Comparative Analysis of Visual Motion and Multimodal Strategies in Depression Recognition

Hemangi Kacha
P P Savani University, Surat, India

Drashti Bhikadiya
P P Savani University, Surat, India

Abhijeetsinh Jadeja
Sankalchand Patel University, Visnagar, India

Kamini Sharma
P P Savani University, Surat, India

Jayashri Patil
P P Savani University, Surat, India

ABSTRACT
The use of computer vision and artificial intelligence (AI) technologies gives people the chance to study automated systems designed to evaluate visual and motion-based signals of depressive behavior. This review assesses the current state of research on facial landmark tracking, head pose estimation, and multimodal feature integration.Motion-based methodologies in the form of kineme modeling, rotation-invariant geometric frameworks, and interpretable motion dynamics explore relationships between motor behavior and depression. The use of visual techniques that combine facial landmarks, temporal geography, and attention-driven deep networks provides high prediction accuracy, although performance is still affected by lighting, pose, and culture. Multimodal systems that use combinations of facial, verbal, and textual streams of data add explanatory power to the diagnosis, but also issues surrounding explainability and temporal imbalance. Together, the studies highlight the diverse range of methodologies that are being employed to develop automated systems to identify depression in individuals.

Keywords
Depression, Depression Detection, Deep Learning, Emotion, Facial Detection, Motion-Visual-Multimodal Approaches

1. INTRODUCTION
Depression is increasingly relevant to modern healthcare given its immense impact on people globally. The methods for diagnosing depression, like interviews or self-assessments, are inefficient and heavily reliant on the subjective opinion of the evaluator, which can cause delays in the diagnosis. Computer vision and automation brought forth by artificial intelligence (AI) can fill the gap by providing scalable and objective methods for identifying signs of depression. It detects and describes patterns in the psychomotor and behavioral expressions people display. Particular attention is paid to faces, gazes, head movements, and other gestures as these capture some of the non verbal signs that are most often overlooked in verbal or auditory interactions. In this review, I focus on the most relevant literature on psychomotor techniques that are movement-based, vision-based, and those that offer a combination of both, to explain integrated approaches using explainable motion features, dynamics of facial landmarks, and integrated multimodal approaches across major depression datasets, AVEC, DAIC-WOZ, BlackDog, and E-DAIC.

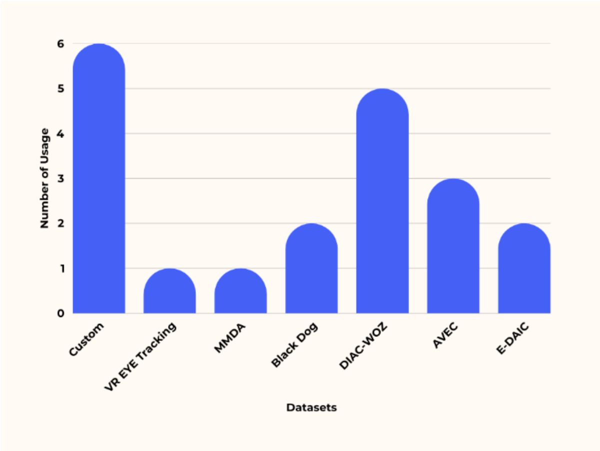

Fig 1: Proportion of Dataset Usage Across 20 Selected Research Papers

2. COMPARATIVE ANALYSIS OF METHODOLOGIES

2.1 Motion-Based Methodologies
A Motion-based approaches powerfully capture psychomotor retardation as a core symptom of depression. These methods examine head pose, micro-movements, eye modulations, and gaze dynamics temporally.

2.1.1 Kineme Models
1st coined the term ‘kinemes’ as head motion biomarkers amenable for depression detection. Their psychomotor approach is based on clinically observed and poorly computed symptoms of depression within motion, irregular or reduced head movement [1]. The study did not utilize opaque deep learning embeddings, rather constructed meaningful and explainable features. It used the BlackDog and AVEC2013 datasets, extracting RNN modeled head movement trajectories. These were segmented and kinematically organized within the RNNs. Regular micro-movements of movement, rather than the static typical visual features were captured. The authors express, “kinemes provide interpretable features that can be directly linked to psychomotor symptoms of depression” [1].

