




Adaptive Scenario Control in Intelligent Training Simulators Using Fuzzy Logic: Integration of Eye-Tracking and Performance Error Analysis

Namig Huseynov 

Department of Artificial Intelligence Technology, National Aviation Academy, Baku AZ1045, Azerbaijan

Corresponding Author Email: namig.huseynov@naa.edu.az

Copyright: ©2026 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/isi.310325>

ABSTRACT

Received: 9 December 2025

Revised: 15 February 2026

Accepted: 20 March 2026

Available online: 31 March 2026

Keywords:

adaptive training, eye tracking, fuzzy inference, intelligent training simulator, performance errors, multimodal analysis

This paper proposes a fuzzy logic-based subsystem integrated into an Intelligent Training Simulator (ITS), designed to optimize training by leveraging real-time eye-tracking data and performance error analysis. The subsystem continuously estimates cognitive load, stress, and fatigue by combining eye-tracking features (e.g., fixation duration, saccade rate, blink rate, and PERCLOS) with error scores derived from training performance. Robust normalization techniques, such as median and Median Absolute Deviation (MAD), address individual variability, followed by temporal smoothing for stable data processing. A Mamdani-type fuzzy inference system maps these inputs to psychophysiological state estimates, which are then used to dynamically adjust training difficulty within predefined boundaries. The system's effectiveness was evaluated in a controlled study involving 20 participants, showing that the fuzzy logic-driven adaptation can maintain stable cognitive load and stress levels while improving task performance in controlled conditions. The results highlight that the approach not only supports stable learning dynamics but also emphasizes interpretability, transparency, and real-time feasibility, making it suitable for adaptive training systems that require explainable decision-making and can be deployed in real-world settings. This work provides a foundation for future developments in adaptive training systems incorporating multimodal data for personalized learning.

1. INTRODUCTION

Intelligent training simulators (ITS) can be viewed as human-in-the-loop information systems that continuously acquire multimodal behavioral data, transform these data into state estimates, and apply decision rules to regulate task conditions. In this setting, adaptation is not limited to final scoring. Instead, the system performs a closed-loop information-processing function in which sensing, integration, estimation, and control are linked in real time under uncertainty and inter-user variability. Immersive simulators support this approach because scenario parameters such as visibility, environmental disturbances, and target behavior can be adjusted in real time while preserving the task structure.

Eye tracking provides a non-invasive stream of behavioral information that can be collected continuously during interaction. Recent reviews describe its use in immersive and VR settings for assessing workload, attention, and fatigue [1-5]. Fixation duration and saccade dynamics capture aspects of information processing and visual search [6-8]. Blink-related metrics and PERCLOS are often associated with vigilance and fatigue and have been validated in applied monitoring contexts [9-13].

A second issue concerns the decision layer. Many adaptive training approaches rely on black-box machine learning models, which can reduce transparency and complicate validation in contexts where interpretability and reliability are

especially important. Interpretable rule-based decision models remain relevant when the system must expose its reasoning and allow expert adjustment. Fuzzy inference systems provide a framework to integrate heterogeneous indicators, represent uncertainty, and encode expert knowledge through linguistic rules [14-16].

In this work, we propose a real-time fuzzy scenario-control loop that combines robustly normalized eye-tracking features with a simulator-derived performance error score. The features are aggregated in sliding windows and processed by a Mamdani-type fuzzy model to estimate cognitive load, stress level, and fatigue. The subsystem combines eye-tracking features aggregated in sliding windows with a simulator-derived performance error score. A Mamdani-type fuzzy model outputs estimates of cognitive load, stress level, and fatigue, which are then used by a scenario manager to regulate difficulty within predefined limits. The approach is evaluated in a controlled study with 20 participants by comparing fixed-scenario training with adaptive fuzzy-driven training. The main contributions of this work are threefold. First, we introduce a task-oriented fusion strategy that combines personalized, robustly normalized eye-tracking indicators with a simulator-derived integrated error score on a common decision scale. Second, we define an interpretable scenario-control logic in which the inferred trainee states are translated into bounded difficulty adaptation through explicit rules and stability constraints. Third, the experimental validation is

designed to assess the practical value of the proposed adaptive control loop, specifically whether it can stabilize inferred user-state trajectories and maintain task performance more effectively than a fixed training policy.

From an information-systems perspective, the proposed framework can be interpreted as a real-time human-in-the-loop architecture for multimodal data integration, latent state estimation, and adaptive decision management in simulator-based training.

2. RELATED WORK

2.1 Eye tracking for assessing cognitive load, attention, and fatigue

Eye tracking has become a widely adopted non-invasive method for inferring learners' internal states in interactive and immersive environments. Recent surveys emphasize its growing relevance in educational and training settings, particularly within virtual and augmented reality applications [1-5].

Pupillometry has been repeatedly shown to be a sensitive indicator of cognitive load, reflecting changes in mental effort during learning and problem solving [17, 18]. Fixation duration and saccadic patterns have been used to characterize perceptual demands and attentional allocation [6-8], whereas eyelid-closure related measures, such as blink rate and PERCLOS, are commonly associated with fatigue and reduced vigilance [9-13]. Together, these measures enable continuous monitoring of attention and alertness without interfering with task execution.

More recently, eye tracking has also been used to assess performance and skill acquisition in complex tasks such as programming comprehension, surgical training, and professional vehicle operation [19-21] demonstrating its applicability beyond controlled laboratory conditions.

2.2 Adaptive learning and intelligent training systems

Adaptive learning systems aim to personalize instruction based on a learner's performance and internal state [22]. Recent work highlights the role of artificial intelligence in enabling dynamic adaptation of content, difficulty, and feedback [23-27].

Several studies have proposed leveraging physiological and behavioral signals to drive adaptation in immersive training environments. For instance, Gehrke et al. [28] demonstrated that gaze tracking can be used to detect cognitive load in real-time and adjust training difficulty accordingly. Other studies [29, 30] investigated gaze-driven adaptive interfaces and attention-training systems, reporting improvements in both efficiency and task safety.

Systematic reviews further suggest that immersive VR training can improve learning outcomes when combined with appropriate instructional design and adaptation mechanisms [31-33]. However, many existing systems rely on opaque machine learning models, which limits transparency and explainability.

2.3 Multimodal approaches to workload and fatigue detection

To improve robustness, a number of studies combine eye

tracking with additional physiological and behavioral signals. In this study [34], eye-movement measures were analyzed alongside performance indicators to estimate pilot workload. Vulpe-Grigorasi et al. [35] and Khan et al. [36] likewise showed that integrating eye tracking with other biosignals can improve the accuracy of cognitive-load estimation.

In safety-critical contexts such as driving, multimodal systems have been proposed for detecting fatigue and drowsiness by integrating facial, ocular, and behavioral features [37-39]. Collectively, these works highlight the value of fusing multiple weak indicators to obtain reliable state estimates in real-world conditions.

2.4 Fuzzy logic and interpretable models for adaptive decision-making

Fuzzy logic provides a natural framework for modeling uncertainty, vagueness, and inter-individual variability in human-centered systems. Recent studies have employed fuzzy-based approaches to integrate heterogeneous signals and support interpretable decision-making.

Fuzzy and hybrid methods are also well suited for adaptive learning, as they allow expert knowledge to be encoded through linguistic rules while remaining flexible to context and individual differences [14, 15]. Compared with "black-box" deep learning approaches, fuzzy systems offer explicit reasoning pathways that are easier to inspect, validate, and refine—an important advantage in educational and safety-critical applications [16].

Despite the relevance of the studies reviewed above, several gaps remain. First, much of the prior work uses eye tracking or other behavioral and physiological signals mainly for monitoring, classification, or offline assessment, rather than for closed-loop scenario control in real time. Second, relatively limited attention has been given to adaptive training frameworks that combine eye-tracking indicators with simulator-derived execution errors in a unified multimodal decision pipeline. Third, although adaptive systems are widely studied, many existing approaches rely on black-box predictive models, which makes the decision layer less transparent and harder to inspect or adjust in training contexts where robust and interpretable adaptation is important. These limitations motivate the present study, which addresses the problem as a human-in-the-loop information system that integrates multimodal behavioral evidence, estimates trainee-relevant latent states, and applies interpretable rule-based decision management for adaptive scenario regulation.

3. ARCHITECTURE OF THE INTELLIGENT TRAINING SIMULATOR WITH A FUZZY SUBSYSTEM

3.1 Overall architecture concept

The proposed ITS is implemented as a closed-loop training system in which the trainee's actions in a physical training room are continuously mapped to the evolution of a virtual Unity scenario and used for adaptive difficulty control. Training is conducted in a dedicated room where the simulator view is projected onto a screen surface, as shown in Figure 1, while the trainee holds a training mock-up that serves as the input device for aiming and triggering a shot.

The system operates on two synchronized data streams. The

first stream is produced by the eye tracking subsystem and captures the dynamics of the trainee’s visual behavior. In real-time, oculomotor measures including fixations, saccades, blinks, and PERCLOS are computed. They are then aggregated in sliding windows and normalized relative to an individualized baseline. The second stream is generated by the Unity simulator and reflects task execution within the current scenario. It includes shot timestamps, grenade impact events, temporal deviations in user response, and attempt outcomes. These data are used to compute an integral error indicator, ErrScore, which serves as a compact behavioral proxy of performance.

