RESEARCH PAPER

Explainable AI for Silent Depression and Mental Fatigue Detection: A Comprehensive Review of Facial Micro-Expressions, Eye Dynamics, and Multimodal Behavioral Analysis

Mr. Rajwardhan Suryakant Todkar[1], Dr. Jaydeep B. Patil[2], Dr. Shrikant D. Bhopale[3], Dr. Sangram T. Patil[4]

[1]Research Scholar, D Y Patil Agriculture & Technical University, Talsande, Kolhapur, India.
Email: rajwardhantodkar57@gmail.com
[2]D Y Patil Agriculture & Technical University, Talsande, Kolhapur, India.
Email: er.jaydeep7576@gmail.com / jaydeep.patil@dyp-atu.edu.in
[3]D Y Patil Agriculture & Technical University, Talsande, Kolhapur, India. Email: shrikant.bhopale@dyp-atu.edu.in
[4]D Y Patil Agriculture & Technical University, Talsande, Kolhapur, India. Email: sangrampatil@dyp-atu.org

ABSTRACT

Due to their impact on emotional health, mental functioning, productivity and quality of life, depression and mental fatigue have become public health problems. Early diagnosis and early intervention are important to the successful treatment of these conditions. Thanks to the recent developments in Artificial Intelligence (AI), Computer Vision (CV), Affective Computing and Explainable Artificial Intelligence (XAI), automatic systems have been developed that are able to analyze behavioral signs associated with psychological conditions. This review provides an overview and detailed analysis of recent research on facial expression recognition, facial micro-expression analysis, monitoring of eye dynamics, depression detection, mental fatigue assessment, multimodal learning and explainable AI. The literature is systematically organized according to application area, data sets, methodology and performance. Comparative, dataset, technology and technique-wise analysis is carried out to find out the trend, strengths and limitations of the existing research. The results show that deep learning-based architectures like CNN, LSTM, CNN-LSTM hybrid, GNN, and transformer models can achieve a significant boost in behavioral analysis performance compared to single-modality models, and multimodal models tend to perform better than their single-modality counterparts. The use of techniques such as SHAP, Grad-CAM, and Integrated Gradients further enhances the transparency and trustworthiness of AI models. Previous research tends to focus on the detection of depression or fatigue separately, but not together with say, facial micro-expressions, eye dynamics, and explainability. Considering the deficiencies in the literature, an Explainable Temporal Vision Framework is proposed to simultaneously detect silent depression and mental fatigue using facial micro-expressions and eye behavior. The review highlights new challenges, future research directions and possible avenues to develop reliable, scalable and clinically viable mental health monitoring systems.

Keywords: Depression Detection, Mental Fatigue Assessment, Facial Micro-Expression Recognition, Eye Dynamics Analysis, Explainable Artificial Intelligence (XAI), Multimodal Mental Health Monitoring.

How to cite this article: Todkar RS, Patil JB, Bhopale SD, Patil ST. Explainable AI for Silent Depression and Mental Fatigue Detection: A Comprehensive Review of Facial Micro-Expressions, Eye Dynamics, and Multimodal Behavioral Analysis. Int J Drug Deliv Technol. 2026;16(57s): 844-858. DOI: 10.25258/ijddt.16.57s.89

Source of support: Nil.

Conflict of interest: None.

1. Introduction

Mental health disorders are a significant concern in the world today, and depression and mental fatigue are a significant factor in emotional health, cognition, productivity and quality of life. Depression is defined as a persistent sadness, loss of interest, emotional instability, and lowered cognitive functioning, while mental fatigue involves cognitive exhaustion, diminished attention, impaired decision making and decreased task performance [4], [36]. Traditional assessment methods like clinical interviews and self-report questionnaires are subjective, time-consuming, and rely on the patient's willingness to report emotional problems [4] and are not suitable for early detection and intervention. Therefore, there is a growing need for objective, automatic and non-invasive systems that can continuously monitor mental health conditions [21], [32].

The development of AI, Computer Vision (CV) and Affective Computing (AC) has allowed for the development of intelligent systems for behavioural analysis. There is huge attention on emotion recognition using facial expression, and the performance of deep learning models has been greatly improved [1], [13], [15], [22]. But hidden emotions may be more evident from the micro-expressions that appear on the face, which are good indicators of stress, depression and emotional suppression [19, 24, 35]. Likewise, eye dynamics have been found to be well

correlated with cognitive workload, stress, and mental fatigue, such as blink rate, gaze behavior, eye closure duration, pupil movement, and saccadic eye movements [20, 29, 30, 34].

There have been a few studies that focus on machine learning and deep learning approaches for detecting depression, monitoring fatigue, recognising emotions, and multimodal mental health assessment. Facial behaviour analysis-based methods have shown good results for detecting depression [18, 21, 33] and eye tracking and blink-based approaches have been successful for detecting fatigue [2, 20, 29, 30]. In addition, multimodal systems combining facial, physiological and behavioural data have been shown to perform better than unimodal systems [8], [14], [31], [32]. In the healthcare sector, Explanatory AI (XAI) methods like SHAP and Integrated Gradients have also enhanced transparency and trustworthiness [17, 21, 23, 25].

However, there are still some challenges to be addressed. Most of the existing studies are limited to the detection of depression or the assessment of fatigue separately, and few studies have been conducted on the joint analysis of depression and fatigue. The study of facial micro-expressions and eye dynamics is generally done independently, limiting the full understanding of behavior [19], [24], [30], [35]. Moreover, many deep learning models are inexplicable [23, 25] and multimodal systems often rely on specific sensors which restricts their deployment in the real world [8, 14, 31, 32]. Hence, this review aims to systematically analyze the recent advances in facial behaviour analysis, eye dynamics monitoring, multimodal learning, and explainable AI, and to pinpoint the key research gaps to address the demand for an integrated approach to early detection of silent depression and mental fatigue.

2. Literature Review

In recent years, AI and Computer Vision technologies have made a substantial leap in the field of automated mental health assessment, with behavioral analysis playing a crucial role. Facial expressions, facial micro-expressions, eye dynamics, multimodal learning, and explainable AI for detecting depression, emotional states, and mental fatigue are explored. This section reviews the literature and discusses the advantages and disadvantages of the approaches that are currently available.

A. Computer Vision Based Mental Health Analysis

Facial expressions and micro-expressions have been used in several studies to recognize emotions and assess mental health. Muniasamy et al. [1] presented a CNN-LSTM model for emotion recognition in real-time while Liu et al. [3] compared the different deep learning architectures and found that the VGG16 model performed best. Sharma et al. [19] further enhanced micro-expression recognition by introducing facial action units, attention mechanisms, graph networks and motion magnification techniques. Zhang et al. [46] also added facial motion and shape to enhance micro-expression recognition, while Ren et al. [55] and Wu et al. [57] introduced facial motion and shape, respectively. Furthermore, Surek et al. [5] and Shana and Christopher [11] showed the usefulness of deep learning-based behaviors analysis of video sequences. Although some studies demonstrated promising results, the majority of studies concentrated primarily on the facial behavior and did not incorporate eye dynamics, fatigue assessment, and explainable AI.

