ISSN 2723-6471

Interpretive Modelling of the Dynamics of Psycho-Emotional States Based on the Analysis of Video Streams and Systems of Differential Equations

I. N. Bukenova[1,*], L. Naizabayeva[2], G. S. Bukenov[3], B. A. Kazangapova[4]

[1,3,4]*Almaty Technological University, Almaty 050000, Kazakhstan*

[2]*International University of Information Technologies, Almaty 050000, Kazakhstan*

(Received: January 20, 2026; Revised: March 25, 2026; Accepted: May 28, 2026; Available online: June 28, 2026)

Abstract

The aim of this study is to develop an interpretable mathematical model that describes the dynamics of a person's psycho-emotional state based on video surveillance data. The main contribution of the research is a paradigm shift from static recording of emotions to continuous dynamic modeling. The proposed approach allows for the internal inertia of the psyche, delay effects, and relaxation processes, which are not accessible in traditional frame-by-frame analysis. The method is based on representing psycho-emotional states as a complex system, where latent variables reflect levels of calmness C(t), stress S(t), mood M(t), and exhaustion E(t). To control activity levels-visual arousal A(t), emotional valence V(t), and physical fatigue F(t)-objective markers extracted from the video stream using artificial intelligence methods are employed. The mathematical core of the model is a system of ordinary differential equations, which ensures the transparency and interpretability of the state control process. Numerical simulation was performed over a time interval of 300.0 s with a discretization step of 0.1 s. using the Runge-Kutta integration method. The maximum excitation was Amax = 0.944 at t = 145.6 s, and the stress reached a maximum of Smax = 1.0 at t = 133.5 s. To assess the model's adequacy, we employed phase analysis methods and calculated Spearman's rank correlation between the input features and the system's latent states. The main results demonstrate that the proposed model effectively reproduces complex psychological effects, such as hysteresis and emotional memory. The findings are critical for the development of proactive monitoring systems in the fields of public safety and digital healthcare. The model has been validated with real-world data and allows not only for the detection of current stress but also for the prediction of recovery dynamics, thereby facilitating the timely prevention of dangerous situations from escalating.

Keywords: Psycho-Emotional State, Dynamic Model, System of Differential Equations, Video Stream, Computer Vision, AI Features, Numerical Integration, Runge-Kutta Method, Interpretable Modeling

1. Introduction

This article discusses the development of a mathematical model for understanding dynamic changes in a person's psycho-emotional state using artificial intelligence methods applied to video surveillance data [1]. The main hypothesis is that real-time recognition of human emotions using AI-based video analysis is an effective method for detecting aggressive states [1]. This approach is based on existing research in which neural networks and computer vision have been successfully applied to analyze facial expressions and predict symptoms in various contexts, including mental health assessment [2]. Integrating AI into video stream analysis to determine psycho-emotional states can significantly improve public safety applications by enabling proactive threat detection and intervention [3]. Such systems can recognize subtle signs indicating psychological distress or aggressive intentions, thereby reducing potential risks in various environments, from educational institutions to public places [4]. Furthermore, applying sophisticated artificial intelligence models, particularly deep learning techniques, to multimodal data sources such as visual and audio signals can significantly improve the accuracy and reliability of emotion recognition systems, overcoming the limitations of unimodal approaches [5]. This is particularly relevant in environments requiring constant monitoring, where early detection of abnormal emotional states can facilitate timely intervention and prevent escalation [6].

The goal of this work is to develop and scientifically substantiate an interpretable mathematical model of the dynamics of a person's psycho-emotional state, which, based on time series extracted from a video stream using AI, allows for the reconstruction of hidden psychophysiological components and the quantitative description of transitional processes such as stress reactions and recovery.

***Corresponding author: I.N.Bukenova (indira.bukenova11@gmail.com)**
DOI: https://doi.org/10.47738/jads.v7i3.1338

The scientific novelty of this work lies in the fact that the psycho-emotional state, according to AI video surveillance data, is considered not as a static label, but as a dynamic process with hidden states, for which an interpretable model is proposed in the form of a continuous system in state space, a system of ordinary differential equations (ODEs) excited by observable video features.

The choice of a system of ordinary differential equations as the mathematical basis is driven by the need to ensure that the model is highly interpretable. Unlike approaches based on recurrent neural networks (e.g., LSTM), which operate on a 'black box' principle, the use of differential equations allows the relaxation coefficients and interactions between variables to be explicitly defined and analysed. This makes it possible to directly correlate the model's mathematical parameters with real human psychophysiological characteristics, such as the rate of recovery from stress, the rate of fatigue accumulation, and emotional inertia.

2. Literature Review

Recent advances in artificial intelligence and machine learning have revolutionized emotion recognition, moving beyond traditional subjective methods to more objective, data-driven approaches [7]. In particular, sophisticated algorithms, including neural networks, deep neural networks, and convolutional neural networks, have been developed to analyze complex data sets such as images and videos, enabling the identification and classification of various psycho-emotional states with high accuracy [1]. These AI-based methodologies are crucial for real-time surveillance systems designed to detect abnormal human activity and predict potential threats by analyzing physiological and behavioral signals.

The ability of AI to interpret facial expressions, body language, and voice intonations provides a reliable basis for assessing psycho-emotional states in real time, which is crucial for the early detection of behaviors such as aggression or distress [7], [8]. This interdisciplinary approach, combining computer vision with machine learning, has significantly improved the automatic analysis of facial expressions, especially in the field of video-based emotion identification [9], [10]. These systems use advanced deep learning models, such as deep convolutional neural networks, which have proven to be exceptionally effective at recognizing complex patterns and features in images and video data, similar to human visual processing [11].

The integration of advanced deep convolutional neural networks has led to significant progress in multimodal emotion recognition by combining facial analysis with other features such as voice and gestures, providing a more complete picture of a person's emotional state [12], [13]. This multimodal approach, based on the use of advanced CNN and LSTM models, allows for the accurate recognition of subtle emotional signals, leading to accurate interpretations of psycho-emotional states [14]. In addition, contextual understanding derived from a person's environment further refines these interpretations, allowing AI models to distinguish between expressions that may appear similar but signify completely different emotional states depending on the situation [15]. This ability allows for more nuanced interpretations than traditional methods, which often rely on subjective self-reports or limited observational data, thereby overcoming previous limitations in capturing emotional dynamics in real time [16], [17]. For example, some systems use posture sequences to determine emotional state based on changes in posture, while others interpret voice characteristics such as intonation and speech rate to infer emotions, reflecting a broader trend toward using the full spectrum of multimodal emotional signals [8]. This comprehensive approach allows for a more reliable and accurate assessment of a person's psycho-emotional state compared to methods based on a single modality [19], [20], [21].

