

Eye-Based Cognitive Overload Prediction in Human-Machine Interaction via Machine Learning

Maria Trigka^a, Elias Dritsas^b and Phivos Mylonas^c

Department of Informatics and Computer Engineering, University of West Attica, Greece

Keywords: Human-Centered Computing, Supervised Learning, User Models, Eye Tracking.

Abstract: Cognitive overload significantly affects human performance in complex interaction settings, making its early detection essential for designing adaptive systems. This study investigated whether gaze-derived features can reliably predict overload states using supervised machine learning (ML). The analysis was conducted on an eye-tracking dataset from a cognitively demanding visual task that incorporated fixations, saccades, and pupil diameter measurements. Five classifiers, namely, Logistic Regression (LR), Naive Bayes (NB), Support Vector Machine (SVM), XGBoost (XGB), and Multilayer Perceptron (MLP), were evaluated using stratified train/test splits and 5-fold cross-validation. XGB achieved the best performance, with an accuracy of 0.902, a precision of 0.958, a recall of 0.821, an F1 score of 0.884, and an area under the ROC curve (AUC) of 0.956. These findings confirm that gaze-derived features alone can reliably distinguish between cognitive overload states. The results also revealed trade-offs between simple models, which are easier to interpret but more conservative, and complex models, such as XGB and MLP, which achieved stronger predictive performance. Future studies should address subject-independent validation, incorporate temporal modeling of gaze dynamics, and explore personalization and cross-task generalization to advance robust and adaptive cognitive monitoring.

1 INTRODUCTION

The ability to monitor users' cognitive states during task execution is increasingly essential in domains such as human-computer interaction (HCI), education, simulation training, and safety-critical operations. When cognitive demand surpasses an individual's capacity, performance degradation is likely to occur, a phenomenon referred to as cognitive overload. Detecting this overload in real time enables systems to adapt their complexity, pacing, or feedback, thereby reducing user frustration and enhancing overall system usability (Kosch et al., 2023).

Recent advances in eye-tracking technology have made it feasible to noninvasively capture detailed gaze behavior, offering insights into attention, information processing, and cognitive effort. Compared with physiological measures such as electroencephalography (EEG) or functional near-infrared spectroscopy (fNIRS), gaze-based features are easier to integrate into practical environments and impose

minimal burden on users. Research has shown that saccade patterns, fixation durations, pupil dilation, and blink rates are modulated by cognitive load, making them useful input signals for classification models (Abbad-Andaloussi et al., 2022), (Gorin et al., 2024).

ML has become the predominant approach for modeling the relationship between gaze behavior and the cognitive state. Classical models, such as SVM and LR, as well as more recent deep learning and ensemble methods, have been applied to various cognitive estimation tasks. However, many studies rely on multimodal inputs or domain-specific datasets, which limit their generalizability (Aksu et al., 2024) (Skaramagkas et al., 2023).

Despite the growing body of literature on cognitive state monitoring, a gap remains in evaluating how well ML models can generalize cognitive overload detection using only gaze data. Existing studies often involve complex sensor setups or focus on specific environments (e.g., virtual reality (VR) or driving), which limits their applicability to more general HCI scenarios (Ghosh et al., 2023).

This study was motivated by the need to support the design of cognitively ergonomic interfaces for professional human-machine interactions. Predict-

^a  <https://orcid.org/0000-0001-7793-0407>

^b  <https://orcid.org/0000-0001-5647-2929>

^c  <https://orcid.org/0000-0002-6916-3129>

ing cognitive overload from eye-tracking data enables system designers to align interface complexity with user capabilities better, thereby minimizing cognitive strain and preserving interaction fluency for optimal performance. By identifying when users experience mental overload, designers can proactively adjust the information flow and visual load, preventing frustration, reducing negative emotional responses, and maintaining effective decision-making. Such predictive insights are critical for ensuring that high-demand operational environments remain user-centered, without hindering human cognition or compromising task performance.

