Techniques for Detecting Depression-related Emotions, Challenges and Future Prospects

Mingyuan Zhang

{ z2900910676@outlook.com }

School of Data Science, Data Science and Big Data Technology, Guangzhou City University of Technology, Guangzhou, Guangdong, China

Abstract. Current depression diagnosis remains subjective. While multimodal HCI methods enable objective assessment, a definitive systematic review remains lacking. The paper will cover the recent changes in technological advances. The article initially reviews depression detection methods using electroencephalographic signals and explains the existing controversies on the dependability of neural markers. It then examines the technology of passive behavior observers based on eye motions and facial recognition, the ways it could be applied in real-life context and the privacy issues involved, and lastly, discusses the multi-source data techniques, which make use of multimodal information integration. These approaches are the trend of mainstream in terms of raising the detection accuracy and robustness; however, it also emerged with novel challenges that include the model complexity, interpretability and privacy ethics. New directions in the lightweight, interpretable and privacy-preserving AI are essential in order to translate those techniques into proactive mental health services.

Keywords: Depression, Multimodal Fusion, Emotion Detection.

1 Introduction

Depression, which is alternatively called the depressive disorder, is one of the most prevalent mental disorders and is a psychological illness that is mainly distinguished by a high degree and prolonged low mood. There are multifaceted and varied clinical manifestations of depression. Principal symptoms are depressed mood, loss of interest and fatigue. It also has some early symptoms like slow response, slow thinking, and loss of memory, these are subject to change depending on the individual.

Depression is triggered by a confluence of factors, which are caused by the complex interactions of biological, psychological and social environmental factors. When it comes to trigger factors, genetic susceptibility constitutes the foundation of the weaknesses in a human being. These triggers are most commonly major setbacks (death of a loved one, career failures in general, etc.), interpersonal conflict or loneliness over a long period of time, etc. Moreover, the symptoms of depression can also be initiated or worsened by some chronic diseases or the intake of specific medications. Physiologically, the mechanism of the development of depression is

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365088

highly complicated as it is the manifestation of the deregulation of a variety of systems. It is directly connected to the disproportion of neurotransmitters within the brain, hyperactivity of the stress system, and alterations both in the morphology and functioning of certain parts of the brain (the hippocampus). The combination of these complicated physiological alterations causes a total disequilibrium of emotions, thought and bodily feelings.

The existing clinical diagnosis is largely based on scales and interviews, which are highly subjective and hard to identify the subtle internal shifts of feelings in the real life scenario, leading to a delay in intervening. Accordingly, there is a dire need to come up with objective and constant emotion monitoring technologies in the context of mental health.

With the human-computer interaction technology at work, scientists have started to consider the use of daily technological devices in developing a natural and non-invasive way of measuring depressive emotions. In this article, the primary achievements of this area will be considered: first, the detection methods are discussed on the basis of unique signals including electroencephalogram, eye movement, and facial images; then multimodal methods that utilize more than one signal and apply intelligent algorithms are discussed, and the drawbacks thereof are outlined. Lastly, it will explain the existing limitations of the research and the development prospects in the future, in the hope of offering a more objective application of emotional health management.

2 Depression detection method based on electroencephalogram signals

EEG is a noninvasive method of detection that records the alterations in the postsynaptic potentials in a group of neurons within the cerebral cortex using electrodes placed on the head. Due to its ability to directly indicate the dynamic neuronal activities in real time, it has become a significant instrument in the analysis of the neural mechanisms of depression. Research has discovered that there are certain frequency ranges (e.g., the increased asymmetry of frontal lobe alpha waves) wherein the electroencephalogram characteristics of depressed patients exhibit deviant characteristics [1]. Thus, the features have the capabilities of being investigated in terms of machine learning to enhance the accuracy and comprehensiveness of depression detection. As an example, one study only used three electrodes to record the resting-state electroencephalogram data and only had 12 time-domain features extracted. It was capable of classifying depressed and healthy control patients and, at the same time, using the Best-First decision tree algorithm to differentiate between the two with a classification accuracy of up to 96.36% [2]. This study outcome was able to integrate a simplified electroencephalogram data acquisition plan with an effective algorithm, with the beneficial effect of screening depression with high precision and opening up the way to the ultimate development of low-loaded wearable schemas that may be utilized in everyday routine. Nevertheless, it is still debatable that the given EEG indicators, like asymmetry of alpha waves in the frontal lobe, can be regarded as biomarkers of depression diagnostics, given the fact that the given features are not as effective as may be expected in the academic community. In 2025, a study conducted by 23 researchers on more than 4500 participants revealed that the total effect value of the indicator was low, and both its diagnosis value alone and that of overall diagnosis were rather weak [3].