The model visual-only approaches the baseline with F1 0.72 Precision 0.70 Recall 0.75 Motion and motion explainability class. Still, the model performed and lacked head movements while the video was played. Reliance on video motion detection data was clearly demonstrated. Nonetheless, the framework laid the groundwork on resolving. Designing AI clinically declared system for depression detection [1].

2.1.2 Feature Fusion Cues

This introduced a dual-feature framework that combines Temporal Dilated Convolution Networks (TDCN) with Feature-Wise Attention (FWA) to analyze visual motion cues for detecting depression [2]. The key motivation was to integrate facial landmark dynamics with head pose trajectories, thereby capturing both subtle expression changes and broader head orientation shifts.

Their model was tested on the DAIC-WOZ dataset, a widely used benchmark for multimodal depression detection. In this method, TDCN processed facial landmarks to analyze temporal variation across different scales, while FWA was applied to head pose features to classify the most relevant signals. This combination enabled the model to assess facial or motion features that indicated depression most strongly.

On DAIC-WOZ, this framework achieved excellent results and surpassed multiple baselines. The attention fused model, in particular, increased F1 score by over 5% relative to single-feature approaches. The attention mechanism, while accurate, highlighted interpretability features, identifying motion and facial movements such as nodding as substantial predictors. The approach, however, was highly data-centric, needing large annotated datasets to train reliable systems [2].

2.1.3 Time-Angle Features

In a similar fashion, motion-centered approach used rotation invariant time angle features derived from facial landmarks[3]. Their thinking was that representing facial motion as geometric angular changes, as opposed to pixel data, would yield stronger features across different poses and lighting conditions.

Testing demonstrated that their approach achieved both accuracy and computational efficiency that surpassed a number of traditional deep learning models. Moreover, real-time analytics capability due to low computational cost of GhostNet design permitted practical use for telehealth. On the downside, the performance of angular dynamics highly correlates with the accuracy of landmark extraction, which could be detrimental to the system due to the inherent problems of landmark detection. In any case, the study opened interesting directions for research and development of real-time detection systems for depression [3].

2.1.4 Cross-Cultural Kinemes

The expandability of the kineme framework with German, Australian and American samples along three datasets of cultural variety [4]. The goal was to assess the applicability of motion- based biomarkers across populations with very different datasets to alleviate depression detection dataset bias.

Assessment of classifier performance by RNN and CNN models probed performance consistency. The authors quoted “kineme-based features generalize better than raw head motion and other visual cues” to illustrate the generality of their claims [4].

The datasets indicated that kineme-based modeling comparative to other motion visual models and baselines was superior across all datasets. These facts indicate the motion biomarkers are cross culturally robust, and thus are strong candidates for clinical deployment. However, the cultural nuances in nonverbal behavior remain the main obstacles, indicating the best performance is likely found in hybrid approaches that use both universal and culture specific features [4].

2.1.5 Interpretable Motion Dynamics

From a biomechanical perspective, it focused on the study of depression through the use of Lie Algebras to represent the facial dynamics of landmarks and head rotation along with their spatial trajectories [5]. The aim of the research was to synthesize clinically relevant features that are both discriminative and clinically relevant through the use of Gaussian Mixture Models and Fisher vectors to encode complex movement trajectories.

Evidently stepwise systems with motion-dynamic features distinguished and classified levels of depression with high accuracy. Thanks to the interpretable Lie Algebra encodings, clinicians were able to follow the logic of the classification. The approach, albeit still computationally expensive, works as a demonstration of the clinically informed mental health AI modeling rigorously predicted as low hanging fruit.

2.1.6 Autoencoding Motion

While the works of multimodal, they still focused on representations of the face and head motion, and thus are tangentially relevant to this line of research. The study employed deep autoencoding networks to concurrently model facial micromovements, head motions, and voice features. The prediction was that motion signals integrated with speech dynamics provided the most stable indicators of the severity of depression.