A supervised learning framework was adopted to infer whether cognitive overload occurred based on gaze-derived features, assuming a unified and efficient pipeline. To position this work within the current research landscape and clarify its methodological scope, the key contributions are as follows:

- Investigation of cognitive overload prediction using gaze-derived features (fixations, saccades, pupil dynamics) in a visually demanding interaction task.
- Application of a structured preparation pipeline, including feature standardization for model training and statistical/visual analysis to confirm the relevance of gaze-based metrics.
- Comparative evaluation of five supervised learning models (LR, NB, SVM, XGB, MLP) with stratified validation and multiple performance metrics.
- Demonstration that gaze-only features can reliably predict overload, with XGB achieving high accuracy and AUC without requiring multimodal inputs.

The remainder of this paper is organized as follows. Section 2 reviews the relevant literature on cognitive workload estimation using gaze-based features and ML techniques. Section 3 details the proposed methodology, including dataset overview, pre-processing, feature analysis, model formulation, and evaluation strategies. Section 4 presents and analyzes the experimental results and provides a comparative assessment of model performance across key metrics. Section 5 discusses the limitations and future directions of this study. Finally, Section 6 summarizes the main findings of this study and outlines key directions for future research.

2 RELATED WORKS

Recent research on cognitive workload estimation has increasingly focused on gaze-based indicators owing to their unobtrusive nature and applicability in real-time applications. Several studies have utilized ML to model the relationship between eye behavior and cognitive demand in various domains, including VR, driving simulations, and tasks that require attention.

A foundational dataset in this area is COLET, which captures gaze behavior under multitasking and time pressure across multiple task conditions (Ktistakis et al., 2022). By training classical classifiers on fixation, saccade, and pupil-related features, the authors reported classification accuracies of nearly 88%, validating gaze signals as effective predictors of cognitive load. To advance generalization in unconstrained settings, the CLERA framework was introduced as a unified deep model for eye-region tracking and load estimation (Ding et al., 2023). It integrates keypoint localization with workload regression in a single trainable architecture and outperforms SVM-based approaches in naturalistic environments. In the context of immersive training, cognitive load was modeled during VR-based disassembly tasks using fixation duration and pupil dilation as inputs for MLP classifiers (Nasri et al., 2024). The results indicated a high F1, underscoring the discriminative value of gaze dynamics as task complexity increased.

Multimodal approaches have also been explored for this purpose. One study combined gaze features with fNIRS signals and driving dynamics within a Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) pipeline, achieving near-perfect classification performance across n-back difficulty levels (Khan et al., 2024). This integration of physiological and behavioral data demonstrates the benefits of signal fusion for robust load inference.

Gaze and pupillary data alone have proven sufficient in low-latency contexts. A CNN-based model was developed to detect stimulus onset using short windows of pupil diameter and gaze vectors across multiple cognitive domains (Dang et al., 2024). Despite domain variation, the models maintained reliable performance, especially for attention-oriented tasks.

Workload prediction in gamified VR environments has also been examined using a combination of ocular and biosignals, such as heart rate and galvanic skin response (GSR) (Szczepaniak et al., 2024). Using SVM and Random Forest (RF) models, the study reported F1 above 0.90, with interpretability analysis highlighting pupil size and blink rate as dominant predictors. Finally, a systematic benchmark evaluated

11 ML algorithms on gaze-derived features extracted under dual-task and time-pressure conditions (Skaramagkas et al., 2021). This study demonstrated that lightweight models, such as RF, can match more complex methods for both binary and multi-class cognitive load classification.

A comparative summary of the aforementioned studies is provided in Table 1, which outlines the core elements, including the domain, modality, feature types, and model classes. As shown in the table, most prior work emphasizes VR or driving contexts and relies on multimodal instrumentation, often including biosignals such as EEG, fNIRS, or GSR. Although such approaches can improve robustness, they require additional sensors, calibration, and user compliance, which limit their scalability and everyday applicability. In contrast, this study demonstrates that fine-grained gaze-only signals suffice for reliable overload prediction, offering a lightweight and unobtrusive alternative particularly suited to adaptive HCI and web-based interaction scenarios.

3 METHODOLOGY

The methodology follows a structured pipeline (Figure 1) consisting of dataset description, preprocessing, model training, and evaluation.

3.1 Dataset Overview

This study utilized a dataset collected using the Gaze-point GP3 eye-tracking system (Mannaru et al., 2017) during simulated print configuration tasks involving complex visual interactions. The task required participants to navigate menus, adjust parameters, and confirm settings under time constraints, thereby inducing either a normal or an overloaded cognitive state. The device recorded continuous streams of fixations, saccades, pupil diameter, and gaze coordinates under ecologically valid conditions. The participants underwent individual calibration procedures to ensure the spatial accuracy of gaze mapping.

