Stress Detection System

Abstract: The objective of this work is to design and implement a real-time, local stress-detection pipeline that estimates human stress levels on a 0–100 scale by fusing physiological and behavioral signals captured through a webcam. A pre-trained DeepPhys rPPG model is employed to extract heart-rate-related physiological information from subtle color changes in the facial region, while blink-rate analysis derived from facial landmarks provides behavioral cues. Both signals are filtered, normalized, and combined through a weighted fusion model to produce a coherent and interpretable stress score. The system adheres to the constraints specified in the challenge specification and demonstrates responsiveness to calm vs. stressed user states in both recorded UBFC videos and live-webcam settings.

1. Problem Statement

Modern computer-vision systems are increasingly expected to interpret human physiological and behavioral states without requiring contact sensors. Detecting stress levels through a webcam is challenging because:

- rPPG signals are extremely noisy and sensitive to motion, illumination, and frame quality.
- Behavioral signals such as blinks, eye closures, or head jitter require robust landmark extraction and noise compensation.
- No single signal is sufficiently stable; therefore, a fusion strategy is needed.
- The model must run locally in real time, avoiding black-box emotion APIs as required by the challenge guidelines.

Thus, the problem addressed is:

How to build a reliable, real-time stress index (0–100) by fusing webcam-based physiological and behavioral signals obtained from pre-trained low-level models?

2. Problem Statement

The proposed system integrates multiple components into a unified stress-detection pipeline:

2.1 Physiological Signal Extraction (rPPG)

- A facial Region of Interest (ROI) is extracted using MediaPipe landmarks.
- A pre-trained DeepPhys model processes RGB frame differences and raw frames to generate a temporal scalar rPPG signal.
- Signals are detrended, band-pass filtered, and spectral-analysed to estimate heart rate (HR) and signal quality.

2.2 Behavioral Signal Extraction (Blink Rate)

- Eye landmarks are used to compute the Eye Aspect Ratio (EAR).
- Temporal EAR patterns determine blink events.
- A sliding window yields smoothed blinks per minute (BPM), which correlate with cognitive load and stress.

2.3 Signal Fusion and Stress-Score Logic

- HR and blink rate are normalized into dimensionless indices.
- Physiological component = 60% weight
- Behavioral component = 40% weight
- rPPG quality modulates the physiological contribution.
- An exponential moving average smooths temporal fluctuations.
- Final output: Stress Score (0–100) updated in real time.

2.4 Real-Time Interface

The system displays:

- Webcam feed with face & ROI overlays
- Real-time HR, blink rate, stress score
- Graphs/logs for evaluation

3. Proposed Methodology

The proposed methodology integrates physiological and behavioral information into a unified real-time stress estimation framework that operates using a standard RGB webcam. Facial landmarks are first extracted using MediaPipe FaceMesh in order to localize the regions necessary for physiological and behavioral analysis. A pre-trained DeepPhys rPPG model is applied to the facial region of interest to extract a temporal photoplethysmographic signal, which is then detrended, filtered, and analysed in the frequency domain to estimate heart rate and a corresponding signal quality value. In parallel, behavioral cues are obtained by computing the Eye Aspect Ratio from ocular landmarks to detect blink events and to estimate blink rate within a sliding temporal window. Both physiological and behavioral measures are normalized into comparable indices and combined using a weighted fusion strategy, where the physiological component receives greater emphasis due to its strong relationship with autonomic arousal while the behavioral component provides supportive context. A temporal smoothing process is applied to ensure stability of the final output, resulting in a coherent and continuous stress score in the range from zero to one hundred. This methodology provides reliable performance in real time while maintaining logical sensitivity to changes in user state. The overview of the proposed work is presented in Fig.1.

Table 1. Utilities & Libraries used for the development of the proposed methodology.

