

Facial Depression Recognition: Challenges and Prospects

Yuchen Pan*

Department of Biomedical Engineering, Renmin University of China, Beijing, China

Corresponding Author*

Yuchen Pan
Department of Biomedical Engineering,
Renmin University of China,
Beijing,
China
E-mail: 23B303004@stu.hit.edu.cn

Copyright: © 2025 Pan Y. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Received: October 19, 2023, Manuscript No. JBTW-23-117374; **Editor assigned:** October 21, 2023, PreQC No. JBTW-23-117374 (PQ); **Reviewed:** November 04, 2023, QC No. JBTW-23-117374; **Revised:** January 13, 2025, Manuscript No. JBTW-23-117374 (R); **Published:** January 20, 2025, DOI: 10.35248/2322-3308.25.14.1.005

Abstract

Depression constitutes a widespread global mental health disorder with far-reaching adverse impacts. In recent times, multiple methodologies have been developed that leverage facial recognition and computer vision technology to identify depression based on facial images. In this paper, the primary framework for depression recognition through facial expressions is introduced, and the key challenges inherent in this task are discussed. The contemporary concern of facial privacy in the context of depression recognition is also examined. Ultimately, the paper concludes by outlining potential future directions for facial depression recognition.

Keywords: Depression recognition • Facial recognition • Explainable artificial intelligence • Privacy protection

Introduction

Depression, recognized as a pervasive global mental health ailment, inflicts substantial harm upon individuals and societies, garnering extensive international scrutiny. Recent data from the World Health Organization (WHO) reveals a concerning 28% growth rate in severe depression cases worldwide, manifesting a notable increase in incidence [1]. Nevertheless, a mere 10% of afflicted individuals actively seek medical intervention, with the principal impediment being the pervasive "stigma" associated with this condition.

In recent years, scholars have shown increasing interest in the fusion of AI healthcare, propelled by the rapid advancement of multimodal algorithms and deep learning-based technologies. The benefits of this emerging trend are palpable: AI-driven solutions possess the potential to facilitate diagnosis, mitigating patient apprehension and societal stigma. Concurrently, AI-enabled digital monitoring equips healthcare providers with more comprehensive patient data, thereby augmenting the quality of treatment.

Despite the presence of numerous depression recognition algorithms relying on facial data, the medical field currently places considerable emphasis on ensuring facial privacy and security. Consequently, there is a growing demand for innovative approaches that can simultaneously execute depression recognition while safeguarding user privacy. The genesis of depression recognition as a substantial research issue can be traced back to the Audio-Visual Emotion Challenge and workshop (AVEC). This competition, held over several years, featured sub-challenges related to a udio-visual

depression recognition in 2013, 2014, 2017, and 2019 [2]. In the earlier editions of AVEC challenges, depression recognition predominantly relied on complete facial images. However, beginning in 2017, the organizers adopted a heightened focus on privacy protection. Consequently, they provided only essential biometric facial information as visual data, thereby substituting full facial images.

Description

Facial depression recognition

In visual depression recognition approaches, the extraction of facial features predominantly encompasses two categories: Handcrafted features and deep features. Handcrafted features are typically crafted by researchers who leverage their expertise and domain knowledge, employing specially devised feature extraction operators to process images [3]. Nevertheless, handcrafted features are often customized for specific patterns or tasks, and when applied to depression recognition, generic feature extraction algorithms may produce suboptimal results. Furthermore, enhancing the design of feature extraction operators to align with depression recognition can present substantial challenges.

Feature extractors for images, trained using deep learning models, possess the capability to autonomously discern particular patterns from extensive datasets. These models can be subsequently fine-tuned on depression-related datasets to assimilate generic image features relevant to the task of depression recognition [4]. However, within the context of depression recognition, relying solely on static single-frame facial images neglects temporal information, and the integration of continuous video frames introduces additional complexities. The development of an efficient spatiotemporal information extraction method, especially considering the limited availability of facial images pertaining to depression, can be a formidable challenge.

In addressing the above challenges, two prevailing approaches have emerged:

- Employing handcrafted spatiotemporal feature extraction operators for feature extraction, subsequently conducting depression recognition through statistical analysis or deep learning model.
- Pretraining deep models equipped with attention mechanisms on dynamic video datasets and subsequently fine-tuning these models using depression-specific datasets [5].

Explainability

The explainability in these approaches has gained prominence in light of the rise of Explainable Artificial Intelligence (XAI). Several studies have endeavored to visualize the facial features within deep learning models more comprehensible, aiming to demonstrate their explainability. These efforts have successfully highlighted significant facial regions or actions associated with depression. Nevertheless, a more nuanced dissection of these regions or actions, facilitating a detailed explanation of the intricate relationship between facial expressions and depression, remains an area of limited development. Such refinement would be highly beneficial for the advancement of explainable depression recognition.

Privacy

With the growing emphasis on safeguarding facial privacy, researchers have begun to explore the utilization of non-image facial structural data, including facial landmarks, Action Units (AUs), and gaze direction, in the context of depression recognition. However, current research predominantly centers on the temporal variations of these facial features, treating them as time-series data for analysis [6]. In contrast, a study has delved into the spatiotemporal characteristics of facial landmarks in the context of depression, considering the dynamic changes in non-privacy-related spatial facial features. Nonetheless, discussions and research regarding privacy protection issues in facial depression recognition remain relatively limited. Of particular note, the quest for a balance between the robustness of privacy protection and the effectiveness of depression recognition is an emerging area that necessitates further exploration and investment from researchers in the future.

Conclusion

The field of depression recognition from facial images has witnessed several years of development, marked by considerable efforts and notable achievements in terms of performance. However, there remain several underdeveloped areas within this domain that warrant further research:

- The refinement and advancement of facial spatiotemporal features hold the potential to enhance the accuracy and robustness of depression recognition.
- There is room for improvement in the realm of model explainability, facilitating a more transparent and comprehensible understanding of the recognition process.
- The critical issue of facial privacy necessitates additional comprehensive study and exploration to address the evolving challenges in the future.

References

1. Niu, M., et al. "Multimodal spatiotemporal representation for automatic depression level detection." *IEEE Trans Affect Comput.* 14.1 (2020): 294-307.
2. He, L., et al. "Automatic depression recognition using CNN with attention mechanism from videos." *Neurocomput.* 422 (2021): 165-175.
3. de Melo, W.C., et al. "Depression detection based on deep distribution learning." 2019 IEEE International Conference on Image Processing, Taipei, Taiwan. *IEEE.* (2019).
4. Valstar, M., et al. "Avec 2013: The continuous audio/visual emotion and depression recognition challenge." Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, Barcelona Spain. *Association for Computing Machinery, New York.* (2013).
5. Pan, Y., et al. "Spatial–Temporal attention network for depression recognition from facial videos." *Expert Syst Appl.* 237 (2024): 121410.
6. Ringeval, F., et al. "Avec 2017: Real-life depression, and affect recognition workshop and challenge." Proceedings of the 7th annual workshop on audio/visual emotion challenge, Mountain View, California, USA. *Association for Computing Machinery, New York.* (2017).