B. Detection of depression from facial videos and behavioural features

Facial videos have been studied extensively for depression diagnosis. Rumahorbo et al. [18] proposed an RNN model to estimate the severity of depression based on the temporal features of the face. Mahayossanunt et al. [21] designed an explainable LSTM model with attention mechanisms and Integrated Gradients and Pan et al. [52] developed STA-DRN for learning depression-related facial dynamics. Lage Cañellas et al. [33] used ResNet-50 based facial texture analysis and emphasized the need for pre-processing. These techniques have been promising, but the use of eye fatigue indicators and multimodal behavioral cues is not fully integrated.

C. The Detection of Eye Movement and Mental Fatigue.

The eye dynamics has been proven to be effective in the non-invasive assessment of mental fatigue. Eye-tracking and facial video were adopted by Hamoud et al. [2] for fatigue detection and Kuwahara et al. [20] introduced the Eye Aspect Ratio Mapping framework for blink-based fatigue estimation. Yamada et al. [30] used eye-tracking data and automated feature selection to achieve high fatigue detection accuracy. Psyllos et al. [42] and Kremer et al. [51] also showed that fixation patterns, pupil dilation, blink rate and PERCLOS could be used for workload estimation, and Marquart et al. [44] showed that pupil dilation could be used for workload estimation. Most studies, however, do not incorporate eye behavior into depression-related indicators and facial micro-expressions.

D. Multimodal Emotion Recognition Based on EEG and Eye Signals.

In order to enhance the ability of emotion recognition, researchers have turned to a multimodal approach that incorporates both physiological and behavioral

signals. To capture the multimodal information of EEG and eye movement, Fu et al. [14, 62] proposed a Multimodal Feature Fusion Neural Network with attention mechanism. In order to incorporate multi-band EEG features and spatio-temporal attention mechanisms, Fang et al. [60] developed a multi-band EEG features based spatio-temporal attention mechanism (MB-MSTFNet). Hosseini et al. [8] created the EmpathicSchool dataset of synchronized facial and physiological data, and Song et al. [32] developed a personalized multimodal mental health monitoring framework. Multimodal fusion has been shown to enhance recognition but most of the methods rely on physiological sensors and offer only limited explainability.

E. Eye Tracking and Human Computer Interaction Systems

There have been multiple studies to integrate eye tracking with facial expression in intelligent interaction systems. Sun and Cai [10] presented a multimodal eye-control system based on gaze intention analysis and facial expression recognition. Menekse Dalveren and Cagiltay [45] used eye-tracking to assess the cognitive workload during neurosurgical simulation training, and Sampei et al. [34] designed a wearable eye-monitoring system to estimate fatigue. Furthermore, Tan et al. [58] and Ma et al. [59] introduced multimodal deep learning architectures for healthcare and behaviour pattern recognition, respectively. But the focus of these systems is mainly on interaction and workload analysis, not mental health monitoring.

F. Explainable Artificial Intelligence for Mental Health Applications

In healthcare, Explainable AI plays a significant role in enhancing transparency and trust. Mahayossanunt et al. [21] used Integrated Gradients to interpret depression predictions based on facial behaviours. Atlam et al. [23] and Monteiro et al. [17] showed that SHAP can be used to interpret machine learning and deep learning models.Atlam et al. [23] and Monteiro et al. [17] proved the effectiveness of SHAP for interpreting machine learning and deep learning models.Atlam et al. [23] and Monteiro et al. [17] proved that SHAP can be used to interpret machine learning and deep learning models. Joyce et al. [25] emphasized explainability in psychiatric AI by introducing the TIFU framework, and Leong et al. [56] pointed out the importance of transparency and interpretability in healthcare decision support systems. Although these developments have been made, explainable AI is still not explored much in video-based depression and fatigue detection.

G. Survey and Review Studies

There are a number of review studies that have compiled progress in the field of mental health monitoring and behavioral analysis. Ghafarfaraji et al. [4] summarized facial and micro-expression recognition techniques for mental disorder diagnosis and Shuai et al. [6] and Li et al. [24] and Zhao et al. [35] reviewed the micro-expression datasets, technologies and deep learning approaches. Poojari et al. [4] discussed the use of AI in mental healthcare and Kunasegaran et al. [36] and Sharma et al. [39] reviewed mental fatigue assessment and sensing technologies. All of these studies highlighted the issues of small data sets, weak generalization, lack of explainability, and lack of integration of facial and eye behavior analysis. Overall, the current research indicates a movement towards deep learning, multimodal learning, and explainable AI, but there are challenges with integration, data availability, and deployment in the real world that need to be addressed. Overall, the literature shows a clear trend towards deep learning, multimodal learning and mental health monitoring systems based on explainable AI. Although major strides have been made, integration, data availability, explainability and deployment issues are not resolved. Thus, a thorough comparative study is required to establish the trends in current research, limitations of existing work and future directions for the development of reliable and scalable systems for the detection of silent depression and mental fatigue.

3. Comparative Analysis of Existing Studies

The comparative analysis shows that there are considerable improvements in emotion recognition, depression detection, mental fatigue assessment, multimodal learning, and explainable AI. Recognitions accuracy has significantly improved in different domains using deep learning models like CNNs, LSTMs, GRUs, Graph Convolutional Networks, and Transformers. Rumahorbo et al. [18] and Mahayossanunt et al. [21] showed that facial behavioral features are effective for the detection of depression. Lage Cañellas et al. [33] proved facial behavioral features for depression detection. Likewise, eye dynamics-based fatigue assessment methods, such as blink rate, gaze patterns, and eye movements, were found to be high performing, with some studies reporting an accuracy of over 90% [20], [29], [30]. Table 1 provides a comparative analysis of representative studies that have been conducted within the following categories: Depression detection, Fatigue assessment, Multimodal learning, and Explainable AI.

Table 1. Comparative Analysis of Existing Studies

Ref.	Year	Application	Dataset	Methodology	Performance / Limitations
[1]	2026	Emotion Recognition	FER-2013, AffectNet, RAF-DB	Attention CNN-LSTM	High accuracy; No depression/fatigue analysis
[2]	2025	Mental Fatigue Detection	Eye Tracking Dataset	TabNet + Vision Model	82% Accuracy; No micro-expression analysis
[18]	2025	Depression Detection	Facial Video Interviews	RNN, LSTM, GRU	MAE=5.04, RMSE=6.03; No explainability
[19]	2024	Mental Health Assessment	CASME II, SAMM	CNN + AU Analysis	92.99% Accuracy; No eye dynamics
[20]	2022	Eye Fatigue Detection	Blink Video Dataset	EARM	High accuracy; No depression assessment
[21]	2023	Explainable Depression Detection	Interview Videos	LSTM + Integrated Gradients	91.67% Accuracy; No micro-expression analysis
[14]	2023	Multimodal Emotion Recognition	SEED-IV	MFFNN + Attention Fusion	87.32% Accuracy; Requires EEG
[29]	2023	Fatigue Detection	Operator Dataset	TICC + KNN	91.83% Accuracy; No emotional analysis
[30]	2018	Mental Fatigue Detection	Eye Tracking Dataset	Feature Selection + ML	91.0% Accuracy; No multimodal info
[31]	2023	Mental Fatigue Detection	Physiological Signals	Random Forest	96% Accuracy; Requires sensors
[32]	2024	Mental Health Monitoring	Multimodal Dataset	Transformer-Based Learning	CCC=0.503; No micro-expressions
[33]	2023	Depression Recognition	AVEC2013/2014	ResNet-50	MSE=5.50; No explainability
[23]	2025	Explainable Mental Health	Mental Health Dataset	SHAP + Ensemble Learning	100% Accuracy; No visual analysis
[35]	2023	Micro-Expression Review	Multiple Datasets	Survey	Comprehensive review; No framework
[40]	2022	Visual Fatigue Detection	VDT Dataset	Blink Feature Analysis	Fatigue correlation; No depression analysis

The comparative analysis reveals a growing trend in the use of multimodal learning and explainable AI for mental health monitoring. Most of the existing systems are based on physiological sensors and wearable devices, and multimodal approaches tend to perform better when they are able to combine complementary behavioral information, but are not easily deployable to large scales. Furthermore, most of the studies only consider the detection of depression or the assessment of mental fatigue separately, and very few studies combine facial micro-expressions, eye dynamics, and explainability in a comprehensive framework. These restrictions highlight the importance of developing an explainable multimodal

system that can detect silent depression and mental fatigue simultaneously by analysing non-invasive behavior.