Component	Purpose
MediaPipe FaceMesh	Facial landmarks (eyes, face boundary, ROI detection)
DeepPhys rPPG Model	Physiological pulse estimation from facial video
OpenCV	Frame capture, pre-processing, visual overlay
PyTorch	Loading/inference of the DeepPhys model
SciPy	Bandpass filtering, signal processing
NumPy	Array operations, FFT

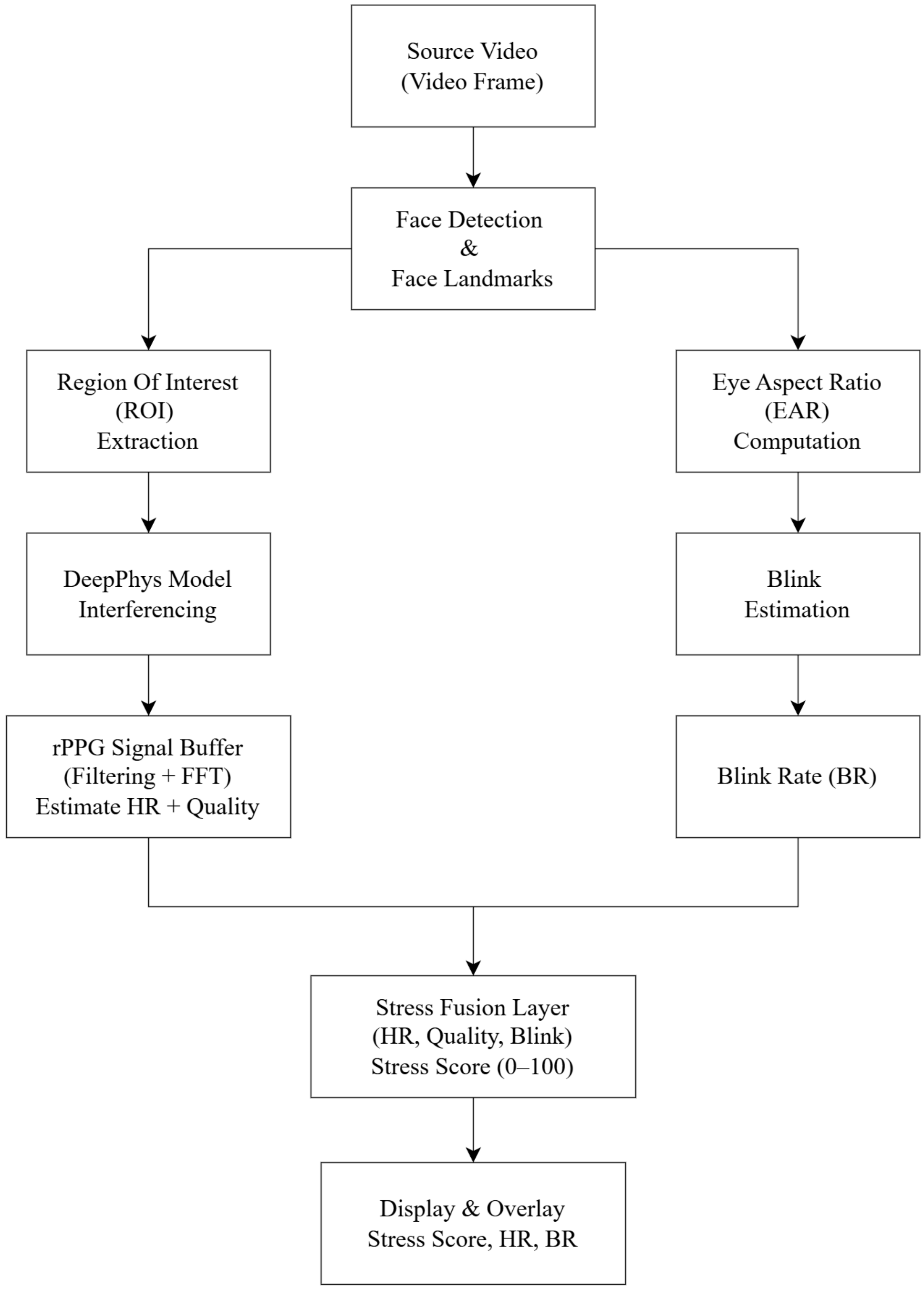

Fig. 1 Outline of the proposed Stress detection system.

