aiMSE: Toward an AI-Based Online Mental Status Examination

Yanchen Liu, Stephen Xia, Jingping Nie, Peter Wei, and Zhan Shu, *Columbia University, New York, NY, 10027-6902, USA*

Jeffrey Andrew Chang, *New York University/VA New York Harbor Healthcare System, New York, NY, 10027-6902, USA*

Xiaofan Jiang, *Columbia University, New York, NY, 10027-6902, USA*

There is a lack of automated tools that utilize artificial intelligence to monitor mental health. The mental status examination (MSE) is an important tool used by mental health providers for assessing mental health. Currently, MSEs are conducted by licensed professionals, which is a barrier for patients in low income and remote areas. We propose an AI-based personal online mental status examination (aiMSE), the first interactive MSE platform. Users can use aiMSE to self-administer MSEs at home through a web browser, using only a camera and microphone. aiMSE uses multimodal image, speech, and natural language processing algorithms to detect signs of abnormalities in mental functioning and recommend them for further examination by a mental health specialist. We conducted a 14-person study, which supports the feasibility of detecting a wide range of signs commonly found in patients with changes in mental or cognitive capacity.

Mental health illnesses affect millions of people each year and are the second leading cause of death for people aged 10–34.[1]

There are many low-cost solutions for people to check their physical health. However, there is no equivalent low cost and widely accessible solution or platform for monitoring and measuring mental health.

The mental status examination (MSE), traditionally administered person-to-person by a licensed mental health provider, is an important tool to assess a patient's mental capacity by observing appearance, behavior, mood, and responses to standard cognitive questions and tasks. The MSE provides clinical information that aides in the diagnosis of mental health disorders. As MSEs require the presence of a licensed mental health professional, patients living in underserved or remote areas with a lack of adequate mental health coverage or a low provider-to-population ratio do not have ready access to this objective assessment for mental health screening.

1536-1268 © 2022 IEEE
Digital Object Identifier 10.1109/MPRV.2022.3172419
Date of publication 2 June 2022; date of current version 23 December 2022.

We propose the AI-based personal online mental status examination (aiMSE), the first web-based and widely accessible system for measuring and detecting characteristics of mental health disorders. Figure 1 shows aiMSE's web interface and examples of how we administer specific types of questions. aiMSE administers an interactive MSE through a web browser using only a camera and microphone, allowing access for anyone with a laptop, computer, or smartphone. aiMSE acts as a widely accessible triaging tool for identifying individuals that may require further examination or a referral to a mental health specialist. There is also potential utility in assisting primary care providers, midlevel mental health providers, nursing specialists, or students by providing an additional, objective measurement tool to assist in medical decision-making.

aiMSE utilizes multimodal image, speech, and natural language processing to assess a person's affect, speech, behavior, cognition, and memory to generate a mental status report of a user's mental functioning. We show, through a 14-person study, that aiMSE can detect these characteristics reasonably well.

We make the following contributions.

- We introduce aiMSE, the first low cost and widely accessible system for monitoring and detecting