Depression features confirmed head motion and facial motion features and single modality models were outperformed by voice systems. However, the limited reliance on large multimodal datasets kept scalability intact. This work supported motion cues as dominant predictors of depression and reinforced the motion cues within multimodal frameworks fusion depression models.

2.1.7 FacialPulse RNN

Along the RNN based architecture to FacialPulse, focused on further speed enhancement and the processing of sequences of facial landmark traces [7]. Capturing the timing dynamics of facial motion rather than the static appearance of features is the reasoning for the shift focus from most conventional visual models.

The model evaluation was on AVEC2014 and the MMDA datasets, designed to capture the optical flow of face regions for long term 3D motion of facial landmark intervals. Fast computation and capture of the dynamic features was achieved with the design that the authors reported, stating that FacialPulse was 21% lower compared to the recognized baseline on the MAE metric, and served double the recognized speed [7].

These findings further validated the accuracy and efficiency of FacialPulse, positioning it as a robust candidate for real-time monitoring of patients with depression during telehealth appointments. FacialPulse still requires improved landmark identification to resolve issues of extreme head motion, poor tracking, and video noise. Ultimately, the model is a good enough in scalable temporal motion depression detection systems [7].

2.1.8 FacePsy Mobile

They studied a novel eye-tracking motion system in virtual reality (vr). The study assumed the depression associated cognitive and attentional deficits could be. captured in oculomotor behavior of screen/fixations and eye saccades.

Subjects used virtual reality headsets and were subsequently tracked as they moved their heads using an algorithm to determine gaze. By tracking head movements, parameters of fixation, saccade, and spatial scanning were taught to machines like XGboost and MLP classifiers. The system got very good results, with MLP classifiers hitting an F1 score of 0.9. 92 and strong correlating with PHQ-9 scores. This suggests VR might be able to track depression and be used in motion based depression tracking systems. This also suggests depression motion tracking systems might need to use VR sensors and centers. [8]

2.2 Visual Methodologies (Facial Landmarks & Expressions)

Visual cue-based methods depend on recognizing facial emotions and action units (AUs) and facial dynamics corresponding to emotional and affective states.

2.2.1 FacePsy Mobile

Islam and Bae designed and developed FacePsy, a system for real-word depression detection via smartphone cameras and mobile devices. Rather than lab settings around depression detection, FacePsy has a principal focus on kinetic authenticity, recording facial expression, eye focus, and head movement in unconstricted settings. Its most major breakthrough revolution is mobile adaptability, which changes the paradigm of depression detection and allows for continuous unmonitored collection of data.

The system is able to work aboard mobile devices because it uses facial behavior metrics of micro expressions, action units (AUs) plus blink frequency, and light cognitive algorithms. These algorithms allow mobile devices to perform real time data processing without incurring computational burdens. Also, the FacePsy application increases the expression and gesture based signals of the mobile device. This is for the unmet need of devices in the assessment of mental health.

The evaluation carried out showed great accuracy in distinguishing the symptoms of depression from normal facial movement in the absence of specific F1, recall, or precision metrics. This system, is easily portable and has close to 100% accuracy in real-world depression detection. FacePsy is still limited to privacy challenges in the field and the camera and light framing placement. Regardless, FacePsy is still very useful regarding the new mobile based approach to depression detection [9].

2.2.2 FacialPulse Landmarks

This variant of FacialPulse also dealt with layered motion in the presence of spatio-temporal visual stimuli, in particular the trajectory of facial landmarks [10]. These researchers developed RNN centered frameworks focusing on moving primary landmarks to overcome the conventional approach of working with still facial images. Such models aim to capture the elusive temporal features of depression facial muscle movements.

The approach consisted of a succession of 68 facial landmarks derived from videos within the AVEC2014 dataset. These sequences were fed into a RNN configured to attend to long intervals of time. The authors point out that capturing the duration and rhythm increases the ability to distinguish target depression users in the dataset.

FacialPulse, in the case of RNN baseline models of expression analysis, obtained a significant increase in recognition accuracy and an increase in processing speed. Also, the architecture is compact enough to enable almost real-time processing speed which is a plus for telemedicine systems.This study also recognizes and appreciates sensitivity to landmark extraction inaccuracies, which is commonplace in low illuminated and occluded cases.