3 Depression detection method based on eye movement signals

The related features of a depressive state are reflected through concrete types of eye movements, including attention to some negative stimuli or a decrease in the total scan processes. Another non-invasive procedure that measures cognitive and emotional states is eye movement tracking. It has the capability of detecting such things as attention bias, shifts in cognitive speed, and slow emotional responses usually found in depression by recording the quantitative measures like fixation points, scanning paths and alterations in pupil diameter. In our time, virtual reality (VR) technology has the capability to make an experimental situation very real and controllable, inducing emotions become more effective, and the data of the behavioral reaction becomes more authentic and closer to real life, making it more useful. Presently, researchers have also started integrating VR with eye movement tracking technology to build more integrated diagnostic assistance systems. The auxiliary system is able to instinctively capture and process the movement of the user's eyes of the user by designing particular task scenarios in VR (e.g., attention and memory interactive tasks), and this has been shown to provide a comprehensive assessment of the user and has been shown to be effective in clinical environments [4]. It is established that the models that are built on eye movement indicators of VR emotional situations (including the time of regional fixation and scanning speed) can, on the one hand, be taken as high-precision biological behavioral indicators of diagnosing depressive moods and, on the other hand, can be implemented in dynamically tracking the change of the symptoms in the process of cognitive behavioral therapy [5]. Accordingly, eye movement characteristics show a lot of potential and practical use as a dynamic in depression monitoring.

4 Depression detection method based on facial images

Facial expressions and movements tend to indicate the emotional condition of depressed patients. Patients will be observed to come in with sluggish faces, reduced smiling, decreased movement of muscle of the face or fixed gaze towards a particular point. These are minor and consistent behavioral traits that can be valuable practical evidence for depression screening through facial images. More studies are trying to transform these facial behavioral characteristics into certain quantitative signals. Indicatively, there are tools like FacePsy, but open-source, which can detect the points of the face features and muscle activity units in real time using smartphones, and has already discovered such indicators as the extent of eye opening, the upward movement of the corners of the mouth of smiling (AU12), and the muscle act of the downward pull of the corners of the smile (AU15), which played a significant role in the recognition of depressive moods [6]. The other project is named MoodCapture, and according to that analysis, one may conduct a passive analysis of the facial images shot by the corresponding smartphone front camera in the everyday life of a user and then cross-link them with the machine learning models of deep forest applications to identify depressed moods [7]. Convenient and continuous emotion assessment tools have been technically prepared; such research has also entailed maintenance of the privacy of users. Thus, the design of the models should also be guided by the strict rules of privacy protection in order to make the process of clinical transformation more feasible.

5 Depression detection method based on multimodal information

Even though each modality has proven its strengths and merits, they all have constraints: EEG mechanisms are vulnerable to noise interference, movement of eyes can be affected by the external environment and the design of a task, and facial expressions can be either deliberately manipulated or hidden. The data in these single forms can only capture a given dimension of the emotional state of a user, and can effortlessly be influenced, leading to a lack of sturdiness and generalization capability in practice.

Thus, to achieve the more stringent and consistent emotional features, the incorporation of multi-source data (including physiological measures, behavioral measurements, etc.) will end up becoming a trend in the future. Multimodal fusion transcends the restrictions imposed by single-modal data in that it combines information in two or more sources and forms a more fully represented and robust model of emotional states and it substantially improves the discriminative characteristics of the model. This multimodal fusion approach has been confirmed to be useful practically by available studies. As an illustration, the study by Kim et al. developed a deep learning model that incorporated the use of facial videos and speech features. The model was very accurate when predicting the measure of depression, which in effect pointed out the benefits of the visual and auditory combination of the modalities [8]. Moreover, the MultiDepNet system designed by Kumar and colleagues is able to process four different types of information at the same time, i.e., videos, physiological signals, audio, and text. The multi-dimensional information cross-intersection enhances the recognition performance and, through the interpretability analysis, one finds out that the visual modality is the most important one and it gives valuable clues that can be used to perceive the work of multimodal models [9]. The foregoing studies have revealed that multimodal fusion cannot be merely viewed as the simple addition of information, as there should be a structured model to predetermine the inherent linking of the various information with the aim of reaching the effect of 1 + 1 = 2.

6 Discussion and Future Prospects

Through the above research, the existing approaches to detecting depressive emotions by means of human-computer interaction primarily have two technical directions: single-mode and multi-mode. These two methods possess their strengths and weaknesses concerning their detection performance, robustness and clinical interpretability.

