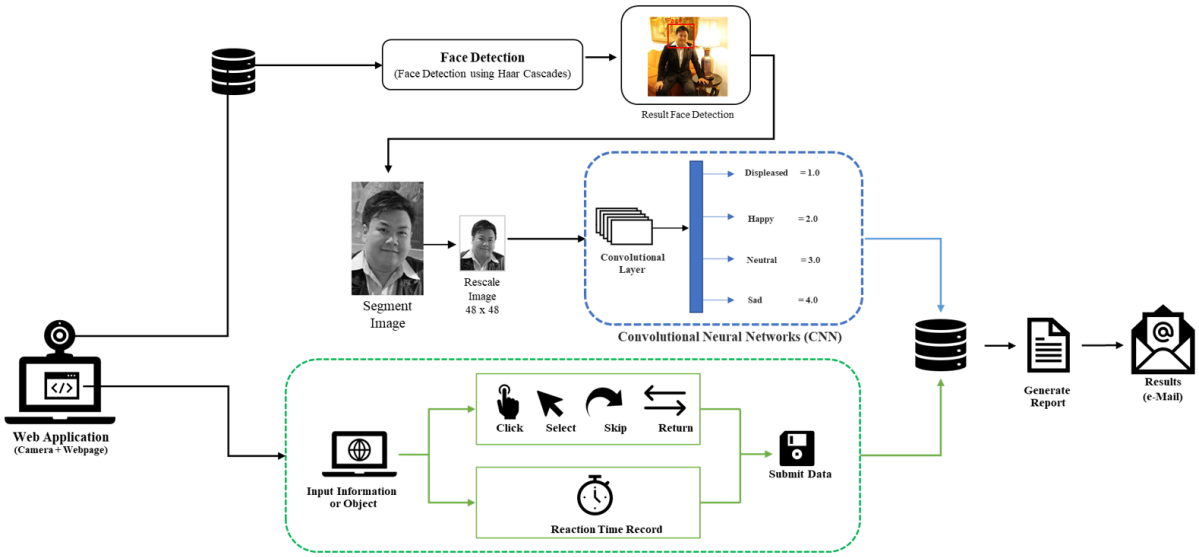

FIGURE 2. Development framework

As shown in Figure 2, the input information or subject (data reception) is set to activate when the camera is turned on, and then the system will command the subsystem to run a parallel operation in the part that does not interact with users but functions as behavior observation during PHQ-9 assessment. Afterward, the data will be submitted to the database and processed for analyzing emotions and tracking users' behaviors. The data are collected in the script for recording on the webpage, which requires data entry.

Next, when the webpage is commanded, the script will transfer the behavior information that interacts with PHQ-9, and then both data will be collected in the database and generated as an experimental result report before being sent to users via email. The processes are detailed in Figure 3.

1) Facial emotion process: The data are saved by embedding the script for facial emotion recognition. The process is set to start when the webcam is turned on. Then, the system will command the subsystem to run a parallel operation on the part that does not interact with users but functions as user behavior tracking while the patients are undergoing PHQ-9 assessment. Next, the data will be saved in the database by face detection (using Haar cascades) and facial emotional analysis (using convolutional neural network) by using the Fer2013 dataset, which contains approximately 30,000 facial RGB images of different expressions on the pre-trained facial emotions model. Then, the images will undergo the color conversion process from RGB to grayscale before being resized to 48×48 pixels for analysis in the final convolutional layer. Next, the images are analyzed in accordance with the determined class to predict four emotions: 1.0 = displeased, 2.0 = happy, 3.0 = neutral, and 4.0 = sad. The analysis will display the results of the prediction and the confidence interval in real time. The numpy.maximum function is used in Python, as shown in Figure 4.