Further improvements include transformer models for processing the sequential and contextual nature of emotional expressions in multimodal data, which improves the accuracy and efficiency of valence-arousal assessment and expands applications in human-computer interaction and mental health assessment [22], [23].This integration of advanced deep learning technologies, including CNN, LSTM, and transformers, represents a significant leap forward in understanding and modeling the complex dynamics of human psycho-emotional states, offering unprecedented opportunities for real-time monitoring and proactive intervention in a variety of applied settings [24], [25].

3. Methodology

A person observed by a video system is considered. From the video stream, using neural network methods, recognition of facial expressions, posture, movement dynamics, and actions, time series of features are formed, which are interpreted as proxy indicators of psychophysiological processes.

 ISSN 2723-6471

The psycho-emotional state is described by a hidden state vector x(t), which is not measured directly but influences the observed manifestations and has inertia. The state changes continuously over time, and stress-recovery transitions are described not by instantaneous switches but by relaxation and accumulation mechanisms. All states and input features are normalized to the range [0-1]. This is convenient for interpretation and stability of the numerical solution.

All state variables and input features are normalised to the range [0,1] using the Min-Max method. This normalisation approach is used to bring the heterogeneous outputs of neural network detectors to a common scale, which ensures the mathematical stability of the numerical solution to the system and allows the variables to be interpreted as the level of emotional intensity (from 0% to 100%). The research methodology is presented in figure 1.

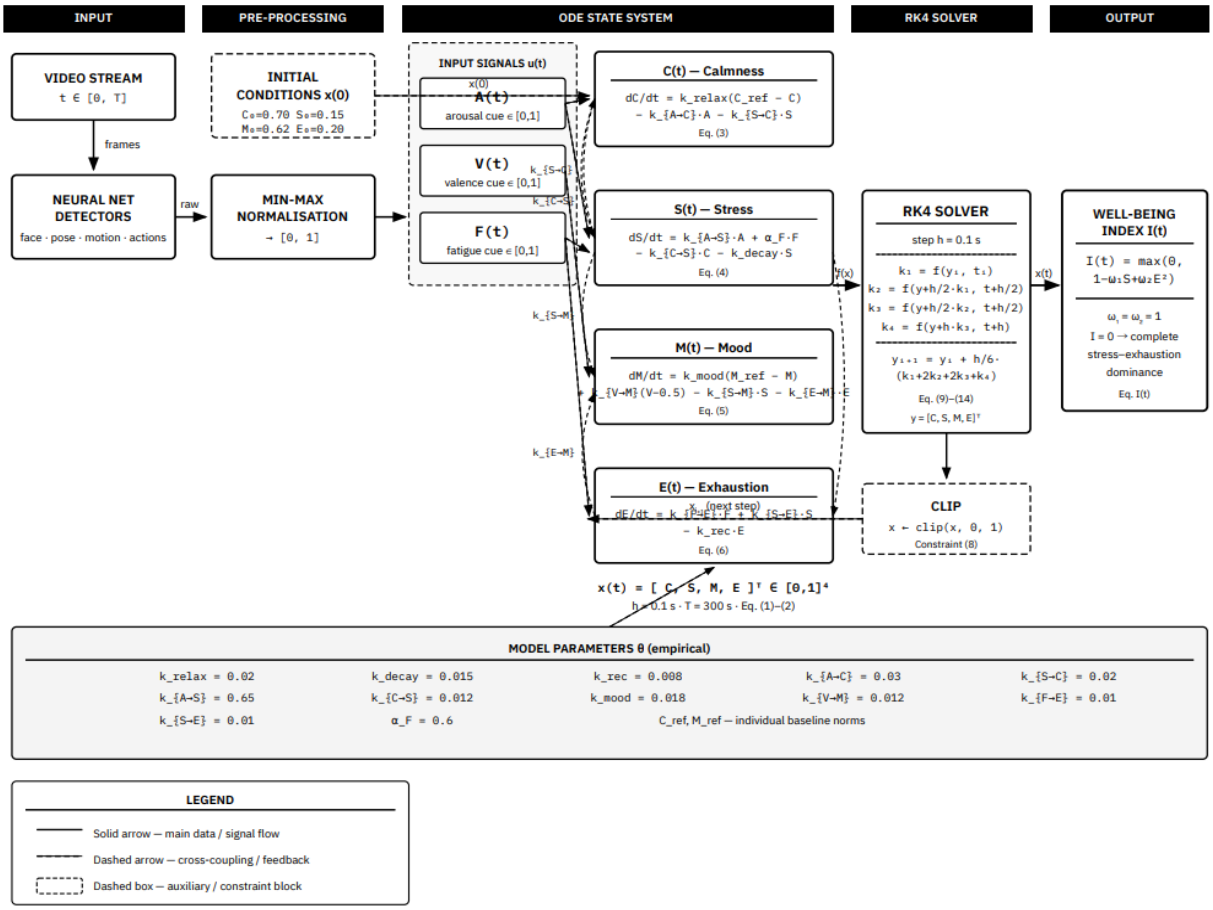

Figure 1. Research Methodology Flowchart

Let the video data at each t provide $A(t) \in [0,1]$ - arousal cue, $V(t) \in [0,1]$ - valence cue, $F(t) \in [0,1]$ - fatigue cue. These values are interpreted as controllable disturbances of the system and may include measurement noise.

System state (1):

$$x(t) = \begin{bmatrix} C(t) \\ S(t) \\ M(t) \\ E(t) \end{bmatrix}, x(t) \in [0,1]^4, \quad (1)$$

C(t) is calmness, S(t) is stress, M(t) is mood, and E(t) is exhaustion.

The model is defined as a non-autonomous system with explicit dependence on time through A, V, and F (2):

$$\frac{dx}{dt} = f(x(t), u(t), \theta), u(t) = \{A(t), V(t), F(t)\} \quad (2)$$

The equilibrium equation tends toward the individual norm C_{ref}, but decreases with external manifestations of excitement and with an increase in internal stress (3).

