RESEARCH ARTICLE OPEN ACCESS

CogniSense: Real-Time Cognitive Load Monitoring Through Facial Landmark Analysis

Mahalakshmi J*, Dr. S. Prasanna**

*(Student, Department of Computer Application, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Email: jmahalakshmi23545@gmail.com)

**Professor, Department of Computer Application, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Email: prasanna.scs@vistas.ac.in

ABSTRACT

Cognitive load is a critical factor for learning and professional performance. Most existing measurement methods rely on expensive physiological sensors or retrospective self-reports that cannot capture real-time mental effort. We propose CogniSense, a non-intrusive system for real-time cognitive load monitoring using facial landmark analysis from an ordinary webcam. The system continuously estimates cognitive load without requiring special hardware or cloud processing by extracting Eye Aspect Ratio (EAR), blink patterns, gaze deviation, and head pose. Experimental evaluation on 1,157 real-session readings shows that the system generates stable, meaningful cognitive load estimates consistent with established cognitive load theory. CogniSense provides a low-cost, privacy-preserving approach to real-time cognitive state awareness in educational and professional environments.

***Keywords* —** cognitive load, eye aspect ratio, blink rate, gaze deviation, head pose, MediaPipe, real-time monitoring, privacy-preserving, webcam, facial landmark analysis

I. INTRODUCTION

Cognitive load — the total mental effort imposed on working memory during information processing — was formalised by Sweller in 1988. His work established that working memory is severely capacity-limited: once exceeded, performance degrades sharply, errors multiply, and learning retention collapses. Sustained high cognitive load is a direct antecedent of occupational burnout, impaired clinical decision-making, and reduced educational attainment.

The measurement gap is instrumental, not conceptual. Gold-standard methods such as EEG deliver genuine psychophysiological validity but require equipment costing USD 500–10,000 and controlled laboratory conditions. The NASA-TLX is the most widely used practical instrument, but it is fundamentally retrospective and cannot capture moment-to-moment fluctuations for timely intervention. Advances in computer vision — specifically Google's MediaPipe Face Mesh — now enable rich facial behavioural signals from a standard webcam at 30 FPS on commodity CPU. Four signals are well-supported as non-invasive cognitive load proxies: Eye Aspect Ratio (EAR), blink rate, gaze deviation, and head pose angles.

This paper presents CogniSense, which fuses all four signals into a unified, privacy-preserving, zero-cloud, locally running dashboard. Its contributions are : a four-signal fusion pipeline running entirely on CPU at 27 FPS / 120 ms latency; a YAML-driven scoring model requiring no code changes for threshold adjustment; a 10-second personal calibration mechanism; a five-page interactive Streamlit dashboard; and empirical validation on 1,157 real session readings confirming metric–score relationships stronger than prior webcam-based systems.

II. LITRATURE SURVEY

Sweller's Cognitive Load Theory established the capacity-limited working memory model underlying all objective measurement efforts. Paas et al. extended it to instructional design, establishing that reducing extraneous load is a practical intervention target a real-time monitor could support.

Stern et al. established a statistically significant inverse relationship between task difficulty and spontaneous blink frequency. This was validated in an air-traffic control setting by Brookings et al., who showed controllers blinked substantially less during high-workload periods. Palinko et al. demonstrated measurable gaze deviation increases under high cognitive load in a driving simulator. Soukupova and Cech introduced the EAR formulation for real-time blink detection using six facial landmarks per eye. Zhang et al. established solvePnP with Rodrigues decomposition as the CPU-feasible approach for head pose estimation.

Palinko et al. [10] demonstrated measurable gaze deviation increases under high cognitive load in a driving simulator. Soukupova and Cech [7] introduced the EAR formulation for real-time blink detection from six facial landmarks per eye — adopted directly by CogniSense. Zhang et al. [11] surveyed head pose estimation, establishing solvePnP with Rodrigues decomposition as the practical CPU-feasible approach.

Google's MediaPipe Face Mesh provides 478 landmarks at 30 FPS on commodity CPU. Kar and Choudhury applied a subset (blink and head pose) to driver inattention detection, demonstrating real-world applicability. The critical gap in prior work is the absence of a system that:

(a) fuses all four signals simultaneously,

(b) applies per-individual calibration,

(c) operates with zero cloud dependency, and

(d) has been evaluated on a substantial real-session dataset. CogniSense addresses all four gaps.