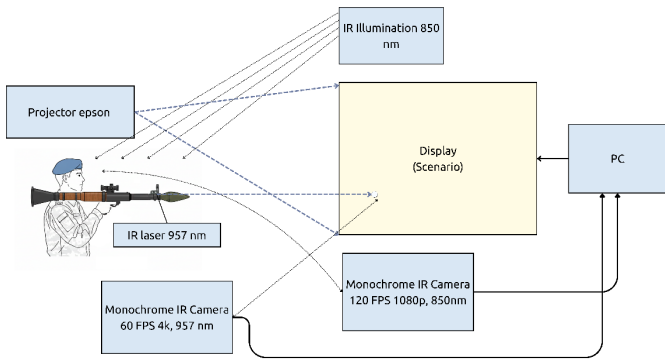


Figure 1. Hardware layout of the training room with infrared eye tracking

Psychophysiological and behavioral information is fused within a joint processing module, where a unified feature representation of the user’s state is formed. This representation is then provided to a Mamdani-type fuzzy subsystem that outputs interpretable estimates of learner-relevant states. The inferred states include cognitive load, stress level, and fatigue. Based on these estimates, the scenario manager adjusts difficulty parameters within predefined bounds. It increases or decreases scenario demands as a function of the current state and execution quality. The architecture therefore supports continuous feedback and enables an adaptive learning trajectory, while preserving decision interpretability through explicit fuzzy rules.

3.2 Architecture of the fuzzy subsystem

Figure 2 illustrates the ITS architecture. The central element is the user, the trainee, who performs tasks within a selected scenario and continuously interacts with the ITS. During operation, the simulator records user actions, registers outcomes, and generates execution events, including errors.

In the diagram, the Scenario corresponds to a specific training mission in which the user practices target skills. The ITS initiates the scenario, supervises its execution, and receives behavioral data from the user. For training adaptation, an oculomotor metrics logging module provides measures such as saccade rate, blink rate, fixation duration, and PERCLOS. These features reflect the temporal dynamics of attention, workload, and fatigue, and are subsequently processed in the interpretation module. In addition, simulator-derived error events are used that is events extracted from ITS logs when the user acts prematurely or too late, skips steps, or responds in an untimely manner.

Eye tracking data and error events are fed into the Eye Tracking and Error Analysis block, where joint analysis is performed. This includes window-based aggregation,

computation of derived indices, and formation of a feature representation of the user’s state. The output of this block is a set of diagnostic features that are forwarded to the Fuzzy Decision Block.

The Fuzzy Decision Block implements a fuzzy decision-making logic: based on the aggregated features, the system infers a user state relevant to the learning process, including cognitive load, attention, and fatigue. The inferred user state is then passed to the Scenario Manager, which is responsible for applying control actions to the scenario component of the training simulator.

Within the proposed model, the Scenario Manager performs one of the following actions:

- 1) Generation of a new scenario and assignment of a difficulty level.
- 2) Regularization of the current scenario parameters.
- 3) No action, if the system determines that the current training mode is optimal.

After the decision is made, the Scenario Manager applies the selected control action to the Scenario for subsequent adaptation.

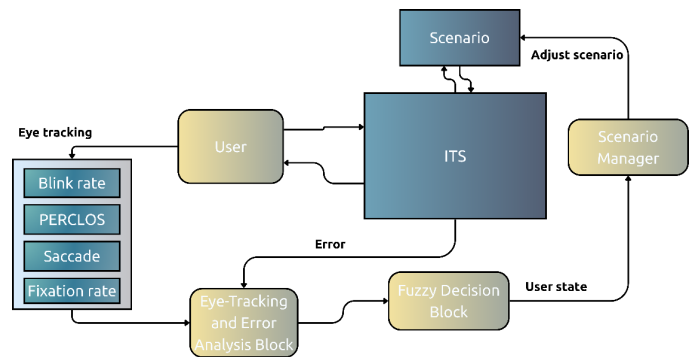


Figure 2. Intelligent training simulator (ITS) architecture with fuzzy subsystem

3.3 Data acquisition layer

Data acquisition is performed in a training room using the simulator’s integrated hardware and software configuration in Figure 3. The Unity visual scene is displayed on a screen via a projector, while the control PC simultaneously runs the simulation and logs all events. The eye tracking subsystem is built around a monochrome IR camera operating at 120 FPS (1080p) with 850 nm IR illumination. This enables robust eye capture that is largely independent of ambient lighting. From the video stream, the following oculomotor measures are computed in real-time: saccade rate, fixation duration, blink rate, and PERCLOS. All measurements are stored as time series with precise timestamps to support subsequent aggregation into windows W_k and alignment with simulator events.

An additional aiming module is based on a 980 nm IR laser mounted on the weapon mock up and a monochrome 60 FPS (4K) IR camera that tracks the laser spot on the projection surface. On the processing side, the spot is converted into normalized coordinates $(U, V) \in [0, 1]$ and transmitted via UDP. Unity interprets these coordinates as the current on-screen aiming point. The shot event is generated through a hardware trigger. A microcontroller detects the press and transmits a signal to Unity, where it is interpreted as the shot initiation time. This provides a direct link between the trainee’s physical action and the virtual process.

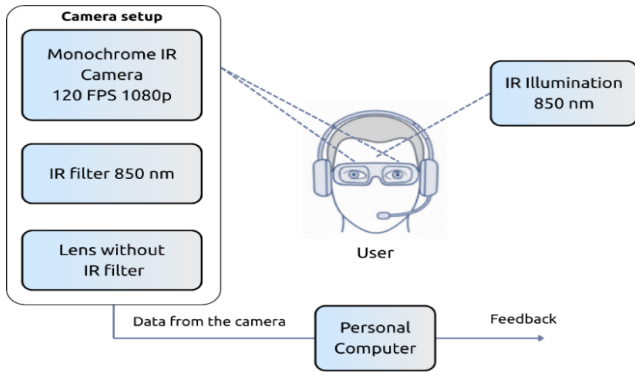


Figure 3. Oculomotor data acquisition setup

After a shot, Unity spawns a grenade that follows a realistic ballistic model and reaches the target object with a delay determined by flight dynamics. The impact event is registered by Unity's collision component. A direct hit is defined as the grenade colliding with the armored target collider. Collision with any other scene object is treated as a miss. In parallel, the system logs the moment a target appears on screen and computes the user's reaction time deviations. This enables discrimination between premature and delayed shots relative to the target detection moment.

To synchronize all data sources, a unified timeline is maintained on the control PC. Oculomotor measurements and simulator events are aligned into a common observation structure using sliding windows W_k with duration $T = 10$ s. On this basis, ErrScore is computed and fuzzy inference is subsequently performed.

4. EYE TRACKING FEATURES AND NORMALIZATION

In the proposed subsystem, eye tracking is used as a source of stable indicators of the trainee's current state during scenario execution. Within each window W_k , the following features are computed.

Fixation Duration is the mean fixation duration in milliseconds. It reflects the depth of information processing. An increase may indicate either sustained focus or emerging difficulty, so it is interpreted jointly with errors and task context. Saccade Rate is the saccade frequency measured in saccades per minute. It is linked to attention switching and interface scanning. An excessive increase can indicate search behavior under uncertainty. Blink Rate is the blink frequency measured in blinks per minute. It is commonly treated as a marker of reduced alertness or changes in cognitive load. PERCLOS is the proportion of time the eyes are closed within the window W_k .

A key practical challenge in applying eye tracking in training systems is pronounced interpersonal variability in oculomotor measures. This variability arises from anatomical and behavioral differences, different baseline blink rates, viewing strategies, and the sensitivity of some metrics to recording conditions. Given this variability, it is considered appropriate to normalize to an individual baseline and to focus on relative changes when comparing participants [40, 41].

To ensure cross user comparability and to improve the robustness of fuzzy rules, features are normalized with respect to an individual baseline and mapped to a common scale from zero to one. The baseline was defined using a two-minute

calibration interval at the beginning of each session under controlled and stable task conditions. During this phase, participants were placed in the baseline training scenario with a single stationary target, favorable environmental conditions (no fog, no wind, no precipitation), and identical visual settings across sessions. Participants were instructed to maintain a relaxed posture, visually explore the scene, and perform a small number of familiarization actions, including aiming and optional test shots, without performance pressure or adaptive difficulty changes. This calibration phase was designed to establish an operational individual reference level for oculomotor behavior under low task demand rather than to capture a state of complete rest.

For each metric $x(t)$ in the baseline sample X_{base} , robust location and scale parameters are estimated.

$$m = \text{median}(X_{base}), \quad s = \text{MAD}(X_{base}) \quad (1)$$

Robust standardization is then applied.

$$z^*(t) = \frac{x(t) - m}{s + \varepsilon} \quad (2)$$

Here ε is a small constant introduced for numerical stability. Compared with the standard z-score, the use of the median and Median Absolute Deviation (MAD) reduces the influence of outliers and is more reliable for real world eye tracking data.

Since the fuzzy model operates on a unified axis from zero to one, values $z^*(t)$ are transformed into a normalized quantity $u(t) \in [0, 1]$ with symmetric anchoring around the baseline.

$$u(t) = \min \left(1, \max \left(0, 0.5 + \frac{z^*(t)}{2z_{max}} \right) \right) \quad (3)$$

The parameter $z^*(t)$ defines the working range of robust deviations. In this work, $z_{max} = 3$.

To increase stability of real-time decisions, the normalized features are additionally smoothed using exponential averaging.

$$\bar{u}_k = \alpha u_k + (1 - \alpha) \bar{u}_{k-1}, \quad (4)$$

Here k is the index of window W_k , and $\alpha \in [0, 1]$ controls the tradeoff between sensitivity and stability. Smoothing reduces the impact of short-term spikes and prevents frequent adaptation toggling. This supports consistent use of linguistic terms Low, Medium, High and improves interpretability under inter-individual differences.