2) Screen recording (behavior tracking) process: For the saving process in screen recording, the script will be embedded for recording the screen of the selected webpage. When

```
Input: Video<Webcam>, Video<Screen>
       Record(Video<Webcam>)
       Record(Video<Screen>)
       var uuid = GenerateUUID()
Function: SaveUuid(uuid)
          SaveToDataBase(uuid)
          return (count of all questionnaire)
EndFunction

Function: UserDoQuestionnaire()
          SaveResultToDataBase()
EndFunction

Input: String<Email>
Function: StopRecord()
  Function SaveVideoToDatabaseAndProcessWebCam()
    return emotes
  EndFunction
  do SaveEmotesToDatabase(emotes)
EndFunction
Output: do GetResultFromDatabaseAndSendEmail(String<Email>)
```

FIGURE 3. User behavior tracking system algorithm

```
Input: Video<Webcam>
var uuid = GenerateUUID()
Function SaveUuid(uuid)
  SaveToDataBase(uuid)
  return (count of all questionnaire)
EndFunction

Function UserDoQuestionnaire()
  SaveResultToDataBase()
  INPUT String<Email>
EndFunction

Function StopRecord()
EndFunction

  Function SaveVideoToDatabaseAndProcessWebCam()
    return emotes
  EndFunction
Output: do SaveEmotesToDatabase(emotes) AndSendEmail(String<Email>)
```

FIGURE 4. Facial emotion process

the webpage's script is commanded, it will send the data of the behaviors that have users' interactions to be saved in the database, and behavior data are collected by jQuery, which is a JavaScript library that tracks every click as determined. When the object is clicked, the data will be saved in the database together with the presently and the previously clicked objects. Therefore, the duration between the present and the previous clicks are calculated to depict the users' decision-making status when they are answering PHQ-9. The process will continue until the users stop taking PHQ-9, as shown in Figure 5.

3.2. **Experimental method.**

1) Representative sample or volunteers: This research was first submitted for the approval of the Human Research Ethics Committee of Silpakorn University and Somdet Phra Phutthaloetla Hospital because this research involves human subjects. The subjects

Input: *Video*<Screen>, List<ScoredPoint>, List<List<Behavior>>,
List<ClickTimeStamp>, List<TimePerIteration>,
do UserDoQuestionnaire()
Function: UserDoQuestionnaire()
SaveResultToDataBase(List<ScoredPoint>,
List<List<Behavior>>)
USER INPUT String<Email>
return {
List<ScoredPoint>,
List<List<Behavior>>,
List<ClickTimeStamp>,
List<TimePerIteration>,
String<Result>
}
EndFunction
Output: do GetResultFromDatabase AndSendEmail(String<Email>)

FIGURE 5. Screen recording (behavior tracking) algorithm

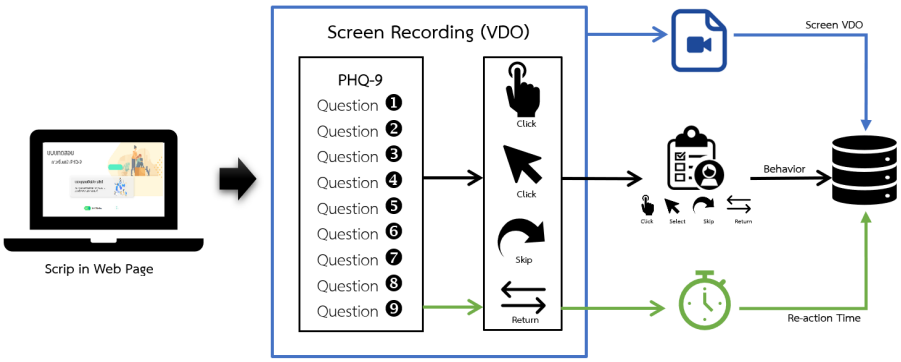

FIGURE 6. Screen recording (behavior tracking) process

who were willing to give their consent to the experiment are selected by a purposive sampling method. On the basis of the selection criteria of the experimental group of 15 cases by diagnosis, the subjects are divided into three groups according to the principles of experimental research under the supervision of a clinical psychologist [20]. All the personal data and identity details used in the depression potentiality analysis system from the recorded videos and users' interaction during PHQ-9 assessment are confidential. The subjects are divided into three groups: 1) nondepressed subjects, 2) undertreated MDD subjects, and 3) undergoing-treatment MDD subjects.

2) Experiment: The experiment uses the system installed in computers with a high-quality webcam, and the experiment site is the Psychiatry and Drug Addiction Subdivision of Somdet Phra Phutthaloetla Hospital. The experiment site is set to have enough exposure and space for effective detection under the conduct of the psychologists. These preparations are determined so that the system can collect videos and images via the webcams set to operate with the system in parallel. The collected data are 1) the answers from PHQ-9, 2) the screen recordings during PHQ-9 assessment, and 3) the subjects' emotional expression during PHQ-9 assessment.