The dataset included recordings from nine users supplemented with demographic and interaction-related metadata, such as age, professional experience, and task familiarity. The features of the dataset are summarized in Table 2. The participant demographics are reported in Table 3, which lists each user's age, total and platform-specific experience (in years), and the proportion of samples per user that were labeled as cognitively overloaded. The dataset comprised 2,510 overload samples (43.9%) and 3,207 normal samples (56.1%).

Class labels distinguishing overload from normal states were assigned based on gaze dynamics, specifically fixation duration and saccade magnitude, which are widely recognized in the literature as behavioral indicators of cognitive load (Rayner, 1998; Holmqvist et al., 2011; Duchowski, 2007). In this dataset, the thresholds were defined relative to the empirical feature distributions, resulting in approximately balanced class proportions. Although this procedure provides consistent labeling across participants, it remains heuristic and may limit generalizability.

This dataset ensures validity through standard preprocessing, including the exclusion of missing or low-confidence gaze samples (based on system confidence scores and pupil validity flags), detection and removal of blinks, reconstruction of fixation events, and derivation of saccade magnitudes from inter-fixation displacement. All timestamps were aligned with the session start time to ensure a consistent temporal reference across participants.

In the next section, the distributions and statistical properties of the gaze-derived features are examined to assess their relationship with cognitive overload.

3.2 Exploratory Feature Analysis

We examined gaze-derived features related to fixation duration, saccadic behavior, pupil size, and gaze distribution to explore the behavioral signatures of cognitive overloading in this study. These features were selected based on the literature and empirical variability. Our aim was not to reduce dimensionality but to evaluate the extent to which feature distributions differed across cognitive states in a statistically and behaviorally meaningful way.

Figure 2 shows the kernel density plots of the distributions of each gaze-derived feature across the normal and overload conditions. The fixation duration and saccade magnitude demonstrated the most distinct separation, with overload samples characterized by longer fixation durations and reduced saccade amplitudes. Pupil diameter measurements are generally elevated under overload conditions, albeit with a moderate overlap in distribution. Features such as blink-constricted pupil size, pupil motion magnitudes, and gaze coordinates exhibited less separability but still reflected subtle class-dependent shifts in the data. These trends are consistent with findings linking prolonged fixation, reduced eye movement, and pupil dilation to cognitive load and sustained attention.

To assess the statistical separability of features across cognitive states, we applied the Mann-Whitney U test to all relevant features in the dataset. This non-parametric test evaluates whether

Table 1: Comparative overview of related works on cognitive workload estimation.

Study	Domain	Input Modalities	Features Used	Model Type	Labels
(Ktistakis et al., 2022)	Visual Search	Eye Tracking	Fixations, Saccades, Pupil, Blinks	RF, SVM, XGB	NASA Task Load Index
(Ding et al., 2023)	Driving (natural)	Eye region video	Keypoints, Pupil, Blinks	Deep multitask CNN	Binary
(Nasri et al., 2024)	VR Training	Eye Tracking	Pupil Dilation, Fixation Duration	MLP, RF	NASA Task Load Index (Binary)
(Khan et al., 2024)	Driving Sim	Eye + fNIRS + Vehicle data	Gaze, HbO2, Vehicle Signals	CNN-LSTM	N-back levels
(Dang et al., 2024)	Multi-domain	Eye Tracking	Pupil, Gaze Vectors	Task-specific CNNs	Stimulus Onset
(Szczeplaniak et al., 2024)	VR Game	Eye + GSR + Heart Rate	Saccades, Pupil, Heart Rate, Electrodermal Activity	SVM, RF	Perceived Load
(Skaramagkas et al., 2021)	Visual + Dual Task	Eye Tracking	29 gaze metrics incl. Blink, Fixation	RF, Extra Trees	NASA Task Load Index (3-class)
This work	Visual Task	Eye Tracking only	Fixation, Saccade, Pupil	LR, NB, SVM, XGB, MLP	Cognitive Overload