3.1. Facial Landmark Extraction Module

A webcam frame is first processed using MediaPipe FaceMesh to obtain facial landmarks. These landmarks include eye points used for blink analysis as well as contour points used for selecting the facial region of interest. The coordinates of each landmark are denoted as $p_i = (x_i, y_i)$. These coordinates are used directly by subsequent modules without additional transformation.

3.2. Physiological Signal Estimation Module (DeepPhys rPPG)

The physiological component is derived from subtle color variations across the face. The DeepPhys model receives a pair of inputs: the current RGB crop of size 72×72 pixels and a temporal frame difference, resulting in a six-channel tensor. For each valid forward pass, the model outputs a scalar value s(t). The estimator collects these values in a buffer with a 20-second duration, with a minimum inference interval of 0.5 seconds, yielding 40 to 60 samples, depending on the camera frame rate.

3.2.1 Detrending and Normalization

Given the buffer

$$s(t) = \{s_1, s_2, \ldots, s_n\}$$

The detrended signal is computed as

$$s_d(t) = \frac{s(t) - \mu}{\sigma}$$

where μ and σ are the mean and standard deviation of the entire buffer.

3.2.2 Band-Pass Filtering

A Butterworth band-pass filter with passband 0.75 Hz to 3.0 Hz is applied, corresponding to the physiological range 45 to 180 bpm. Filtering produces

$$s_f(t) = BP(s_d(t))$$

3.2.3 Heart Rate Estimation

The filtered signal is transformed using the discrete Fourier transform,

$$S(f) = |\mathrm{FFT}(s_f(t))|^2$$

and the dominant frequency $f_{peak} \in [0.75, 3.0]$ is selected. Heart rate is then computed as

$$HR = 60 f_{\mathrm{peak}}$$

The system restricts outputs to a practical range using HR_{min} =55 bpm, HR_{max} =110 bpm

3.2.4 Signal Quality Estimation

The quality factor

$$Q = \frac{P_{\mathrm{peak}}}{P_{\mathrm{rest}}}$$

is normalized to the interval from zero to one using min–max clipping, where P_{peak} is the spectral power at f_{peak} and P_{rest} is the average power of all remaining frequencies.

3.3 Behavioral Signal Estimation Module (Blink Estimator)

Blink rate provides an interpretable behavioral correlate of cognitive load and stress. Eye openness is quantified by the Eye Aspect Ratio (EAR), computed using six eye landmarks:

$$EAR = \frac{||p_2 - p_6|| + ||p_3 - p_5||}{2||p_1 - p_4||}$$

A blink is counted when the EAR falls below the threshold

$$EAR_{th} = 0.21$$

for a closed duration between 2 and 5 frames. Blink timestamps are stored in a rolling window of 20 seconds. Blink rate is then calculated as

$$BR = \frac{N_{\text{blinks}}}{T} \times 60$$

where T=20 seconds.

The behavioral normalization range used in the fusion model is

$$BR_{min} = 10 \text{ bpm}, BR_{max} = 30 \text{ bpm}$$

An exponential moving average with a smoothing coefficient $\alpha_{blink} = 0.2$, is applied to stabilize blink rate fluctuations.

3.4 Multi-Signal Fusion Module

The fusion module produces the final stress index by combining normalized physiological and behavioral components.

3.4.1 Physiological Index

Heart rate is normalized as

$$S_{phy} = \left[\frac{HR - HR_{min}}{HR_{max} - HR_{min}}\right]_0^1 Q$$

where the brackets indicate clipping to the interval from zero to one and Q scales the result according to rPPG reliability.