4. Research Trend Analysis

4.1 Publication Trend Analysis

Recent studies analysis shows that the research on the detection of depression, mental fatigue, facial micro-expression recognition, and explainable artificial intelligence has been significantly increased.

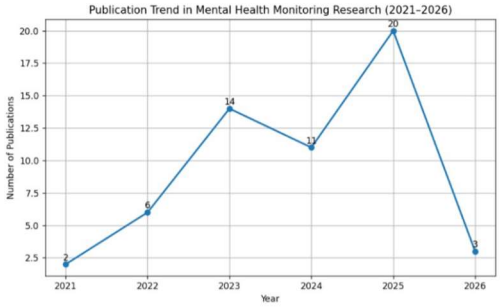

Fig. 1. Publication Trend of Studies(2021–2026).

The number of published works as in fig. 1 has grown significantly between 2021 and 2026, as the research field of deep learning, computer vision, wearable sensing technologies, and mental healthcare applications has progressed. Previous research mainly centered on the traditional facial expression recognition and physiological signal analysis. Recent studies have focused on multimodal learning, temporal behavioural analysis, eye dynamics monitoring, and explainable AI techniques to enhance the accuracy and transparency of predictions.

4.2 Taxonomy of Existing Studies

The reviewed studies can be grouped into six main areas: facial expression analysis, facial micro-expression recognition, depression detection, eye dynamics based fatigue assessment, multimodal mental health monitoring and explainable artificial intelligence. The taxonomy of the existing studies on depression detection and mental fatigue assessment can be seen in Figure 2.

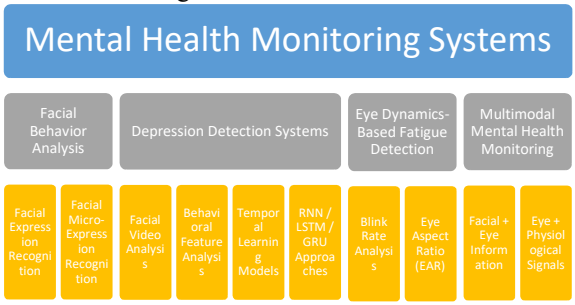

Fig. 2. Taxonomy of Existing Studies in Mental Health Monitoring

As can be seen in fig. 2, the majority of previous research has concentrated on a single modality, facial expressions, facial micro-expressions or eye dynamics. A variety of deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, and attention-based models, are currently employed in emotion recognition, depression detection, and fatigue assessment. There has been a growing number of studies in recent years that use multimodal approaches involving facial behavior, eye movements, physiological signals and speech features to enhance performance. There are many, however, that require specialized sensors and are not explainable. Thus, further research is needed on the explanation of multimodal systems that collectively process facial micro-expressions and eye movements for accurate and reliable assessment of mental health.

5. Dataset and Technology Analysis

5.1 Dataset Analysis

Datasets are essential for development and evaluation of the automated depression detection, mental fatigue assessment, emotion recognition, and facial micro-expression analysis systems. The performance and generalization ability of machine learning and deep learning models are directly affected by the quality, size, diversity and modality of the datasets. The reviewed studies utilize a variety of benchmark datasets ranging from facial image datasets to multimodal datasets containing video, eye-tracking, and physiological signals. Table 2 presents the most common datasets in the literature.

Table 2. Dataset Analysis

Dataset	Year	Application	Samples	Modality
CASME II	2014	Micro-Expression Recognition	247	Video
SAMM	2018	Micro-Expression Recognition	159	Video
FER2013	2013	Emotion Recognition	35,887	Image
AffectNet	2017	Emotion Recognition	1M+	Image
AVEC2013	2013	Depression Detection	150+	Video

SEED-IV	2018	Emotion Recognition	EEG + Eye Signals	Multimodal

According to the dataset analysis, CASME II and SAMM are the most popular datasets for facial micro-expression recognition, and FER2013 and AffectNet are the popular datasets for emotion recognition. AVEC is broadly used for detecting depression, while SEED-IV is used for multimodal emotion analysis. The datasets have enabled a vast amount of research on behavioral and mental health, but many suffer from small sample sizes, class imbalance, and lack of multimodal annotations. Moreover, most of the available datasets only deal with single tasks but not with integrated data of depression, mental fatigue, facial micro-expressions and eye dynamics. The restrictions demonstrate the need for larger and more extensive multimodal data for future systems of mental health assessment.

Traditional machine learning methods have given way to sophisticated deep learning and explainable AI systems in the world of mental health monitoring technologies. Recent studies are increasingly using CNNs, LSTMs, GRUs, Graph Neural Networks and Transformer-based models to decipher facial behaviour, eye movement and emotional patterns. CNNs are very useful for extracting spatial features, and recurrent models for temporal behavioral changes. Recently, Transformers and GNNs have shown great power in capturing complex relationships and long-range dependencies. Furthermore, explainable AI methods like SHAP, Grad-CAM, and Integrated Gradients are being incorporated to enhance transparency and trust in models for clinical use. The most significant technologies, their applications, benefits, and drawbacks are summarized in Table 3 for the more recent mental health monitoring studies.

Table 3. Technology Analysis

Category	Techniques Used	Application Area	Advantages	Limitations
Traditional Machine Learning	Random Forest, SVM, KNN	Fatigue Detection, Classification	Fast training and interpretable	Limited feature learning capability
CNN-Based Models	VGG16, VGG19, ResNet50, DenseNet, EfficientNet	Facial Expression Recognition	Strong spatial feature extraction	Limited temporal understanding
Recurrent Models	RNN, LSTM, GRU	Depression Detection, Sequence Analysis	Temporal pattern learning	Higher computational complexity
Hybrid Deep Learning	CNN-LSTM, CNN-GRU	Emotion and Behavioral Analysis	Spatial and temporal learning	Increased model complexity
Graph Neural Networks	GCN, Graph Attention Networks	Facial Landmark Analysis	Captures facial relationships effectively	Complex architecture
Attention-Based Models	Self-Attention, Attention-CNN-LSTM	Emotion Recognition	Focuses on important regions	Computationally expensive
Transformer Models	Vision Transformer (ViT), Video Transformer	Mental Health Monitoring	Long-range dependency learning	Requires large datasets
Multimodal Fusion Models	MFFNN, Attention Fusion Networks	Emotion and Mental Health Analysis	Improved robustness and accuracy	Data synchronization challenges
Explainable AI Methods	SHAP, Grad-CAM, Integrated Gradients	Model Interpretation	Improves transparency and trust	Additional computational overhead

The field has progressed from classic machine learning to explainable multimodal architectures of deep learning, as illustrated in Fig. 3.