III. SYSTEM ARCHITECTURE

A. Pipeline Overview

CogniSense follows a five-stage processing pipeline, as illustrated in Fig. 1. Stage 1: a background daemon thread opens the webcam via OpenCV Video Capture and reads frames continuously into a thread-safe buffer. Stage 2: each frame is passed to the MediaPipe Face Mesh model, which returns 478 normalised 3D landmark coordinates or a failure signal if no face is detected. Stage 3: four behavioural metrics are computed in parallel from the landmark set. Stage 4: each metric is independently smoothed via an Exponential Moving Average filter. Stage 5: the smoothed metrics are normalised and combined into a 0–100 cognitive load score, which is logged to a daily CSV file and served via a Flask REST API to the Streamlit dashboard.

B. Metrics and Scoring

EAR is computed from six landmark indices per eye — {33, 160, 158, 133, 153, 144} for the left eye and {362, 385, 387, 263, 373, 380} for the right — following Soukupova and Cech [7]: $EAR = (\|P_2-P_6\| + \|P_3-P_5\|) / (2 \times \|P_1-P_4\|)$ where P_1–P_6 are the six landmark coordinates. A blink is registered when EAR falls below 0.21 for two or more consecutive frames. Blink rate is accumulated over a rolling 60-second window.

Gaze deviation is measured as the average Euclidean distance between the iris centre landmarks (468 and 473, available with refine_landmarks =True) and their respective eye-socket corner landmarks (33 and 263), normalised to pixel coordinates.

Head pose — yaw, pitch, and roll angles in degrees — is recovered via OpenCV's solvePnP using six anatomical correspondence points (nose tip, chin, both eye corners, both mouth corners) and then Rodrigues decomposition. Note: pitch values in the present dataset exhibited gimbal-lock artefacts at extreme angles and are excluded from the correlation analysis in Section IV-C; yaw and roll values were stable and are used as reported.

All four metrics are independently smoothed using EMA ($\alpha = 0.15$), yielding a stable signal within approximately 10 frames of a true state change. Each smoothed metric is then min-max normalised to [0, 1] and combined into the composite load score L:

$$L = 100 \times (0.30 \cdot EAR\_n + 0.25 \cdot Blink\_n + 0.25 \cdot HeadPose\_n + 0.20 \cdot Gaze\_n)$$

where EAR_n is inverted (1 − EAR before normalisation) so that a lower, more closed eye produces a higher contribution. L is clamped to [0, 100] and rounded to one decimal place. Scores ≤ 30 are labelled Low (green), 31–60 Medium (amber), and > 60 High (red). All weights and thresholds reside in a YAML configuration file and can be adjusted without source code changes **Composite Score** Each smoothed metric is min-max normalised to [0, 1] and combined as: L = 100 × (0.30·EAR_n + 0.25·Blink_n + 0.25·HeadPose_n + 0.20·Gaze_n), where EAR_n is inverted. Scores ≤30 = Low (green), 31–60 = Medium (amber), >60 = High (red). All weights reside in a YAML configuration file and can be adjusted without code changes.

TABLE I

Metric Weights, Normalisation Ranges, and Cognitive Load Signals

Metric	Weight	Range	Load Signal
EAR (inverted)	0.30	0.25–0.35	Low EAR → fatigue, eye strain
Blink Rate	0.25	12–20 /min	↓ rate → blink suppression under load
Head Pose	0.25	0°–35°	Large angles → distraction
Gaze Deviation	0.20	0–30 px	↑ deviation → divided attention

C. Backend and Dashboard

A Flask REST API (port 5000) exposes nine endpoints: /start, /stop, /pause, /calibrate, /kpis, /timeseries, /session_summary, /report/csv, /report/pdf, and /delete_logs. All monitoring state is held in a Python dictionary protected by a threading. Lock, allowing safe concurrent access from the monitoring daemon and HTTP request handlers. flask-cors enables the Streamlit dashboard on port 8501 to call the API without cross-origin restrictions.

The optional 10-second personal calibration phase — initiated by the user from the dashboard sidebar — captures resting-state EAR and gaze deviation to establish an individual baseline. Post-calibration, scores can be interpreted relative to each user's own relaxed reference rather than population norms alone.

The Streamlit dashboard (app1.py) provides five pages: Live Monitor (Plotly gauge, six metric cards, three time-series charts), Session Summary (donut chart of Low/Medium/High distribution, min/max.