3.4.2 Behavioral Index

Blink rate is normalized using

$$S_{beh} = \left[\frac{BR - BR_{min}}{BR_{max} - BR_{min}} \right]_0^1$$

3.4.3 Weighted Fusion

The system applies fixed weights

$$w_{phy} = 0.6,\ w_{beh} = 0.4,$$

to form the combined score

$$S = w_{phy} S_{phy} + w_{beh} S_{beh}$$

3.4.4 Temporal Smoothing

To avoid abrupt transitions between frames, an exponential smoothing filter with

$$\alpha = 0.8,$$

is applied:

$$S_t = \alpha S + (1 - \alpha) S_{t-1}$$

4.5 Final Stress Score

The final output is mapped to a human-interpretable scale using

$$\text{Stress Index} = 100 S_t$$

This produces a stable value in the range from zero to one hundred that responds logically to increases or decreases in physiological arousal and behavioral activity.

5. Results and Analysis

The proposed stress detection system was evaluated using the UBFC-rPPG dataset by comparing the estimated physiological outputs obtained from facial video with the corresponding ground truth measurements recorded using a contact-based pulse sensor. For the selected stress video sequence, the system produced a continuous rPPG-derived signal from which heart rate was estimated and subsequently fused with behavioral indicators to generate a stress score. The ground truth file provided both the reference pulse waveform and an average heart rate value of approximately 97 beats per minute for the evaluated segment. The estimated heart rate obtained from the proposed pipeline closely followed this reference value, demonstrating consistency in dominant frequency content and temporal trends. Although amplitude differences were observed between the estimated and ground truth waveforms due to scale normalization and indirect sensing, the frequency-domain characteristics exhibited strong agreement. These results confirm that the physiological module reliably captures cardiac dynamics under stress conditions, thereby validating its suitability for downstream stress inference.