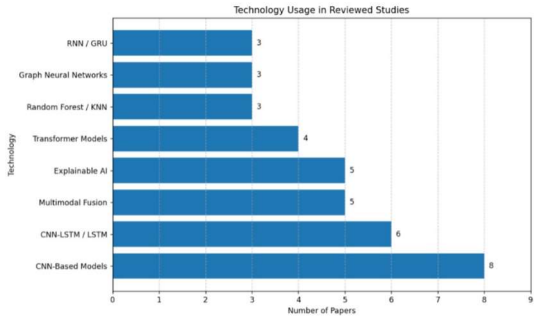

Fig. 3. Technology Evolution Timeline

The technology analysis reveals a clear trend towards deep learning and multimodal approaches, a contrast with traditional machine learning methods. Previous approaches like Random Forest, SVM, and KNN were based on manually engineered features and lacked the capacity to represent intricate behavioral patterns. CNNs were used to enhance facial behavior analysis by using automatic facial feature extraction, whereas the LSTMs and GRUs were used to learn temporal behavior. Recently, Transformers, Graph Neural Networks and multimodal fusion methods have been used to further boost performance using a variety of behavioral cues. In addition, Explainable AI techniques are increasingly being adopted to improve transparency and clinical trust in mental health monitoring systems. The main results are CNN is the most popular architecture for facial behavior analysis and LSTM and GRU models are good at learning the temporal behavior pattern. Recent developments in sequence modeling are based on transformer architectures. Moreover, multimodal fusion techniques tend to outperform single-modality techniques, and Explainable AI has emerged as a critical feature in healthcare applications to ensure transparency and trustworthiness. Future systems are expected to combine multimodal learning with explainable Transformer-based architectures for more accurate and trustworthy mental health assessment.

6. Technique-wise Performance Analysis

The analysed studies reveal a significant transition from conventional machine learning techniques to more sophisticated deep learning and multimodal approaches in mental health monitoring. The handcrafted features used by conventional techniques like KNN, Random Forest, and SVM were less effective at capturing complex behavioral patterns. However, CNNs showed better performance in facial expression and micro-expression recognition, and CNN-LSTM models showed better performance by learning spatial and temporal information [19], [46]. Recent research on depression detection experiments showed that RNNs, LSTMs and GRUs are effective in modeling temporal behavioral patterns [18, 21]. Likewise, eye dynamics-based approaches, such as blink rate, gaze behavior, EARM, and PERCLOS, have been shown to be accurate in mental fatigue assessment [20, 29, 30, 31]. More recent works have focused on the benefits of multimodal methods that integrate facial expression, eye-gaze, physiological data and speech [14], [32] which typically achieve better performance than unimodal methods. Furthermore, Explainable AI (XAI) methods like SHAP and Integrated Gradients are being increasingly used to enhance transparency and trust in health care systems [21], [23], [25]. Overall, multimodal learning, hybrid deep learning models, and explainable AI seem to offer the best potential avenues for future mental health assessment systems. Table 4 shows the comparison of the major approaches reported in the reviewed studies using a technique-wise approach.

Table 4. Performance Comparison of Different Techniques

Technique	Studies	Application Area	Best Reported Performance	Advantages	Limitations
CNN	[3], [19], [22]	Facial Expression, Micro-expression	92.99 %	Strong spatial feature extraction	Limited temporal modeling
RNN	[18]	Depression Detection	RMSE = 6.03	Captures temporal patterns	Vanishing gradient problem
LSTM	[1], [18], [21]	Emotion & Depression Detection	91.67 %	Effective sequence learning	Computationally expensive
GRU	[18]	Depression Detection	Comparable to LSTM	Faster training	Less expressive than LSTM

CNN-LSTM	[1], [46]	Video-based Emotion Analysis	>90%	Spatial + Temporal learning	Higher complexity
ResNet	[33]	Depression Recognition	MSE = 5.50	Deep feature extraction	Large computational cost
Transformer	[32]	Mental Health Monitoring	CCC = 0.503	Long-range dependency learning	Data hungry
Graph Neural Network	[57]	Micro-expression Recognition	High recognition rate	Captures facial relationships	Complex architecture
Random Forest	[31]	Fatigue Detection	96%	Fast and interpretable	Limited deep representation
TabNet	[2]	Mental Fatigue Detection	82%	Feature selection capability	Lower accuracy
SHAP-Based Models	[23]	Explainable Mental Health	100%*	Highly interpretable	Dataset-specific

Figure 4 shows the development of machine learning techniques in mental health monitoring systems, with traditional machine learning methods giving way to deep learning, multimodal learning, Transformer-based methods, and explainable AI frameworks.

Fig. 4. Evolution of Techniques

The analysis of the performances reveals a significant transition from traditional machine learning approaches to deep learning and transformer-based architectures. CNNs achieved better recognition of facial expression and micro-expression, and CNN-LSTM, RNN, LSTM and GRU models performed better in temporal behavioral analysis for depression detection, respectively [18, 19, 21, 46]. Recently, Transformer-based models have been found to have great potential for modeling long-range dependency and multimodal interaction [32]. The multimodal approaches usually yield the best performance when combining facial behavior, eye dynamics and physiological signals, but are typically dependent on special hardware [14], [31], [32]. Furthermore, the use of Explainable AI methods, like SHAP and Integrated Gradients, is growing to enhance transparency and trustworthiness in mental health applications [21, 23, 25].

7. Results and Discussion

7.1 Overall Findings

The review highlights advancements in mental health monitoring using AI technology. The use of deep learning models like CNNs, LSTMs, CNN-LSTM, GNNs, and Transformers has enhanced emotion, depression, and mental fatigue detection, with numerous studies achieving accuracies exceeding 90%.

7.2 Performance Comparison Discussion

The sensor based approach is the most accurate but is only possible with specialist equipment, which is limiting for practical use. However, camera-based systems that utilize facial expressions, micro-expressions and eye movements offer a more scalable and non-invasive solution for real-world applications.

7.3 Dataset Analysis Discussion

Some of the most popular databases for micro-expression recognition are CASME II and SAMM, and for depression and emotion analysis are AVEC, FER2013, and AffectNet. But many of the datasets have small sample sizes, class imbalance and lack of combined depression and fatigue annotations.

7.4 Technique Analysis Discussion

The research has progressed from the traditional machine learning to deep learning and multimodal approaches. CNNs are well suited for spatial feature extraction, while LSTM, GRU, and Transformer models are capable of capturing temporal behavioral patterns. Most multimodal frameworks work better because they use several information sources.

7.5 Explainability Discussion

SHAP, Grad-CAM, and Integrated Grad-Ints are examples of explainable AI techniques that enhance the transparency and trust within healthcare systems. But, they have not been widely used in the field of depression and mental fatigue detection, which limits their clinical acceptance.

8. Research Challenges

While AI-driven mental health monitoring has made significant strides recently, there are still several hurdles that need to be overcome to create systems that are reliable and clinically applicable. Some of the challenges are limited datasets, privacy concerns, explainability, real-time deployment, and model generalization.