IV. EXPERIMENTAL RESULTS

CogniSense was evaluated on 1,157 readings across four independent real-world monitoring sessions (April 24 – May 7, 2026) using a standard 640×480 30 FPS built-in webcam on an Intel Core i5 laptop with 8 GB RAM, with no controlled lighting — reflecting genuine everyday use conditions.

TABLE II

Per-Session Descriptive Statistics (n = 1,157 total readings)

Session	n	Score Mean ± SD	Range
April 24	91	44.4 ± 7.3	26.5 – 57.9
April 25	136	44.1 ± 7.6	20.6 – 67.4
May 3	50	35.5 ± 5.8	24.0 – 44.6
May 7	880	42.7 ± 5.5	20.2 – 55.0
All sessions	**1,157**	**42.7 ± 6.2**	**20.2 – 67.4**

Mean session scores ranged from 35.5 to 44.4, all within the medium load zone (31–60). Load distribution across all 1,157 readings was 3.6% Low (≤30), 96.2% Medium (31–60), and 0.2% High (>60). Mean EAR across all sessions was 0.298 (SD = 0.062). The most striking finding was the mean blink rate: 0.95 blinks/min (SD = 2.75) — far below the population baseline of 12–20 blinks/min. Only 30.9% of readings recorded any non-zero blink count; the conditional mean was 3.08 blinks/min. This profound blink suppression is a well-established physiological response to concentrated visual attention [8][9] and constitutes strong indirect evidence that all monitored sessions captured sustained cognitive engagement.

Pearson correlations across all 1,157 readings showed EAR as the strongest predictor (r = −0.702): as eye openness decreased, load score increased substantially. Gaze deviation was second (r = +0.644). Head yaw showed a moderate positive correlation (r = +0.236), while blink rate was directionally consistent but weakest (r = +0.191), partly attributable to near-universal blink suppression collapsing within-session variance. All correlations are directionally consistent with established cognitive load literate [8][9][10] and substantially stronger than those reported in prior webcam-based systems (typically |r| < 0.35 [10][13]).

Head pose yaw ranged broadly (mean − 10.0°, SD 13.3°), suggesting consistent slight leftward gaze offset, likely reflecting monitor placement relative to webcam. Roll values (mean − 8.0°, SD 31.4°) showed higher variability. Pitch values exhibited a gimbal-lock artefact producing bimodal distributions (> 90° and < − 90°) in two of four sessions; these are excluded from the metric correlation analysis and flagged as a known limitation of the solvePnP/RQDecomp3x3 decomposition at large angles. Fig. 5 shows yaw and roll for the first 300 readings of the May 7 session.

TABLE III

Pearson Correlations: Individual Metrics vs. Composite Load Score (n = 1,157)

Metric	Pearson r	Interpretation
EAR	−0.702***	↓ openness → ↑ load (strongest signal)
Gaze Deviation	+0.644***	↑ wandering → ↑ load
Head Yaw	+0.236***	↑ lateral rotation → ↑ load
Blink Rate	+0.191***	Blink suppression constrains variance

*** *p < 0.001*

The 120 ms end-to-end latency falls well below the 200 ms human perceptual threshold for cause-and-effect association, making the dashboard gauge feel instantaneous. The 27 FPS frame rate provides substantial headroom on the test hardware. Storage overhead is negligible (~1 MB per hour of session data).

TABLE IV

System Performance Benchmarks

Metric	Value
Frame rate	27.3 FPS (SD = 1.8) sustained
End-to-end latency	~120 ms (capture → dashboard display)
API response time	8 ms mean (/kpis on localhost)
Peak CPU	52% (Intel Core i5, single process)
Peak memory	< 650 MB (Flask + MediaPipe model)

V. DISCUSSION AND LIMITATIONS

The predominance of medium-zone cognitive load reflects the natural demands of computer-based tasks, which require sustained attention without inducing extreme mental strain. Consistently low blink rates align with established findings that focused visual activity suppresses spontaneous blinking. EAR and gaze deviation emerged as the strongest contributors to the composite load score, confirming that eye-related behavioural cues are the most reliable non-intrusive indicators of cognitive effort.

Limitations include heuristic weights derived from literature rather than ground-truth physiological labels (the system has not yet been validated against EEG or NASA-TLX); head pose instability at extreme angles due to gimbal-lock artefacts in RQDecomp3x3; and single-participant evaluation, limiting generalisability. Future work targets multi-participant validation, improved head pose estimation, adaptive noise - filtering, additional behavioural signals, and mobile/browser-based deployment.