Because the initial calibration phase may include transient novelty or adaptation effects, all baseline windows were excluded from subsequent analysis and performance comparison. Baseline statistics were used solely for individualized normalization, ensuring that any residual initial arousal did not directly influence the reported Fixed versus Adaptive results.

5. ERROR ANALYSIS IN THE SIMULATOR

The error analysis module evaluates task execution quality

using Unity simulator events and shot telemetry from the RPG-7V2 training mock-up. The system records the shot outcome, hit or miss, and the temporal deviation of the user response, rushing or being late. All metrics are computed in sliding windows and synchronized by timestamps with the oculomotor features so that they can be jointly used in fuzzy inference.

In the setup, the aiming direction is determined by an IR beam that produces an aiming spot on the projection screen. An IR camera detects the spot and transmits normalized coordinates $(U, V) \in [0,1]$ via UDP. On the PC, these values are mapped to screen coordinates $x^{aim}(t) = V(t) \cdot (Screen\ width)$, $y^{aim}(t) = U(t) \cdot (Screen\ height)$.

These coordinates are treated as the observed aiming point at the moment of firing. In the software implementation, the firing moment is registered by a trigger press event sent by the microcontroller and recorded in Unity.

Each shot $shot_i$ is characterized by the trigger time t_i^{fire} and the grenade impact event time t_i^{impact} , generated by the projectile hit handler. A target hit is defined strictly by a direct collision with the armored target collider. We set $hit_i = 1$ if the grenade collides with the target object, the tank, and $hit_i = 0$ if the collision occurs with any other scene object. The outcome error is therefore defined as

$$E_i^{miss} = \begin{cases} 1, & hit_i = 0 \\ 0, & hit_i = 1 \end{cases} \quad (5)$$

To estimate reaction time, a target visibility event is introduced. For each target j is recorded as the first moment when the target becomes observable on the screen. Given the current target position $p_j(t)$, the screen projection $s_j(t) = (x_j(t), y_j(t))$ is computed using the standard Unity transformation `WorldToScreenPoint`. The visibility time t_j^{vis} is defined as the first t for which the target lies in front of the camera and $s_j(t)$ falls within the screen region.

Since multiple targets can be present simultaneously and the order of engagement is not fixed, each shot is associated with the target that the user was effectively aiming at when the trigger was pressed. Let $a(t_i^{fire})$ denote the aiming point at firing time obtained from the IR spot, and let $s_j(t_i^{fire})$ be the screen projection of target j at the same time. The associated target is then selected as

$$j^*(i) = \arg \min_{j \in V(t_i^{fire})} \|a(t_i^{fire}) - s_j(t_i^{fire})\| \quad (6)$$

where, $V(t)$ denotes the set of targets visible on the screen at time t . After selecting the associated target, reaction time is computed as

$$RT_i = t_i^{fire} - t_{j^*(i)}^{vis} \quad (7)$$

Based on RT_i , two classes of temporal errors are defined, with thresholds depending on scenario difficulty D .

$$E_i^{late} = \begin{cases} 1, & RT_i > \tau_{late}(D) \\ 0, & else \end{cases} \quad (8)$$

$$E_i^{rush} = \begin{cases} 1, & RT_i < \tau_{rush}(D) \\ 0, & else \end{cases}$$

The thresholds $\tau_{late}(D)$ and $\tau_{rush}(D)$ are specified as a function of difficulty level. As difficulty increases, the admissible reaction time range becomes narrower, reflecting higher demands on decision speed.

Within each window W_k , relative error frequencies are computed over shots falling into the window

$$\begin{aligned} MissRate_k &= \frac{\sum_{shot_i \in W_k} E_i^{miss}}{N_k + \varepsilon}, \\ LateRate_k &= \frac{\sum_{shot_i \in W_k} E_i^{late}}{N_k + \varepsilon}, \\ RushRate_k &= \frac{\sum_{shot_i \in W_k} E_i^{rush}}{N_k + \varepsilon}, \end{aligned} \quad (9)$$

where, N_k is the number of shots in the window.

A single integrated error indicator is then formed and used as the only linguistic variable derived from the simulator.

$$\begin{aligned} ErrScore_k &= \\ w_m \cdot MissRate_k + w_l \cdot LateRate_k + w_r \cdot RushRate_k, & \quad (10) \\ w_m + w_l + w_r &= 1. \end{aligned}$$

Since the error rates lie in $[0,1]$ and the weights are normalized, $ErrScore_k \in [0,1]$. In the baseline configuration, expert-defined weights are used with $w_m = 0.6$, $w_l = 0.25$, $w_r = 0.15$, where misses have the largest contribution as an outcome error. To confirm robustness, a sensitivity analysis is performed. When weights vary within plus or minus 0.1 while preserving the priority constraint $w_m \geq w_l, w_r$, the qualitative conclusions in the comparison of fixed and adaptive modes remain unchanged. If there are no shots in the window, $N_k = 0$, the error rates are set to zero and $ErrScore_k$ is not increased.

For real-time stability, $ErrScore_k$ is additionally smoothed using exponential averaging

$$\overline{ErrScore}_k = \alpha ErrScore_k + (1 - \alpha) \overline{ErrScore}_{k-1} \quad (11)$$

The smoothed value $ErrScore_k \in [0,1]$ is then provided to the fuzzy subsystem input and interpreted using the linguistic terms Low, Medium, and High. The smoothed quantity $ErrScore_k$ is updated only as new events become available.

6. FUZZY MODEL INPUTS OUTPUTS AND MEMBERSHIP FUNCTIONS

6.1 Input and output linguistic variables

This section describes a Mamdani-type fuzzy model that transforms normalized oculomotor features and an integrated execution error score into interpretable estimates of the trainee state. All input features are first mapped to a common scale from 0 to 1 using individualized robust normalization and window-based smoothing over sliding windows W_k .

Inputs are constructed from the preprocessed eye tracking features and the error indicator. To improve robustness, all inputs are expressed on the normalized scale from 0 to 1. They are then interpreted as linguistic variables with the terms Low, Medium, and High, as summarized in Table 1.

Table 1. Types and descriptions of linguistic variables

Variable	Type	Description
FixDur	Input	Aggregated fixation duration within window W_k
SaccRate	Input	Saccade rate
BlinkRate	Input	Blink rate
PERCLOS	Input	Proportion of eye closure time within the window
ErrScore	Input	Integrated execution error score combining miss, late response, and rush response
CognitiveLoad	Output	Estimated cognitive load of the trainee
StressLevel	Output	Estimated stress level during mission execution
Fatigue	Output	Estimated fatigue and reduced alertness

6.2 Membership functions

To support interpretability and stable real-time operation, triangular and trapezoidal membership functions were defined on the normalized input and output domain $[0,1]$. After individualized robust normalization based on the median and MAD and subsequent mapping to a unified scale, all features represent baseline-relative deviations, where $u = 0.5$ corresponds to typical baseline behavior. Values closer to 0 and 1 indicate sustained negative and positive deviations relative to the individual norm.

For most input variables and for all output variables, a shared set of membership parameters was used to preserve a consistent interpretation of the linguistic terms Low, Medium, and High across the fuzzy inference system. Under this representation, the linguistic labels are not tied to absolute physical units, but to the magnitude of a robust baseline-relative deviation. This choice reduces the number of free parameters, simplifies expert inspection of the rule base, and supports transparent operation in training applications where interpretability is important.

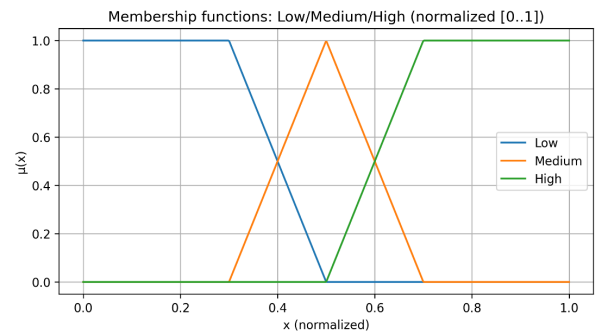
Despite the shared parametrization, not all oculomotor features have identical physiological semantics. In particular, PERCLOS is a fatigue-related indicator with a more direct interpretation than other eye-tracking measures. Therefore, a slightly more sensitive mapping was applied to PERCLOS by shifting the onset of the High term toward lower normalized values. Specifically, for PERCLOS the trapezoidal High membership function was set to $(0.50, 0.70, 1.00, 1.00)$, while fixation duration, saccade rate, blink rate, and all outputs used the shared parameters in Table 2. Fixation duration and saccade-related metrics were retained under the common parametrization because their interpretation is more task-dependent and can be ambiguous when considered in isolation.

Table 2. Parameters of input and output variables

Variable	Type	Description
Low	trapezoidal	$(0.00, 0.00, 0.25, 0.45)$
Medium	triangular	$(0.30, 0.50, 0.70)$
High	trapezoidal	$(0.55, 0.75, 1.00, 1.00)$

To verify that controller behavior is not overly sensitive to this adjustment, we conducted a limited robustness check by varying the PERCLOS membership breakpoints within ± 0.05 on the normalized axis while keeping all other membership functions, fuzzy rules, and control parameters unchanged. The

qualitative behavior of the fuzzy outputs and the Fixed versus Adaptive comparison remained unchanged, indicating that the results do not depend on a single hand-tuned configuration. Figure 4 illustrates the reference membership functions.

