International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

"Visual Emotion Analysis for Depression Detection on Social Media: A Review"

Ritika Verma, Prof. Balram Yadav

Abstract:

Depression continues to be one of the most prevalent mental health challenges, affecting millions of individuals across the globe. Early identification is crucial for ensuring timely intervention and psychological support. In recent years, the widespread use of social media has opened up new avenues for understanding user behaviour and emotional patterns. These digital platforms, when analysed using artificial intelligence (AI) techniques, offer promising potential in detecting early signs of depression.

While earlier research primarily focused on analysing textual content from social media, recent progress in computer vision has enabled the exploration of visual data—particularly facial expressions, body posture, and image content—for depression detection. This review paper presents a comprehensive overview of image-based approaches, with an emphasis on the application of machine learning (ML) and deep learning (DL) models. It further discusses popular methodologies, standard benchmark datasets, commonly used evaluation metrics, and prevailing challenges in this growing area of research. The paper concludes by outlining potential future directions, including multi-modal data fusion, ethical implications, and the scope for real-world implementation.

Keywords: Computer vision, Depression detection, deep learning, facial expression recognition, image analysis, mental health monitoring.

1. Introduction:

Depression has emerged as a significant global health concern, currently impacting more than 280 million individuals, as reported by the World Health Organization (WHO, 2023). Traditionally, the diagnosis of depression has been reliant on structured clinical interviews and self-reported assessments. While these methods are effective in controlled environments, they often lack scalability and may not capture the day-to-day emotional fluctuations experienced by individuals.

In recent years, the rapid expansion of digital content—particularly on social media platforms—has introduced new possibilities for mental health monitoring. Visual content such as selfies, photographs, and digital art shared online may carry subtle psychological cues. Elements like facial expressions, eye contact, body posture, lighting, and even colour tones used in images can reflect underlying emotional states. Unlike textual inputs, which require deliberate user interaction, images offer a passive and non-intrusive way to observe mental well-being. This has opened up a promising avenue for image-based depression detection using advanced computer vision and deep learning techniques.

The role of language in mental health analysis is already well-established. For instance, Rhanoui et al. (2019) used the Linguistic Inquiry and Word Count (LIWC) tool to detect suicide risk and emotional distress based on social media posts, highlighting the power of language in uncovering hidden psychological issues. Similarly, Cheng et al. (2017) demonstrated how machine learning algorithms can assess suicide risk by analysing online user behaviour, thereby validating the use of AI in early mental health screening.

In this context, models such as attention-based CNN-BiLSTM have proven effective in capturing emotional cues from digital content. While most prior research has focused on textual analysis, there is now growing interest in visual-based models that can process image data shared on platforms like Instagram, Facebook, and Reddit. These platforms often serve as digital diaries, providing unfiltered glimpses into individuals' emotional journeys.

Compared to traditional methods, image-based depression detection offers several advantages: it is scalable, real-time, and non-invasive. Moreover, it enables proactive mental health monitoring without waiting for clinical consultations, making it highly suitable for implementation in low-resource settings and online wellness platforms.

Real-time tracking: It allows continuous observation of a person's mood or behavior without time lags.

Broad reach: It enables large-scale mental health studies without physical limitations.

Low invasiveness: Detection systems can work in the background without requiring active participation from individuals.

2. Fundamentals of Image-Based Depression Detection

With the growing dominance of visual content on digital and social media platforms, image-based approaches for mental health monitoring have gained considerable attention. Unlike textual data, visual cues such as facial expressions, body posture, and even artistic representations provide a silent yet powerful medium to convey a person's emotional and psychological state. In the case of depression, subtle but consistent visual patterns often emerge—many of which can be detected more reliably by machines than by the human eye.

2.1 Visual Cues of Depression

Depression can manifest through a variety of physical signs, many of which are reflected in a person's facial appearance, expressions, and overall posture. While these changes may go unnoticed by casual observers, advanced computer vision and machine learning techniques are capable of identifying these fine-grained details with high consistency.