2.2.3 LSTM with Attention

This study employed deep learning techniques with LSTM layers and attention mechanisms for detecting depression from facial expression datasets [11]. The authors emphasize depression's core relevance to facial dynamics but highlight that the models must also offer transparency to be clinically relevant. Hence, the coherence fusions layer feature was added to the structures to increase their interpretability.

The methods used to derive these facial action units and spatio-temporal expression features were the stacked LSTMs with attention mechanisms which then multiplied the specific elements in the output that were deemed important. The model was previously evaluated on benchmark datasets consisting of neutral and emotional expression data. The authors quote that “attention weighting enabled the model to emphasize clinically relevant features, such as less smiling or more expressive face breathing” [11].

The system was also found to offer competitive recognition rates and the accuracy and F1 scores were found to exceed the benchmark baseline CNNs. The clinical nature of the system was more attractive because of its interpretability and facility to feature which features of the data contributed more to the predictions. The main shortcomings were the need of high quality facial videos to be used in the system and also the diminished performance in natural and uncontrolled conditions. Still, it showed the first signs of interpretability in the facial-expression based depression detection systems [11].

2.2.4 Emotion Deficit Meta

Unlike the studies on single models, this meta-analysis focused on the accumulation of evidence that individuals with Major Depressive Disorder (MDD) have a considerable and unchanging deficit in the recognition of facial emotional expressions [12]. People suffering from depression demonstrated a tendency to struggle the most with recognition of positive emotions, especially happiness, while negative emotions were recognized with greater accuracy.

The depression-emotion recognition deficit linkage becomes stronger with each subsequent study; meta analyses, as the one used here, provide the cross study methodology to assess bias emotion processing model phenomena. Depression bias has been shown to profoundly impact the way faces are perceived, which allows such bias to be modeled. In this way the depression context acts to provide a cognitive basis for the negative bias in processing emotions, thus the model of the depression profoundly bias outlook serves to prop the computational basic for the model depression.

The lack of positive emotions recognition fully justifies the absence of positive emotions in produced systems concerning sadness or the over accentuated, stereotypical emotions of sadness. There is a ‘clinical bottom line’ that such systems ought to be clinically positioned, exercising greater adherence to the psychiatric foundations of the issue to the evidence posed. [12].

2.2.5 Micro-Expression Real-Time

They developed a real-time depression detection framework leveraging facial micro expressions [13]. Those micro-expressions of emotion tend to escape one’s attention as they are the instantaneous, reflexive movements of the face, underlying sadness, or depression, that appear to be more reliable than the voluntary, exaggerated facial poses one is trained to use. There is an effort to move beyond the use of photographs to an engagement with the very rapid and subtle to detect systems.

The system utilized facial action coding and micro expression recognition algorithms along with high temporal dynamics to process high-frame-rate videos. Features were then classified using deep neural networks streamlined for efficient processing. The authors stayed focused on the balance between real-time processing and speed remarking that the framework was designed for active surveillance situations.

The research showed high levels of accuracy in detection, especially in detecting mild and moderate depressive symptoms. However, the actual deployment was difficult to achieve, as micro expressions are difficult to capture in free ranging settings. Nonetheless, the model demonstrated micro-expression recognition could be used as a depression biomarker and proved the concept necessary for field application. [13].

2.2.6 *Fuzzy-CNN Hybrid*

This integrates the fuzzy logic systems with deep learning CNNs to extract and classify depressive features from facial expressions. This line of reasoning coupled with deep learning sought to sidestep the boundary rigid classification problems of fuzzy systems that is emblematic of emotion expression.

The system architecture is structured such that facial landmarks and expression features are interleaved throughout shallow and deep structures in the CNN. This forms layered hierarchies of features that classify emotion into upper and lower depressions for easy interpretation and smoother processing. This interplay in the model sought to balance the emotional and cognitive aspects of the system.

The results indicate that the hybrid model CNN - fuzzy systems had higher F1 scores and accuracy than the CNN - only models, particularly in cases that were difficult to express. The best part about the model is that it was easy to understand, but also meant that the model was heavily reliant on the tuned rules for the fuzzy sets, which was unlikely to be transferable to other datasets. It demonstrates the value of combining deep learning and symbolic AI [14].