**Figure 4.** Reference membership functions for Low, Medium, and High on the normalized $[0,1]$ axis

While a shared membership parametrization simplifies expert inspection and reduces the number of free parameters, not all oculomotor features have identical physiological semantics. In the proposed system, most input variables are interpreted as robust deviations from an individualized baseline rather than as absolute physical quantities. However, PERCLOS is a well-established fatigue-related indicator with a more direct physiological interpretation than other eye-tracking features.

6.3 Design rationale for membership functions, rule base, and ErrScore weighting

The fuzzy subsystem was designed as an expert-guided and pilot-informed interpretable controller rather than as a purely data-driven model. Accordingly, the membership functions, rule base, and integrated error-score weights were specified using a combination of domain knowledge, practical simulator requirements, and observations from preliminary pilot sessions.

The common Low/Medium/High parameterization used for most input variables and for all outputs was introduced to maintain a consistent linguistic interpretation on the normalized $[0,1]$ axis and to reduce the number of free design parameters. After individualized robust normalization relative to baseline, all inputs represent baseline-relative deviations rather than absolute physiological quantities. Under this representation, a shared parameterization simplifies expert inspection of the rule base and improves interpretability in real-time operation. PERCLOS was treated differently because it has a more direct and established association with fatigue and vigilance degradation than fixation duration, saccade rate, or blink rate considered in isolation. For this reason, the onset of the High membership region for PERCLOS was shifted slightly toward lower values in order to increase sensitivity to early fatigue-related changes.

The rule base was constructed in two stages. First, an initial set of monotonic rules was formulated from expert knowledge about the expected qualitative relationship between eye-tracking variables, execution errors, and the latent states of cognitive load, stress level, and fatigue. For example, sustained increases in PERCLOS and blink-related measures were linked to fatigue, while elevated error score together with intensive visual search behavior was linked to stress and

higher cognitive demand. Second, the compact 25-rule set reported in the paper was refined during pilot observations of simulator sessions in order to better handle mixed cases and avoid unstable responses in ambiguous situations. In particular, the interaction rules were introduced to cover cases in which single indicators could not be interpreted reliably on their own.

The final rule base was reviewed for interpretability and consistency. Rules with clearly redundant consequences were removed, and the remaining rules were checked to ensure that no single antecedent combination produced logically incompatible outputs without aggregation. In situations where multiple rules with different consequents were activated simultaneously, conflicts were resolved through the standard Mamdani max-min aggregation and centroid defuzzification procedure, yielding a smooth compromise output rather than a hard switch. Because the rule base was intentionally kept compact and interpretable, no formal automated rule-reduction algorithm was applied.

The integrated execution error score, ErrScore, was also specified using expert-guided weighting. Misses were assigned the largest weight because they represent the most direct task-outcome failure, whereas late and rushed responses were treated as secondary temporal errors. The baseline weights ($w_m = 0.6$, $w_l = 0.25$, $w_r = 0.15$) were selected before the main experiment and then examined in a limited sensitivity analysis. Specifically, each weight was varied within a range of ± 0.1 while preserving the priority relation $w_m \geq w_l, w_r$ and the normalization constraint $w_m + w_l + w_r = 1$. Under these variations, the qualitative comparison between Fixed and Adaptive modes remained unchanged.

A similar robustness check was performed for the PERCLOS membership breakpoints by varying them within ± 0.05 on the normalized axis while leaving the remaining fuzzy-system components unchanged. The controller behavior and the overall Fixed versus Adaptive conclusions remained qualitatively stable. These checks do not replace future data-driven optimization, but they indicate that the reported results do not depend on a single arbitrarily chosen hand-tuned configuration.

7. RULE BASE AND MAMDANI FUZZY INFERENCE

Mamdani fuzzy inference is used as an interpretable mechanism for transforming observed oculomotor metrics and error indicators into estimates of trainee states, namely cognitive load, stress level, and fatigue.

Below, a compact rule set is provided that is sufficient for real-time model operation. Notation is as follows. FixDur is denoted as FD, SaccRate as SR, BlinkRate as BR, and ErrScore as ES.

The presented rule base yields consistent model behavior under typical simulator operating regimes. An increase in PERCLOS and blink rate raises F. An increase in ES strengthens both CL and SL. A high visual search intensity SR under unstable performance increases SL. Rules 22 to 25 target mixed cases where individual features may yield ambiguous interpretations, and they enforce consistent coordination among model outputs. The rule base was constructed to cover dominant monotonic and interaction effects observed during pilot experiments. Table 3 presents the compact fuzzy rule set used for real-time model operation.

For each output variable, after all rules are applied and

aggregated, a resulting membership function $\mu_{out}(u)$. The crisp output value is then computed using the centroid method.

$$y_k = \frac{\int_0^1 \mu_{out}(u) u du}{\int_0^1 \mu_{out}(u) du} \quad (12)$$

where, $y_k \in \{CL_k, SL_k, F_k\}$. The resulting values are subsequently used to generate user feedback and to adapt scenario parameters within the training difficulty control module.

Table 3. Fuzzy rules

No	IF Antecedent	THEN Consequent
1	PERCLOS is High	F is High
2	PERCLOS is Medium AND BR is High	F is High
3	PERCLOS is Medium AND BR is Medium	F is Medium
4	PERCLOS is Low AND BR is Low	F is Low
5	PERCLOS is High AND ES is High	F is High AND SL is High
6	FD is High AND SR is Medium	CL is High
7	FD is High AND SR is High	CL is High AND SL is High
8	FD is Medium AND SR is Medium	CL is Medium
9	FD is Low AND SR is Low	CL is Low
10	ES is High AND PERCLOS is Low	CL is High AND SL is High
11	ES is High AND PERCLOS is Medium	CL is High AND SL is High
12	ES is Medium AND FD is High	CL is Medium
13	ES is Low AND FD is Low	CL is Low
14	ES is Low AND PERCLOS is Low	SL is Low
15	ES is High AND SR is High	SL is High
16	ES is High AND SR is Medium	SL is High
17	ES is Medium AND SR is High	SL is Medium
18	ES is Medium AND SR is Medium	SL is Medium
19	SR is High AND BR is High	SL is High
20	SR is Medium AND BR is Medium	SL is Medium
21	PERCLOS is High AND SR is Low	SL is Medium
22	ES is High AND FD is Low AND SR is High	SL is High AND CL is Medium
23	ES is High AND FD is High AND SR is Low	CL is High AND SL is Medium
24	ES is Medium AND PERCLOS is High	F is High AND CL is Medium
25	ES is Low AND PERCLOS is High	F is High AND CL is Low

8. FUZZY MECHANISM FOR SCENARIO SELECTION

The set of admissible simulator scenarios is represented as a set S , where each scenario is defined by the trained skill type and a set of difficulty parameters. Within the experiment, two scenario settings are used. The baseline scenario includes a single target and favorable conditions. The advanced scenario includes multiple targets and degraded conditions.

Adaptation is applied to the advanced scenario by adjusting parameters within predefined ranges. The number of targets is

$N_{tgt} \in [1,5]$. Fog is $Fog \in [0,1]$. Rain is $Rain \in [0,1]$. Wind speed is $Wind \in [0, W_{max}]$. Target speed and maneuverability are $V_{tgt} \in [V_{min}, V_{max}]$.

A discrete difficulty level $D \in \{1,2,3\}$ is introduced for control. It determines a parameter configuration. Table 4 specifies the baseline parameter values for each difficulty level. The values are reported in normalized form.

Table 4. Advanced scenario parameter configuration by difficulty level

Level (D)	N_{tgt}	Fog	$Rain$	$Wind$	V_{tgt}
Low	1	0.2	0.0-0.2	0.2	low
Medium	2	0.5	0.3-0.5	0.5	medium
High	3	0.8	0.6-0.8	0.8	high

Adaptation aims to keep the system within a region where learning is most effective. In the proposed formulation, the target region corresponds to the following qualitative conditions. F_k does not exceed Medium, which indicates no

pronounced fatigue. CL_k stays around Medium, meaning the task requires effort but remains feasible. SL_k does not remain persistently High, meaning stress does not dominate execution. The smoothed error indicator $\overline{ErrScore}_k$ remains within Low to Medium, so errors occur at a reasonable rate and serve as a learning signal rather than indicating breakdown. When the system leaves this region, difficulty parameters are adjusted to steer the trajectory back toward stable operation.

Difficulty updates are discrete and limited to at most a one level change in D per control cycle, which prevents abrupt shifts in task conditions. A minimum time interval between substantial difficulty changes is also enforced with $\Delta t_{min} = 45$ seconds. This allows the trainee to adapt and avoids frequent mode switching. Decisions are based on smoothed estimates of CL_k, SL_k, F_k , and the smoothed error indicator $\overline{ErrScore}_k$, which reduces the influence of short-term fluctuations and occasional outliers. Table 5 presents a compact rule set that specifies how the difficulty level D is updated based on the current state estimates.