8.4 Real-Time Deployment

Analysis of facial videos, eye movements, and behavioral patterns in real time demands a significant amount of computational resources. Lightweight and efficient models are needed for deployment on mobile and edge devices.

8.5 Explainability and Clinical Trust

Most deep learning models are black boxes, which can make their predictions hard to understand. While techniques like SHAP, Grad-CAM, and IG have enhanced the transparency of XAI, these methods are not widely used.

8.6 Generalization and Robustness

Models can be very successful on a set of data but not so successful when it comes to real life because lighting, camera quality, facial appearance and user demographics vary.

8.7 Multimodal Integration Complexity

The fusion of these different data sources is still difficult and can benefit the performance when facial behavior, eye dynamics, speech and physiological signals are combined.

8.8 Clinical Validation

Most systems are tested on research data sets and not on actual clinical data sets. Reliability and usefulness for large-scale clinical studies need to be established.

To ensure reliable and real-world deployment, future mental health monitoring systems must tackle critical issues of data availability, privacy, explainability, scalability, and clinical validation.

9. Research Gaps

The literature surveyed shows that the field of facial expression recognition, micro-expression analysis, depression detection, eye tracking based fatigue assessment and multimodal emotion recognition has come a long way. Most of the existing studies however only consider a single modality and not combine facial micro-expressions and eye dynamics into a single framework. Most of the depression detection systems do not include mental fatigue assessment, and many multimodal systems require EEG signals or wearable sensors, which are not feasible in real-world applications. Moreover, explainable AI methods are not well studied in video-based mental health tracking systems. Thus, it is necessary to have a non-invasive, explainable and multimodal approach to detect silent depression and mental fatigue from facial micro-expressions and eye dynamics.

10. Proposed Conceptual Framework

The proposed framework is introduced in this section.This section provides an overview of the proposed framework.

In order to address the aforementioned research gaps, a novel Explainable Temporal Vision Framework is proposed for the simultaneous detection of silent depression and mental fatigue from facial micro-expressions and eye dynamics as illustrated in Fig. 5. The framework offers a non-invasive, automatic, and explainable mental health assessment solution. It combines facial micro-expression analysis, eye behavior monitoring, temporal deep learning, multimodal feature fusion, and Explainable AI in a single architecture. Emotional and fatigue related features are extracted from facial video sequence, and depression and fatigue patterns are extracted from fused behavioral features. Lastly, Explainable AI methods produce clear and understandable forecasts for doctors and end-users.

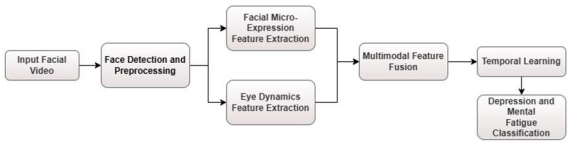

Fig. 5. Proposed Conceptual Framework

10.2 Video Acquisition and Preprocessing

The proposed framework starts by capturing facial video sequences from a regular camera or webcam. The captured videos offer insights into behavioral patterns, such as facial expressions, eye movement, blink frequency, gaze patterns and subtle emotional cues. The framework is non-invasive, meaning that it does not require any special physiological monitoring device or wearable sensor. After acquisition, the video frames are preprocessed, including face detection, face alignment, illumination normalization, noise removal, and facial landmark localization. These operations help to ensure the accuracy of the data and facilitate

the extraction of behavioral characteristics in the following steps.

10.3 Facial Micro-Expression Analysis

Facial micro-expressions are involuntary brief facial movements that express hidden emotions. Micro-expression analysis can be useful in detecting underlying psychological states because people who are depressed tend to hide their emotions. The framework includes facial action unit (FAU) analysis, optical flow patterns, facial landmark motion, muscle activation patterns, and temporal motion descriptors to capture subtle differences in emotion. The discriminative facial representation and behavioral patterns associated with depression can be learned using deep learning models like CNNs.

10.4 Eye Dynamics Analysis

Eye-based behavioral patterns have significant correlations with mental fatigue, cognitive workload, and emotion. Thus, the proposed framework extracts multiple eye dynamics features such as blink frequency, blink duration, Eye Aspect Ratio (EAR), Percentage of Eye Closure (PERCLOS), gaze direction, saccadic eye movements and eye closure patterns. Changes in these parameters offer valuable insights into attention levels, fatigue progression, stress, and cognitive exhaustion. Eye behavior analysis is incorporated into the framework to improve its ability to identify mental fatigue along with depression symptoms.

10.5 Multimodal Feature Fusion

A multimodal fusion mechanism is used to combine the extracted facial micro-expression features and eye dynamics features. The aim of this step is to merge complementary behavioral information from both modalities into a single representation. Different fusion strategies including early feature fusion, late feature fusion, attention-based fusion, and transformer-based fusion can be used to achieve the best information utilization. Multimodal integration allows for more powerful and precise behavioral analysis than can be achieved with single modality systems.

10.6 Temporal Behavioral Learning

10.6 Temporal Behavioral Learning

Depression and mental fatigue are chronic illnesses that take time to develop. Thus, temporal analysis is necessary to comprehend behavior throughout video sequences. The fused behavioral features are then processed with temporal learning architectures like LSTM, Bi-LSTM, GRU, Temporal CNN, Vision Transformer or Video Transformer models. These models are able to capture long-term dependencies and temporal relationships between the behavioral cues, which allows for the accurate identification of sustained depression and fatigue-related patterns.

Depression and Mental Fatigue Classification is a classification for depression and mental fatigue.

The learned temporal representations are subsequently used for behavioral condition classification. The framework is used to predict the presence and severity of depression and mental fatigue based on the extracted patterns. Depending on the requirements for the application, the classification stage can be backed up by both a binary and a multi-class prediction scheme. This stage is the fundamental decision making part of the framework.

10.8 Explainable AI-Based Interpretation

The framework integrates explainable AI techniques to enhance transparency and clinical trustworthiness. Various methods like SHAP, Grad-CAM, Integrated Gradients and Attention Visualization are used to determine the most influential facial regions, eye movements and temporal features of behavior that lead to the final prediction. These explanations help to understand the model's reasoning process and are useful for clinical decision making.

10.9 Decision Support and Mental Health Assessment

The last stage produces outputs that are interpretable and behavioral assessment reports. Depression risk scores, mental fatigue scores, attention heatmaps, feature importance visualizations and recommendation alerts can be given by the system. These outputs can support the early mental health assessment and intervention of clinicians, psychologists, health workers and people. The proposed framework integrates predictive performance and interpretability, enabling reliable and transparent mental health monitoring.

11. Future Research Directions

Future studies should aim to create more precise, explainable, scalable, and clinically valid mental health monitoring systems. Key directions include:

11.1 Integration of Facial Micro-Expressions and Eye Dynamics

Facial micro-expressions and eye dynamics can be combined for a more complete understanding of emotional and cognitive states, which can help to better detect depression and fatigue.

11.2 Large-Scale Multimodal Datasets

A variety of multimodal data sets with facial videos, gaze data, depression scores and fatigue labels are

required to enable comprehensive model development and assessment.

11.3 Explainable AI
The potential inclusion of XAI methods like SHAP, Grad-CAM, and Integrated Gradients in future systems to enhance transparency and clinical trust.