Some of the commonly observed visual indicators include:

- **Reduced facial activity**: Diminished movement in facial muscles, often leading to a flat or blank expression.
- **Eye-related cues**: Downward gaze, reduced blinking, or avoidance of eye contact.
- **Sadness indicators**: Subtle frowning, drooping of the eyelids or corners of the mouth.
- **Fatigue and stress symptoms**: Presence of dark circles, pale complexion, and slouched or withdrawn body posture.
- **Emotionally muted selfies**: Rare smiling or frequent posting of blank-faced, low-expression images on social media platforms.

These signs may present individually or in combinations, and their interpretation can also be influenced by cultural and demographic factors. Therefore, while image-based depression detection holds substantial potential, it also poses challenges in terms of generalisability and fairness.

2.2 Applications in Clinical and Social Media Settings

The utility of visual-based depression detection spans both clinical and non-clinical environments. Two of the most prominent application areas are:

A. Clinical Screening and Diagnosis

- Facial expression analysis during therapy sessions or psychological interviews can offer clinicians deeper insights into a patient's emotional state—often without any verbal input.
- Machine learning tools, when used as supportive systems, can track visual patterns over time, helping in early diagnosis and continuous monitoring of depressive symptoms.

B. Social Media-Based Monitoring

- Platforms like Instagram and Facebook act as visual diaries where users regularly share their selfies, photos, and artwork. These visual posts can reflect mood, energy levels, and behavioural patterns.
- Research suggests that individuals suffering from depression tend to post images that are visually darker, grayer, and lower in saturation, potentially signalling emotional withdrawal or blunting.
- Beyond selfies, even images drawn or shared in the form of digital art or memes can contain affective cues. Emotion detection in profile pictures, cover photos, or sketches is also being explored by researchers.

2.3 Role of Facial Expression Analysis in Depression Detection

Facial expressions remain one of the most informative and studied visual indicators for depression. Automated systems for facial expression analysis typically follow a structured pipeline:

1. **Face Detection**: The first step involves identifying and isolating the face region from an image using algorithms like Haar Cascades, Histogram of Oriented Gradients (HOG), or more advanced deep convolutional neural networks (CNNs).
2. **Facial Landmark Detection**: Next, the system pinpoints critical facial points such as eyes, eyebrows, nose, and mouth corners, which help in tracking micro-movements and expressions.
3. **Feature Extraction**: This involves computing relevant features that describe facial patterns—either using handcrafted methods like Local Binary Patterns (LBP) and HOG, or using feature embeddings obtained from pre-trained deep learning models.
4. **Emotion Classification or Depression Scoring**: Finally, classifiers such as Support Vector Machines (SVM), Random Forests, or deep regression models are employed to predict emotional states or assign a depression score to the individual.

These systems continue to evolve, with newer architectures incorporating temporal dynamics, context-awareness, and cross-modal fusion for enhanced accuracy.

2.4 Advantages over Text-Based Detection

Feature	Text-Based Methods	Image-Based Methods
Language Dependent	Yes	No
Cultural Expression Variability	High	Moderate
Passive Sensing Possible	No	Yes (via images/selfies)
Emotion Leakage	Less visible	More visually apparent
Applicability in Children	Limited	High (visual cues more available)

3. Machine Learning and Deep Learning in Depression Detection

The evolution of ML and DL methods has significantly enhanced the ability to analyze online behavior for mental health evaluation. These tools classify users based on patterns in how they express emotions and interact with others online. This section covers key algorithms, model architectures, and evaluation practices in this growing field.

3.1 Classical Machine Learning Techniques

Before deep learning became mainstream, traditional ML methods formed the foundation of automated mental health analysis. These models required explicit feature extraction before classification.

Facial Landmark Detection: Detect key points (e.g., eyebrows, lips) and measure deviations.

Hand-crafted Features: Histogram of Oriented Gradients (HOG), Local Binary Patterns (LBP) to describe textures.

Support Vector Machines (SVM), Random Forests: Classical classifiers used with engineered features.

3.1.2 Feature Engineering Strategies

ML models rely on crafted features such as:

- Lexical Features: Word frequencies, n-grams, part-of-speech tags.
- Sentiment Analysis: Scoring of emotional tone.
- Psycholinguistic Indicators: Use of self-focused language or emotional words.
- Behavioral Metrics: Posting habits, interaction levels, and timing.

While effective, ML approaches fall short in capturing the richness and variability of human language, paving the way for deep learning.