Table 5. Difficulty update rules

Condition within W_k	Action
F is High	$\max(D - 1, 1)$
F is Medium AND ES is High	$\max(D - 1, 1)$
F is Low AND ES is High AND SL is High	$\max(D - 1, 1)$
ES is Medium AND CL is Medium AND SL is Medium AND F is Low	D
ES is Low AND F is Low AND CL is Low	$\min(D + 1, 3)$
ES is Low AND F is Low AND SL is Low	$\min(D + 1, 3)$
ES is Medium AND F is Low AND CL is Low	$\min(D + 1, 3)$
ES is Low AND CL is High AND SL is High	D
F is Medium AND SL is High	$\max(D - 1, 1)$

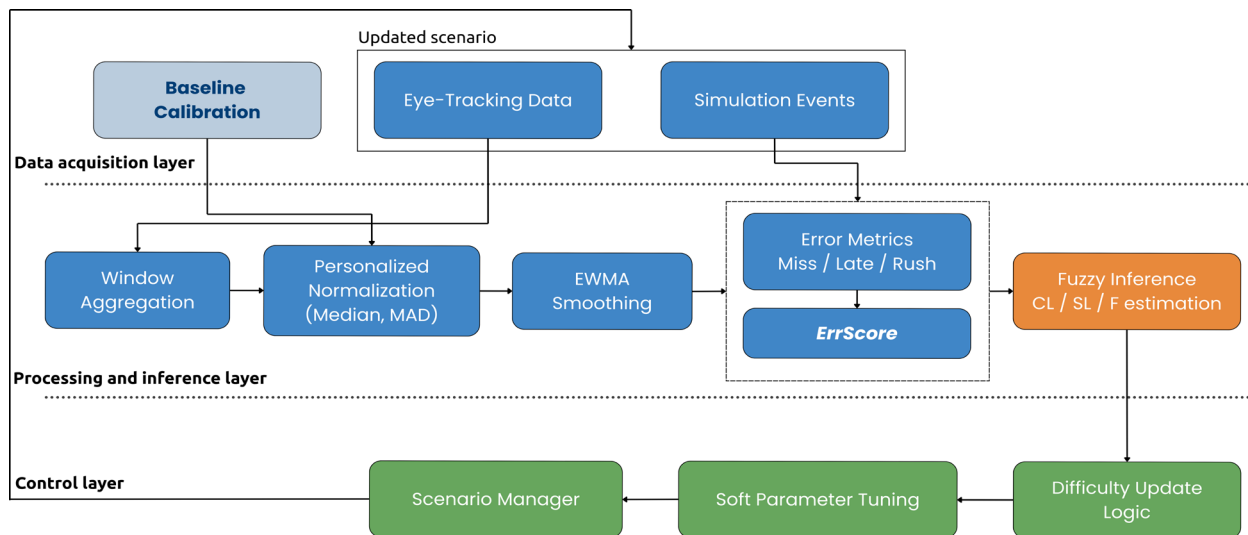


Figure 5. End-to-end online workflow of the proposed adaptive fuzzy scenario-control subsystem

In addition to the discrete level D , the system can apply soft adjustments to individual parameters within the current level while preserving the overall scenario regime. This is useful when a minor simplification or complication is required rather than a level change. The following parameter adjustment priorities are applied after selecting the action for D .

- 1) When ES and SL increase, it is preferable to reduce Wind and V_{tgt} because they increase aiming uncertainty.
- 2) When ES and CL increase, it is preferable to reduce

Fog because reduced visibility amplifies visual search load.

- 3) When F increases, it is preferable to reduce N_{tgt} and decrease Fog and Rain to shorten sustained effort without abruptly changing task structure.
- 4) When ES remains low and CL and SL are low, it is preferable to increase N_{tgt} first, then increase Wind and Fog.

This ordering helps keep training within a productive region without abrupt switching. Conflicting consequents are

resolved by Mamdani aggregation using max and min operators and centroid defuzzification, which yields a smooth compromise output.

To improve readability and provide a single end-to-end systems view, the complete online workflow of the proposed adaptive subsystem is summarized in Figure 5. The figure consolidates the stages described in Sections 4-7, including baseline calibration, window-based feature aggregation, personalized normalization, temporal smoothing, simulator-event processing, integrated error-score formation, fuzzy inference, latent state estimation, difficulty update, and soft scenario parameter tuning.

As shown in Figure 5, the subsystem operates as a closed adaptive loop in which the updated scenario influences subsequent simulator events and user behavior in the next observation windows. This end-to-end pipeline can be interpreted as an information-processing system consisting of sensing, state estimation, and decision control layers.

Because several real-time settings are introduced across different sections, the principal implementation parameters and control rules used in the proposed subsystem are summarized in Table 6 for compact reference.

Table 6. Key implementation and control parameters of the proposed subsystem

Parameter	Value	Description
Baseline calibration duration	2 min	Initial low-demand calibration interval used to compute personalized reference statistics
Baseline scenario	Single stationary target; no fog, no wind, no precipitation	Controlled low-demand setup for individualized normalization
Observation window	10 s	Sliding window W_k used for feature aggregation and event alignment
Eye-tracking features	FixDur, SaccRate, BlinkRate, PERCLOS	Features aggregated within each observation window
Robust baseline statistics	median, MAD	Personalized normalization parameters computed from baseline
Numerical stability constant	ϵ	Small constant used in normalization and rate computation to avoid division instability
Normalized operating range	[0,1]	Common scale for fuzzy inputs and outputs
Robust deviation range	$z_{max} = 3$	Symmetric anchoring range used in baseline-relative normalization
Temporal smoothing	EWMA	Applied to normalized eye features and ErrScore for real-time stability
EWMA coefficient	$\alpha \in [0,1]$	Controls sensitivity-stability trade-off
Error components	MissRate, LateRate, RushRate	Event-based execution errors computed per window
ErrScore weights	$w_m = 0.6, w_l = 0.25, w_r = 0.15$	Weighted combination of miss, late, and rushed responses
No-shot handling	If $N_k = 0$, rates are set to zero	Prevents artificial error inflation when no shots occur in a window
Target association rule	Nearest visible target to aiming point at trigger time	Associates each shot with the target effectively engaged by the trainee
Reaction-time thresholds	$\tau_{late}(D), \tau_{rush}(D)$	Difficulty-dependent thresholds for temporal error classification
Difficulty levels	$D \in \{1,2,3\}$	Low, Medium, High control regime
Maximum difficulty changes per cycle	1 level	Prevents abrupt difficulty switching
Minimum interval between major changes	$\Delta t_{min} = 45$ s	Anti-oscillation constraint for stable adaptation
Soft parameter tuning priority	Reduce Wind/ V_{tgt} , then Fog; reduce N_{tgt} , Fog, Rain under fatigue	Fine-grained adjustment within the current difficulty level
Shared membership parameterization	Low / Medium / High on [0,1]	Common linguistic mapping for most inputs and outputs
PERCLOS high-membership setting	(0.50,0.70,1.00,1.00)	More sensitive fatigue-oriented mapping for PERCLOS
Defuzzification	Centroid	Produces crisp outputs for CL, SL, and F

This summary is intended to make the runtime configuration easier to inspect without repeating the detailed derivations already provided in the corresponding sections.

9. EXPERIMENTAL SETUP AND TRAINING SCENARIOS

The experimental study aimed to evaluate the effectiveness of the proposed fuzzy subsystem for adaptive training difficulty control based on eye tracking data and execution error analysis within an ITS. Twenty participants took part in the experiment. A counterbalanced within subject design was used. Each participant completed two conditions, fixed and adaptive, and the order of conditions was balanced across

participants. Within each experimental session, the scenario structure and session duration were kept identical. The only difference was the control policy used to manage the parameters of the advanced scenario. This design was selected to minimize potential order and learning effects.

The experiment followed a within subject design in which each participant completed two consecutive conditions, fixed scenario parameter control and adaptive control driven by the fuzzy subsystem. The order of conditions was balanced across participants, which reduced the impact of learning and order

effects. A short rest period was included between conditions to reduce fatigue carryover. The scenario structure and session duration were kept identical, and only the control policy for the advanced scenario parameters differed. Figure 6 illustrates the ITS training process.

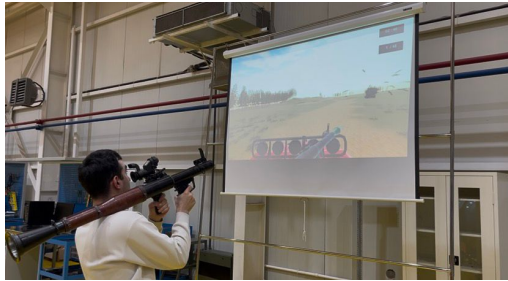
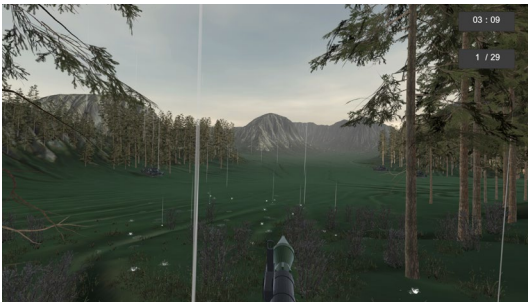


Figure 6. Experimental setup of an immersive simulator for RPG operator training



(a) Baseline training scenario with a single armored target under favorable environmental conditions



(b) Multi target scenario with increased environmental complexity

Figure 7. Training scenarios used in both conditions

In the first scenario, shown in Figure 7(a), a single armored target was present. The target was controlled by an agent that was pre-trained using deep reinforcement learning. The agent performed continuous maneuvering to minimize the probability of being hit. Environmental conditions were favorable, with no fog and no wind, and clear weather. A realistic ballistic model was used, and the weapon was a virtual analogue of the RPG-7V2.

In the second scenario, shown in Figure 7(b), three armored targets were present. They were controlled by a multi-agent system that supported coordinated evasive behavior. Environmental conditions were substantially more challenging due to fog, precipitation, and wind. This reduced visibility and increased aiming difficulty while preserving the same realistic ballistics and the same weapon type.

The transition to the second scenario occurred only after all targets in the first scenario had been hit. Throughout task execution, oculomotor data were continuously recorded as described in Section 4, along with performance measures

including hit accuracy as described in Section 5. These data were used for timestamp-based alignment with oculomotor features and subsequent analysis, and for computation of ES in both conditions.