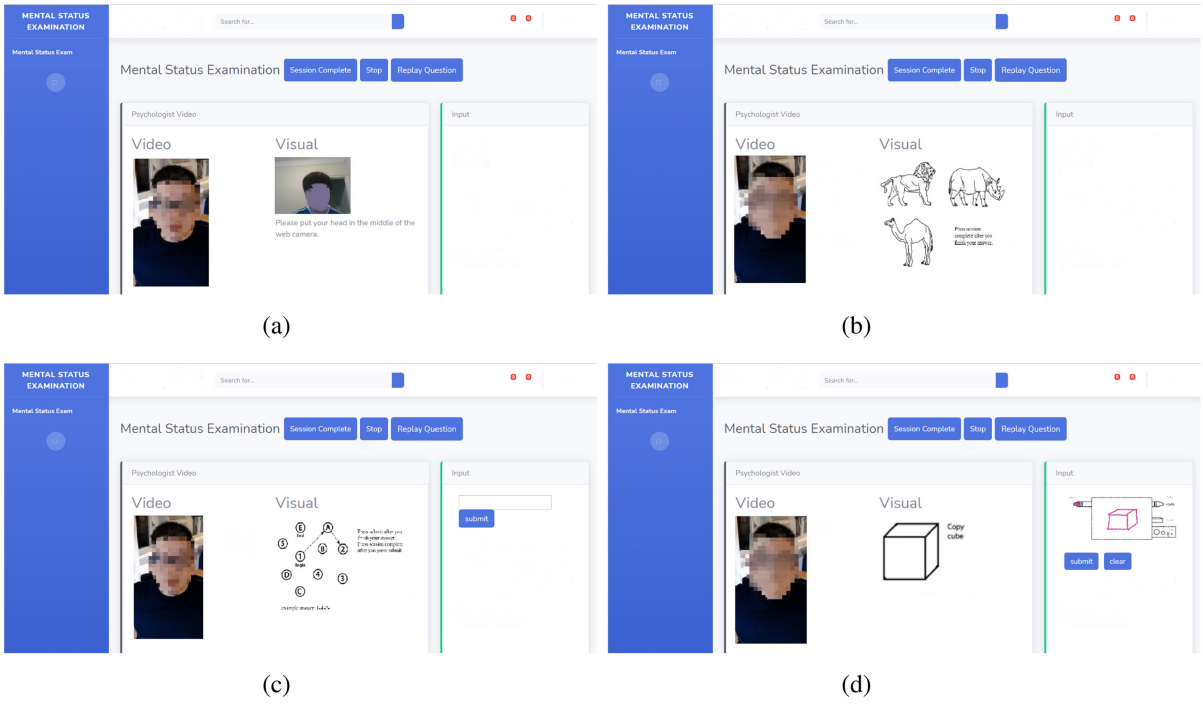

FIGURE 1. Examples of aiMSE's user interface for different types of questions. (a) Video of the psychiatrist introducing the aiMSE platform; the user is asked to place the head in the camera view. (b) Video of the psychiatrist asking a question that requires a verbal response from the user. (c) Video of the psychiatrist asking a question that requires text input from the user. (d) Video of the psychiatrist asking a question that requires drawing input from the user.

characteristics indicative of mental or cognitive impairment. aiMSE allows anybody to self-conduct an MSE with only a camera, microphone, and internet-enabled device.

- We introduce novel multimodal image, speech, and natural language processing algorithms and architectures to detect relevant mental status characteristics related to affect speech, behavior, and cognition.
- We evaluate aiMSE in a 14-person study and show that aiMSE can detect a wide range of characteristics consistent with that of a psychiatrist's observations. Overall, aiMSE can detect six indicators of mental health characteristics with 68.9% accuracy.

RELATED WORKS

To the best of our knowledge, there is no system that completely automates the MSE. In contrast to prior works, we simulate the process of a video interview by incorporating facial expression detection in video, sentiment analysis of audio, and cognition assessment. In this work, we study the autonomous administration of an MSE, including a cognitive assessment based on the Montreal Cognitive Assessment (MoCA).[2] MoCA is one of several available tools that mental health providers commonly administer to patients, face-to-face, to detect signs of cognitive impairment and assess various aspects of cognition, including, but not limited to, memory, language, and attention. Questions on the MoCA have patients perform basic cognitive tasks, such as repeating back spoken sentences, naming pictures of animals, and copying a picture of a cube. Each question has an objective correct response, and the entire test is scored out of 30 points, where a score of 25 or lower suggests cognitive impairment. MoCA has been shown to be effective in the assessment of various mental conditions, such as Alzheimer's.[3] In comparison with older mental status assessments, MoCA can screen for changes in executive functioning.

There are existing computer-based works that aim to provide low cost, self-administered mental status assessments. One such assessment is SATURN,[4] which has been shown to be highly correlated with MoCA results. SATURN and other recently developed assessments utilize different tasks than MoCA and have yet to be studied through a large clinical trial.[4]