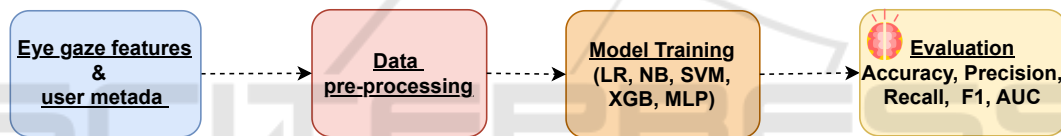


Figure 1: Pipeline for eye gaze-based overload prediction.

the values in the two classes originate from distinct distributions, without assuming a normal distribution. Features were then grouped based on whether they exhibited statistically significant differences at the $p < 0.01$ level:

- **Significant Features ($p < 0.01$):** Most gaze-derived features showed strong evidence of distributional divergence between cognitive states. These included FPOGD (fixation duration), SAC_MAG, LPD, RPD (left/right pupil diameter), LPMM, RPMM, BKDUR, BKPMIN, and gaze position CX and CY. Other significant features are AGE, TOT_EXP, EXP_PLAT, TIME, CNT, FPOGS, FPOGID, BKID, and CS. These results are consistent with the literature linking these features to visual attention and the cognitive load. The statistical significance of these variables supports their inclusion in subsequent analyses and model development.
- **Non-Significant Features ($p \geq 0.01$):** A small number of features did not show statistically significant differences. These include i) eye-specific gaze coordinates LPCX, LPCY, RPCX, RPCY, and ii) pupil validity flags LPV, RPV. The lim-

ited separability of these features is likely due to their dependence on external factors, such as display layout or signal quality, rather than the internal cognitive state. While retained for modeling purposes, these features were excluded from interpretative and visual analyses because of their minimal relevance to behavior analysis.

Figure 3 presents the exploratory correlations between the workload proportion and the selected participant-level variables. A weak negative association was observed between overload status and both total and platform-specific experience, suggesting that greater familiarity with the task environment may reduce cognitive strain. In contrast, pupil diameter and fixation duration tended to increase with workload, which is consistent with the established psychophysiological markers of increased mental effort. Saccade magnitude showed an inverse trend, indicating more localized gaze behavior under higher cognitive load.

Table 2: Structured summary of extracted features from eye-tracking data.

Feature(s)	Type	Description
Participant Metadata		
UID	Nominal	Participant identifiers for grouping or stratified sampling, not predictive.
AGE	Numeric	Age of the participants in years.
TOT_EXP	Numeric	Total professional experience; reflects overall expertise.
EXP_PLAT	Numeric	Experience specific to the simulated platform.
Fixation and Saccade Features		
CNT	Numeric	Frame/sample index; useful for computing fixation order or timing.
TIME	Numeric	Time elapsed since the session start; used for temporal analysis.
FPOGID	Nominal	Identifier for each fixation event.
FPOGS	Numeric	Fixation onset time, marking the start of fixation.
FPOGD	Numeric	Fixation duration (ms): key indicator of cognitive effort.
SAC_MAG	Numeric	Saccade magnitude: amplitude of movement between fixations.
SAC_DIR	Nominal	Saccade direction used in visual-scanning analysis.
Pupil Metrics and Motion		
LPD, RPD	Numeric	Left and right pupil diameters.
LPV, RPV	Nominal	Validity flags for pupil diameter measurements.
LPM, RPM	Numeric	Eye motion magnitude: may reflect fatigue or stress.
LPMV, RPMV	Numeric	Pupil motion velocity: complementary to LPM, RPM.
Blink Features		
BKID	Numeric	Blink ID grouping samples during the same blink.
BKDUR	Numeric	Blink duration (ms).
BKPMIN	Numeric	Minimum pupil diameter recorded during a blink.
Gaze Coordinates and Confidence		
CX, CY	Numeric	Central gaze coordinates on screen.
CS	Numeric	System-provided confidence score for gaze sample validity.
LPCX, LPCY	Numeric	Left eye gaze X/Y screen coordinates.
RPCX, RPCY	Numeric	Right eye gaze X/Y screen coordinates.
BPOGX, BPOGY	Numeric	Raw base point of gaze coordinates; system-derived, not used directly.

Table 3: Participant demographics and overload proportion.