The overall training structure was preserved. The participant completed the baseline scenario and then proceeded to the advanced scenario. However, unlike the fixed condition, the advanced scenario parameters were not predefined and were automatically updated at control intervals based on smoothed estimates of CL, SL, F, and ES. At each step, the Scenario Manager produced a control action by updating the difficulty level D , by no more than one level per cycle, or by applying soft parameter tuning within the current level inside the allowed ranges. As a result, the adaptive condition produced a personalized variant of the advanced scenario that was consistent with the user's current psychophysiological and behavioral characteristics. This ensured smooth adaptation and reduced frequent difficulty switching.

10. RESULTS

This section reports results from a counterbalanced within-subject study in which each participant completed both conditions, Fixed and Adaptive ($N = 20$). Metrics were computed in 10-s sliding windows after excluding the initial 2-min calibration interval. Eye-tracking features and simulator-derived variables were normalized using the personalized robust pipeline described earlier and smoothed with identical EWMA parameters in both conditions. Data from the initial two-minute calibration phase were excluded from the analysis.

We report: mean \pm SD across participants within each condition, mean paired difference $\Delta = \text{Adaptive} - \text{Fixed}$ with a 95% confidence interval, Wilcoxon p-values, FDR-adjusted q-values using Benjamini-Hochberg across the four metrics and rank-biserial effect size r_{rb} , where $|r_{rb}| = 1$ indicates a consistent direction of change across all paired observations. Confidence intervals for Δ were computed using nonparametric bootstrap resampling over participants.

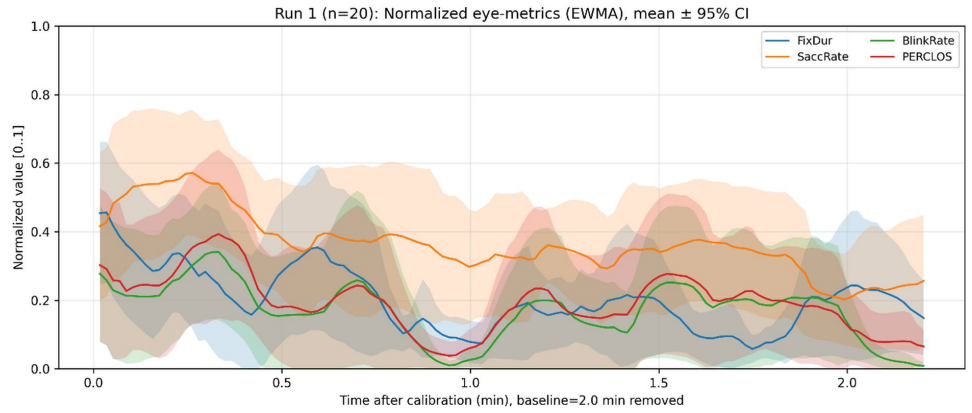
Figures 8(a) and (b) show the group level evolution of the normalized eye tracking metrics for the fixed and adaptive training conditions, respectively. Results are presented as mean trajectories with 95 percent confidence intervals.

In the fixed scenario condition, Figure 8(a), the eye tracking metrics exhibit pronounced temporal fluctuations throughout the session. During the early phase following calibration, an increase in saccade rate and PERCLOS is observed, which is consistent with intensive visual exploration and increased attentional demand at task onset. As the session progresses, fixation duration shows intermittent decreases, while blink rate and PERCLOS fluctuate irregularly. The confidence intervals remain relatively wide over time, indicating substantial inter-individual differences in visual behavior and attentional regulation.

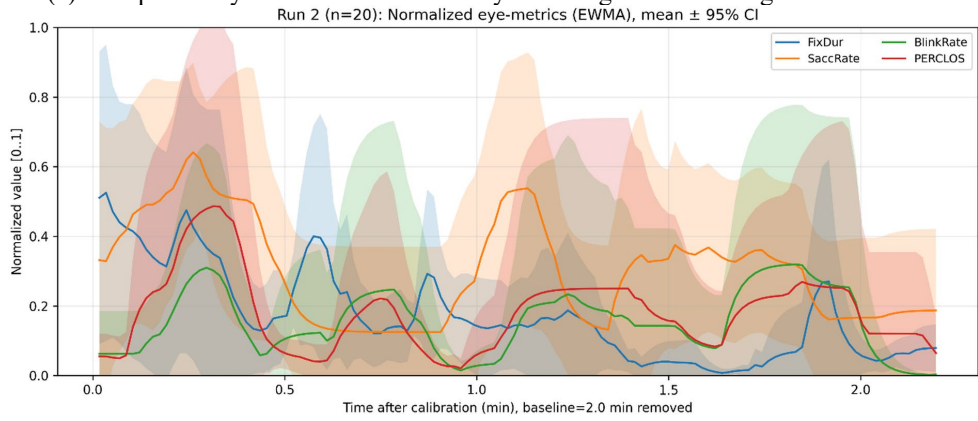
In contrast, the adaptive condition, Figure 8(b), is characterized by a more regular and stable evolution of eye tracking metrics. After an initial adjustment period, saccade rate remains within a narrower range and does not exhibit abrupt peaks. Fixation duration, blink rate, and PERCLOS display smoother trajectories with reduced dispersion across participants. The confidence intervals in this condition are generally narrower than those observed in the fixed condition,

particularly in the later phases of the session. This pattern suggests that adaptive scenario adjustment contributes to more

consistent visual behavior across users, despite ongoing changes in task conditions.

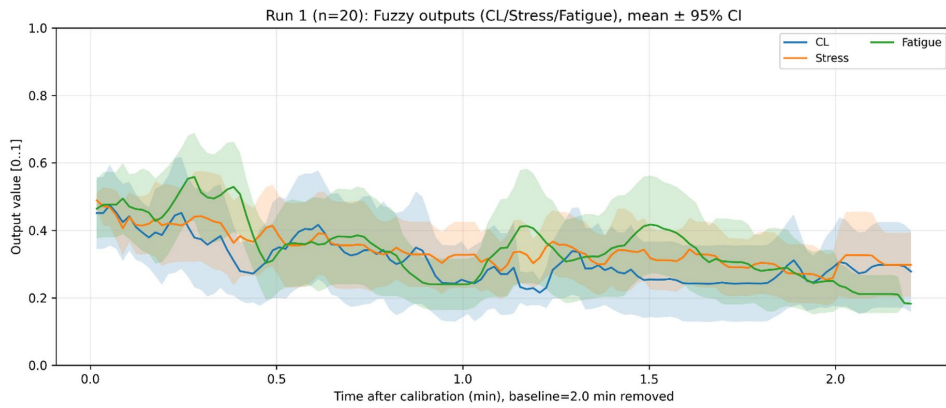


(a) Group level dynamics of normalized eye tracking metrics during fixed condition

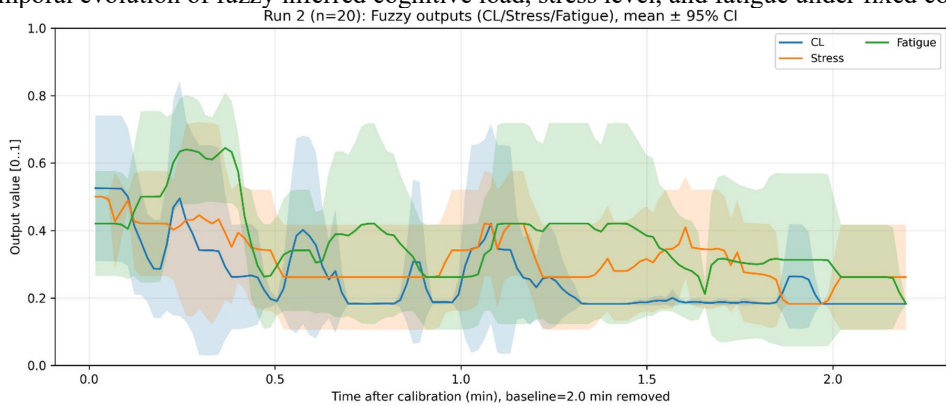


(b) Group level dynamics of normalized eye tracking metrics during adaptive condition

Figure 8. Group averaged normalized eye tracking metrics



(a) Temporal evolution of fuzzy inferred cognitive load, stress level, and fatigue under fixed condition



(b) Temporal evolution of fuzzy inferred cognitive load, stress level, and fatigue under adaptive condition

Figure 9. Group level fuzzy state estimates of CL, SL, and F

Figures 9(a) and (b) present the outputs of the fuzzy inference model for cognitive load, stress level, and fatigue during the fixed and adaptive conditions, respectively. The trajectories represent group means with corresponding 95 percent confidence intervals.

In the fixed scenario condition, Figure 9(a), the inferred cognitive load and stress level show noticeable oscillations over time. Periodic increases occur, particularly during segments associated with higher task demands. Fatigue tends to increase during the early part of the session and remains elevated for a subset of participants, as reflected by relatively wide confidence intervals. These patterns indicate heterogeneous responses to the fixed training protocol, with individual differences in workload tolerance and fatigue development becoming more pronounced over time.

Under adaptive control, Figure 9(b), the fuzzy estimates

follow a different temporal pattern. Cognitive load converges toward a relatively stable low to medium range after the initial phase and remains there for most of the session. Stress level exhibits moderate transient increases but does not show prolonged excursions to high values. Fatigue does not increase monotonically. Instead, it demonstrates phase-like variations, which may reflect gradual adaptation to the task rather than cumulative overload. The confidence intervals for cognitive load and stress level are notably narrower than in the fixed condition, indicating reduced interpersonal variability.