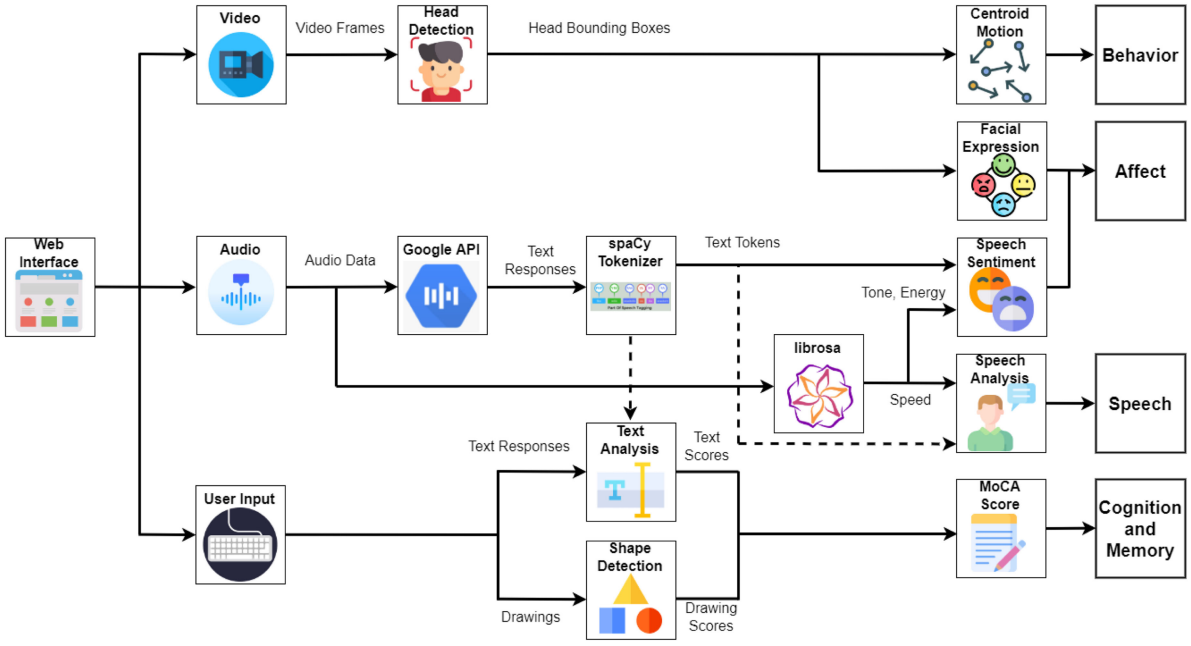

FIGURE 2. aiMSE system architecture block diagram. Three sensing inputs (video, audio, and user input) are used to assess the patient's behavior, affect, speech, and cognition.

Works, such as Wallace *et al.*'s work,[5] show the differences between the standard written MoCA and an online MoCA version. Though the authors observed some differences, they concluded that the online versions can still be effective. In addition, these works only utilize online interfaces to perform an assessment, while aiMSE utilizes other indicators, including affect, behavior, and speech.

Recently, there have been several works developing wearable systems to monitor mental health that utilize wrist sensors,[6] wireless body area networks,[7] and wearable EEG sensors.[8] In contrast to these works, aiMSE can be easily accessed through a simple web interface.

AIMSE SYSTEM

Mental Status Indicators

There are 10 major points and indicators that are observed during an MSE to determine a person's mental state.[9] aiMSE is capable of identifying four indicators of mental functioning. To sense these different facets, aiMSE administers a video-interview-based online examination similar to an in-person MSE. As shown in Figure 2, aiMSE uses three sensing inputs: video, audio, and user input, to assess these five aspects and administers the MoCA in a portion of the examination. These four were chosen because they can be reliably estimated from an audio and video recording. We discuss challenges in incorporating the other six indicators in the "Future Work" section. We introduce the four indicators aiMSE senses before discussing how aiMSE senses these indicators in the "Sensing" section.

Affect

Affect is the description of an individual's capacity to exhibit changes in outwardly displayed emotional states. Abnormalities may indicate depression, manic, or psychotic state.

aiMSE can characterize the range and nature of expressed affects by detecting and analyzing user speech for tone, energy, and rate, as well as analyzing video for facial expressions. Atypical findings include the maintenance of a constricted affect throughout the entire encounter, or conversely, a rapidly changing, expanded affect. The general emotional state based on detected expressions can also be estimated.

Speech

The speech indicator is a characterization of an individual's vocal patterns, including the rate and rhythm as well as appropriateness and clarity of responses. Abnormalities may indicate word-finding difficulties, subcortical changes, or a manic process.

The irregularity of speech rhythm or tone is another mental status characteristic. aiMSE estimates

tone, energy, and the average word rate that each subject speaks during the examination.

Behavior

The behavior indicator is an observation of an individual's capacity to engage appropriately with the interviewer. Abnormalities may indicate cognitive changes related to executive functioning, a psychotic process, an affective process, or malingering.

aiMSE can estimate the presence of either psychomotor agitation or retardation through measuring the amount of head movement a person makes during the examination using video data as described in the "Video" section.

IN AIMSE, WE EXAMINE ERRATIC HEAD MOVEMENT, WHICH IS ONE POTENTIAL INDICATOR OF AGITATED BEHAVIOR.

Cognition and Memory

Cognition is a general assessment relating to an individual's ability to gain knowledge and demonstrate comprehension. Abnormalities may indicate general cognitive impairment and dementia, or a more acute impact on cognitive functioning.

Memory is a description of an individual's short-term memory. Impairment may indicate difficulty with memory formation and retrieval.