3.2 Deep Learning Approaches

Deep learning models excel at uncovering intricate patterns in text without manual intervention. They learn directly from raw data and often outperform traditional ML in both accuracy and adaptability.

Applications in Depression Detection:

Majid et al. (2022) introduced attention mechanisms in CNN models to enhance feature focus. Babu et al. (2022) highlighted hybrid models like CNN-BiLSTM as powerful tools in sentiment analysis. Ombabi et al. (2020) demonstrated the success of CNN-LSTM models in analyzing Arabic text, proving these methods are adaptable across languages.

Following are Popular DL Architectures

Convolutional Neural Networks (CNNs): Automate feature extraction and achieve high accuracy.

Pre-trained Networks: ResNet, VGGNet, and Inception architectures fine-tuned for depression recognition.

Facial Expression Recognition Models: Trained on datasets like AffectNet and adapted for depression-specific cues.

Table 1: Reviewed Models and Approaches with Their Advantages and Limitations

Model / Approach	**Advantages**	**Limitations**
CNN (Convolutional Neural Network)	Effective in detecting local spatial patterns like facial features.	Struggles with long-range dependencies and context.
ResNet / VGG (Pre-trained CNNs)	High accuracy due to transfer learning and large-scale training.	Computationally expensive; may overfit small datasets.
Facial Action Units (FAUs)	Interpretable indicators tied to specific muscle movements.	Requires detailed facial annotation and expert-labeled data.
FER with Emotion Classification	Simple architecture; effective for basic emotion detection tasks.	Limited performance on subtle or compound emotions.
Multimodal Integration (Image + Text)	Combines multiple cues for more accurate detection and context understanding.	Requires synchronized multimodal data; increases complexity.

3.4 How Models Are Evaluated

To ensure reliable results, models are judged using metrics like:

Accuracy: Overall correctness.
Precision: Accuracy of positive (depressed) predictions.
Recall: Ability to detect all actual depressed cases.
F1-Score: Balances precision and recall.
ROC-AUC: Measures discrimination power between classes.

4. Benchmark Datasets

Effective depression detection using image data relies heavily on access to rich and diverse datasets that capture relevant visual cues. Unlike video or audio datasets, the following collections focus exclusively on still images, such as facial expressions, user profile pictures, or posed emotional portraits. These datasets serve as critical training and evaluation resources for machine learning and deep learning models targeting visual signs of depression.

Table 2: Benchmark Datasets for Image-Based Depression Detection

Dataset	Type	Labels
AffectNet	Facial images from web (1M+)	Basic emotions, valence-arousal
EmotiW (Image Subset)	Still images from movies, real-world scenes	Basic emotions (e.g., sadness, neutral)
IMFDB	Indian movie face images (34k+)	Emotion, age, gender
RAF-DB	Real-world affective faces (30k)	Basic & compound emotions
FER+	Facial expression dataset (35k+)	8 emotion categories including sadness

5. Challenges and Future Directions

Although image-based depression detection offers a promising and non-invasive approach to mental health screening, it is still a maturing research area with several significant challenges. These issues span technical limitations, ethical concerns, data diversity, and practical deployment. Addressing these gaps is vital to ensure the development of reliable, ethically sound, and clinically acceptable solutions.

5.1 Key Challenges

A. Data Privacy and Ethical Concerns

One of the foremost concerns in using image data for mental health analysis is user consent and privacy. Many users share photographs on social media platforms without being fully aware that these images could be utilised in AI-based emotion or depression detection studies. This raises important questions around informed consent, data transparency, and data usage rights.

Stigmatization Risk: Incorrect classification or unintended exposure of an individual's emotional condition can lead to psychological distress or social stigma.

Ownership Ambiguity: Platforms like Instagram and Facebook have complex data usage policies, making it legally and ethically challenging to use such images for model training or deployment.

B. Data Quality and Demographic Diversity

- **Bias in Datasets**: Most existing datasets (e.g., FER+, AffectNet) are skewed towards Western facial features and expressions, resulting in poor generalisability to other ethnic groups and age ranges.
- **Low-Quality Visuals**: Images from social media are highly variable in terms of lighting, resolution, and clarity, which affects the accuracy of facial detection and analysis.
- **Unreliable Labels**: Unlike clinically validated datasets, many image corpora do not come with professionally annotated depression diagnoses, leading to noisy or ambiguous learning data.