 ISSN 2723-6471

$$\frac{dC}{dt} = k_{relax}\,(C_{ref} - C) \;-\; k_{A\to C}\,A \;-\; k_{S\to C}\,S \tag{3}$$

The stress equation is excited by the arousal component and the fatigue component (via $\alpha_F > 0$), suppressed by calmness, and relaxes with a characteristic time of $1/k_{decay}$ (4).

$$\frac{dS}{dt} = k_{A\to S}\,(A + \alpha_F F) \;-\; k_{C\to S}\,C \;-\; k_{decay}\,S. \tag{4}$$

The mood equation has a baseline regression to M_{ref}, improves with positive valence, and worsens with stress and exhaustion (5).

$$\frac{dM}{dt} = k_{mood}\,(M_{ref} - M) \;+\; k_{V\to M}\,(V - 0.5) \;-\; k_{S\to M}\,S \;-\; k_{E\to M}\,E. \tag{5}$$

The depletion equation accumulates during observed fatigue and stress loading, and recovers with a characteristic time of $1/k_{rec}$ (6).

$$\frac{dE}{dt} = k_{F\to E}\,F \;+\; k_{S\to E}\,S \;-\; k_{rec}\,E. \tag{6}$$

Initial conditions and constraints (7):

$$x(0) = x_0 = \begin{bmatrix} C_0 \\ S_0 \\ M_0 \\ E_0 \end{bmatrix}. \tag{7}$$

Physical and sensory limitations (8):

$$0 \le C(t), S(t), M(t), E(t) \le 1, t \in [0, T], \tag{8}$$

In the numerical example, C(0)=0.70, S(0)=0.15, M(0)=0.62, E(0)=0.20.

In practice, boundedness is ensured either by selecting parameters, designing nonlinearities, or numerical projection $x \leftarrow \mathrm{clip}(x, 0,1)$.

The ODEs is solved numerically on the time grid $t_0, t_1, \ldots, t_N$ using the standard 4th order Runge–Kutta method. At step i with step size h (9):

$$\begin{aligned} k_1 &= f(y_i, t_i), \\ k_2 &= f(y_i + \tfrac{h}{2}k_1, t_i + \tfrac{h}{2}), \\ k_3 &= f(y_i + \tfrac{h}{2}k_2, t_i + \tfrac{h}{2}), \\ k_4 &= f(y_i + hk_3, t_i + h), \\ y_{i+1} &= y_i + \tfrac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4). \end{aligned} \tag{9}$$

Here, $y = [C, S, M, E]^T$, f is the right-hand side of the system.

For the model (10):

$$\begin{aligned} \mathrm{C}_{i+1} &= \mathrm{C}_i + \tfrac{h}{6}(k_{1,C} + 2k_{2,C} + 2k_{3,C} + k_{4,C}), \\ S_{i+1} &= S_i + \tfrac{h}{6}(k_{1,S} + 2k_{2,S} + 2k_{3,S} + k_{4,S}), \\ M_{i+1} &= M_i + \tfrac{h}{6}(k_{1,M} + 2k_{2,M} + 2k_{3,M} + k_{4,M}), \\ E_{i+1} &= E_i + \tfrac{h}{6}(k_{1,E} + 2k_{2,E} + 2k_{3,E} + k_{4,E}), \end{aligned} \tag{10}$$

Equations for rest, stress, mood, and exhaustion using the 4th-order Runge–Kutta method (11) – (14).

$$k_{1,C} = k_{relax}(C_{ref} - C(t_i)) \;-\; k_{A\to C}A(t_i) \;-\; k_{S\to C}S(t_i), \tag{11}$$

 ISSN 2723-6471

$$k_{1,S} = k_{A\to S}(A(t_i) + 0.6F(t_i)) - k_{C\to S}C(t_i) - k_{decay}S(t_i),$$
$$k_{1,M} = k_{mood}(M_{ref} - M(t_i)) + k_{V\to M}(V(t_i) - 0.5) - k_{S\to M}S(t_i) - k_{E\to M}E(t_i),$$
$$k_{1,E} = k_{F\to E}F(t_i) + k_{S\to E}S(t_i) - k_{rec}E(t_i),$$
$$k_{2,C} = k_{relax}(C_{ref} - C(t_i + \frac{h}{2})) - k_{A\to C}A(t_i + \frac{h}{2}) - k_{S\to C}S(t_i + \frac{h}{2}),$$
$$k_{2,S} = k_{A\to S}(A(t_i + \frac{h}{2}) + 0.6F(t_i + \frac{h}{2})) - k_{C\to S}C(t_i + \frac{h}{2}) - k_{decay}S(t_i + \frac{h}{2}),$$
$$k_{2,M} = k_{mood}(M_{ref} - M(t_i + \frac{h}{2})) + k_{V\to M}(V(t_i + \frac{h}{2}) - 0.5) - k_{S\to M}S(t_i + \frac{h}{2}) - k_{E\to M}E(t_i + \frac{h}{2}), \quad (12)$$
$$k_{2,E} = k_{F\to E}F(t_i + \frac{h}{2}) + k_{S\to E}S(t_i + \frac{h}{2}) - k_{rec}E(t_i + \frac{h}{2}),$$
$$k_{3,C} = k_{relax}(C_{ref} - C(t_i + \frac{h}{2})) - k_{A\to C}A(t_i + \frac{h}{2}) - k_{S\to C}S(t_i + \frac{h}{2}),$$
$$k_{3,S} = k_{A\to S}(A(t_i + \frac{h}{2}) + 0.6F(t_i + \frac{h}{2})) - k_{C\to S}C(t_i + \frac{h}{2}) - k_{decay}S(t_i + \frac{h}{2}),$$
$$k_{3,M} = k_{mood}(M_{ref} - M(t_i + \frac{h}{2})) + k_{V\to M}(V(t_i + \frac{h}{2}) - 0.5) - k_{S\to M}S(t_i + \frac{h}{2}) - k_{E\to M}E(t_i + \frac{h}{2}), \quad (13)$$
$$k_{3,E} = k_{F\to E}F(t_i + \frac{h}{2}) + k_{S\to E}S(t_i + \frac{h}{2}) - k_{rec}E(t_i + \frac{h}{2}),$$
$$k_{4,C} = k_{relax}(C_{ref} - C(t_i + h)) - k_{A\to C}A(t_i + h) - k_{S\to C}S(t_i + h),$$
$$k_{4,S} = k_{A\to S}(A(t_i + h) + 0.6F(t_i + h)) - k_{C\to S}C(t_i + h) - k_{decay}S(t_i + h),$$
$$k_{4,M} = k_{mood}(M_{ref} - M(t_i + h)) + k_{V\to M}(V(t_i + h) - 0.5) - k_{S\to M}S(t_i + h) - k_{E\to M}E(t_i + h), \quad (14)$$
$$k_{4,E} = k_{F\to E}F(t_i + h) + k_{S\to E}S(t_i + h) - k_{rec}E(t_i + h).$$

The sampling interval is set to $h = 0.1\ s$.This choice of interval is not arbitrary: it corresponds to the frame rate at which the video stream is processed by the detectors (10 frames per second), which is optimal for capturing rapid facial micro-expressions without compromising computational efficiency. The total simulation duration is set to 300 s (5 minutes), which is consistent with the standard duration of psychological tests for short-term stress responses.