2.3 Multimodal Methodologies.

Multimodal approaches seek to capture and leverage visual, motion, and audio attributes for complementary information retrieval. Systems designed with multimodal information analytics and retrieval capabilities accomplish these tasks, and subsequently, they also need to address a plethora of complexities, including but not restricted to interpretability.

2.3.1 *PHQ-8 Fusion*

Team members have tie microphones and video monitors to build a Brown-and-Person model estimating PHQ-8 scores and gauging depression from facial audio-visual feeds.

They worked under the idea that expressions on the face in a video alone could be used as unreliable information to make and that it was better used in conjunction with the information from the speaker's voice from the video in addition to the verbal components.

The model had a computer with spatiotemporal frame strengths with a composition of other computers surrounded with long short term memories. They extracted visual components from the face with sets of markers to see facial information and then bought sound from the mic figures that controlled pitch and speech over rhythm and timing. The computer with the Graph Convolution performed facial area spatial relations and the computer with the other graphics performed on the other side of the computer to controlling timing. When looking at the E-DAIC dataset, it was apparent that the depressed model had more sleeping and gained better PHQ-8 scores than people relying on one of the systems alone and that using the Brown model improved score estimated mean error. The investigation showed that multimodal fusion is in line with how clinicians detect depression: via listening and observing. Its drawback was computational cost and sensitivity to absence of modality data [15].

2.3.2 *Lie Algebra Motion*

Designed an interpretable multimodal system for depression detection using facial landmark dynamics and head motion features. Most Depression Detection systems usually focus on interpretability, to assure that the outcomes align with the clinically recognized psychomotor symptoms patients exhibit.

Along these phases, the methods which attempt to identify the psycho-motor diagnoses have largely been the ones which describe facial landmark trajectories and head rotations using Lie Algebra Frameworks, and then encode them using Gaussian Mixture Models (GMMs) and Fisher Vectors. the features being tracked permit motion and irregular dynamics to be interpretable as passive behavioral indicators of depression.

The models were accurate to the extent that the interpretable features accurately predicted the severity of depression. indicators of motion instead of vague guesses. This was an attempt at defining explainable multimodal AI. [16].

2.3.3 *Autoencoding Fusion*

They created and proposed a unique approach aimed at combining the use of facial, head, and voice analysis in assessing depression severity [17]. Their assumption was that working on a micro level of different ways of expressing depressive behavior and combining them would result in a better representation of depressive behavior.

The framework incorporated the use of autoencoders which learned specific features specific for each modality, which then were compressed into a singular representation. The resulting representation was used for severity estimation. The facial features were expressive, the head movements expressive of psychomotor retardation, and the voice occupied the features of expressive tone.

Results showed that feature fusion greatly exceeded the performance of individual features, particularly head + face to audio alone. The paper maintained that weak predictors across multiple modalities, when combined, can produce a stronger classifier. The difficulties mentioned included lack of feature and component fusion, particularly audio and the overall background noise. Yet, the approach demonstrated the value of multi deep learning for depression assessment [17].

2.3.4 *Cross-Attention Fusion.*

This study outlines a hybrid multi-head cross-attention network for recognizing depression across modalities [18]. Their work argues that depression is multi-faceted and that facial appearance provides coarse contextual cues, while facial motion offers fine-grained behavioral information. Similar to prior research that utilized linguistic and behavioral features for depression analysis on social media [19], this model integrates complementary cues to enhance recognition performance. The architecture employed feature extraction using pre-trained CNNs (ResNet-50, GoogleNet) and incorporated recurrent layers to encode temporal dynamics. Cross-attention mechanisms were used to effectively merge appearance and motion features from multiple modalities, aligning with multimodal fusion principles observed in other NLP-driven affective computing frameworks [19]-[20].

Performance improvements in accuracy and recall on the AVEC dataset significantly surpassed unimodal baselines. The model's main strength was its ability to dynamically balance motion and static facial cues. However, like other studies in personality and behavioral analysis [22]-[25], it faced limitations related to computational complexity and reliance on large-scale annotated datasets. This work reinforces the growing importance of attention-based multimodal fusion approaches in depression detection, echoing insights reported in earlier affective computing and mental-health analysis research [19], [23].