11.4 Transformer-Based Models
The vision Transformers and multimodal Transformer architectures are promising for capturing complex temporal behavioral patterns.

11.5 Privacy-Preserving Frameworks
To safeguard sensitive mental health data, techniques like federated learning, differential privacy, and secure aggregation should be incorporated.

11.6 Real-Time Deployment
Lightweight models, model compression, pruning, and edge AI solutions are needed for efficient real-time monitoring on mobile and wearable devices.

11.7 Personalized Assessment
Future systems should be flexible enough to learn and adapt to the individual's behavior to increase the accuracy of prediction and decrease the number of false alarms.

11.8 Clinical Validation
For the practical adoption and healthcare deployment, large-scale clinical studies and integration with telemedicine platforms are crucial.

Multimodal learning, explainability, privacy preservation, personalization, and clinical validation are key elements that should be integrated within future mental health monitoring systems to ensure reliable and real-world depression and fatigue assessment.

12. Threats to Validity
This review offers a thorough overview of the latest progress in the fields of depression detection, mental fatigue assessment, facial micro-expression analysis, eye dynamics monitoring, multimodal learning, and explainable artificial intelligence, but there are several factors that could impact the validity of the findings and conclusions.

12.1 Dataset Heterogeneity
The studies reviewed employ a broad range of datasets with different characteristics, such as sample size, demographic makeup, acquisition conditions, annotation methods, and evaluation protocols. These variations can make it difficult to compare results across different studies and can affect the results reported.

12.2 Evaluation Metric Variability
Various evaluation measures like accuracy, precision, recall, F1 score, RMSE, MAE, MSE and Concordance Correlation Coefficient (CCC) are used in different studies. The metrics measure different aspects of the model's performance; direct quantitative comparisons may not be appropriate and should be used with caution.

12.3 Publication Bias
Most of the reviewed literature is from peer-reviewed journals and conference proceedings. Studies with positive or significant results are more likely to be published than studies with negative or inconclusive results. Thus, the overall impression of the effectiveness of the existing approaches could be affected by publication bias.

12.4 Rapid Technological Evolution
AI and Computer vision technologies are still evolving at a rapid pace. New architectures, datasets, explainability methods and multimodal learning strategies could develop following this review. Thus, some of the findings and trends in this paper may need to be updated in the future as the field progresses.

12.5 Generalization of Reported Results
Numerous studies reviewed assess the models on limited data sets that are gathered under controlled conditions. So reported performance is not necessarily representative of actual deployment situations with diverse populations, environmental conditions, and behavioral variations. Thus, care should be taken in extrapolating experimental data to practical use.

12.6 Scope Limitation
The main areas of focus in this review are facial micro-expression, eye movements and dynamics, depression detection, mental fatigue assessment, multimodal learning and explainable AI. Other related modalities, such as, speech analysis, physiological signal processing, social media behaviour analysis, and clinical biomarkers were not covered in detail. Consequently, the review may not be comprehensive of all mental health monitoring strategies.

13. Conclusion
Depression and mental fatigue are important mental health issues that need timely and accurate identification. The recent developments in facial expression recognition, facial micro-expression analysis, eye dynamics monitoring, depression detection, mental fatigue assessment, multimodal

learning and explainable artificial intelligence were analyzed in this review. The results show that deep learning models like CNNs, LSTMs, CNN-LSTM, GNN, and Transformers have shown remarkable improvements in recognition performance, and multimodal models generally outperform single-modality models. Explainable AI techniques further contribute to transparency and trust in healthcare applications. The review also highlighted several research gaps, such as the lack of integration of depression and fatigue detection, few multimodal datasets, explainability problems, and deployment issues in real-world scenarios. To overcome these drawbacks, an Explainable Temporal Vision Framework is proposed, which integrates facial micro-expression, eye movement, temporal behavioral learning, multimodal feature fusion, and explainable AI. Further studies are needed to create scalable, privacy-preserving, and clinically valid solutions to detect early signs of silent depression and mental fatigue.

REFERNECES

[1] A. Muniasamy, R. Abbas, G. Ybytayeva, A. Alasmari, N. Aldahwan, and H. K. Alkahtani, "Attention-enhanced CNN-LSTM framework for real-time video-based emotion recognition," The Visual Computer, vol. 42, p. 229, 2026, doi: 10.1007/s00371-026-04396-z.

[2] B. Hamoud, W. Othman, N. Shilov, and A. Kashevnik, "Deep-Learning-Based Human Activity Recognition: Eye-Tracking and Video Data for Mental Fatigue Assessment," Electronics, vol. 14, no. 19, p. 3789, 2025, doi: 10.3390/electronics14193789.

[3] R. Liu, "Comparison of CNN-Based Models in Facial Micro-Expression Classification," Highlights in Science, Engineering and Technology, vol. 124, pp. 368–376, 2025.

[4] S. Ghafarfaraji, "AI-based recognition of facial and micro-expressions for the diagnosis of mental and neurological disorders: a systematic review," BMC Psychiatry, vol. 26, p. 78, 2026, doi: 10.1186/s12888-025-07739-7.

[5] G. A. S. Surek, L. O. Seman, S. F. Stefenon, V. C. Mariani, and L. S. Coelho, "Video-Based Human Activity Recognition Using Deep Learning Approaches," Sensors, vol. 23, no. 14, p. 6384, 2023.DOI: 10.3390/s23146384.

[6] V. J. Poojari, "AI for Mental Health Diagnosis," 2025. [Online]. Available as preprint/manuscript.

[7] T. Shuai, S. Beng, F. B. Khalid, and R. W. O. K. Rahmat, "Advances in Facial Micro-Expression Detection and Recognition: A Comprehensive Review," Information, vol. 16, no. 10, p. 876, 2025, doi: 10.3390/info16100876.

[8] R. Jayaswal, M. A. Ansari, M. Dixit, D. K. Singh, and S. Ahmad, "Advances in facial expression recognition technologies for emotion analysis," Discover Computing, vol. 28, p. 203, 2025, doi: 10.1007/s10791-025-09699-8.

[9] M. Hosseini, F. Sohrab, R. Gottumukkala, R. T. Bhupatiraju, S. Katragadda, J. Raitoharju, A. Iosifidis, and M. Gabbouj, "A multimodal stress detection dataset with facial expressions and physiological signals," Scientific Data, vol. 12, p. 1844, 2025, doi: 10.1038/s41597-025-05812-0.

[10] X. Sun and Z. Cai, "Research on an Eye Control Method Based on the Fusion of Facial Expression and Gaze Intention Recognition," Applied Sciences, vol. 14, no. 22, p. 10520, 2024, doi: 10.3390/app142210520.

[11] L. Shana and C. Seldev Christopher, "A Deep Learning Behavior Analysis Model for Efficient Video Surveillance Using Multi Pose Features," Ain Shams Engineering Journal, vol. 16, p. 103245, 2025.DOI: 10.1016/j.asej.2024.103245.

[12] J. Bhanbhro, A. A. Memon, B. Lal, S. Talpur, and M. Memon, "Speech Emotion Recognition: Comparative Analysis of CNN-LSTM and Attention-Enhanced CNN-LSTM Models," Signals, vol. 6, no. 2, p. 22, 2025, doi: 10.3390/signals6020022.