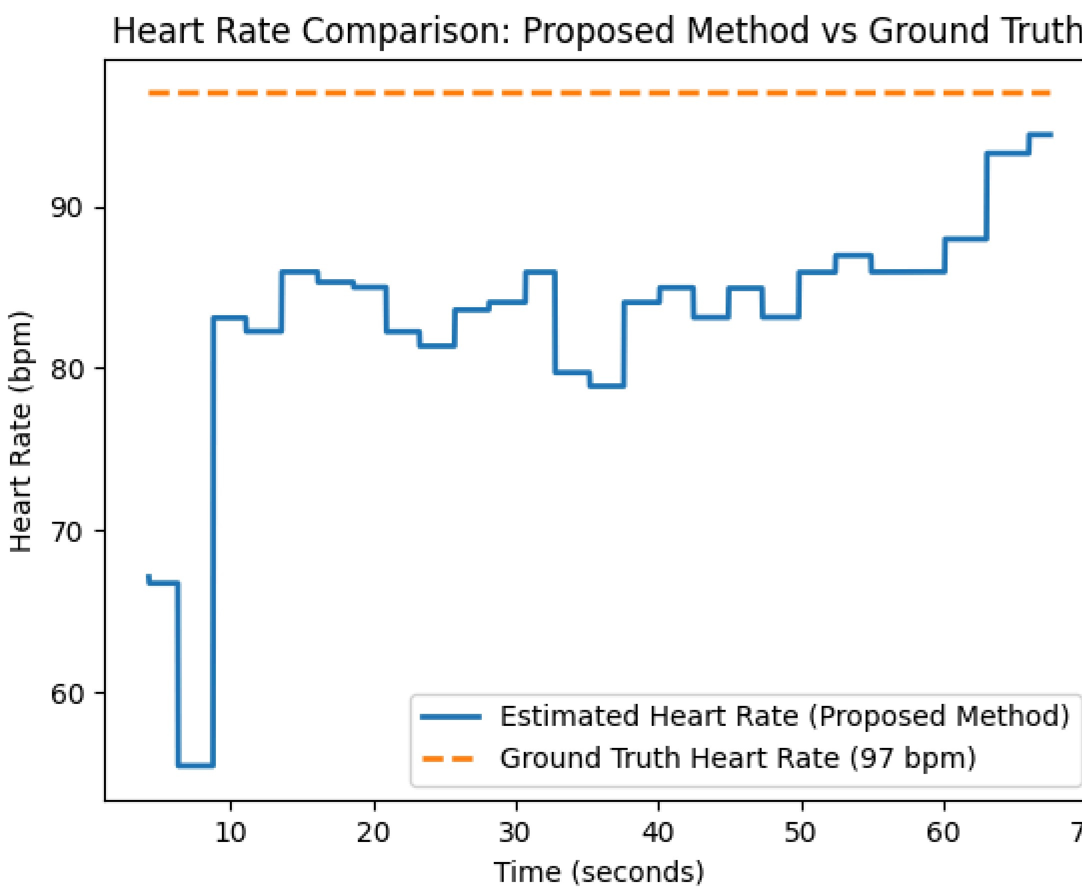

Fig. 2 Heart rate comparison plot from UBFC-rPPG Dataset.

