PEER-REVIEW REPORT

Name of journal: *World Journal of Psychiatry*

Manuscript NO: 90160

Title: Automatic Recognition of Depression Based on Audio and Video: A Review

Provenance and peer review: Invited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 04482726

Position: Peer Reviewer

Academic degree: PhD

Professional title: Associate Professor

Reviewer's Country/Territory: Thailand

Author's Country/Territory: China

Manuscript submission date: 2023-11-25

Reviewer chosen by: AI Technique

Reviewer accepted review: 2023-11-27 02:52

Reviewer performed review: 2023-12-01 09:45

Review time: 4 Days and 6 Hours

Scientific quality	[ ] Grade A: Excellent [Y] Grade B: Very good [ ] Grade C: Good [ ] Grade D: Fair [ ] Grade E: Do not publish
Novelty of this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No novelty
Creativity or innovation of this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No creativity or innovation

Scientific significance of the conclusion in this manuscript	[ ] Grade A: Excellent [ ] Grade B: Good [Y] Grade C: Fair [ ] Grade D: No scientific significance
Language quality	[Y] Grade A: Priority publishing [ ] Grade B: Minor language polishing [ ] Grade C: A great deal of language polishing [ ] Grade D: Rejection
Conclusion	[ ] Accept (High priority) [Y] Accept (General priority) [ ] Minor revision [ ] Major revision [ ] Rejection
Re-review	[ ] Yes [Y] No
Peer-reviewer statements	Peer-Review: [ ] Anonymous [Y] Onymous
	Conflicts-of-Interest: [ ] Yes [Y] No

SPECIFIC COMMENTS TO AUTHORS

Automatic Recognition of Depression Based on Audio and Video: A Review This paper summarizes a recent literature survey on automatic depression estimation (ADE) methods. The inclusion focuses on those based on extracting and classifying relevant features from audio and video data, by means of deep learning. It was hypothesized that such schemes would alleviate the problems found in traditional evaluation by human physicians. Its content was divided into datasets, current limitations and prospectives. The manuscript was well written and structured. The studied area would be of interest to the World Journal of Psychiatry readers. The references list is relevant and up-to-date. The narratives, insights, and discussions on the topic, as presented by the authors, are appropriate and scientifically sound. That being said, the paper could benefit from improvement, addressing concerns and following suggestions listed as follow: 1) Graphical representation of various elements, e.g., facial action units (AU) and their characterization, could help the readers in broader fields to grasp how FACS operates. Please consider. 2) General overview of prominent ADE methods (in general, e.g., biological measurements and classifications, and not limited to deep learning) and their

conventional counterparts, could be systematically grouped. Subsequently, the authors could clarify what types of ADE were focused here. 3) Prior to discussing the datasets, a section describing how deep learning-based ADE plays its part in psychiatric diagnosis and how they were incorporated into modern practice, should be added. Please consider. 4) Short paragraphs pertaining detailed insight and authors' own critiques/ opinions on existing audio-based and video-based, and their fusion depression estimation methods should be given at the end of the respective sections. 5) Relevant citations of related works should be added in appropriate places in the Discussion section. 6) In conclusion and discussion, the "lack of exploration of the body expressions of individuals with depression," as stated by the authors is rather superficial. Are recognizing and classifying "body expressions" not already the prime areas of investigations in the most recent research? Please elaborate.