MoCA is a standardized test used to screen for cognitive dysfunction. aiMSE can administer the set of questions asked in the MoCA to score each user. A MoCA score of 18–25 suggests mild cognitive impairment, a score of 10–17 suggests moderate cognitive impairment, and a score less than 10 suggests severe cognitive impairment. aiMSE can provide an interpretation of this scoring and provide this information to the user.

Sensing

Existing online cognition assessments lack behavioral and emotional feedback. While these online tools are designed to assess cognitive abilities, they are not able to assess visual or audio characteristics. aiMSE is designed to capture additional information to improve assessment.

Users interact with aiMSE through a web interface. aiMSE captures video and audio streams using personal facing cameras and microphones available on all modern laptops, computers, and tablets. These streams allow aiMSE to monitor other facets of mental state, such as affect and behavior, which is not possible in existing works.

Video

During typical MSEs, face-to-face interaction provides critical information, such as a person's affect or behavior. aiMSE utilizes video of the user's face instead.

aiMSE uses video data to estimate expressions from a patient's face. Expression estimation takes place in two steps: 1) head detection, and 2) expression classification. First, a bounding box of the head is extracted from the raw video frame using a Haar cascade filter.[10] This is used as input to a ResNet-based emotion classification network, which estimates one of seven classes: 1) fear, 2) anger, 3) happiness, 4) disgust, 5) sad, 6) surprise, and 7) neutral. We selected these seven expressions because they cover a wide range of expressions that people make on a regular basis. We retrained ResNet[11] by freezing the baseline model and replacing the output layer with a five-neuron layer, to represent each emotion class. Although we use a ResNet-based emotion classification network, we note that any other algorithm or neural network that classifies facial expressions is suitable to be used here. We chose ResNet because it is commonly used and performs well in a wide range of image classification tasks.

Certain characteristics relevant to the MSE can also be detected through video. In aiMSE, we examine erratic head movement, which is one potential indicator of agitated behavior. We use the same bounding box of the head extracted for emotion detection. We compute head movements for consecutive frames by noting the changes in the center of the head bounding box. Typical head movements, such as nodding or head shaking, last for around 0.5–1 seconds, are generally less than 4 Hz in frequency, and generally move less than 160°.[12] If at any point, the head movement greatly exceeds these thresholds, then this is interpreted as being unusual.

Audio

Some questions in MoCA are recorded aurally by a clinician. For example, the patient may be asked to repeat a spoken sentence, which clinicians can use to identify irregular features in speech, tone, and rhythm. Audio feedback is typically nonexistent in current online assessments.

aiMSE uses the microphone to allow patients to respond orally during the examination. aiMSE analyzes patient speech in three ways. First, the *librosa*[13] library is used to analyze speech speed, tone, and energy, which are important for recognizing when a patient may be agitated, have an unusual speech rhythm or have delayed speech.

TABLE 1. aiMSE seven-class expression detection results.

		Target						
		Anger	**Disgust**	**Fear**	**Happy**	**Sad**	**Surprise**	**Neutral**
Predicted	Anger	**282 (57.4%)**	14 (25.5%)	64 (12.1%)	19 (2.2%)	46 (7.7%)	12 (2.9%)	28 (4.5%)
	Disgust	11 (2.2%)	**27 (49.1%)**	7 (1.3%)	0 (0.0%)	2 (0.3%)	1 (0.2%)	3 (0.5%)
	Fear	48 (9.8%)	4 (7.3%)	**213 (40.3%)**	11 (1.3%)	70 (11.8%)	56 (13.5%)	20 (3.2%)
	Happy	16 (3.3%)	2 (3.6%)	19 (3.6%)	**778 (88.5%)**	32 (5.4%)	22 (5.3%)	39 (6.2%)
	Sad	70 (14.3%)	5 (9.1%)	113 (21.4%)	27 (3.1%)	**320 (53.9%)**	7 (1.7%)	96 (15.3%)
	Surprise	10 (2.0%)	1 (1.8%)	56 (10.6%)	15 (1.7%)	6 (1.0%)	**307 (73.8%)**	11 (1.8%)
	Neutral	54 (11.0%)	2 (3.6%)	56 (10.6%)	29 (3.3%)	118 (19.9%)	11 (2.6%)	**429 (68.5%)**

Notes: Entries bolded are the portion of samples correctly identified in each of the classes.

Second, speech content is analyzed for expected responses to questions in the MoCA. To accomplish this, aiMSE processes audio through the Google Speech-to-Text API[a] to convert the patient's speech to text, and then through a tokenizer[14] to break down the speech into tokens useful for natural language processing.