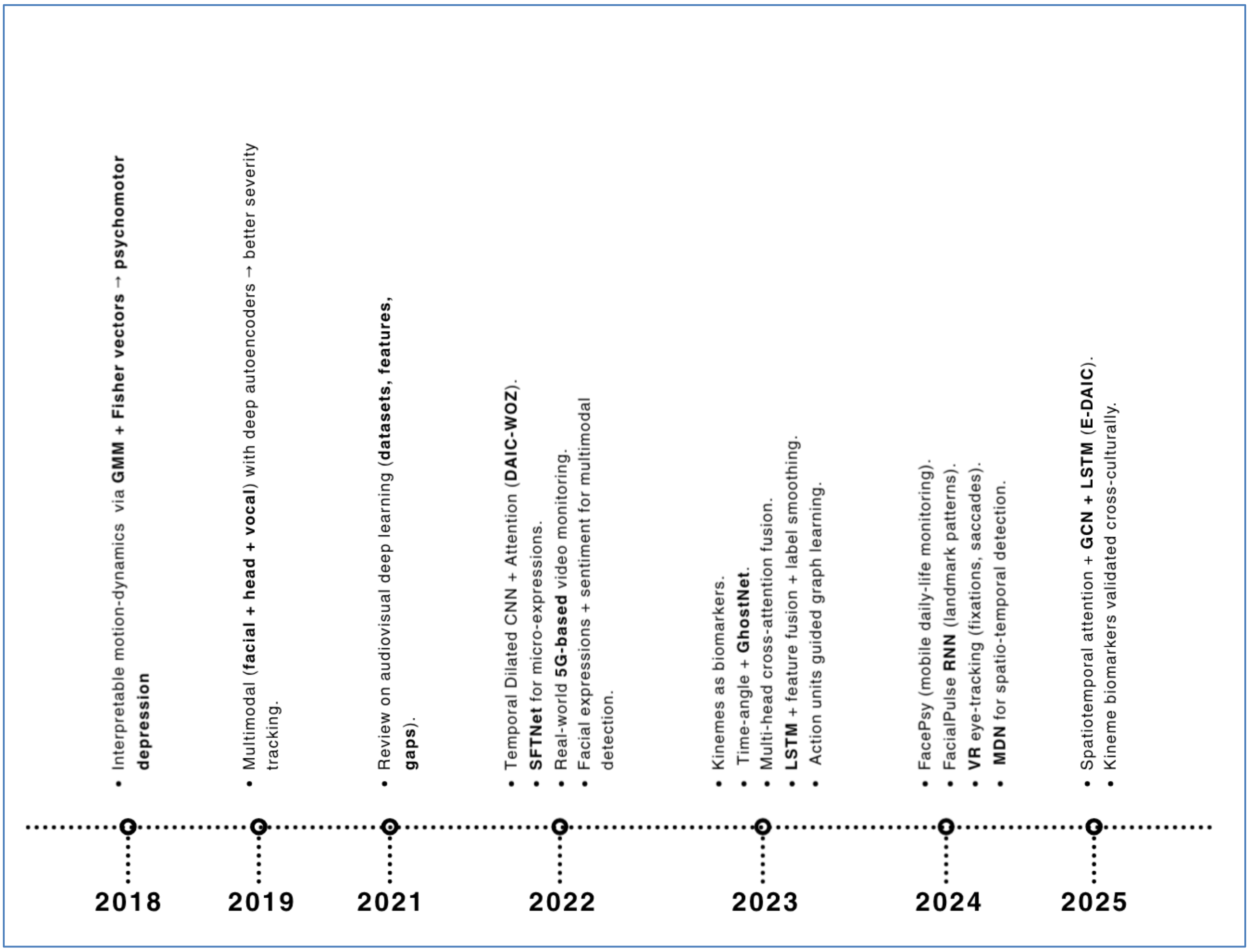

Fig 2: Timeline of Advances in Motion, Visual and Multimodal depression detection.