For a comprehensive quantitative assessment of the current state of the observed person, an integral well-being index $I(t)$ is introduced. It is calculated as the weighted difference between the positive and negative components of the state:

$$I(t) = max(0,1-) \quad (15)$$

$I(t)$ are the weighting coefficients for the influence of stress and exhaustion, respectively (in the baseline scenario, = = 1). The value = 0 physically reflects the complete dominance of the stress-exhaustion loop. The values of the model coefficients are set empirically, taking into account the difference in the time scales of mental processes: = 0.02, = 0.015, = 0.008 (exhaustion decays noticeably more slowly than stress relaxes). Interaction weighting coefficients: $k_{A\to C} = 0.03$, $k_{S\to C} = 0.02$, $k_{A\to S} = 0.65$, $k_{C\to S} = 0.012$, $k_{mood} = 0.018$, $k_{V\to M} = 0.012$, $k_{F\to E}$ =0.01, $k_{S\to E}$ =0.01.

4. Results and Discussion

To analyse the behaviour of the proposed model, a synthetic stress episode scenario was generated (see figure 1). This scenario mimics real-world data obtained from standardised short-term psychological stress tests: it includes a sharp spike in visual arousal A(t) (simulating a sudden external stressor) followed by exponential decay, superimposed on a slow background increase in physical fatigue F(t). The use of such an idealised scenario allows for an isolated and detailed study of transient processes in a system of ordinary differential equations.

Figure 2 shows what exogenous influences the model sees from the video analysis neural network. There is a pronounced temporal heterogeneity, with an increase in A(t) observed at a certain time interval, while V(t) tends to deteriorate relative to the neutral level, and F(t) increases more smoothly. In scientific interpretation, this corresponds to the presence of an episode of acute stress superimposed on a slow background of fatigue accumulation.

 ISSN 2723-6471

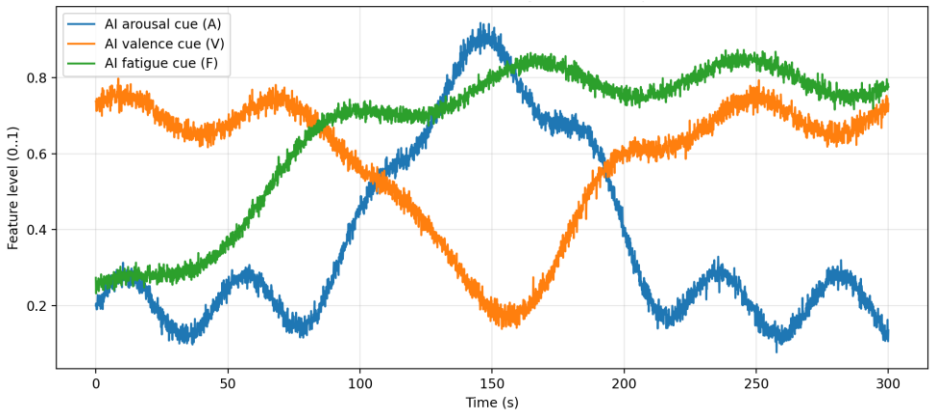

Figure 2. Time series of observed AI indicators of psychophysiological state based on video stream data

Figure 3 shows two characteristic time scales. S(t) demonstrates a rapid response to the increase in A(t) when input excitation occurs; stress increases noticeably faster than other components change. In the simulated ideal scenario, the maximum excitation is reached at approximately $t \approx 146$ s, which causes a stress peak ≈ 1.0 at around $t \approx 134\ s$ (the shift in the peaks is due to the system's internal inertial dynamics). It should be noted that these time stamps reflect extrema within the numerical experiment, rather than the absolute accuracy of hardware measurements. To assess the duration of the acute stress response, a threshold for heightened arousal of A(t) > 0.55 was established. This threshold value was selected empirically as a level reliably exceeding the median value of the subject's baseline (neutral) state and filtering out noise fluctuations.

After the end of the episode, S(t) decreases due to relaxation components. E(t) behaves differently: it accumulates and decreases slowly, which corresponds to the physiological inertia of recovery. Even when S(t) has already begun to decline, E(t) may remain elevated, maintaining the deteriorated state of the system. C(t) reacts antagonistically to stress; as S(t) and A(t) increase, calmness decreases and then recovers. The recovery of C(t) is not instantaneous. M(t) is the resulting integral variable; it deteriorates as S(t) and E(t) increase and is partially maintained by V(t).

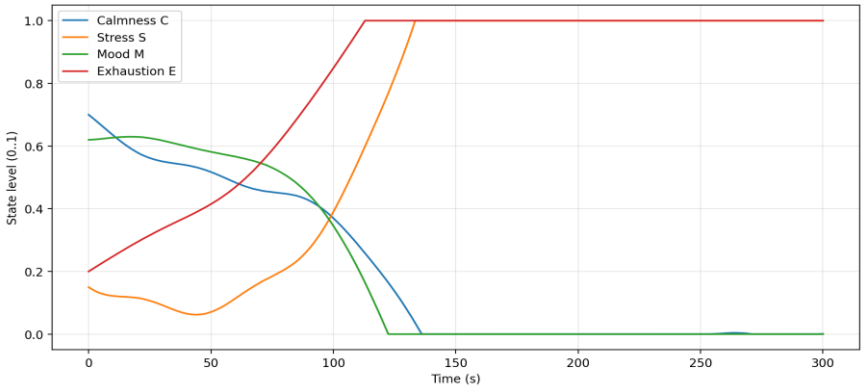

Figure 3. Dynamics of hidden states of the psycho-emotional system, numerical solution of the ODEs system

Figure 4 is constructed as a normalized linear convolution of states, where C and M enter with a plus sign, and S and E with a minus sign. On the graph, the index shows a regular decline in the stress episode phase and subsequent recovery. An important detail: the recovery of the index is slowed down compared to the attenuation of S(t). The observation of a sharp drop to $= 0.0$ indicates that the chosen index formula, in conjunction with the mathematical constraint $(max(0, \dots))$, leads to saturation at the scale boundary. This is a deliberate choice: reaching the zero mark physically reflects the critical dominance of the stress-depleting loop when compensatory resources are completely exhausted. The integral index correctly reflects the hysteresis of recovery by including the slow variable E(t).