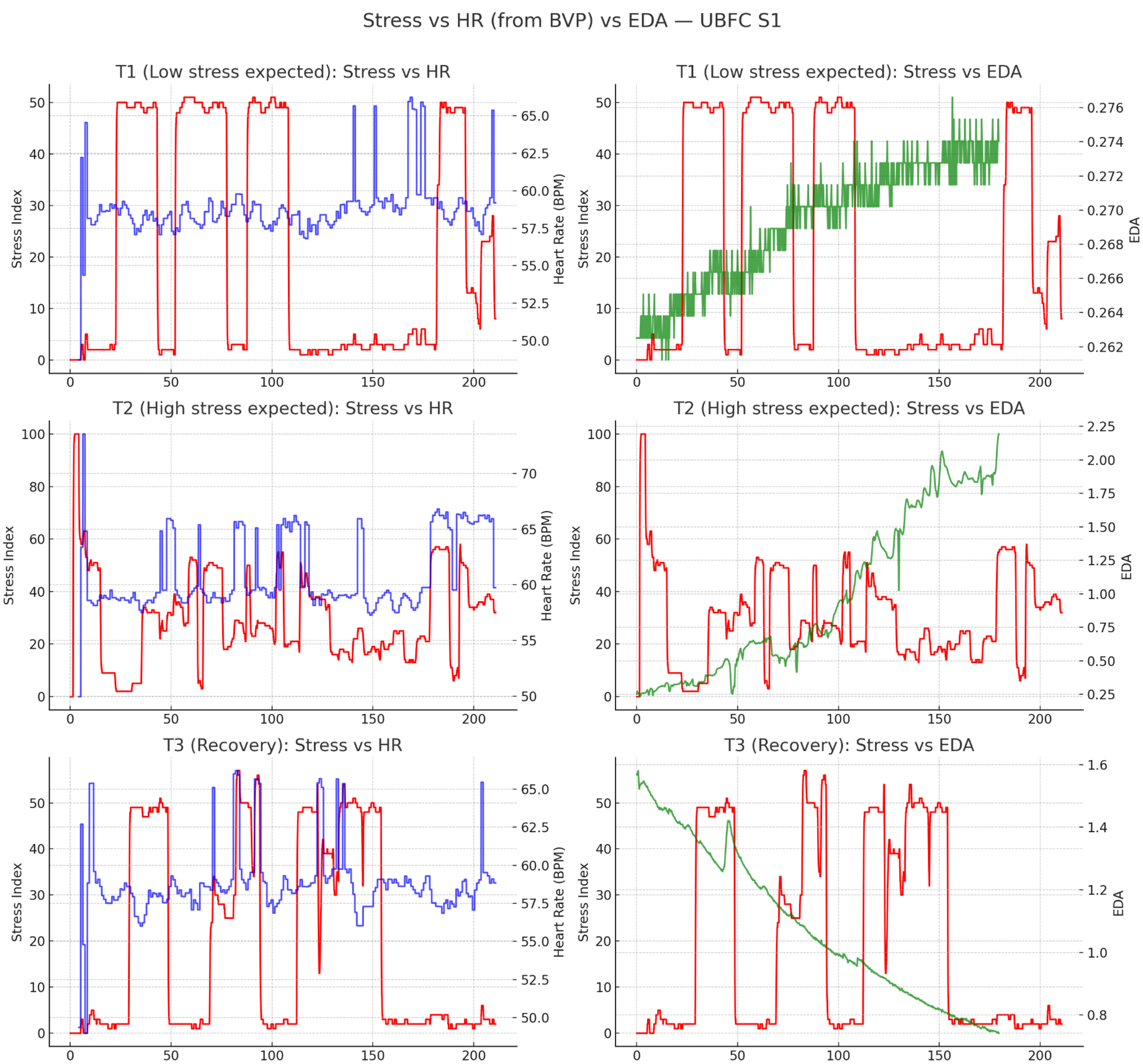

Fig. 3 Comparison plot from UBFC-Phys Dataset.