Table 7 summarizes log-derived performance outcomes at the participant level. Adaptive control was associated with directionally consistent changes across most participants: HitRate tended to increase, whereas LateRate, RushRate, and ErrScore tended to decrease.

Table 7. Participant-level performance comparison between fixed and adaptive conditions

Metric	Fixed (Mean ± SD)	Adaptive (Mean ± SD)	$\Delta (A - F)$	95% CI for Δ (Bootstrap)	Wilcoxon p	BH-FDR q	r_{rb}	Cohen's d_z
HitRate	0.201 ± 0.071	0.218 ± 0.082	+0.017	[+0.008, +0.026]	0.00315	0.0126	+0.474	+0.702
LateRate	0.514 ± 0.066	0.508 ± 0.075	-0.006	[-0.013, +0.001]	0.0328	0.0328	-0.263	-0.446
RushRate	0.444 ± 0.066	0.431 ± 0.071	-0.013	[-0.021, -0.006]	0.0240	0.0319	-0.321	-0.528
ErrScore	0.715 ± 0.065	0.708 ± 0.075	-0.007	[-0.014, 0.000]	0.0319	0.0319	-0.268	-0.453

Participant-level performance outcomes for Fixed and Adaptive conditions (N = 20). Values are participant means aggregated over retained 10-s windows (calibration excluded). Δ denotes within participant paired difference (Adaptive - Fixed). Confidence intervals are bootstrap 95% CI over participants. Wilcoxon signed-rank tests were used for paired inference. q-values are Benjamini-Hochberg FDR corrected across the four metrics. r_{rb} is rank-biserial correlation.

At the participant level, HitRate increases in the Adaptive condition by an average of + 0.017 (bootstrap 95% CI [+ 0.008, + 0.026]), with a medium paired effect (Cohen's d_z = 0.702) and FDR-corrected significance (q = 0.0126). RushRate decreases by -0.013 (95% CI [-0.021, -0.006], d_z = -0.528, q = 0.0319), and ErrScore decreases by -0.007 with a small to moderate paired effect d_z = -0.453, q = 0.0319). LateRate shows a small average decrease (-0.006) and remains close to zero within the uncertainty interval, although the Wilcoxon test indicates a modest paired shift (q = 0.0328). Overall, Adaptive control is associated with a small improvement in hit performance and modest reductions in rushed actions and the composite error indicator, while late responses remain broadly similar between conditions.

11. DISCUSSION

The reported findings suggest that the proposed adaptive controller supports smoother task-related eye-tracking trajectories and more stable inferred state dynamics than the fixed-condition policy under the controlled conditions of the present study. At the group level, the Adaptive condition was associated with narrower confidence intervals for several normalized eye-tracking measures and for the fuzzy outputs, suggesting reduced between-participant dispersion during task execution. In this sense, the adaptive subsystem appears to support more stable regulation of the training process under the controlled conditions of the present study.

At the same time, the interpretation of the inferred state variables requires caution. The variables CL, SL, and F are

model-derived latent state estimates produced by the fuzzy inference system rather than independently validated physiological ground-truth measures. Therefore, the reduced variability observed in these trajectories should be interpreted primarily as stabilization of controller-estimated user states. This pattern is consistent with the intended effect of adaptive regulation, but it does not by itself establish a verified reduction in true cognitive load, stress, or fatigue. In addition, part of the observed smoothing may reflect the structure of the inference and control pipeline itself.

The participant-level performance results also indicate that the observed effects are not uniform in magnitude across individuals. Table 7 shows generally consistent directional changes between the Fixed and Adaptive conditions, but with noticeable heterogeneity in effect size across participants. This is consistent with the personalized design of the normalization and control framework, in which a perfectly uniform response across all users would not be expected. Accordingly, the main result should be interpreted not as an identical participant-level benefit for every trainee, but as a tendency toward more stable individualized trajectories under adaptive regulation.

The practical size of the observed performance gains remains modest. HitRate increased slightly, while RushRate and ErrScore decreased, and LateRate changed only minimally. Thus, statistical significance should be distinguished from practical significance. Nevertheless, in an adaptive simulator context, even small but stable gains may still be meaningful when they are accompanied by improved control stability and do not come at the expense of task execution quality. From this perspective, the present findings are better understood as evidence of incremental but consistent benefit, rather than a large operational performance shift.

The present evaluation should also be interpreted as a proof-of-concept validation rather than as a fully developed comparative study. Although the fixed-versus-adaptive comparison is sufficient to examine whether the proposed framework can produce stable and directionally favorable changes under controlled conditions, it does not isolate the individual contribution of its main components. In particular,

the current design does not yet distinguish the relative role of personalized normalization, temporal smoothing, the fuzzy inference layer, or adaptive control more generally. Therefore, the observed gains should be interpreted at the level of the integrated system rather than as evidence for the unique effect of any single module. A more comprehensive comparative analysis, including ablation studies and stronger alternative baselines, remains an important direction for future work.

From an information-systems perspective, the proposed framework functions as a human-in-the-loop architecture that integrates multimodal behavioral inputs, transforms them into interpretable latent state estimates, and supports adaptive decision management within a closed control loop. In this sense, the contribution of the study is not limited to simulator control alone, but also to the design of a real-time information-processing structure for sensing, state estimation, and transparent rule-based adaptation in training environments. This distinguishes the present approach from prior work focused mainly on monitoring, offline assessment, single-modality analysis, or less interpretable adaptive decision layers.

Several limitations should also be considered. The sample size was modest, and the scenario set and session structure were limited, which constrains external validity and broader generalization. The study was conducted under controlled experimental conditions and therefore does not yet establish that the proposed controller will remain equally reliable across different user groups, longer sessions, noisier sensing conditions, or a wider variety of operational tasks. Accordingly, the present results should be interpreted as evidence of potential rather than established deployment suitability in high-risk or safety-critical training settings. In addition, the membership functions, fuzzy rules, and error-score weights were specified using expert knowledge; although qualitative conclusions were stable under reasonable parameter variations, more systematic parameter identification may improve robustness and portability. Eye-tracking data quality also remains a practical constraint, particularly under head motion, occlusions, and transient tracking loss.

Another important limitation is that the latent variables CL, SL, and F were not externally validated against independent physiological or subjective reference measures. As a result, the present study cannot determine to what extent the observed stabilization of these trajectories corresponds to true changes in cognitive load, stress, or fatigue beyond the controller's internal state representation. The study also employed a single initial baseline for normalization. While this approach is common in real-time adaptive systems, future work should examine alternative strategies such as multi-stage baselines, explicit familiarization periods, or adaptive baseline updates across the session. Broader validation with larger cohorts, more diverse scenarios, and longitudinal protocols will also be necessary to assess learning efficiency, retention, and transfer to more complex missions. An especially important next step will be an ablation-based evaluation that separately tests the contribution of personalized normalization, temporal smoothing, and the fuzzy decision layer against simpler adaptive and non-adaptive alternatives.

12. CONCLUSIONS

This study presented a real-time fuzzy logic subsystem for adaptive scenario control in an ITS that combines eye-tracking

signals with simulator-derived performance errors. The proposed framework integrates personalized robust normalization, multimodal feature fusion, Mamdani-type fuzzy inference, and bounded scenario adaptation within a closed-loop training architecture.

In a controlled within-subject study with 20 participants, the adaptive condition was associated with smoother eye-tracking trajectories, more stable model-estimated user-state dynamics, and modest improvements in selected log-derived performance metrics relative to a fixed-parameter policy. These findings support the feasibility of the proposed approach as an interpretable human-in-the-loop framework for multimodal data integration, state estimation, and adaptive decision management in simulator-based training.

Overall, the results suggest that transparent rule-based adaptive control can support smoother and more consistent training trajectories under controlled conditions. Future work should extend validation to broader user groups, longer sessions, richer scenario sets, and independent external reference measures for the inferred user-state variables.

REFERENCES

- [1] Shadiev, R., Li, D. (2023). A review study on eye-tracking technology usage in immersive virtual reality learning environments. *Computers & Education*, 196: 104681. <https://doi.org/10.1016/j.compedu.2022.104681>
- [2] Souchet, A.D., Philippe, S., Lourdeaux, D., Leroy, L. (2022). Measuring visual fatigue and cognitive load via eye tracking while learning with virtual reality head-mounted displays: A review. *International Journal of Human-Computer Interaction*, 38(9): 801-824. <https://doi.org/10.1080/10447318.2021.1976509>
- [3] Adhanom, I.B., MacNeilage, P., Folmer, E. (2023). Eye tracking in virtual reality: A broad review of applications and challenges. *Virtual Reality*, 27(2): 1481-1505. <https://doi.org/10.1007/s10055-022-00738-z>
- [4] Ndiaye, Y., Lim, K.H., Blessing, L. (2023). Eye tracking and artificial intelligence for competency assessment in engineering education: A review. *Frontiers in Education*, 8: 1170348. <https://doi.org/10.3389/educ.2023.1170348>
- [5] Tóthová, M., Rusek, M. (2025). Eye tracking in science education research: Comprehensive literature review. *Science & Education*, 34(6): 4583-4626. <https://doi.org/10.1007/s11191-025-00644-1>
- [6] Katona, J. (2022). Measuring cognition load using eye-tracking parameters based on algorithm description tools. *Sensors*, 22(3): 912. <https://doi.org/10.3390/s22030912>
- [7] Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8): 1457-1506. <https://doi.org/10.1080/17470210902816461>
- [8] Unema, P.J.A., Pannasch, S., Joos, M., Velichkovsky, B.M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3): 473-494. <https://doi.org/10.1080/13506280444000409>
- [9] Jackson, M.L., Kennedy, G.A., Clarke, C., Gullo, M., et al. (2016). The utility of automated measures of ocular metrics for detecting driver drowsiness during extended wakefulness. *Accident Analysis & Prevention*, 87: 127-