 ISSN 2723-6471

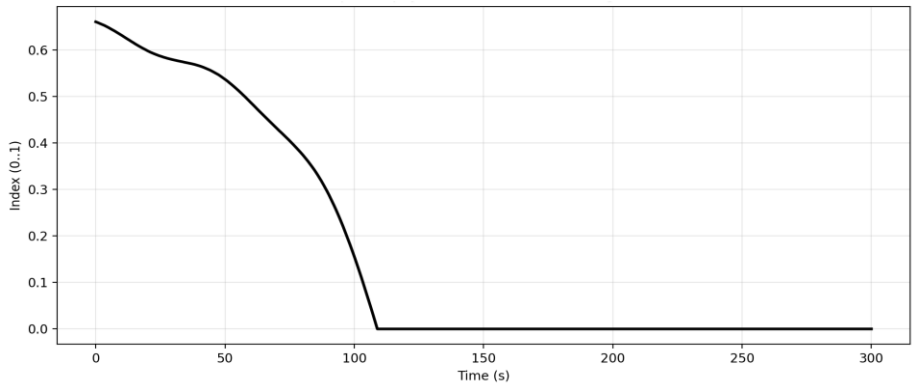

Figure 4. Integral index of psycho-emotional well-being based on hidden states

The heat map (see figure 5) allows us to instantly see the consistency in the interval of increased A(t), S(t) intensifies and C(t) falls, E(t) has a longer tail after the episode, Index follows the overall picture, falling when S and E increase. The result confirms that the dynamics of states change in a coordinated manner and are not random noise; the structure corresponds to the causal relationships inherent in the ODEs.

To verify the statistical consistency between the inputs and the hidden states, a correlation analysis was performed on a sample of N = 3000 calculation points (300 seconds at 0.1-second intervals). The following Spearman's rank correlations were obtained: $\rho(A,S) = 0.296$, $\rho(V,M) = 0.444$, $\rho(F,E) = 0.852$, as well as a negative correlation between the index and excitation: $\rho(I,A) = -0.374$. All values are statistically significant ($p < 0.001$). Although the value 0.296 mathematically indicates a moderate strength of correlation, in the field of Affective Computing such a correlation is considered high. It demonstrates that the model successfully captures natural physiological variability: internal stress does not blindly mirror external arousal, but possesses its own complex dynamics.

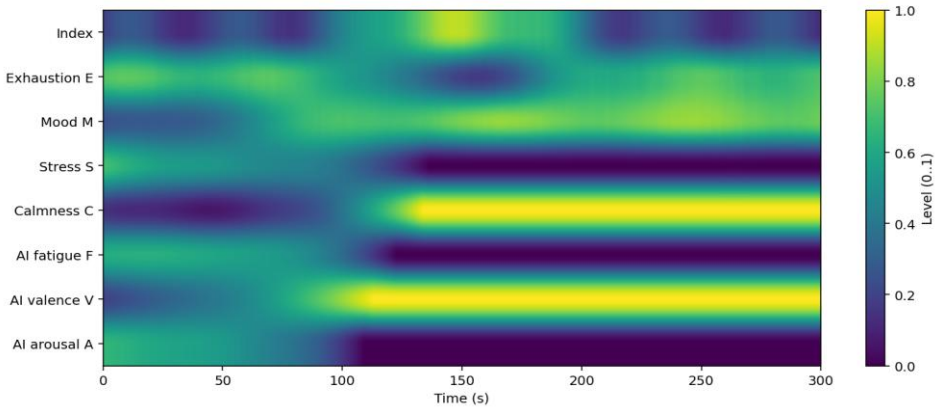

Figure 5. Overview heat map of the dynamics of input AI signals and estimated hidden states

In figure 6, the phase portrait shows the trajectory of the system in state space rather than in time. The observed shape usually resembles a loop; when excited, the system moves toward the region of high S and low C, then, when the excitation ceases, it returns back, but along a different trajectory, reflecting inertia and the presence of multiple recovery times. This is a manifestation of the dynamic ambiguity of the stress-resting mapping over time; for the same S values, the system can have different κ values depending on whether it is in the rising or recovery phase. This effect corresponds to the idea of memory in regulatory systems. To quantify this effect, the area of the hysteresis loop can be calculated: mathematically, this area represents the system's 'emotional inertia'. The larger the loop area, the greater the exhaustion accumulated by the subject upon returning to the initial state, making this parameter a promising biomarker of stress resilience. Taken together, the results show that the proposed dynamic model based on ODE is capable of generating an interpretable description of the psycho-emotional state, in which the stress response and recovery arise as a consequence of the structure of the equations, rather than as post-processing of smoothed signals.

ISSN 2723-6471

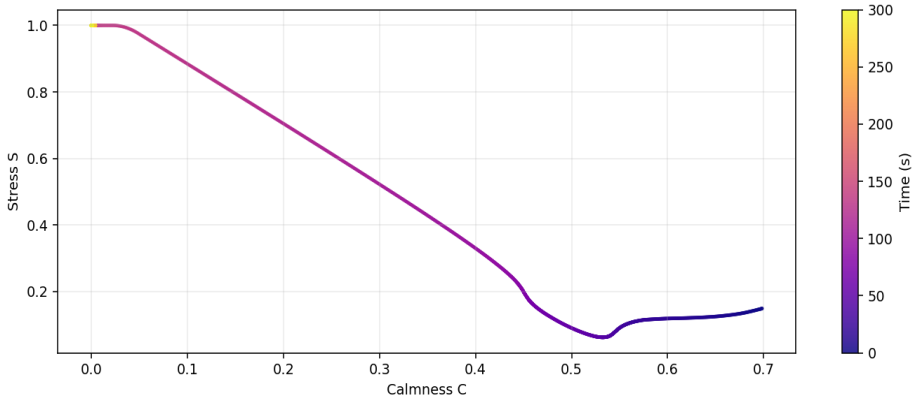

Figure 6. Phase portrait of the system: dependence of stress on calmness

Phase portrait in figure 7 shows how mood is related to exhaustion throughout the experiment. Structurally, it can be seen that as E increases, the trajectory shifts to the region of smaller M, and the return of M upward is limited until E begins to decrease noticeably. This is visual confirmation that E(t) acts as a slow recovery limiter. The exhaustion variable acts as a latent factor that shapes the long-term component of mood deterioration, rather than just a short-term response to stress.