3) The data processing and the result report: After the subjects submit their answers, the system will process and send them via email as reports that contain the following information: emotions, click times, reaction time, behaviors, and group test. The information is displayed without revealing any personal or identity data of the subjects.

4) Correlation analysis: The experiment results are analyzed to find the correlation (1) to investigate the relationship between the experiment results, PHQ-9 questions, and subjects' emotional expressions by calculating with Statistical Package for Social Science.

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{1}$$

5) The comparison between the system functions and experts' abilities: The system functions (research tools) and the abilities of the three experts who have more than seven years of experience in MDD are compared to analyze the relationship between the subjects' facial expressions and PHQ-9 assessment by using the focus group method.

4. **Experimental Results.** To analyze depression potentiality from the videos recorded during PHQ-9 assessment, the confirmatory test of psychological results in this study is divided into psychometric terms according to psychologists and facial emotion model efficiency measurement.

Further, we considered the experiment process on the subjects, in which the data are collected in the Psychiatry and Drug Addiction Subdivision of the Somdet Phra Phutthaloetla Hospital under the control of clinical psychologists and summarized to compare the results from the website and the MDD experts' opinions (psychometric). The experiment is implemented on three sample groups of five subjects. The experiment processes, the statistical analysis, and the comparison between the experiment results and the psychometric are as follows.

1) The results of the correlation depict that "neutral" has relationships with every question. The responses to questions (1), (2), (6), and (9) have an opposite relationship (−), whereas the participants agreed with "sad" (+), thus affecting the relationship of PHQ-9 the most. Question (9) influences the "sad" emotion the most, followed by question (6) and then the other questions. This result is also aligned with "displeased". The results illustrate that "happy" almost does not have any relationship with PHQ-9, or even if it does, it is at a low level and presents the opposite relationship (−). These outcomes indicate that emotions are the factor that reflects the relationship between PHQ-9 and emotions. Table 2 shows and compares the relationship between the real subjects (Experimental group) and the fake subjects (Experimental excluded group).

2) To obtain the weight and the frequency of the emotions, all the emotions that occur during the analysis and prediction of the system in experiment process are compared with the MDD psychologists' opinions on each PHQ-9 question from their more than seven years of experience in behavior observation and reaction time (psychometric). The results of the study are as follows.

- In terms of emotion, the undertreated MDD subjects (Depression: D) and the undergoing-treatment MDD subjects (Patient: P) indicate good agreement between the experiment results. "Sad", "displeased", and "neutral" emotions are likely to occur and show cumulative frequency, whereas "happy" does not occur in these two groups of subjects. However, with regard to the normal subjects (Normal: N), they show "neutral" and "happy" without any other emotions, indicating MDD. These outcomes agree with the opinions of the MDD psychologists.
- The experiment results are derived more from behavior observation than from naked-eye observation. The behaviors during the PHQ-9 assessment can be observed as follows: change of answering (Changed) in the undergoing-treatment MDD subjects (Patient: P), which implies their intention to distort their inner feelings; and question skipping (Skip) in the undertreated MDD subjects (Depression: D), which represents that they might feel frustrated or do not want to answer some of PHQ-9 questions. All the aforementioned behaviors are the aspects that the subjects (respondents)

Table 2. Correlation analysis between experimental and excluded group of subjects

PHQ-9	Subjects experimental	Correlation (r)					
		Displeased	Happy	Neutral	Sad	Reaction time	Behavior
(1) Feel bored and hard to be entertained	Experimental group	.859**	−.355	−.853*	.802*	.710*	–
	Excluded group	–	.542**	.435*	–	.091	–
(2) Uncomfortable, depressed or downhearted	Experimental group	.283	−.222*	−.686*	.882*	.423*	–
	Excluded group	–	.682*	.478*	–	.106	–
(3) Insomnia, dogsleep or hypersomnia	Experimental group	.675*	−.172	.446**	.617**	.652*	–
	Excluded group	–	.752*	.350**	–	.162	–
(4) Easily tired or weary	Experimental group	.276	.117	.124	.487	.748*	–
	Excluded group	–	.365*	.381*	–	.072	–
(5) Anorexia or overeating	Experimental group	.406	.028	.569*	.779*	.691*	–
	Excluded group	–	.741*	.441*	–	.234	–
(6) Suffering from self-pity or fear of failure	Experimental group	.890*	−.060*	−.454*	.931**	.874**	.864**
	Excluded group	–	.612**	.503*	–	.152	–
(7) Attention deficit in various activities	Experimental group	.874*	.194	.149*	.841*	.600*	.531*
	Excluded group	–	.556*	.445*	–	.430	–
(8) Hypoactive or hyperactive	Experimental group	.823*	.297	.442*	.888**	.804**	.822**
	Excluded group	–	.654*	.566*	–	.311	–
(9) Prone to self-harm or suicidal ideation	Experimental group	.962**	−.055*	−.817**	.963**	.708*	.647*
	Excluded group	–	.509*	.875*	–	.380	–