Third, aiMSE analyzes the audio for speech sentiment. The tokens outputted from the tokenizer are processed through a binary classification network based on BERT.[15] The network outputs whether the speech is positive or negative, which is used to estimate the affect on the patient.

User Input

Some of the questions on the MoCA require text input or drawings. Responses to questions that require text input are compared directly to expected answers. Each question has an objectively correct set of answers, so it is simple to make this comparison. The web interface also incorporates a drawing space.

To analyze drawings, aiMSE utilizes contour detection from OpenCV[16] with vertex counting to detect shapes.

EVALUATION

In this section, we evaluate our expression classification network before evaluating the entire system in a study of 14 participants.

Expression Detection Evaluation

To train our ResNet-based expression classification network, we use the fer2013 dataset[17] containing 35,886 frames of each expression (anger, fear, disgust, happy, sad, surprise, and neutral) and use 3589 samples for testing and the rest for training. Table 1 shows the confusion matrix metrics for the seven-class classification problem. We bold the diagonal, as these entries show the percentage of correctly identified expressions. Across all classes, our expression detection classifier has a 66% accuracy, which is much higher than random guessing.

Study Setup

Before beginning the study, the Institutional Review Board of Columbia University at Morningside Campus approved all human subjects experimental procedures in this study listed in protocol AAAT7906. To evaluate the effectiveness of aiMSE, we conduct a 14-person study with nine males and five female participants with no known mental impairments. All participants were students at Columbia University. The average age of all participants recruited is 25.09 years, where the minimum and max age is 21 and 29, respectively. All subjects participated voluntarily and reported having no history of mental health issues and neurological disorders. We recruited our subjects through word of mouth and obtained informed consent from each participant in accordance with the terms of the IRB for this project. During the study, we instructed participants to repeat the examination seven times, each time acting out one of the six behaviors that would indicate an abnormal mental status exam finding or acting normally. There was a five-minute break between each examination session. One of the authors was present at all times to answer questions on the side. The purpose of this study is to show that aiMSE can accurately detect a wide range of characteristics relevant to mental health. We do not claim that this is a clinical study showing that aiMSE can correctly identify real patients with mental health impairments. We plan on conducting full studies showing

[a][Online]. Available: https://cloud.google.com/speech-to-text/docs/apis

TABLE 2. Comparison between characteristics of abnormal mental statuses observed by the psychiatrist and aiMSE.

		Psychiatrist observation					
		Affect:Neg	**Affect:Rapid**	**Speech:Fast/Slow**	**Behavior**	**Cognition**	**Normal**
Predicted	Affect:Neg	**109 (66.1%)**	46 (30.9%)	0	4 (16.7%)	0	1 (7.1%)
	Affect:Rapid	42 (25.5%)	**92 (61.7%)**	0	2 (8.3%)	0	0
	Speech:Fast/Slow	6 (3.6%)	0	**21 (87.5%)**	0	0	0
	Behavior	1 (0.6%)	2 (1.3%)	0	**18 (75.0%)**	0	0
	Cognition	7 (4.2%)	4 (2.7%)	0	0	**24 (100%)**	0
	Normal	0	5 (3.4%)	3 (12.5%)	0	0	**13 (92.9%)**

Notes: *Normal* refers to not observing any characteristics that would suggest an abnormal mental status. We bold the portion of scenarios where aiMSE's assessment aligns with the psychiatrist's.

clinical effectiveness in future work. The six behaviors we instruct our subjects to act out are as follows:

1) Affect:Neg—exhibiting negative emotions with unchanging expressions;
2) Affect:Rapid—exhibiting rapidly changing emotional states;
3) Speech:Fast—speaking with rapid pace;
4) Speech:Slow—speaking with slow pace;
5) Behavior—increased psychomotor agitation and poor eye contact;
6) Cognition—provide incorrect responses to questions and tasks from MoCA.

To obtain ground truth, a licensed psychiatrist analyzed the audio and video responses of the test subjects, generated his own report, and characterized their mental status based on the above mentioned indicators. Although all scenarios were acted, the psychiatrist analyzed each scenario as any real scenario, giving us how he would characterize each scenario as if it were a real patient.

In total, we recruited 14 participants, who each performed seven different scenarios, yielding a total of 98 different samples. To increase the number of samples, we augment our dataset as follows. We divide the recorded audio and video of each scenario at the halfway point of the examination. To create a new scenario, we randomly select a video component and an audio component for each half of the examination and then combine both halves together. In each of the recorded scenarios, we only instruct each user to exhibit one of the MSE indicators; by combining different scenarios together, we create new scenarios that not only increase the total number of scenarios we evaluate but also create new scenarios containing multiple indicators. We take this approach in augmenting our dataset because it is difficult for a typical person to exhibit multiple behaviors concurrently. In total, we increased the total number of scenarios we evaluate to 402.