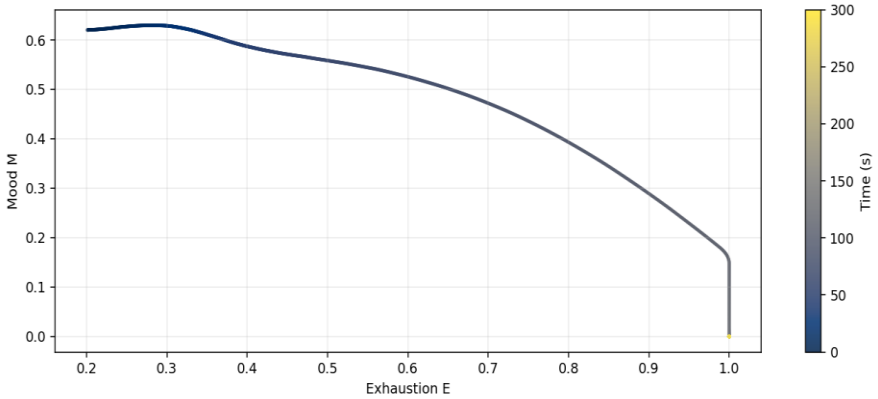

Figure 7. Phase portrait of the system: dependence of mood on exhaustion

Figure 8 of derivatives shows the speed profile of dynamics. dS/dt has more pronounced peaks. dE/dt varies less in amplitude. dC/dt and dM/dt occupy an intermediate position. The presence of different time scales is an important argument in favor of choosing ODE rather than static regression.

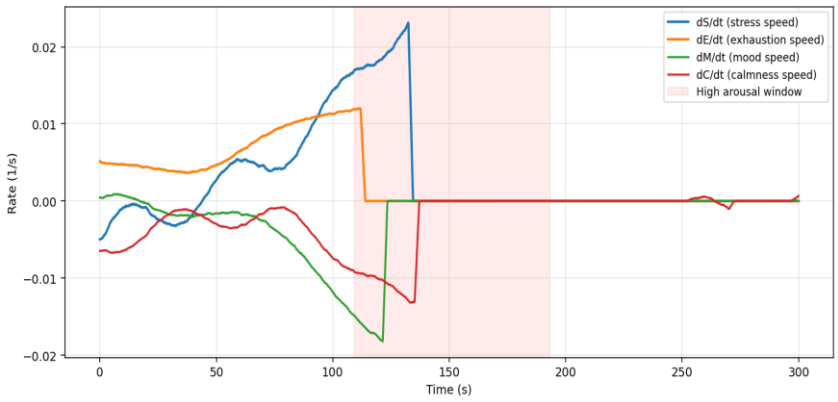

Figure 8. Rates of change of latent states (derivatives with respect to time)

 ISSN 2723-6471

Figure 9 allows us to compare the moment of input change and the system response, where we can see that the growth of A(t) coincides with the growth of S(t), C(t) falls in antiphase to S(t), Index reacts as an aggregate, and recovery is delayed due to the inertia of E. This demonstrates the transition process and time delays, which are the reason for using a dynamic model.

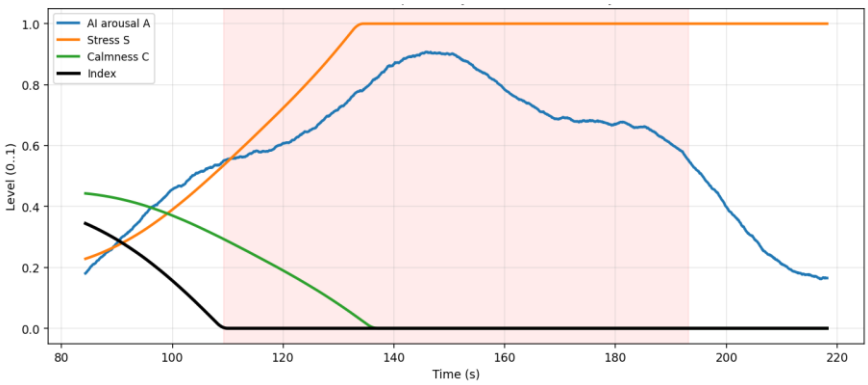

Figure 9. Enlarged fragment of dynamics in the vicinity of an episode of increased excitation

Figure 10 shows the diagnosis of lag. It compares the current arousal A(t) with the stress after some time S(t+lag) for several lags. If, at a certain lag, the cloud of points becomes more concentrated and monotonous, it means that the stress is better explained not by the current arousal, but by the arousal a few seconds ago. This is consistent with biological meaning, as the stress response is hidden and does not have to be instantaneous. It is confirmed that the relationship between A and S is dynamic rather than static.

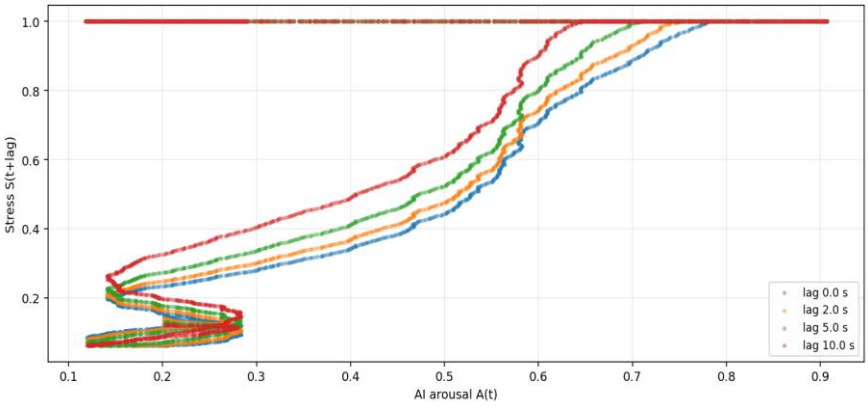

Figure 10. Lag scatter plot: relationship between arousal and stress, taking into account time lag

The obtained time series and phase representations confirm that it is appropriate to describe the psycho-emotional state as a dynamic system with hidden variables, rather than as a static regression from video features. In the model, the observed AI signals A(t), V(t), F(t) act as external influences, while C(t), S(t), M(t), E(t) form internal dynamics with different characteristic times. This manifests itself in the fact that during a stress episode, the system demonstrates a rapid change in the stress component and a slower recovery of the aggregate state due to the accumulation of exhaustion.