Monomodal techniques (including EEG or eye movements) are based on one single signal dimension and their primary benefit is evident features and effective processing. They can have a high level of detection performance in conditions that are appropriate. As an example, the simplified EEG scheme with a temporal feature analysis will be able to reach 96 percent of the classification accuracy [2]. Eye movement analysis using VR also confirmed the usefulness of behavioral indicators [5]. Its weaknesses are also evident, however. This approach not only lacks the ability to generalize due to its inability to capture the full picture of complex and heterogeneous depressive states, but is also highly susceptible to certain interferences, e.g., motion artifacts and environmental light, thus not being adaptable to real and dynamic situations in application.

This led to the existence of multimodal fusion methods. It might be possible to combine information contained in the sources and use their synergistic relationship to create more robust and high-performance through their complementary and redundant nature. As an example, neuroimaging and clinical evidence have been shown to aid in the distinction of depression subtypes with an accuracy of 98.12% in the MAMF-GCN model [10], which illustrates the power of multimodal means to resolve issues. Nonetheless, in order to achieve this, it is important to create a fusion architecture that possesses real knowledge about the complementary nature of modalities and does not just tie and stack data. Meanwhile, other problems with multimodal fusion include the complexity of the model, high standards of data privacy, and additional expenses for calculations. Such matters present greater demands of clinical credibility and feasibility.

Table 1. Comparison between Unimodal and Multimodal Approaches for Depression Detection

Comparison dimension	Monomodal methods (such as EEG, eye movement)	Multimodal fusion methods (as EEG + eye movement + clinical data)
Core logic	Fine-tuned in one dimension of signal, and focused on specific cases.	Combine several information sources and use complementarity and redundancy to get synergy and efficiency.
Testing performance	With appropriate parameters and under controlled conditions, super accuracy is attainable (in the EEG scheme, this is more than 96%) [2].	It wants to reach greater and more constant total performance and is competent in addressing complicated problems (subtype accuracy more than 98) [10].
Information dimension and generalization ability	The Information dimension is single and generalization ability is impossible to take the whole picture of depression.	The information is detailed and allows forming a more thorough emotional representation and possesses a greater generalization capacity.
Robustness	Robustness Extremely delicate to certain perturbations (e.g., the EEG will be subject to motion artifacts, eye movement is subject to ambient light), etc.	The strong side is achieved through incorporating a variety of sources of information. One modality can depend on the other modalities when anywhere there is the influence of one modality.
Clinical interpretability	The source of the features is well defined, and the decision-making process of the model is fairly simple that can be easily articulated and traced.	It has a complicated model architecture and a black box decision-making process, which is difficult to interpret and has an impact on clinical trust.

As per the analysis above (Table 1), it is observable that the technology in the above field continues to suffer numerous challenges whenever it comes to transitioning between the research prototype phase and real implementation.

The major bottleneck is in data. Multimodal models require large-scale, high-quality high-quality and more precisely labeled clinical datasets to be able to be trained and validated effectively. These data are, however, sensitive to privacy, have high costs of collection, and lack standards, they are hard to get. Thus, the vast majority of the current studies are restricted by small, nonhomogeneous datasets, which eventually results in an inadequate generalization capacity of the models and makes the question of their clinical validity questionable.

Secondly, privacy and ethical issues represent another major core challenge. Passive sensing technologies like MoodCapture, which continuously collect sensitive data such as users' facial expressions, voices, and physiological conditions, directly touch upon the core of privacy. Balancing the need to ensure data security, obtain genuine informed consent, and achieve seamless and natural monitoring is a dual challenge in both technological and ethical design.

In addition, in terms of applied implementation, the costs of computing and deployment are of considerable weight. Multimodal models are often complex and might demand resource-intensive computing such that they cannot be implemented on portable and real-time terminals (like mobile phones). Consequently, lowering the implementation threshold is the priority of designing model compression and lightweighting technologies that have been specifically developed in terminals to realize low-power real-time computing.

The next step in the technological process should not be to continue obtaining an increased number of correctness rates in the laboratory, but to resolve the fundamental contradictions of the technology that do not allow its implementation in practice. The technology is to come up with light and flexible technology in regard to integration. These systems are capable of executing themselves on a limited-resource device, can serve to act as intelligent agents, dynamically reconfiguring their techniques of computing to the real-time context, the quantity and condition of the data received and autonomously determining the optimal balance between accuracy, efficiency and robustness. Simultaneously, the critical point to attain clinical trust and overcome ethical dilemmas is the establishment of the intensive absorption of interpretable AI and privacy computing technologies. Building an explainable and open decision-making process and finalizing model training and optimization without exchanging data beyond the local region can earn the trust of the doctors and users. It is assumed that, as significant as it ensures the security and the privacy, it will be able to capture the value of large-scale, high-quality multimodal data, which will ultimately change how debilitating feelings are identified in the latest studies into the provision of a publicly available, reliable, and known health service that is accessible to everyone and everyone puts their trust in it. As the core technologies will become mature enough, and the practical challenges will be resolved, depression emotion detection will not be a one- tool solution any longer. It is capable of becoming a subset and natural component of the digital health ecosystem. Its functions will not be limited to its present screening and auxiliary diagnosis, but a long-term monitoring and personal intervention and holistic health management taking the form of life-long health services. This implies that the upcoming systems can be intricately embedded into the everyday terminals of smartphones and wearables, which would do the constantly daily observations of emotional health rather than relying on active collaboration between the user and the systems. More significantly, it will not only surpass the attributes of a particular tool but it will also attain organic synergy with digital therapy, psychological service platform and clinical pathways, eventually creating a smart, dynamic closed-loop health management ecosystem. This growth transforms customary passive interventions in mental health care to active and preventive interventions. Simultaneously, it also defines the essential and distinctive worth and tenet of the human-computer interaction technology in advancing mental health.