Note: *Sig. < 0.05, and **Sig. < 0.01

want to conceal from the observers so that they might not be recognized. Applying the system of this research for observing respondents' behaviors might be able to detect MDD symptoms in a more obvious and detailed manner.

- For the reaction time, the results derived from the experiment indicate the details of the reaction regarding the time in PHQ-9 assessment more clearly than observing by general time tracking. Hence, the system is more effective in terms of time tracking. To illustrate, the calculation results from each question to find the average value show that question (9) has the longest reaction time. Question (6) has the second longest reaction time, which agrees with psychologists' opinions that these two questions (numbers 6 and 9 in Table 2) require the longest reaction time. Nevertheless, the system is able to specify more detailed and obvious average reaction times, as shown in Table 3.

Some samples were excluded. Some experimental results show symptoms of depression, but the behavioral characteristics are not consistent with those of depression. The samples were excluded according to the following criteria: 1) The test result had a high depressive

TABLE 3. Comparison between the experiment results and clinically psychological experiences (psychometric)

Question	Type	Experiment (Frequency, $n = 15$)						Psychometric (Experience)					
		Displeased	Happy	Neutral	Sad	Behavior	Reaction time (Average)	Displeased	Happy	Neutral	Sad	Behavior	Reaction time (Average)
(1)	N	–	2	**5**	–	–	14.56 Sec.	–	✓	✓	–	–	10 Sec.
	D	1	–	3	**4**	–		✓	–	✓	✓	–	
	P	4	–	2	**5**	–		✓	–	✓	✓	–	
(2)	N	–	3	**5**	–	–	10.38 Sec.	–	✓	✓	–	–	10 Sec.
	D	–	–	–	**5**	–		–	–	–	✓	–	
	P	2	–	1	**5**	–		✓	–	✓	✓	–	
(3)	N	–	2	**5**	–	–	6.86 Sec.	–	✓	✓	✓	–	10 Sec.
	D	4	–	2	**5**	–		✓	–	✓	✓	–	
	P	3	–	3	4	–		✓	–	–	✓	–	
(4)	N	–	2	5	–	–	10.34 Sec.	–	✓	–	–	–	10 Sec.
	D	2	–	1	3	–		–	–	✓	✓	–	
	P	1	–	5	2	–		✓	–	✓	✓	–	
(5)	N	–	2	5	–	–	7.22 Sec.	–	✓	✓	–	–	10 Sec.
	D	1	–	–	4	–		–	–	✓	✓	–	
	P	2	–	5	2	–		✓	–	✓	✓	–	
(6)	N	–	5	5	–	–	18.65 Sec. (2nd)	–	✓	✓	–	–	20+ Sec. (Long time)
	D	4	–	–	5	Cry/Skip		✓	–	–	✓	Cry	
	P	3	–	–	5	–		✓	–	✓	✓	Cry	
(7)	N	–	1	5	–	–	13.05 Sec.	–	✓	✓	–	–	10 Sec.
	D	3	–	4	4	Cry		✓	–	✓	✓	–	
	P	1	–	5	3	Changed		✓	–	✓	–	–	
(8)	N	–	2	5	–	–	17.46 Sec. (3rd)	–	✓	–	–	–	10 Sec.
	D	–	–	2	5	Cry/Skip		–	–	✓	✓	–	
	P	2	–	4	2	–		–	–	✓	✓	–	
(9)	N	–	5	5	–	–	19.35 Sec. (1st) (Long time)	–	✓	–	–	–	20+ Sec. (Long time)
	D	4	–	–	5	Cry/Changed		✓	–	–	✓	–	
	P	2	–	–	5	–		✓	–	✓	✓	–	

Note: N = Normal, D = Depression, P = Patient

score; 2) the detected emotion was not indicative of depression; 3) quick replies were given during the PHQ-9 assessment or the responses to all questions were similar; and 4) the time average was less than 2 seconds per question as shown in Table 4.