[13] H. Kim, J.-H. Lee, and B. C. Ko, "Facial Expression Recognition in the Wild Using Face Graph and Attention," IEEE Access, vol. 11, pp. 59774–59789, 2023, doi: 10.1109/ACCESS.2023.3286547.

[14] B. Fu, C. Gu, M. Fu, Y. Xia, and Y. Liu, "A Novel Feature Fusion Network for Multimodal Emotion Recognition from EEG and Eye Movement Signals," Frontiers in Neuroscience, vol. 17, p. 1234162, 2023, doi: 10.3389/fnins.2023.1234162.

[15] B. Chandra, G. O. Kindy, K. S. Gunawan, G. P. Satria, I. S. Edbert, and D. Suhartono, "Deep Learning Architectures for Facial Emotion Analysis," in Proc. International Conference on Sustainable Information Engineering and Technology (SIET), Bali, Indonesia, 2023, pp. 1–9, doi: 10.1145/3626641.3627492.

[16] S. Attrah, "Emotion Estimation from Video Footage with LSTM," Commun. Math. Biol. Neurosci., 2025. Available: https://huggingface.co/papers/2501.13432.

[17] G. O. A. Monteiro, G. S. Difante, D. B. Montagner, V. P. B. Euclides, M. Castro, J. G. Rodrigues, M. G. Pereira, L. C. V. Ítavo, J. A. Campos, A. B. da Costa, et al., "Interpreting Machine Learning Models with SHAP Values: Application to Crude Protein Prediction in Tamani Grass Pastures," Agronomy, vol. 15, no. 12, p. 2780, 2025, doi: 10.3390/agronomy15122780.

[18] B. N. Rumahorbo, G. N. Elwirehardja, and B. Pardamean, "Depression Severity Detection from Facial Expressions in Videos Using Recurrent Neural

Networks," Communications in Mathematical Biology and Neuroscience, vol. 2025, Article 134, 2025, doi: 10.28919/cmbn/9590.
[19] D. Sharma, J. Singh, S. S. Sehra, and S. K. Sehra, "Demystifying Mental Health by Decoding Facial Action Unit Sequences," Big Data and Cognitive Computing, vol. 8, no. 7, p. 78, 2024, doi: 10.3390/bdcc8070078.
[20] A. Kuwahara, K. Nishikawa, R. Hirakawa, H. Kawano, and Y. Nakatoh, "Eye Fatigue Estimation Using Blink Detection Based on Eye Aspect Ratio Mapping (EARM)," Cognitive Robotics, vol. 2, pp. 50–59, 2022, doi: 10.1016/j.cogr.2022.01.003.
[21] Y. Mahayossanunt, N. Nupairoj, 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.
[22] T.-D. Pham, M.-T. Duong, Q.-T. Ho, S. Lee, and M.-C. Hong, "CNN-Based Facial Expression Recognition with Simultaneous Consideration of Inter-Class and Intra-Class Variations," Sensors, vol. 23, no. 24, p. 9658, 2023, doi: 10.3390/s23249658.
[23] E.-S. Atlam, M. Rokaya, M. Masud, H. Meshref, R. Alotaibi, A. M. Almars, M. Assiri, and I. Gad, "Explainable Artificial Intelligence Systems for Predicting Mental Health Problems in Autistics," Alexandria Engineering Journal, vol. 117, pp. 376–390, 2025, doi: 10.1016/j.aej.2024.12.120.
[24] Y. Li, J. Wei, Y. Liu, J. Kauttonen, and G. Zhao, "Deep Learning for Micro-Expression Recognition: A Survey," IEEE Transactions on Affective Computing, vol. 15, no. 1, pp. 1–23, 2024, doi: 10.1109/TAFFC.2022.3205170.
[25] D. W. Joyce, A. Kormilitzin, K. A. Smith, and A. Cipriani, "Explainable Artificial Intelligence for Mental Health Through Transparency and Interpretability for Understandability," npj Digital Medicine, vol. 6, no. 1, p. 6, 2023, doi: 10.1038/s41746-023-00751-9.
[26] R. Wahyudi, A. A. Dawai, D. Stiawan, A. Pranolo, Y. Mao, and A. H. Bagdade, "SHAP-Based Interpretable Deep Learning Framework for Phishing Website Detection," Science in Information Technology Letters, vol. 6, no. 2, pp. 66–88, 2025, doi: 10.31763/sitech.v6i2.2353.
[27] A. V. Ponce-Bobadilla, V. Schmitt, C. S. Maier, S. Mensing, and S. Stodtmann, "Practical Guide to SHAP Analysis: Explaining Supervised Machine Learning Model Predictions in Drug Development," Clinical and Translational Science, vol. 17, no. 12, p. e70056, 2024, doi: 10.1111/cts.70056.
[28] T.-N. T. Nguyen, T.-D. T. Nguyen, and P. T. Bao, "Micro-Expression Recognition Based on the Fusion Between Optical Flow and Dynamic Image," in Proc. 5th International Conference on Machine Learning and Soft Computing (ICMLSC), Virtual Event, Vietnam, 2021, pp. 1–6, doi: 10.1145/3453800.3453821.
[29] L. Dai, Y. Li, and M. Zhang, "Detection of Operator Fatigue in the Main Control Room of a Nuclear Power Plant Based on Eye Blink Rate, PERCLOS and Mouse Velocity," Applied Sciences, vol. 13, no. 4, p. 2718, 2023, doi: 10.3390/app13042718.
[30] Y. Yamada and M. Kobayashi, "Detecting Mental Fatigue from Eye-Tracking Data Gathered While Watching Video: Evaluation in Younger and Older Adults," Artificial Intelligence in Medicine, vol. 91, pp. 39–48, 2018, doi: 10.1016/j.artmed.2018.06.005.
[31] C.-A. Cos, A. Lambert, A. Soni, H. Jeridi, C. Thieulin, and A. Jaouadi, "Enhancing Mental Fatigue Detection through Physiological Signals and Machine Learning Using Contextual Insights and Efficient Modelling," Journal of Sensor and Actuator Networks, vol. 12, no. 6, p. 77, 2023, doi: 10.3390/jsan12060077.
[32] M. Song, Z. Yang, A. Triantafyllopoulos, Z. Zhang, Z. Nan, M. Tang, H. Takeuchi, T. Nakamura, A. Kishi, T. Ishizawa, K. Yoshiuchi, B. Schuller, and Y. Yamamoto, "Empowering Mental Health Monitoring Using a Macro-Micro Personalization Framework for Multimodal-Multitask Learning: Descriptive Study," JMIR Mental Health, vol. 11, p. e59512, 2024, doi: 10.2196/59512.
[33] M. Lage Cañellas, C. Álvarez Casado, L. Nguyen, and M. Bordallo López, "Depression Recognition from Facial Videos: Preprocessing and Scheduling Choices Hide the Architectural Contributions," Electronics Letters, vol. 59, no. 22, pp. 1–4, 2023, doi: 10.1049/ell2.12996.
[34] K. Sampei, M. Ogawa, C. C. Cortes Torres, M. Sato, and N. Miki, "Mental Fatigue Monitoring Using a Wearable Transparent Eye Detection System," Micromachines, vol. 7, no. 2, p. 20, 2016, doi: 10.3390/mi7020020.
[35] G. Zhao, X. Li, Y. Li, and M. Pietikäinen, "Facial Micro-Expressions: An Overview," Proceedings of the IEEE, vol. 111, no. 10, pp. 1–21, 2023, doi: 10.1109/JPROC.2023.3275192.
[36] K. Kunasegaran, A. M. H. Ismail, S. Ramasamy, J. V. Gnanou, B. A. Caszo, and P. L. Chen, "Understanding Mental Fatigue and Its Detection: A Comparative Analysis of Assessments and Tools," PeerJ, vol. 11, p. e15744, 2023, doi: 10.7717/peerj.15744.
[37] R. Beri and P. Sachdeva, "Hidden Signals: An Automated Human Emotion Detection via Micro-Expression Analysis," International Journal of Advanced Research in Computer Science and Engineering, 2025.