Table 1 : Comparision Analysis of Motion, Visual Cues and Multimodal Approches

Aspect	Top	In-between	Bottom
Input Features	Head motion patterns, kinemes, geometric angles, movement velocity/acceleration	Facial landmarks, AUs, micro/macro expressions, temporal dynamics	Combination of facial, vocal, textual, or physiological signals
Model Architectures	RNNs, Lie Algebra models, GMMs, Fisher vectors, interpretable motion encoders	CNNs, 3D CNNs, LSTMs, transformers, MIL, attention-based fusion	GCNs, LSTMs, GNNs, autoencoders, dual-stream CNNs, cross-modal attention

Strengths	High interpretability (link to psychomotor symptoms), lower computational cost, generalizable across datasets	Rich visual detail, direct link to emotion recognition deficits, effective with highquality video	Captures complementary cues across modalities, achieves highest accuracy, clinically more reliable
Limitations	Sensitive to head tracking errors, limited expressive range,weaker with static/lowmotion subjects	Strongly affected by lighting, occlusions, cultural bias in expressions, higher data needs	High computational cost, synchronization challenges, data imbalance (missing modalities), harder to deploy in real-world
Best Use-Case	Explainable,lightweight clinical screening, crosscultural robustness	Controlled lab settings, emotion-driven tasks, video interviews	Telehealth, in-depth diagnosis, severity prediction, robust but resource-intensive

3. CONCLUSION

This review examined approaches for detecting depression: motion-based, visual, and multimodal methods. Motion techniques provide interpretability. However, they encounter problems when individuals do not move much. Systems that rely on visual cues are quite successful, but their analyses are highly influenced by video quality and culture. Despite needing a lot of resources, multimodal systems provide the greatest diagnostic accuracy. Future research should target explainable multimodal models that incorporate principles of cross-cultural validation and are designed for scalability to enable efficient depression screening in clinic-suitable environments.

4. ACKNOWLEDGMENTS

Our thanks to the experts who have contributed towards development of the template.

5. REFERENCES

[1] S. Gahalawat et al., “Explainable Depression Detection via Head Motion Patterns,” Proc. 25th ACM Int. Conf. Multimodal Interaction (ICMI), 2023, doi : 10.1145/3577190.3614141.

[2] H. Guo et al., “Automatic Depression Detection via Learning and Fusing Features from Visual Cues,” arXiv preprint arXiv:2203.11983, 2022.

[3] Y. Ding et al., “Depression Detection Using Time-Angle Features from Video,” Appl. Sci., vol. 13, no. 16, p. 9230, 2023, doi: 10.3390/app13169230.

[4] S. Gahalawat et al., “On the Validity of Head Motion Patterns as Generalisable Depression Biomarkers,” arXiv preprint arXiv:2502.04938, 2025.

[5] M. Kacem et al., “Detecting Depression Severity by Interpretable Representations of Motion Dynamics,” Proc. 13th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), pp. 739–745, 2018, doi: 10.1109/FG.2018.00116.

[6] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, “Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding,” IEEE J. Biomed. Health Inform., vol. 22, no. 2, pp. 525–536, 2018, doi: 10.1109/JBHI.2017.2676878.

[7] R. Wang et al., “FacialPulse: An Efficient RNN-based Depression Detection via Temporal Facial Landmarks,” Proc. ACM Multimedia (MM), pp. 311–320, 2024, doi: 10.1145/3664647.3681546.

[8] T. Zheng, Y. Wang, M. Wang, Y. Guan, P. Chen, and G. Mao, “Diagnosing and Tracking Depression Based on Eye Movement in Virtual Reality,” Front. Psychiatry, vol. 15, art. 1280935, 2024, doi: 10.3389/fpsyt.2024.1280935.

[9] M. R. Islam and S. W. Bae, “FacePsy: An Open-Source Affective Mobile Sensing System—Analyzing Facial Behavior and Head Gesture for Depression Detection in Naturalistic Settings,” Proc. ACM Hum. Comput. Interact. (PACMHCI), vol. 8, art. 260, 2024, doi: 10.1145/3676505.