UID	Age	TotExp	PlatExp	Overload
1	45	21.0	15.00	0.44
2	46	20.0	14.00	0.37
3	24	0.0	0.67	0.45
4	32	10.0	0.00	0.39
5	45	22.0	15.00	0.41
6	21	0.0	0.67	0.42
7	27	0.2	1.00	0.46
8	59	30.0	16.00	0.09
9	42	18.0	12.00	0.26

3.3 Machine Learning Models

Let $\mathbf{x} \in \mathbb{R}^d$ denote the feature vector, concatenating fixation, saccadic, and pupil metrics with participant metadata, and let $y \in \{0, 1\}$ represent the binary cognitive state label. Five supervised classifiers were employed for overload detection: LR, SVM, NB, XGB, and MLP. This configuration yields a compact representation of gaze dynamics and user traits, enabling comparative evaluation across diverse inductive biases without requiring multimodal inputs.

LR (Das, 2024) provides a linear and interpretable baseline by mapping features to class probabilities.

SVMs (Pisner and Schnyer, 2020) identify the decision boundary that maximizes the separation between classes and can capture nonlinear relationships using kernel functions. NB (Chen et al., 2020b) offers a probabilistic approach based on conditional independence assumptions and remains efficient even when these assumptions are violated moderately. XGB (Chen et al., 2020a) is an ensemble method that sequentially combines decision trees, achieving strong predictive accuracy while sacrificing interpretability. Finally, the MLP (Cinar, 2020) represents a neural network-based model, where layered nonlinear transformations enable the learning of complex input-output mappings.

3.4 Pre-Processing, Training & Evaluation

All numerical features were standardized using z-score normalization to ensure optimization stability in gradient-based models. To minimize sampling bias, stratified 5-fold cross-validation was applied within an initial 80/20 training/testing split, which preserved class balance across the folds. The hyperparameters were tuned using the training folds, and the final mod-