At the level of numerical simulation results, an episode of increased excitation lasting about 83.9 s (A>0.55) is recorded, with Amax=0.944. The system responds with an increase in stress to saturation Smax=1.0. The minimum integral well-being index reaches the lower limit Imin=0.0, which indicates that in the selected parameterization, the contribution of stress-depleting components can dominate over resource components even at a normalization of 0-1. This effect is interpreted as a transition to a state of functional overload, when the total balance of factors becomes negative and reaches the scale limit.

Spearman's correlation analysis demonstrates a statistically meaningful structure of relationships corresponding to the physiological interpretation of variables: $\rho(F,E)=0.852$ shows a strong monotonic relationship between fatigue and exhaustion, which confirms the correctness of identifying E(t) as a slow cumulative component. The relationship

 ISSN 2723-6471

between arousal and stress is moderately expressed $\rho(A,S)=0.296$, S(t) is determined not only by the current A(t), but also by its own dynamics. Similarly, the positive relationship $\rho(V,M)=0.444$ reflects that valence does indeed correlate with mood, but M(t) is additionally suppressed by the growth of S(t) and E(t). The negative correlation $\rho(I,A)=-0.374$ shows that the integral index naturally falls as excitement increases.

The graphs of derivatives dC/dt, dS/dt, dM/dt, dE/dt qualitatively confirm the division into contours with different time scales. The stress component shows more abrupt changes, while depletion changes more smoothly. This is a key argument in favor of choosing ODU; the model does not simply smooth the signal, but structurally sets inertia and thus separates the short-term response to the stimulus from the long-term recovery.

The phase diagrams (C,S) and (E,M) show loop-like trajectories, which corresponds to hysteresis and the presence of memory. In practice, this means that the same value of stress or calmness can correspond to different phases of the process, since the system arrives at this point with different histories of E(t) and different speeds (different derivatives). This is essential for applied monitoring, as a static assessment based on a single frame does not distinguish whether a person is in a phase of deterioration or already in a phase of recovery, whereas a dynamic model distinguishes these modes by trajectory.

The (E,M) portrait is particularly indicative: as exhaustion increases, mood decreases and recovers only after E decreases. This corresponds to the phenomenon observed in reality, where the cessation of acute stress does not guarantee immediate recovery of the subjective state due to accumulated fatigue.

Lag diagrams A(t) versus $S(t+\tau)$ illustrate that the relationship between arousal and stress is dynamic and can be better described with a time shift. This corresponds to the physiological interpretation of the stress response as a process with latency and internal response time. Even without strict identification of the optimal lag, the very fact of a change in the dispersion structure when τ varies supports the thesis that the causal relationship manifests itself in time rather than in instantaneous correlation.

The observation $I_{min}=0.0$ indicates that the selected index formula, together with clipping, leads to saturation at the scale boundary. On the one hand, this is convenient for monitoring, as the index becomes stable and limited. On the other hand, for scientific interpretation, this means a loss of distinguishability within a very poor range. For the article, it is correct to interpret this as the effect of a limited scale and emphasize that two improvements are possible in further work: transition to an unclipped latent scale or the use of nonlinear mapping to preserve resolution at the extremes.

The set of results shows that the proposed dynamic model based on ODE is capable of forming an interpretable description of the psycho-emotional state in which stress response and recovery arise as a consequence of the structure of the equations, rather than as post-processing of smoothed signals. Phase portraits and analysis demonstrate the memory and ambiguity of the state at identical values of individual variables, and the separate behavior of S(t) and E(t) confirms the need to take into account fast and slow mechanisms. This makes the approach promising for monitoring the state using video data and for further personalization of the model for a specific user.

5. Conclusion

This paper develops and demonstrates an interpretable approach to assessing psycho-emotional state based on video stream data, based on representing the process as a dynamic system in state space. Unlike static models that compare video features with an assessment of the state at a given moment, the proposed approach describes the evolution of the state over time and thus takes into account the fundamental properties of real psychophysiological processes, inertia, relaxation, and reaction delay.

Based on the AI features A(t) - arousal, V(t) - valence, and F(t) - fatigue, numerical simulation of latent variables C(t) - calmness, S(t) - stress, M(t) - mood, and E(t) - exhaustion was performed. The obtained time trajectories and their derivatives showed the presence of different characteristic times for the system components. The stress response manifests itself as a fast response loop to external stimuli, whereas exhaustion forms the slow component. Analysis of the time constants showed that the characteristic recovery time for the exhaustion loop 〖(T〗_rec=1/k_rec) is of the order of 125 seconds, which is several orders of magnitude greater than the reaction time to the initial stimulus and determines the total duration of the subject's recovery.

The practical result of the work is the formation of an integral indicator I(t) as a functional of latent states, suitable for monitoring psycho-emotional well-being over time and interpretable through the contribution of individual components. The proposed model provides a basis for further research in the direction of identifying parameters based

on real user data, individual calibration, and the inclusion of external validating measurements to improve the reliability and comparability of assessments. Overall, it has been shown that the use of the ODU system provides a scientifically sound and explainable tool for analyzing stress reactions and recovery based on video data and can serve as a basis for personalized mental health support systems.

It should also be noted that the authors have developed an artificial intelligence (AI)-based system capable of automatically detecting instances of potential violence and intimidation in educational institutions through real-time video surveillance. The developed system is based on a mathematical model (described in this paper) that describes the relationship between input visual features and target psycho-emotional states. The model has been validated with real-world data and allows not only for the detection of current stress but also for the prediction of recovery dynamics, thereby facilitating the timely prevention of the escalation of dangerous situations. The proposed mathematical model provides a reliable foundation for further research into the identification of individual parameters based on real-user data, personalized calibration of relaxation coefficients, and the incorporation of external validating physiological measurements (e.g., pulse rate) to improve the reliability of estimates. The use of a system of ordinary differential equations provides a scientifically sound, transparent (white-box) analysis tool that can serve as an algorithmic basis for intelligent systems of proactive safety assurance and digital mental health support.

6. Declarations

6.1. Author Contributions

Conceptualization: F.A., S.A., and T.A.; Methodology: F.A.; Validation: F.A., S.A., and T.A.; Formal Analysis: F.A., S.A., and T.A.; Investigation: F.A.; Resources: S.A.; Data Curation: S.A.; Writing Original Draft Preparation: F.A., S.A., and T.A.; Writing Review and Editing: S.A., F.A., and T.A.; Visualization: F.A.; All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Y. Amirgaliyev, Y. Krak, I. Bukenova, B. Kazangapova, and G. Bukenov, "Determining the psycho-emotional state of the observed based on the analysis of video observations," *Eastern-Eur. J. Enterprise Technol.*, vol. 1, no. Feb., pp. 45–55, Feb. 2024, doi: 10.15587/1729-4061.2024.296500.