133. <https://doi.org/10.1016/j.aap.2015.11.033>
- [10] Wilkinson, V.E., Jackson, M.L., Westlake, J., Stevens, B., et al. (2013). The accuracy of eyelid movement parameters for drowsiness detection. *Journal of Clinical Sleep Medicine*, 9(12): 1315-1324. <https://doi.org/10.5664/jcs.m.3278>
- [11] Abe, T. (2023). PERCLOS-based technologies for detecting drowsiness: Current evidence and future directions. *Sleep Advances*, 4(1): zpad006. <https://doi.org/10.1093/sleepadvances/zpad006>
- [12] McIntire, L.K., McKinley, R.A., Goodyear, C., McIntire, J.P. (2014). Detection of vigilance performance using eye blinks. *Applied Ergonomics*, 45(2): 354-362. <https://doi.org/10.1016/j.apergo.2013.04.020>
- [13] Stern, J.A., Boyer, D., Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *Human Factors*, 36(2): 285-297. <https://doi.org/10.1177/001872089403600209>
- [14] Troussas, C., Chrysaftadi, K., Virvou, M. (2019). An intelligent adaptive fuzzy-based inference system for computer-assisted language learning. *Expert Systems with Applications*, 127: 85-96. <https://doi.org/10.1016/j.eswa.2019.03.003>
- [15] Gacto, M.J., Alcalá, R., Herrera, F. (2011). Interpretability of linguistic fuzzy rule-based systems: An overview of interpretability measures. *Information Sciences*, 181(20): 4340-4360. <https://doi.org/10.1016/j.ins.2011.02.021>
- [16] Cao, J., Zhou, T., Zhi, S., Lam, S., et al. (2024). Fuzzy inference system with interpretable fuzzy rules: Advancing explainable artificial intelligence for disease diagnosis—A comprehensive review. *Information Sciences*, 662: 120212. <https://doi.org/10.1016/j.ins.2024.120212>
- [17] Rodemer, M., Karch, J., Bernholt, S. (2023). Pupil dilation as cognitive load measure in instructional videos on complex chemical representations. *Frontiers in Education*, 8: 1062053. <https://doi.org/10.3389/educ.2023.1062053>
- [18] Ayres, P., Lee, J.Y., Paas, F., Van Merriënboer, J.J.G. (2021). The validity of physiological measures to identify differences in intrinsic cognitive load. *Frontiers in Psychology*, 12: 702538. <https://doi.org/10.3389/fpsyg.2021.702538>
- [19] Cheng, G., Zou, D., Xie, H., Wang, F.L. (2024). Exploring differences in self-regulated learning strategy use between high- and low-performing students in introductory programming: An analysis of eye-tracking and retrospective think-aloud data from program comprehension. *Computers & Education*, 208: 104948. <https://doi.org/10.1016/j.compedu.2023.104948>
- [20] Tien, T., Pucher, P.H., Sodergren, M.H., Sriskandarajah, K., Yang, G.Z., Darzi, A. (2014). Eye tracking for skills assessment and training: A systematic review. *Journal of Surgical Research*, 191(1): 169-178. <https://doi.org/10.1016/j.jss.2014.04.032>
- [21] Alberti, C.F., Gamberini, L., Spagnoli, A., Varotto, D., Semenzato, L. (2012). Using an Eye-tracker to assess the effectiveness of a three-dimensional riding simulator in increasing hazard perception. *Cyberpsychology, Behavior, and Social Networking*, 15(5): 274-276. <https://doi.org/10.1089/cyber.2010.0610>
- [22] Osmanli, T. (2025). AI-enhanced predictive modelling of virtual laboratory microlearning in online distance education. *Ingénierie des Systèmes d'Information*, 30(9): 2461-2471. <https://doi.org/10.18280/isi.300920>
- [23] Tan, L.Y., Hu, S., Yeo, D.J., Cheong, K.H. (2025). Artificial intelligence-enabled adaptive learning platforms: A review. *Computers and Education: Artificial Intelligence*, 9: 100429. <https://doi.org/10.1016/j.caeai.2025.100429>
- [24] Hariyanto, Kristianingsih, F.X.D., Maharani, R. (2025). Artificial intelligence in adaptive education: A systematic review of techniques for personalized learning. *Discover Education*, 4(1): 458. <https://doi.org/10.1007/s44217-025-00908-6>
- [25] Essa, S.G., Celik, T., Human-Hendricks, N.E. (2023). Personalized adaptive learning technologies based on machine learning techniques to identify learning styles: A systematic literature review. *IEEE Access*, 11: 48392-48409. <https://doi.org/10.1109/ACCESS.2023.3276439>
- [26] Wang, S., Christensen, C., Cui, W., Tong, R., Yarnall, L., Shear, L., Feng, M. (2023). When adaptive learning is effective learning: Comparison of an adaptive learning system to teacher-led instruction. *Interactive Learning Environments*, 31(2): 793-803. <https://doi.org/10.1080/10494820.2020.1808794>
- [27] Gligorea, I., Cioca, M., Oancea, R., Gorski, A.T., Gorski, H., Tudorache, P. (2023). Adaptive learning using artificial intelligence in e-learning: A literature review. *Education Sciences*, 13(12): 1216. <https://doi.org/10.3390/educsci13121216>
- [28] Gehrke, L., Koselevs, A., Klug, M., Gramann, K. (2025). Neuroadaptive haptics: A proof-of-concept comparing reinforcement learning from explicit ratings and neural signals for adaptive XR systems. *Frontiers in Virtual Reality*, 6: 1616442. <https://doi.org/10.3389/frvir.2025.1616442>
- [29] Novák, J.Š., Masner, J., Benda, P., Šimek, P., Merunka, V. (2024). Eye tracking, usability, and user experience: A systematic review. *International Journal of Human-Computer Interaction*, 40(17): 4484-4500. <https://doi.org/10.1080/10447318.2023.2221600>
- [30] Sesin, A. (2008). Adaptive eye-gaze tracking using neural-network-based user profiles to assist people with motor disability. *Journal of Rehabilitation Research and Development*, 45(6): 801-818. <https://doi.org/10.1682/JRRD.2007.05.0075>
- [31] Tusher, H.M., Mallam, S., Nazir, S. (2024). A systematic review of virtual reality features for skill training. *Technology, Knowledge and Learning*, 29(2): 843-878. <https://doi.org/10.1007/s10758-023-09713-2>
- [32] Won, M., Ungu, D.A.K., Matovu, H., Treagust, D.F., et al. (2023). Diverse approaches to learning with immersive virtual reality identified from a systematic review. *Computers & Education*, 195: 104701. <https://doi.org/10.1016/j.compedu.2022.104701>
- [33] Muzata, A.R., Singh, G., Stepanov, M.S., Musonda, I. (2024). Immersive learning: A systematic literature review on transforming engineering education through virtual reality. *Virtual Worlds*, 3(4): 480-505. <https://doi.org/10.3390/virtualworlds3040026>
- [34] Hebbar, P.A., Bhattacharya, K., Prabhakar, G., Pashilkar, A.A., Biswas, P. (2021). Correlation between physiological and performance-based metrics to estimate pilots' cognitive workload. *Frontiers in Psychology*, 12: 555446. <https://doi.org/10.3389/fpsyg.2021.555446>
- [35] Vulpe-Grigorasi, A. (2023). Cognitive load assessment based on VR eye-tracking and biosensors. In *Proceedings*

- of the 22nd International Conference on Mobile and Ubiquitous Multimedia, Vienna, Austria, pp. 589-591. <https://doi.org/10.1145/3626705.3632618>
- [36] Khan, R., Vernooij, J., Salvatori, D., Hierck, B.P. (2025). Assessing cognitive load using EEG and eye-tracking in 3-D learning environments: A systematic review. *Multimodal Technologies and Interaction*, 9(9): 99. <https://doi.org/10.3390/mti9090099>
- [37] Kim, D., Park, H., Kim, T., Kim, W., Paik, J. (2023). Real-time driver monitoring system with facial landmark-based eye closure detection and head pose recognition. *Scientific Reports*, 13(1): 18264. <https://doi.org/10.1038/s41598-023-44955-1>
- [38] Priyanka, S., Shanthi, S., Saran Kumar, A., Praveen, V. (2024). Data fusion for driver drowsiness recognition: A multimodal perspective. *Egyptian Informatics Journal*, 27: 100529. <https://doi.org/10.1016/j.eij.2024.100529>
- [39] Fonseca, T., Ferreira, S. (2025). Drowsiness detection in drivers: A systematic review of deep learning-based models. *Applied Sciences*, 15(16): 9018. <https://doi.org/10.3390/app15169018>
- [40] Seropian, L., Ferschneider, M., Cholvy, F., Michey, C., Bidet-Caulet, A., Moulin, A. (2022). Comparing methods of analysis in pupillometry: application to the assessment of listening effort in hearing-impaired patients. *Heliyon*, 8(6): e09631. <https://doi.org/10.1016/j.heliyon.2022.e09631>
- [41] Pauszek, J.R. (2023). An introduction to eye tracking in human factors healthcare research and medical device testing. *Human Factors in Healthcare*, 3: 100031. <https://doi.org/10.1016/j.hfh.2022.100031>