[10] S. Hemrungrojn and P. Vateekul, “Explainable Depression Detection Based on Facial Expression Using LSTM on Attentional Intermediate Feature Fusion with Label Smoothing,” Sensors, vol. 23, no. 23, p. 9402, 2023, doi: 10.3390/s23239402.

[11] M. N. Dalili, I. S. Penton-Voak, C. J. Harmer, and M. R. Munafò, “Facial Emotion Recognition in Major Depressive Disorder: A Meta-Analysis,” Psychol. Med., vol. 45, no. 6, pp. 1135–1144, 2015, doi: 10.1017/S0033291715001104.

[12] G. Gilanie et al., “An Automated and Real time Approach of Depression Detection from Facial Micro-expressions,” Comput. Mater. Continua, 2022, 10.32604/cmc.2022.028229.

[13] A. S. Rajawat et al., “Fusion Fuzzy Logic and Deep Learning for Depression Detection Using Facial Expressions,” Procedia Comput. Sci., vol. 218, pp. 2795–2805, 2023, doi: 10.1016/j.procs.2023.01.251.

[14] N. Jin, R. Ye, and P. Li, “Diagnosis of Depression Based on Facial Multimodal Data,” Front. Psychiatry, vol. 16, art. 1508772, 2025, doi : 10.3389/fpsyt.2025.1508772.

[15] A. Kacem, Z. Hammal, M. Daoudi, and J. F. Cohn, “Detecting Depression Severity by Interpretable Representations of Motion Dynamics,” Proc. 13th IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), pp. 739–745, 2018, doi: 10.1109/FG.2018.00116.

[16] H. Dibeklioglu, Z. Hammal, and J. F. Cohn, "Dynamic Multimodal Measurement of Depression Severity Using Deep Autoencoding," IEEE J. Biomed. Health Inform., vol. 22, no. 2, pp. 525–536, Mar. 2018, doi: 10.1109/JBHI.2017.2676878.

[17] Z. Zhu, J. Liu, L. Zhang, and Y. Zhao, "Facial Depression Recognition via Hybrid Multi Head Cross Attention," Front. Psychol., vol. 14, art. 1211598, 2023,doi:10.3389/fpsyg.2023.121159.

[18] M. Gahalawat et al., "On the Validity of Head Motion Patterns as Generalisable Depression Biomarkers," arXiv preprint arXiv:2505.23427, 2025.

[19] Patil, J., Patil, V., Prajapati, K., Patel, D., Trivedi, S., & Patel, R., "Enhanced Depression Detection on Social Media Using Advanced Machine Learning and Linguistic Analysis Techniques", In International Conference on Intelligent Computing and Communication (pp. 263-275), 2024, Singapore: Springer Nature Singapore.

[20] Verma, S., Patil, J. A., & Tamhankar, I., "Integrating Natural Language Processing with CGANs to Generate Customized, Realistic Traffic Scenarios for Autonomous Vehicle Training", In 2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE) (pp. 1-5), 2024, IEEE.

[21] Patil, M. J. A., & Godhwani, M. P. B., "Review of Name Entity Recognition in Marathi Language", IJSART, 2(6), 2016.

[22] Patil, J., & Sheth, J., "Deep Learning and Machine Learning Approaches for the Classification of Personality Traits. In Rising Threats in Expert Applications and Solutions", Proceedings of FICR-TEAS 2022 (pp. 139-146), 2022, Singapore: Springer Nature Singapore.

[23] Patil, J., & Sheth, J., Comparative study of data sources, features, and approaches for automatic personality classification from text. Int. J. Comput. Appl, 174, 0975-8887, 2021.

[24] Patil, J., & Sheth, J., "Data Preparation and Quality Challenges for the Personality Recognition in Indian Languages using Machine Learning and Deep Learning Approaches. Journal of IoT in Social, Mobile, Analytics, and Cloud, 4(1), 33-40, 2022.

[25] Patil, J., & Sharma, K., Personality Recognition from Text in Indian Languages : Challenges, Progress, and Future Directions", International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2456-3307, Vol. 11, no. 5, pp 398-401, 2025, doi:https://doi.org/10.32628/CSEIT251117136