C. Generalisation and Overfitting

Machine learning models trained to identify sadness or fatigue may misinterpret neutral expressions as depressive, especially in culturally reserved individuals. Conversely, those with smiling depression may present cheerful images despite experiencing severe symptoms internally—leading to false negatives.

D. Cultural Sensitivity and Interpretation

Facial expressions and emotional cues are interpreted differently across cultures. A facial gesture that signifies sadness in one region may be considered neutral or polite in another. The current over-reliance on Western datasets highlights the need for culturally inclusive and region-specific training data to improve fairness and effectiveness.

5.2 Future Research Directions

To overcome these limitations and make image-based depression detection more robust and applicable, the following future directions are recommended:

A. Multimodal Fusion Techniques

Combining visual data with textual (captions, comments), audio (tone, pitch), or physiological data (EEG, heart rate) can provide a more complete picture of an individual's emotional state. Hybrid models like CNN-RNN architectures and Vision Transformers with text embeddings are particularly promising in this regard.

B. Explainable AI (XAI)

Interpretability is crucial, especially in healthcare applications. Techniques such as Grad-CAM and attention heatmaps can help visualise the specific facial regions influencing a model's decision, thus improving trust and transparency for both clinicians and users.

C. Real-World Deployment and Validation

Most models are currently evaluated on pre-collected datasets in lab settings. There is a growing need to test these models in real-world environments, such as mobile mental health apps, chatbots, or teletherapy platforms. This includes longitudinal validation to assess consistency over time.

D. Culturally Inclusive Dataset Development

To promote fairness and representation, researchers must create datasets from underrepresented regions, including South Asia, Africa, and Latin America. Institutional support, ethical oversight, and local partnerships are necessary to ensure culturally sensitive data collection and annotation.

E. Policy and Regulatory Compliance

Depression detection models must comply with regulations such as GDPR and HIPAA, especially when deployed in clinical settings. Collaboration with psychiatrists, psychologists, legal experts, and public health policymakers will be essential for successful integration into existing healthcare systems.

6. Conclusion

Image-based depression detection marks a transformative advancement in the landscape of mental health diagnostics. By moving beyond traditional methods such as interviews and self-reported scales, this approach harnesses visual indicators - like facial expressions, emotional tone, and shared images on social media - to provide passive, scalable, and real-time emotional assessment.

This review has explored the evolution of image-based models, highlighting the application of Convolutional Neural Networks (CNNs), attention mechanisms, and emotion recognition algorithms in detecting depression-related cues such as emotional flatness, facial asymmetry, and reduced expression. Datasets like AffectNet, RAF-DB, and FER+ have provided a foundation for benchmarking and training these models.

However, several challenges persist. Ethical concerns, cultural generalisation, labelling ambiguities, and practical deployment barriers must be addressed to make these systems truly impactful. Going forward, integrating multimodal signals, developing culturally diverse datasets, and ensuring clinical alignment will be critical in shaping the next generation of AI-based mental health tools.

References:

[1] M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali, “A CNN-BiLSTM Model for Document-Level Sentiment Analysis,” *Mach. Learn. Knowl. Extr.*, vol. 1, pp. 832–847, 2019, doi: 10.3390/MAKE1030048.

[2] N. J. Prottasha, A. A. Sami, M. Kowsher, S. A. Murad, A. Bairagi, M. Masud, and M. Baz, “Transfer Learning for Sentiment Analysis Using BERT Based Supervised Fine-Tuning,” *Sensors (Basel)*, vol. 22, 2022, doi: 10.3390/s22114157.

[3] M. Hasan, E. A. Rundensteiner, and E. Agu, “Automatic emotion detection in text streams by analyzing Twitter data,” *Int. J. Data Sci. Anal.*, vol. 7, pp. 35–51, 2019, doi: 10.1007/s41060-018-0096-z.

[4] G. Chandrasekaran, T. N. Nguyen, and J. H. D., “Multimodal sentimental analysis for social media applications: A comprehensive review,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 11, 2021, doi: 10.1002/widm.1415.