Study Results

Table 2 shows the aggregate performance metrics across all different scenarios for each of the behaviors we instructed the participants to act out. We compare aiMSE's identified behavior with the observations from the psychiatrist. For scenarios created by combining multiple scenarios together, we included them into each of the characteristics that the psychiatrist identified. We see from this table that aiMSE can identify a large portion of scenarios and indications consistent with the psychiatrist's findings. In total, aiMSE correctly identified 68.9% of scenarios with abnormal findings, which is much higher than randomly guessing one of six abnormalities. Next, we analyze how aiMSE identifies scenarios with indications of abnormal mental characteristics.

Affect Indicators

There were two scenarios we instructed participants to act out involving affect. The first scenario is to express predominantly negative expressions (anger, fear, sad, or disgust). For samples where a negative affect was identified as a primary indication, on average aiMSE detected a negative affect 66.0% of the examination, while aiMSE detected a negative affect only 25.7% of the time in other scenarios. In the second scenario, participants were instructed to rapidly change their expressions. In samples where a rapidly changing affect is a primary indication, aiMSE detected on average 3.4 changes in affect per second, while in other scenarios, aiMSE only detected 1.3 changes in affect per second.

Speech Indicators

A typical speaking rate for English is 150 words per minute.[18] For all scenarios where subjects were

TABLE 3. Ablation study.

		Performance					
		Affect:Neg	**Affect:Rapid**	**Speech:Fast/Slow**	**Behavior**	**Cognition**	**Normal**
Indicator	Step 1: Affect	79.6%	67.7%	0	0	0	92.9%
	Step2: Behavior	79.6%	67.7%	0	87.5%	0	92.9%
	Step 3: Speech	79.6%	67.7%	90.0%	87.5%	0	85.7%
	Step 4: Cognition	79.6%	67.7%	90.0%	87.5%	100%	85.7%

Notes: In each step, we add in one additional indicator summarized in Figure 2 and see significant improvements in detection performance for corresponding scenarios (e.g., cognition).

instructed to speak faster than normal (Speech:Fast), aiMSE observed a speaking rate of 181 words per minute, which is much higher than this typical rate. In all scenarios where subjects were instructed to speak slower than normal (Speech:Slow), aiMSE observed a speaking rate of 65 words per minute, which is much slower than the typical rate. In all other scenarios, aiMSE observed an average speaking rate of 143 words per minute.

Behavior Indicators

To evaluate behavior, aiMSE detects the maximum head movement speed (deg/s) and the movement frequency (Hz), and if aiMSE observes a head movement speed and frequency that is too high (greater than 160°/s and 4 Hz, respectively), then aiMSE indicates this as an abnormal finding. Of the scenarios where participants were instructed to rapidly move their head, aiMSE observed on average a maximum head movement speed and frequency of 176°/s and 5.1 Hz, respectively. These values are much greater than the typical values observed for normal head movements. In all other scenarios, aiMSE detected on average a maximum head movement speed of 41°/s and a frequency of 0.9 Hz, which is in line with typical behavior.

Cognition and Memory Indicators

To evaluate cognition and memory, we instruct participants to incorrectly answer most of the questions during the MSE. aiMSE observed that in scenarios where cognition was a primary indicator of the abnormal mental state, subjects scored on average 21.8 points, indicating cognitive dysfunction. In all other scenarios, aiMSE observed an average score of 28.5, which is well above the cutoff.

Ablation Study

In this section, we perform an ablation study to show how each indicator improves aiMSE. Table 3 shows the performance of aiMSE in each scenario after incorporating one indicator at a time. For instance, in step one, we only utilize the affect indicators (negative and/or rapidly changing expressions) before adding in the behavior indicators (head movements) in step two. When we start with only the affect indicators, only the scenarios where affect are the primary indicators were identified with high efficacy. By adding in more components, aiMSE begins to perform better across all scenarios. The performance metric refers to the percentage of samples considered by the psychiatrist to be exhibiting the corresponding class (e.g., Affect:Neg, where participants are predominantly showing a negative affect) that are correctly identified by aiMSE using the incorporated indicators at each step.

FUTURE WORK

There are 10 major points that mental health providers typically observe during an MSE to determine a person's mental state,[9] of which aiMSE observes four. The following indicators are factors we leave for future work.