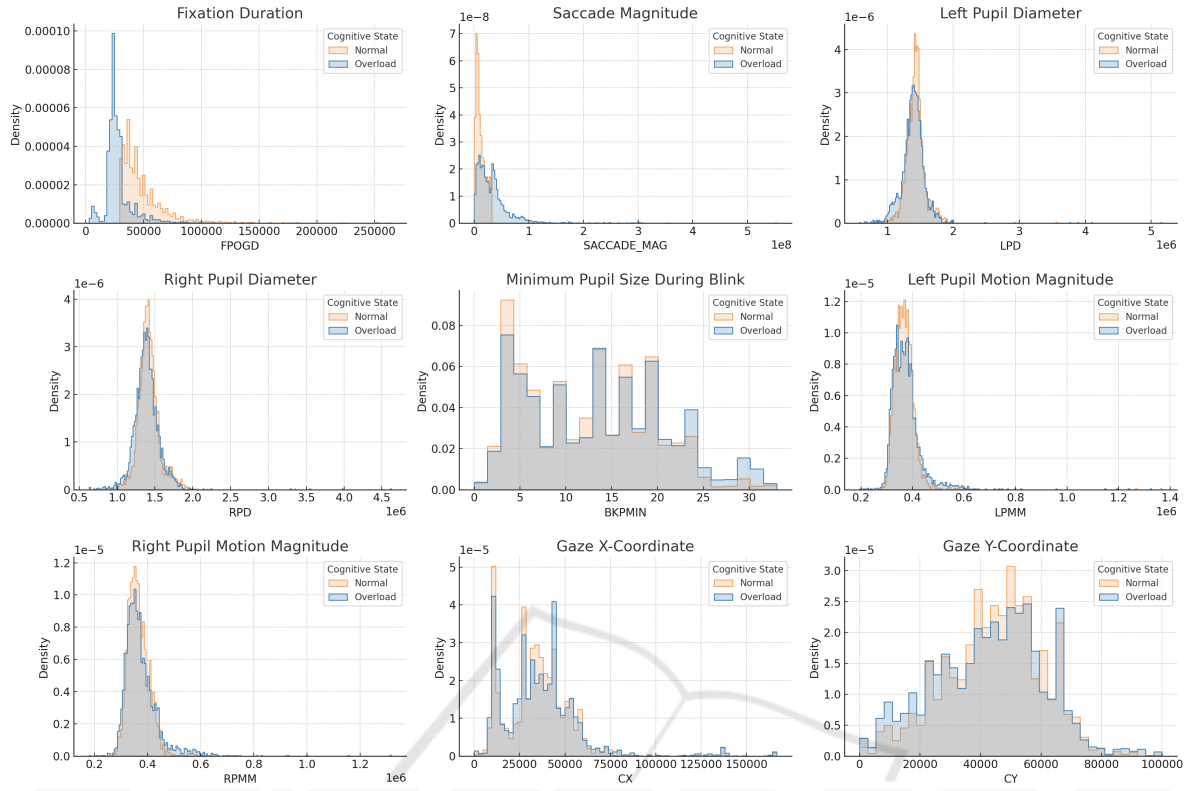


Figure 2: Distributions of 9 gaze-derived features across cognitive states (Normal vs. Overload), capturing fixation, saccade, pupil, and spatial attention dynamics.

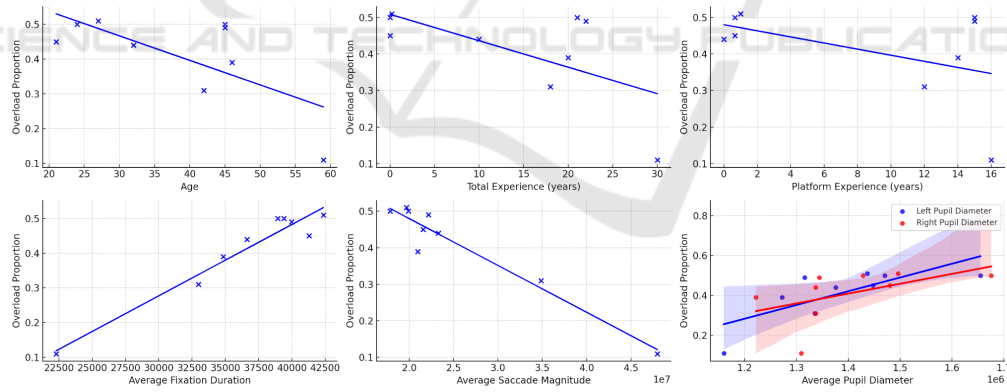


Figure 3: Participant-level correlations between overload proportion and gaze-derived features.

els were retrained on the full training subset before being evaluated on a held-out test set.

Each algorithm was trained with standard configurations: LR using the L-BFGS optimizer with ℓ_2 -regularization, SVM with RBF kernel (default regularization parameter $C = 1.0$ and $\gamma = \text{scale}$), NB with closed-form Gaussian estimates, XGB with 100 boosted trees (learning rate of 0.1), and MLP with a single hidden layer (100 units, ReLU activations, and Adam optimizer). Performance was as-

sessed through accuracy, precision, recall, F1 score, and AUC, defined in terms of the confusion matrix (TP, TN, FP, FN) as

- Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$
- Precision = $\frac{TP}{TP+FP}$, Recall = $\frac{TP}{TP+FN}$
- F1 = $2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$

The AUC was used to quantify threshold-independent discrimination. The models were implemented in

scikit-learn 1.3.0 and XGB 1.7.6, executed on Ubuntu 22.04 with an Intel i7 CPU and 32GB RAM.

4 RESULTS ANALYSIS

The five ML models were evaluated on a stratified 20% test set using accuracy, precision, recall, F1, and AUC. These metrics capture both overall performance and sensitivity to cognitive overload. Table 4 summarizes the performance of all evaluated models. XGB consistently outperformed the other models across all metrics, achieving the highest accuracy (0.902), F1 (0.884), and AUC (0.956), highlighting its ability to effectively model non-linear dependencies. The MLP also yielded strong results, particularly in terms of recall and F1, indicating its capacity to learn complex interactions among gaze-based features. LR and SVM demonstrated comparable but more conservative behavior, with high precision but lower recall, while NB trailed slightly due to its simplifying independence assumptions.