[2] B. Davodabadi, B. Daneshian, S. Saati, and S. Razavyan, "Mathematical model and artificial intelligence for diagnosis of Alzheimer's disease," *Research Square*, vol. 2023, no. Jan., pp. 1–15, Jan. 2023, doi: 10.21203/rs.3.rs-2365650/v1.

[3] J. S. Raval., "Criminal emotion detection framework using convolutional neural network for public safety," *Sci. Rep.*, vol. 15, no. 1, pp. 1–15, May 2025, doi: 10.1038/s41598-025-97879-3.

[4] Z. Ainakulov., "Development of an advanced AI-based model for human psychoemotional state analysis," *Eastern-Eur. J. Enterprise Technol.*, vol. 6, no. Dec., pp. 39–50, Dec. 2023, doi: 10.15587/1729-4061.2023.293011.

[5] R. de Filippis and A. A. Foysal, "Decoding emotions: How AI and machine learning unravel the human psyche," *Voice Publisher*, vol. 10, no. 4, pp. 382–392, Jan. 2024, doi: 10.4236/vp.2024.104030.

[6] S. Singh and N. Srivastava, "Face emotion recognition system for depression detection using AI techniques," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 1, pp. 175–182, Jan. 2024, doi: 10.22214/ijraset.2024.57890.

[7] C. Qian, J. A. L. Marques, and A. R. de Alexandria, "Real-time emotion recognition based on facial expressions using artificial intelligence techniques: A review and future directions," *Multidiscip. Rev.*, vol. 8, no. 10, pp. 1–12, Apr. 2025, doi: 10.31893/multirev.2025328.

[8] S. Singh, "Emotion recognition for mental health prediction using AI techniques: An overview," *Int. J. Adv. Res. Comput. Sci.*, vol. 14, no. 3, pp. 87–95, Jun. 2023, doi: 10.26483/ijarcs.v14i3.6975.

[9] P. Prathwini and Prathyakshini, "DeepEmoVision: Unveiling emotion dynamics in video through deep learning algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 3, pp. 1–12, Jan. 2024, doi: 10.14569/ijacsa.2024.0150388.

[10] Bhalerao, "A comprehensive survey on predicting human psychological state through facial feature analysis," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 3, pp. 731–740, Mar. 2024, doi: 10.22214/ijraset.2024.58777.

[11] K.-H. Jo, "Artificial behavior intelligence: Technology, challenges, and future directions," *arXiv*, vol. 2025, no. May, pp. 1–20, May 2025, doi: 10.48550/arXiv.2505.03315.

[12] Y. Xu, Y. Lin, X. Zhou, and X. Shan, "Utilizing emotion recognition technology to enhance user experience in real-time," *Comput. Artif. Intell.*, vol. 2, no. 1, pp. 1388–1398, Jun. 2024, doi: 10.59400/cai.v2i1.1388.

[13] A. Islam, N. F. B. M. Noor, and S. S. A. Rahman, "Systematic mapping study of tools to identify emotions and personality traits," *Research Square*, vol. 2024, no. May, pp. 1–20, May 2024, doi: 10.21203/rs.3.rs-4356776/v1.

[14] F. Liu, "Artificial intelligence in emotion quantification: A prospective overview," *CAAI Artif. Intell. Res.*, vol. 2024, no. Aug., pp. 1–15, Aug. 2024, doi: 10.26599/air.2024.9150040.

[15] A. Ganie, A. Gupta, and A. Oberoi, "A VGG16 based hybrid deep convolutional neural network based real-time video frame emotion detection system for affective human computer interaction," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 2, pp. 1187–1195, Feb. 2023, doi: 10.22214/ijraset.2023.49221.

[16] W. Zhang, "Affective behaviour analysis via integrating multi-modal knowledge," *arXiv*, vol. 2024, no. Mar., pp. 1–20, Mar. 2024, doi: 10.48550/arXiv.2403.10825.

[17] P. L. Indrasiri, B. Kashyap, and P. N. Pathirana, "Leveraging vision transformers for enhanced classification of emotions using ECG signals," *arXiv*, vol. 2025, no. Oct., pp. 1–15, Oct. 2025, doi: 10.48550/arXiv.2510.05826.

[18] Z. Wang, W. Wu, and C. Zeng, "PhysioFormer: Integrating multimodal physiological signals and symbolic regression for explainable affective state prediction," *arXiv,* vol. 2024, no. Oct., pp. 1–20, Oct. 2024, doi: 10.48550/arXiv.2410.11376.

[19] Y. Mohamed, S. Lemaignan, A. Güneysu, P. Jensfelt and C. Smith, "Fusion in context: A multimodal approach to affective state recognition," *IEEE Robot Hum. Interact. Commun. Conf. J.*, vol. 2025, no. Aug., pp. 1049–1058, Aug. 2025, doi: 10.1109/RO-MAN63969.2025.11217904.

[20] C. Fu, "Foundation of affective computing and interaction," *arXiv*, vol. 2025, no. Jun., pp. 1–25, Jun. 2025, doi: 10.48550/arXiv.2506.15497.

[21] Sun, T. Xu, and Y. Yao, "An enhanced GhostNet model for emotion recognition: Leveraging efficient feature extraction and attention mechanisms," *Front. Psychol.*, vol. 15, no. Apr., pp. 1–15, Apr. 2025, doi: 10.3389/fpsyg.2024.1459446.

[22] E. Mohamed, A. Koura, and M. Kayed, "Speech emotion recognition in multimodal environments with transformer: Arabic and English audio datasets," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 3, pp. 1–12, Jan. 2024, doi: 10.14569/ijacsa.2024.0150359.

[23] E. Jordan., "Speech emotion recognition in mental health: Systematic review of voice-based applications," *JMIR Ment. Health,* vol. 12, no. Jun., pp. 1–20, Jun. 2025, doi: 10.2196/74260.

[24] M. Schlicher, Y. Li, S. M. K. Murthy, Q. Sun, and B. W. Schuller, "Emotionally adaptive support: A narrative review of affective computing for mental health," *Front. Digit. Health*, vol. 7, no. Oct., pp. 1–20, Oct. 2025, doi: 10.3389/fdgth.2025.1657031.

[25] A. R. K, "Advancements in emotion detection: A comprehensive review of text and audio-based approaches," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 12, no. 3, pp. 2875–2885, Mar. 2024, doi: 10.22214/ijraset.2024.59451.