7 Conclusion

The article critically examines the technological advancement of emotion detection techniques of depression under human-computer interaction or multi-modal and single-modal fusion. The conventional forms of the depression diagnosis process have their severe drawbacks; they are very subjective and very noticeable in latency. Introducing emotion detection models into smart devices may passively gather physiological and behavioral data that has become a new direction of the attack on detecting depressed emotions.

The initial technology that was detailed by the study was the single-modal detection technology. EEG and eye movement analysis were employed as objective indicators of detection in the experimental setting, and the results obtained in the experiment could already be very high-precision. Nevertheless, the single modality continues to possess its own inherent issues in the information dimension and strength. Multi-modal Information fusion is thus a trend that is inevitable. The analysis presented in this paper presents the view that the complementary information across modalities can be mined by combining heterogeneous data sets using advanced computer models and in this way, the limitations of a single modality can be overcome and the effectiveness of the detection process and robustness of the system are greatly enhanced.

Despite the fact that multimodal information fusion is an unstoppable trend, there are numerous issues with the evolution of the technology. These are primarily the purchase of multimodal data, the inability to interpret a highly intricate, huge model, and user privacy and ethical considerations of recalling passive constant detection. The current research must proceed to lightweight and adaptive fusion technology, explainable technology in AI and privacy computing technology, etc., to transform the technology to something other than high-precision prototypes but a reliable service.

Conclusively, the advancement of human-computer interaction technology offers effective technical opportunities to detect and constantly monitor depressive moods. With the study of the perception of physiological cues, one can create and be creative in creating multi-modal information fusion directions. One hopes that going forward, a more complete, more accurate, more universal mental health management plan would be erected, and eventually the management of mental health would cease to be a passive, inaccurate, intermittent treatment of the population, it would be one that is continuous and all-round and active.

References

[1] Boby, K., Veerasingam, S.: Depression diagnosis: EEG-based cognitive biomarkers and machine learning. Behavioural Brain Research. 478, Art. no. 115325 (2025)

[2] Khan, S., Umar Saeed, S. M., Frnda, J., et al.: A machine learning based depression screening framework using temporal domain features of the electroencephalography signals. PLoS One. 19(3), Art. no. e0299127 (2024)

[3] Luo, Y., Tang, M., Fan, X.: Meta-analysis of resting frontal alpha asymmetry as a biomarker of depression. npj Mental Health Research. 4(1), Art. no. 2 (2025)

[4] Wuji, Z., Jianli, Y., Xiaoru, Y., et al.: Assisted diagnosis system for depression based on eye-movement virtual scenes. Journal of Hebei University (Natural Science Edition). 44(5), p. 535 (2024)

[5] Zheng, Z., Liang, L., Luo, X., et al.: Diagnosing and tracking depression based on eye movement in response to virtual reality. Frontiers in Psychiatry. 15, Art. no. 1280935 (2024)
[6] Islam, R., Bae, S. W.: Facepsy: An open-source affective mobile sensing system—analyzing facial behavior and head gesture for depression detection in naturalistic settings. Proceedings of the ACM on Human-Computer Interaction. 8(MHCI), pp. 1–32 (2024)
[7] Nepal, S., Pillai, A., Wang, W., et al.: MoodCapture: Depression detection using in-the-wild smartphone images. In Proceedings of the CHI Conference on Human Factors in Computing Systems. pp. 1–18 (2024)
[8] Jin, N., Ye, R., Li, P.: Diagnosis of depression based on facial multimodal data. Frontiers in Psychiatry. 16, Art. no. 1508772 (2025)
[9] Kumar, P., Misra, S., Shao, Z., et al.: Multimodal interpretable depression analysis using visual, physiological, audio and textual data. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 5305–5315. IEEE (2025)
[10] Wang, S., Zhang, J., Zhang, R., et al.: MAMF-GCN model for anxious and non-anxious depression classification and neuroimaging marker recognition. International Journal of Psychophysiology. Art. no. 113301 (2025)