- *Appearance:* A poorly groomed individual may be suffering from physical ailments, self-harm, or alcohol abuse that may affect the mental status. aiMSE would require a view of the entire body, but typical webcams only capture the face.
- *Mood:* This is the underlying subjective emotion of a person during the MSE. aiMSE can only judge what is apparent through a person's voice and face.
- *Thoughts:* A person showing illogical thought processing and disorganized thoughts could have an underlying thought disorder. This is a difficult factor to measure since a person's thought process is often unique to the person and the situation.
- *Perception:* This measures the process in how a person becomes aware of stimuli presented to the body from all senses, whether internal or

external, and also whether there is the presence of auditory or visual hallucinations. aiMSE can only effectively evaluate responses to external auditory and visual cues.

- *Insight:* This is the extent to which a person recognizes and understands their mental condition. This is very difficult to objectively measure and would require us to design a new set of algorithms.
- *Judgement:* This measures a person's problem-solving skills within an ethical framework. This is also difficult to objectively measure.

In addition, we plan to conduct extensive clinical trials to show the effectiveness of aiMSE on mental health patients with psychiatric and cognitive disorders in the future. Next, we plan to explore the effectiveness of aiMSE as a screening tool, to be followed by a clinician assessment, or to be used together with a clinician assessment rather than completely relying on aiMSE to accurately distinguish all cases of mental impairment. Finally, we plan to incorporate cutting-edge deep learning, such as utilizing video transformers, to improve the overall accuracy of aiMSE and move toward a more fully automated system for detecting indicators of mental impairment.

CONCLUSION

Nearly 20% of all adults in the United States experience some type of mental health illness, and this percentage has only increased due to the ongoing effects of COVID-19. Thus, a widely accessible tool for early detection of mental health illnesses is critical for bringing down the overall cost of mental healthcare. In this work, we present aiMSE, a proof-of-concept illustrating the potential feasibility of online systems for detecting indicators of mental impairment. With only the standard video and audio sensors on a laptop computer or mobile device, aiMSE can evaluate characteristics in affect, speech, behavior, and cognition to quickly identify and triage patients who may be exhibiting abnormal mental health statuses.

ACKNOWLEDGMENTS

This work involved human subject in its research. Approval of all ethical and experimental procedures and protocols was granted by the Institutional Review Board of Columbia University at Morningside Campus under protocol number AAAT7906. This work was supported by the National Science Foundation under Grants CNS-1704899, CNS-1815274, CNS-11943396, and CNS-1837022. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Columbia University, NSF, or the U.S. Government or any of its agencies.

REFERENCES

1. National Institute of Mental Health, "Suicide," May 2021. [Online]. Available: https://www.nimh.nih.gov/health/statistics/suicide
2. S. Ziad *et al.*, "The montreal cognitive assessment, MoCA: A brief screening tool for mild cognitive impairment," *J. Amer. Geriatrics Soc.*, vol. 53, no. 4, pp. 695–699, 2005.
3. E. M. Fine *et al.*, "Identifying the 'source' of recognition memory deficits in patients with Huntington's disease or Alzheimer's disease: Evidence from the CVLT-II," *J. Clin. Exp. Neuropsychol.*, vol. 30, no. 4, pp. 463–470, 2008.
4. D. Bissig, J. Kaye, and D. Erten-Lyons, "Validation of saturn, a free, electronic, self-administered cognitive screening test," *Alzheimer's Dementia: Transl. Res. Clin. Interv.*, vol. 6, no. 1, 2020, Art. no. e12116.
5. S. E. Wallace *et al.*, "A comparison of electronic and paper versions of the montreal cognitive assessment," *Alzheimer Dis. Assoc. Disord.*, vol. 33, no. 3, pp. 272–278, 2019.
6. H. Jebelli, B. Choi, H. Kim, and S. H. Lee, "Feasibility study of a wristband-type wearable sensor to understand construction workers' physical and mental status," in *Proc. Construction Res. Congr.*, 2018, pp. 367–377.
7. M. G. R Alam, E. J. Cho, E.-N. Huh, and C. S. Hong, "Cloud based mental state monitoring system for suicide risk reconnaissance using wearable bio-sensors," in *Proc. 8th Int. Conf. Ubiquitous Inf. Manage. Commun.*, 2014, pp. 1–6.
8. S. Egawa, T. Hifumi, H. Nakamoto, Y. Kuroda, and Y. Kubota, "Diagnostic reliability of headset-type continuous video EEG monitoring for detection of ICU patterns and NCSE in patients with altered mental status with unknown etiology," *Neurocritical Care*, vol. 32, no. 1, pp. 217–225, 2020.
9. F. Hufton, J. Petch, and S. Rege, "Ten point guide to mental state examination (MSE) in psychiatry," Nov. 2020. [Online]. Available: https://psychscenehub.com/psychinsights/ten-point-guide-to-mental-state-examination-mse-in-psychiatry/
10. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1, 2001.