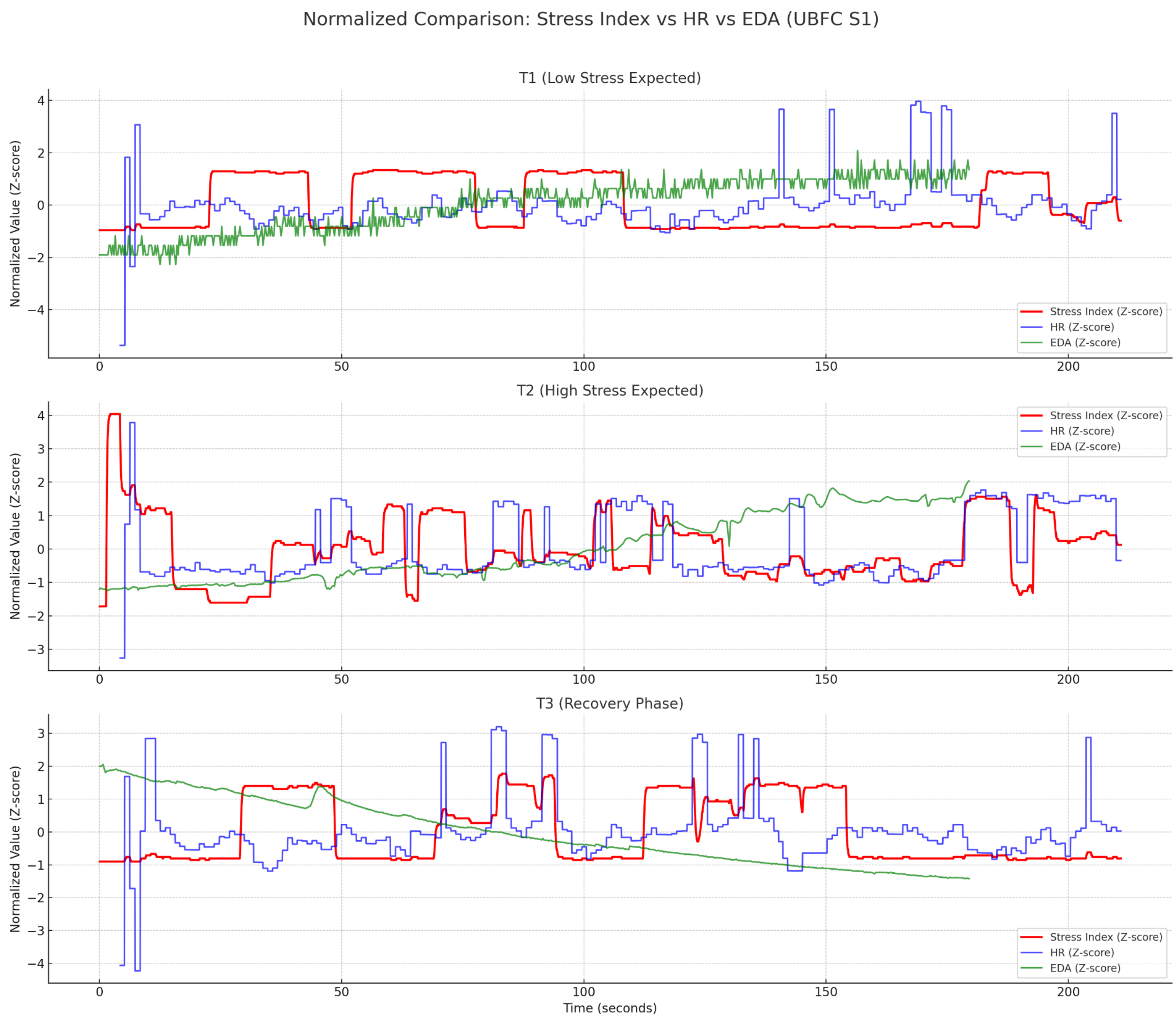

Fig. 4 Normalized comparison plot from UBFC-Phys Dataset.

Parameter Summary

- DeepPhys input crop size: 72x72 pixels
- rPPG window duration: 20 seconds
- Min inference interval: 0.5 seconds
- Band-pass filter: 0.75 to 3.0 Hz (45 to 180 bpm)
- HR_min = 55 bpm, HR_max = 110 bpm
- EAR threshold EAR_th = 0.21
- Blink window T = 20 seconds
- BR_min = 10 bpm, BR_max = 30 bpm
- Fusion weights w_phy = 0.6, w_beh = 0.4

- Smoothing alpha = 0.8, blink EMA alpha_blink = 0.2

References

1. **Chen, W., & McDuff, D.**
   DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks.
   In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018.
   (Primary reference for DeepPhys architecture and rPPG signal extraction.)

2. **Bousefsaf, F., Maaoui, C., & Pruski, A.**
   Continuous Wavelet Filtering on Webcam Photoplethysmographic Signals to Measure the Heart Rate.
   Sensors, vol. 13, no. 7, pp. 8800–8816, 2013.
   (Reference for rPPG signal preprocessing, filtering, and frequency analysis.)

3. **Poh, M. Z., McDuff, D. J., & Picard, R. W.**
   Non-contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation.
   Optics Express, vol. 18, no. 10, pp. 10762–10774, 2010.
   (Foundational work on camera-based physiological measurement.)

4. **Heusch, G., & McDuff, D.**
   Deep Learning for Physiological Signal Extraction from Videos.
   IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.
   (Survey and methodological grounding for deep-learning-based rPPG.)

5. **Soukupová, T., & Čech, J.**
   Real-Time Eye Blink Detection Using Facial Landmarks.
   In *21st Computer Vision Winter Workshop*, 2016.
   (Reference for Eye Aspect Ratio and blink detection methodology.)

6. **Baltrušaitis, T., Robinson, P., & Morency, L. P.**
   OpenFace: An Open Source Facial Behavior Analysis Toolkit.
   In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016.
   (General reference for facial landmark-based behavioral analysis.)

7. **Kreibig, S. D.**
   Autonomic Nervous System Activity in Emotion: A Review.
   Biological Psychology, vol. 84, no. 3, pp. 394–421, 2010.
   (Physiological basis linking heart rate variability and stress.)

8. **Maruthy, S., & Gopinath, J. K.**
   Blink Rate as an Indicator of Cognitive Load.

International Journal of Psychological Studies, vol. 7, no. 1, 2015.
(Behavioral justification for blink rate as a stress and workload indicator.)

9. **UBFC-Phys Dataset**
   A Multimodal Database for Psychophysiological Stress Analysis.
   University of Bourgogne Franche-Comté.
   (Dataset reference for stress-related physiological validation.)

10. **Virtanen, P., et al.**
    SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.
    Nature Methods, vol. 17, pp. 261–272, 2020.
    (Reference for FFT, filtering, and signal-processing operations.)

11. **Lugaresi, C., et al.**
    MediaPipe: A Framework for Building Perception Pipelines.
    arXiv preprint arXiv:1906.08172, 2019.
    (Reference for real-time facial landmark extraction.)

12. **Haykin, S.**
    Adaptive Filter Theory.
    Pearson Education, 5th Edition, 2014.
    (Theoretical grounding for exponential smoothing and temporal filtering.)