5. **Discussion and Recommendations.** Cross-validation for applying and combining various research methodologies from various fields of science according to psychologists' opinions on clinical manifestations and behavioral congruence is applied in this study to setting the weighted score from the assessment results. The results can be categorized into three groups (referring to each subject group) as follows.

1) Nondepressed subjects (normal subjects): The experiment result of this group indicates agreement with psychologists' experiences (psychometric), including questions (1), (2), (5), (6), and (7), which is congruent to the study by Dede et al. [12] on emotional recognition of nonverbal language in MDD patients. The study points out that nonverbal expressions will present via facial expression instead. All the questions are possibly easily observed by the system and the psychometric. However, the results illustrate that questions (3), (4), (8), and (9) represent the difference between the system and the

Table 4. Excluded experiment samples

Samples excluded	Depression level (Score)	Emotion				Behavior	Reaction time (Average)
		Displeased	Happy	Neutral	Sad		
❶	27**	–	–	✓	–	Quick reply/ all same answer	1.15 Sec.
❷	27**	–	✓	✓	–	Quick reply/ all same answer	1.05 Sec.
❸	18*	–	✓	✓	–	Quick reply/ all same answer	1.63 Sec.
❹	27**	–	–	✓	–	Quick reply/ all same answer	0.30 Sec.
❺	17*	–	–	✓	–	Quick reply	0.90 Sec.
❻	25**	–	✓	✓	–	Quick reply	0.20 Sec.

Note: $\geq$ 20 Have severe symptoms of depression**, 15-19 Have quite severe depression*

psychometric. Although the system can detect behaviors, emotions, and reaction time during PHQ-9 assessment, some complexities remain, one of which is that its prediction is not on par with the psychometric. This finding agrees with the work of the Faculty of Medicine Ramathibodi Hospital [11], which states that the treatments for clinical MDD symptoms require professional experience and skills.

2) MDD subjects without any treatment (undertreated MDD subjects): The experiment results of this group agree with the psychologists' experiences in questions (1), (2), (3), (6), (7), (8), and (19). The system and the observation by psychologists' experiences have familiar outcomes: They detect emotions in the same direction. In terms of behaviors, the system is able to detect behaviors clearly and provide more details in the reaction time. This finding aligns with clinical symptoms, which vary in each patient, according to the data from the Faculty of Medicine Ramathibodi Hospital [10], which points out that MDD patients consult their doctor on various clinical symptoms. Regardless, the experiment results indicate that questions (4) and (5) present different outcomes between the system and the psychometric. The system can detect behaviors, emotions, and reaction time, which are neurovegetative symptoms in patients that manifest as physical behaviors, such as fatigue, anorexia, or overeating. These symptoms also vary per individual, thus generating different results.

3) MDD subjects receiving treatments (undergoing-treatment MDD subjects): The experiment results for this group are similar to those of the psychometric. Most of the system and psychometric results for questions (1), (2), (4), (5), and (6) are consistent. However, the emotions, behaviors, and reaction time are different. This finding accords with the study conducted by Van Vleet et al. [17], who found that MDD patients are more depressed as they show facial expressions related to their emotions while taking the PHQ-9 assessment. Nonetheless, the experiment results also indicate that the reactions of the subjects to questions (3), (7), (8), and (9) are dissimilar because they might have taken PHQ-9 before. Therefore, they are probably used to the assessment in terms of behaviors and reaction time. This condition may cause different reactions. The naked-eye observation conducted by the psychologists are able to detect behaviors better because the observers are familiar with MDD patients to the extent that they can understand patients' behaviors. Despite this situation, however, the system still performs better by providing more obvious and detailed reaction time tracking, which agrees with the information from Ramathibodi Hospital [10], stating that experts' experiences and skills are required to distinguish the similarities of clinical symptoms.