11. K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
12. M A. Gresty and G. M. Halmagyi, "Abnormal head movements," *J. Neurol., Neurosurg. Psychiatry*, vol. 42, no. 8, pp. 705–714, 1979.
13. B. McFee *et al.*, "librosa: Audio and music signal analysis in python," in *Proc. 14th Python Sci. Conf.*, 2015, vol. 8, pp. 18–25.
14. M. Honnibal, I. Montani, S. Van Landeghem, and A. Boyd, "spaCy: Industrial-strength natural language processing in python," vol. 1212303, no. 6, 2020, doi: 10.5281/zenodo
15. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," 2018, *arXiv:1810.04805*.
16. G. Bradski, "The OpenCV library. Dr Dobb's," *J. Softw. Tools Prof. Prog.*, vol. 25, no. 11, pp. 120–123, 2000.
17. I. J. Goodfellow *et al.*, "Challenges in representation learning: A report on three machine learning contests," in *Proc. Int. Conf. Neural Inf. Process.*, 2013, pp. 117–124.
18. The National Center for Voice and Speech, "Voice qualities." Accessed: May 1, 2021. [Online]. Available: http://ncvs.org/e-learning/tutorials/qualities.html

YANCHEN LIU is a Ph.D. student in the Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA, supervised by Prof. Xiaofan Jiang. His research interests include artificial intelligence, Internet of Things, and connected health. Liu received his M.Sc. degree in electrical engineering from Columbia University. Contact him at yl4189@columbia.edu.

STEPHEN XIA's research interests include designing intelligent systems for mobile/wearable computing, connected health, and the Internet of Things. Xia received his Ph.D. degree in electrical engineering from Columbia University, New York, NY, USA. Contact him at sx2194@columbia.edu.

JINGPING NIE is a Ph.D. student in the Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA, studying under the supervision of Prof. Xiaofan Jiang and Prof. Matthias Preindl. Her research interests are at the intersection of artificial intelligence, machine learning, cyber-physical systems, and data analytics, designing the next generation of human-centric intelligent wearable devices and Internet of Things platforms to improve our lives by sensing the physical world in its many forms. Nie received her M.Sc. degree (Honor student) in electrical engineering from Columbia University. Contact her at jn2551@columbia.edu.

PETER WEI is a machine learning research scientist at Meta, Menlo Park, CA, 94025-1456, USA. His research focuses on design of intelligent systems for energy optimization in commercial buildings. Wei received his Ph.D. degree in electrical engineering from Columbia University, New York, NY, USA. Contact him at wei.peter@columbia.edu.

ZHAN SHU is currently working toward the B.Sc. degree in computer and systems engineering from Rensselaer Polytechnic Institute, Troy, NY, 10027-8 6902, USA. His research interests include artificial intelligence and machine learning. Contact him at zhanshu1129@gmail.com.

JEFFREY ANDREW CHANG is a clinical assistant professor in the Department of Psychiatry, NYU Grossman School of Medicine, New York, NY, 10027-6902, USA, and a site director for residency training at the VA NY Harbor Healthcare System Manhattan campus, New York. His research interests include treatment of PTSD and MDD in veterans using neuromodulatory and psychedelic modalities. Chang received his medical degree from the National Yang Ming University School of Medicine, Taipei, Taiwan, and completed his residency training in psychiatry at the SUNY Downstate Health Sciences University, Brooklyn, NY, in 2016, serving as a chief resident of outpatient services. He is ABPN Board Certified in psychiatry, a member of the American Psychiatric Association and licensed to practice medicine in the state of New York. Contact him at jeffrey.chang2@va.gov.

XIAOFAN JIANG is an associate professor in the Department of Electrical Engineering, Columbia University, New York, NY, 10027, USA, and chair of the Smart Cities Center, Data Science Institute. His research interests are at the intersection of systems and data, with a focus on intelligent embedded systems and their applications in mobile and wearable computing, intelligent built environments, Internet of Things, and connected health. Jiang received his Ph.D. degree in computer science from the University of California Berkeley, Berkeley, CA, USA. He is the corresponding author of this article. Contact him at jiang@ee.columbia.edu.