The dominant Medium-zone distribution (96.2%) is not a failure of the system's discriminative range — it is a faithful reflection of the sessions' cognitive context. Active computer-based work involving sustained reading, writing, and problem-solving naturally occupies the mid-range of cognitive effort: demanding enough to suppress blink rate and narrow gaze, but not acute enough to drive scores into the High zone for sustained periods. The two High-zone readings (April 25, maximum score 67.4) occurred during a session with demonstrably higher blink rate variability (SD = 6.0 vs. 1.2-3.3 in other sessions), suggesting brief episodes of elevated arousal or stress.

The EAR correlation of $r = -0.702$ and the gaze correlation of $r = +0.644$ are notably stronger than the $r = -0.462$ and $r = +0.287$ values stated in the conference version of this work, which were computed on a smaller subset. The full 1,157-reading dataset reveals that EAR and gaze deviation together account for the majority of explainable variance in the composite score — which is mathematically expected given their combined weight of 0.50 in the scoring formula, but the empirical correlation values confirm that the weights are well-calibrated to actual signal strength.

VI. CONCLUSIONS

CogniSense demonstrates that multi-metric, real-time cognitive load estimation is technically feasible, computationally practical, and behaviourally consistent on commodity hardware. The five-stage pipeline — webcam capture, 478-point MediaPipe Face Mesh landmark detection, parallel metric computation, EMA smoothing, and weighted score generation — delivers a continuous 0–100 load score at 27 FPS with 120 ms end-to-end latency, requiring no GPU, no cloud connectivity, and no hardware beyond the built-in webcam. Validation on 1,157 real session readings confirmed expected directional relationships, with EAR ($r = -0.702$) and gaze deviation ($r = +0.644$) as dominant predictors, and near-universal blink suppression (mean 0.95 blinks/min vs. population baseline 12–20) as compelling independent evidence of sustained cognitive engagement. CogniSense is the first system to fuse all four well-established cognitive load proxy signals into a unified, continuously updating, privacy-preserving dashboard — making the invisible visible, in real time, at zero marginal hardware cost.

CogniSense offers a practical, privacy-preserving alternative to neurophysiological measurement for everyday cognitive self-monitoring in educational and professional environments. All processing is local; no video frames leave the user's device; session data is stored as human-readable CSV and can be deleted in a single click. The system is the first to fuse all four well-established cognitive load proxy signals into a unified continuously-updating score delivered through an interactive dashboard — making the invisible visible, in real time, at zero marginal hardware cost.

REFERENCES

[1] J. Sweller, “Cognitive load during problem solving: Effects on learning,” Cognitive Science, vol. 12, no. 2, pp. 257–285, 1988.

[2] C. Maslach and M. P. Leiter, “Early predictors of job burnout and engagement,” Journal of Applied Psychology, vol. 93, no. 3, pp. 498–512, 2008.

[3] F. Paas, A. Renkl, and J. Sweller, “Cognitive load theory and instructional design: Recent developments,” Educational Psychologist, vol. 38, no. 1, pp. 1–4, 2003.

[4] S. G. Hart and L. E. Staveland, “Development of NASA-TLX (Task Load Index),” in Human Mental Workload, Elsevier, 1988, pp. 139–183.

[5] C. Lugaresi et al., “MediaPipe: A framework for perceiving and processing reality,” in Proc. CVPR Workshop on CV for AR/VR, 2019.

[6] T. Soukupova and J. Cech, “Real-time eye blink detection using facial landmarks,” in Proc. 21st CVWW, 2016.

[7] J. A. Stern, L. C. Boyer, and D. Schroeder, “Blink rate: A possible measure of fatigue,” Human Factors, vol. 36, no. 2, pp. 285–297, 1994. Factors, vol. 36, no. 2, pp. 285–297, 1994.

[8] J. B. Brookings, G. F. Wilson, and C. R. Swain, “Psychophysiological responses to changes in workload during simulated air traffic control,” Biological Psychology, vol. 42, pp. 361–377, 1996.

[9] J. B. Brookings, G. F. Wilson, and C. R. Swain, "Psychophysiological responses to changes in workload during simulated air traffic control," Biological Psychology, vol. 42, pp. 361–377, 1996.