[38] Z. Dong, G. Wang, S. Lu, J. Li, W. Yan, and S.-J. Wang, "Spontaneous Facial Expressions and Micro-Expressions Coding: From Brain to Face," Frontiers in Psychology, vol. 12, p. 784834, 2022, doi: 10.3389/fpsyg.2021.784834.

[39] P. Sharma, J. C. Justus, and G. R. Poudel, "Sensors and Systems for Monitoring Mental Fatigue: A Systematic Review," Sensors and Systems Review, 2023.

[40] Z. Yin, B. Liu, D. Hao, L. Yang, and Y. Feng, "Evaluation of VDT-Induced Visual Fatigue by Automatic Detection of Blink Features," Sensors, vol. 22, no. 3, p. 916, 2022, doi: 10.3390/s22030916.

[41] F. Qu, S.-J. Wang, W.-J. Yan, and X. Fu, "CAS(ME)²: A Database of Spontaneous Macro-Expressions and Micro-Expressions," in Human-Computer Interaction. Theory, Design, Development and Practice, Lecture Notes in Computer Science, vol. 9733, 2016, pp. 48–59. DOI: 10.1007/978-3-319-39513-5_5

[42] A. Psyllos, J. Matulewski, A. Falkowski, A. Łukasik, I. Grzankowska, J. Karłowska-Pik, B. Godek, and D. Pietrykowski, "Eye-Movement Patterns Reflecting Mental Fatigue While Working in Augmented Reality," Procedia Computer Science, vol. 270, pp. 2630–2639, 2025. DOI: 10.1016/j.procs.2025.09.385

[43] P. V. K. Borges, N. Conci, and A. Cavallaro, "Video-Based Human Behavior Understanding: A Survey," IEEE Transactions on Circuits and Systems for Video Technology, vol. 23, no. 11, pp. 1993–2008, Nov. 2013. DOI: 10.1109/TCSVT.2013.2270402

[44] G. Marquart, C. Cabrall, and J. de Winter, "Review of Eye-Related Measures of Drivers' Mental Workload," Procedia Manufacturing, vol. 3, pp. 2854–2861, 2015. DOI: 10.1016/j.promfg.2015.07.783

[45] G. G. Menekse Dalveren and N. E. Cagiltay, "Using Eye-Movement Events to Determine the Mental Workload of Surgical Residents," Journal of Eye Movement Research, vol. 11, no. 4, 2018. DOI: 10.16910/jemr.11.4.3

[46] H. Zhang, B. Liu, J. Tao, and Z. Lv, "Facial Micro-Expression Recognition Based on Multi-Scale Temporal and Spatial Features," in Companion Proceedings of the 2021 International Conference on Multimodal Interaction (ICMI Companion '21), Montréal, QC, Canada, 2021, pp. 485–489. DOI: 10.1145/3461615.3491107

[47] X. Zhu, Z. He, L. Zhao, Z. Dai, and Q. Yang, "A Cascade Attention Based Facial Expression Recognition Network by Fusing Multi-Scale Spatio-Temporal Features," Sensors, vol. 22, no. 4, p. 1350, 2022. DOI: 10.3390/s22041350

[48] W. Chen, T. Sawaragi, and T. Hiraoka, "Comparing Eye-Tracking Metrics of Mental Workload Caused by NDRTs in Semi-Autonomous Driving," Transportation Research Part F: Traffic Psychology and Behaviour, vol. 89, pp. 109–128, 2022. DOI: 10.1016/j.trf.2022.05.004

[49] Y. Ran, "Adaptive Spatio-Temporal Attention Neural Network for Cross-Database Micro-Expression Recognition," Virtual Reality & Intelligent Hardware, vol. 5, no. 2, pp. 142–156, Apr. 2023. DOI: 10.1016/j.vrih.2022.03.006

[50] Y. Zheng and E. Blasch, "Facial Micro-Expression Recognition Enhanced by Score Fusion and a Hybrid Model from Convolutional LSTM and Vision Transformer," Sensors, vol. 23, no. 12, p. 5650, 2023. DOI: 10.3390/s23125650

[51] K. Lisanne, J. Gehrmann, R. Röhrig, and B. Breil, "Investigation of eye movement measures of mental workload in healthcare: Can pupil dilations reflect fatigue or overload when it comes to health information system use" Applied Ergonomics, vol. 114, p. 104150, 2024.DOI: 10.1016/j.apergo.2023.104150

[52] Y. Pan, Y. Shang, T. Liu, Z. Shao, G. Guo, H. Ding, and Q. Hu, "Spatial-Temporal Attention Network for Depression Recognition from Facial Videos," Engineering Applications of Artificial Intelligence, vol. 126, p. 107192, 2023.DOI: 10.1016/j.engappai.2023.107192

[55] Y. Ren, R. Lu, G. Yuan, D. Hao, and H. Li, "Attention-Based Spatiotemporal-Aware Network for Fine-Grained Visual Recognition," Applied Sciences, vol. 14, no. 17, p. 7755, 2024.
DOI: 10.3390/app14177755

[56] S.-M. Leong, R. C.-W. Phan, and V. M. Baskaran, "Emotion-Specific AUs for Micro-Expression Recognition," Multimedia Tools and Applications, vol. 83, pp. 22773–22810, 2024.
DOI: 10.1007/s11042-023-16326-5

[57] F. Wu, Y. Xia, T. Hu, B. Ma, J. Yang, and H. Li, "Facial Micro-Expression Recognition Based on Motion Magnification Network and Graph Attention Mechanism," Heliyon, vol. 10, no. 16, p. e35964, 2024.DOI: 10.1016/j.heliyon.2024.e35964

[58] C. W. Tan, T. Du, J. C. Teo, D. X. H. Chan, W. M. Kong, and B. L. Sng, "Automated Pain Detection Using Facial Expression in Adult Patients with a Customized Spatial Temporal Attention Long Short-Term Memory (STA-LSTM) Network," Scientific Reports, vol. 15, p. 13429, 2025.DOI: 10.1038/s41598-025-97885-5

[59] J. Ma, H. N. Chaudhry, F. Kulsoom, Y. Guihua, S. U. Khan, S. Biswas, Z. U. Khan, and F. Khan, "MULTICAUSENET Temporal Attention for Multimodal Emotion Cause Pair Extraction," Scientific Reports, vol. 15, p. 19372, 2025.DOI: 10.1038/s41598-025-01221-w

[60] C. Fang, S. Liu, and B. Gao, "MB-MSTFNet: A Multi-Band Spatio-Temporal Attention Network for

EEG Sensor-Based Emotion Recognition," Sensors, vol. 25, no. 15, p. 4819, 2025. DOI: 10.3390/s25154819