Recommendations and Future Work. This study focuses only on the application program for analyzing depression potentiality from videos recorded during PHQ-9 assessment. For effective assessment in the future, the FER used should be more unique, or the model used should be trained to maximize the optimum efficiency and enhance the practical tools for supporting MDD screening.

REFERENCES

[1] World Health Organization, *Depression*, https://www.who.int/news-room/fact-sheets/detail/depression, Accessed on August 21, 2021.

[2] T. J. Legg, *Situational Depression or Clinical Depression?*, https://www.medicalnewstoday.com/articles/314698, Accessed on August 21, 2021.

[3] Costantini et al., Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): A systematic review, *Journal of Affective Disorders*, vol.279, no.15, pp.473-483, 2020.

[4] M. Schuler et al., Assessment of depression before and after inpatient rehabilitation in COPD patients: Psychometric properties of the German version of the Patient Health Questionnaire (PHQ-9/PHQ-2), *Journal of Affective Disorders*, vol.232, pp.268-275, 2018.

[5] O. Scoppetta et al., Validity of the patient health questionnaire-2 for the detection of depression in primary care in Colombia, *Journal of Affective Disorders*, vol.278, pp.576-582, 2021.

[6] V. Lehtinen, E. Riikonen and E. Lahtinen, *Promotion of Mental Health on the European Agenda*, National Research and Development Centre for Welfare and Health, 1997.

[7] C. N. Rivera-Bonet, R. M. Birn, C. O. Ladd, M. E. Meyerand and H. C. Abercrombie, Cortisol effects on brain functional connectivity during emotion processing in women with depression, *Journal of Affective Disorders*, vol.287, no.15, pp.247-254, 2021.

[8] Y. Sun, Z. Fu, Q. Bo et al., The reliability and validity of PHQ-9 in patients with major depressive disorder in psychiatric hospital, *BMC Psychiatry*, vol.20, no.1, 474, DOI: 10.1186/s12888-020-02885-6, 2020.

[9] K. Kroenke, R. L. Spitzer and J. B. W. Williams, The PHQ-9 validity of a brief depression severity measure, *Journal of General Internal Medicine*, vol.16, pp.606-613, 2001.

[10] Faculty of Medicine Ramathibodi Hospital, *Ramathibodi Essential Psychiatry*, Mahidol University Press, Bangkok, 2015.

[11] Faculty of Medicine Ramathibodi Hospital, *Depression in Detail*, https://med.mahidol.ac.th/ramamental/generalknowledge/general/09042014-1017, Accessed on September 19, 2021.

[12] B. Dede, L. Delk and B. A. White, Relationships between facial emotion recognition, internalizing symptoms, and social problems in young children, *Personality and Individual Differences*, vol.171, DOI: 10.1016/j.paid.2020.110448, 2021.

[13] K. Hümmer et al., Impaired recognition of nonverbal expressions of emotions in depressed individuals, *Psychiatry Research*, vol.302, 114031, DOI: 10.1016/j.psychres.2021.114031, 2021.

[14] S. Meher et al., Automatic depression level analysis, *International Journal of Research and Analytical Reviews (IJRAR)*, no.2, pp.1-6, 2020.

[15] Q. Wang, H. Yang and Y. Yu, Facial expression video analysis for depression detection in Chinese patients, *Journal of Visual Communication and Image Representation*, vol.57, pp.228-233, 2018.

[16] J. K. Bone et al., Variation in recognition of happy and sad facial expressions and self-reported depressive symptom severity: A prospective cohort study, *Journal of Affective Disorders*, vol.257, pp.461-469, 2019.

[17] T. Van Vleet et al., Biases in processing of mood-congruent facial expressions in depression, *Psychiatry Research*, vol.275, pp.143-148, 2019.

[18] N. Kanraweekultana, S. Waijanya, N. Promrit and A. Korsanan, The algorithms for the prototyping of the depression analysis from behavioural observation while conducting the PHQ-9 assessment via the web application, *Sripatum Review of Science and Technology*, vol.13, no.1, pp.65-82, 2021.

[19] S. Z. Jumani et al., Facial expression recognition with histogram of oriented gradients using CNN, *Indian Journal of Science and Technology*, vol.12, no.24, pp.1-9, 2019.

[20] E. John, Sample size estimation: How many individuals should be studied?, *Radiology*, vol.227, no.2, pp.309-313, 2003.