[5] L. Khan, A. Amjad, K. M. Afaq, and H.-T. Chang, “Deep Sentiment Analysis Using CNN-LSTM Architecture of English and Roman Urdu Text Shared in Social Media,” *Appl. Sci.*, 2022, doi: 10.3390/app12052694.

[6] Semantic Scholar, “A CNN–BiLSTM Model for Document-Level Sentiment Analysis,” [Online]. Available: https://www.semanticscholar.org/paper/8ffefa0acd2d23e9a3b8315aea0aa6d7a3292205.

[7] A. H. Ombabi, W. Ouarda, and A. Alimi, “Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks,” *Soc. Netw. Anal. Min.*, vol. 10, 2020, doi: 10.1007/s13278-020-00668-1.

[8] R. Sawhney, H. Joshi, S. Gandhi, and R. Shah, “A Time-Aware Transformer Based Model for Suicide Ideation Detection on Social Media,” in *Proc. EMNLP*, pp. 7685–7697, 2020, doi: 10.18653/v1/2020.emnlp-main.619.

[9] J. Deng, L. Cheng, and Z. Wang, “Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification,” *Comput. Speech Lang.*, vol. 68, 2021, Art. no. 101182, doi: 10.1016/j.csl.2020.101182.

[10] Q. Cheng, T. M. H. Li, C.-L. Kwok, T. Zhu, and P. Yip, “Assessing Suicide Risk and Emotional Distress in Chinese Social Media: A Text Mining and Machine Learning Study,” *J. Med. Internet Res.*, vol. 19, 2017, doi: 10.2196/jmir.7276.

[11] A. Bello, S.-C. Ng, and M.-F. Leung, “A BERT Framework to Sentiment Analysis of Tweets,” *Sensors (Basel)*, vol. 23, 2023, doi: 10.3390/s23010506.

[12] S. Majid, F. S. Alenezi, S. Masood, M. Ahmad, E. S. Gündüz, and K. Polat, "Attention based CNN model for fire detection and localization in real-world images," *Expert Syst. Appl.*, vol. 189, Art. no. 116114, 2022, doi: 10.1016/j.eswa.2021.116114.

[13] N. V. Babu, E. Grace, M. Kanaga, T. Mandl, S. J. Modha, and E. Grace, "Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review," *SN Comput. Sci.*, vol. 3, 2021, doi: 10.1007/s42979-021-00958-1.

[14] Y. Shen, H. Yang, and L. Lin, "Automatic Depression Detection: an Emotional Audio-Textual Corpus and A Gru/Bilstm-Based Model," in *Proc. ICASSP 2022 - IEEE Int. Conf. Acoustics, Speech and Signal Processing*, pp. 6247–6251, 2022, doi: 10.1109/icassp43922.2022.9746569.

[15] K. Wu et al., "An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system," *Int. Trans. Electr. Energy Syst.*, 2020, doi: 10.1002/2050-7038.12637.

[16] N. J. Carson et al., "Identification of suicidal behavior among psychiatrically hospitalized adolescents using natural language processing and machine learning of electronic health records," *PLoS ONE*, vol. 14, 2019, doi: 10.1371/journal.pone.0211116.

[17] J. Camacho-Collados et al., "TweetNLP: Cutting-Edge Natural Language Processing for Social Media," *arXiv preprint*, arXiv:2206.14774, 2022, doi: 10.48550/arXiv.2206.14774.

[18] A. Hassan, J. Hussain, M. Hussain, M. Sadiq, and S. Lee, "Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression," in *Proc. 2017 Int. Conf. Inf. Commun. Technol. Converg. (ICTC)*, pp. 138–140, 2017, doi: 10.1109/ICTC.2017.8190959.

[19] Y. Zhang and H. Wang, "Multi-head attention-based probabilistic CNN-BiLSTM for day-ahead wind speed forecasting," *Energy*, 2023, doi: 10.1016/j.energy.2023.127865.

[20] L. Mou et al., "Driver stress detection via multimodal fusion using attention-based CNN-LSTM," *Expert Syst. Appl.*, vol. 173, Art. no. 114693, 2021, doi: 10.1016/j.eswa.2021.114693.