These results illustrate the value of using complementary metrics beyond accuracy, as F1 and AUC capture the trade-offs between precision and recall under class imbalance and subtle cognitive effects. Such trade-offs are particularly important for deployment, where factors such as interpretability, responsiveness, and tolerance to false negatives must be considered. XGB was the most reliable model, consistently achieving the best scores for all metrics. The MLP also performed strongly, particularly in terms of recall and F1 scores, reflecting its ability to capture complex feature interactions. LR and SVM showed higher precision but lower recall, indicating a more conservative classification behavior, while NB lagged behind due to its simplifying independence assumptions.

Table 4: Experimental results on the test set.

Model	Accuracy	Precision	Recall	F1	AUC
XGB	0.902	0.958	0.821	0.884	0.956
MLP	0.870	0.925	0.765	0.837	0.918
LR	0.851	0.910	0.723	0.805	0.894
SVM	0.846	0.902	0.714	0.797	0.889
NB	0.828	0.872	0.690	0.770	0.871

5 DISCUSSION

The results demonstrate that gaze-derived features alone can provide meaningful indicators of cognitive overload, supporting lightweight and noninvasive monitoring for adaptive human-machine inter-

actions. By relying only on fixations, saccades, and pupil dynamics, this study complements prior work that depends on multimodal or domain-specific data and shows that robust detection is feasible without intrusive instrumentation.

This study had some limitations. Although the dataset contains thousands of samples, they originate from only nine participants. This yields sufficient sample-level power but restricts generalizability because the models may capture user-specific traits rather than universal markers of overload. Moreover, training and testing on the same participants increased the risk of overfitting because cross-subject validation was not performed. Although the dataset was approximately balanced between the overload and normal samples, the confidence intervals could not be reliably estimated because of the limited number of participants. This underscores the need for larger and more diverse cohorts in future studies to enable robust statistical inferences. Class labels were also derived heuristically from fixation duration and saccade magnitude distributions. While grounded in established cognitive science, such labeling remains an indirect proxy of cognitive state and may limit construct validity. Therefore, future studies should validate this approach on datasets in which independent measures, such as task performance and multimodal markers, define the labels.

The present analysis relied on feature-based classifiers that ignored gaze sequential dynamics. Incorporating temporal models, such as recurrent or transformer architectures, can exploit this structure. Error analysis through confusion matrices would further clarify systematic misclassifications, whereas benchmarking against multimodal baselines would help quantify the trade-offs of unimodal gaze input. Assessing latency and computational costs is also necessary to establish the feasibility of real-time and adaptive interface systems. In addition, calibration analysis, which evaluates how well the predicted probabilities reflect the actual outcomes, would strengthen deployment readiness in settings where system actions depend on confidence thresholds. Overall, these limitations open concrete directions for advancing gaze-based cognitive state modeling in more realistic settings.

6 CONCLUSIONS

This study examined cognitive overload detection using only gaze-derived features and applied five ML models to data from a cognitively demanding visual task. Among them, XGB delivered the best per-

formance, achieving an accuracy of 0.902, precision of 0.958, recall of 0.821, F1 of 0.884, and AUC of 0.956. These results demonstrate that eye-based metrics, including fixations, saccades, and pupil diameter, are sufficient for reliable binary classification, thereby eliminating the need for multimodal inputs. Beyond predictive performance, the findings highlight the feasibility of deploying lightweight gaze-based models in real-time HCI systems. Unlike multimodal approaches, this method offers a focused and interpretable solution based solely on ocular behavior. As an exploratory study based on nine participants, the findings provide initial evidence of the discriminative power of gaze-only features. Larger and more diverse datasets are necessary to confirm the generalizability and establish statistical reliability. Future work should also extend the evaluation to subject-independent scenarios, incorporate temporal modeling of gaze dynamics, and explore personalization and cross-task generalization to advance robust and adaptive cognitive monitoring systems.

REFERENCES

- Abbad-Andaloussi, A., Sorg, T., and Weber, B. (2022). Estimating developers' cognitive load at a fine-grained level using eye-tracking measures. In *Proceedings of the 30th IEEE/ACM international conference on program comprehension*, pages 111–121.
- Aksu, Ş. H., Çakıt, E., and Dağdeviren, M. (2024). Mental workload assessment using machine learning techniques based on eeg and eye tracking data. *Applied Sciences*, 14(6):2282.
- Chen, J., Zhao, F., Sun, Y., and Yin, Y. (2020a). Improved xgboost model based on genetic algorithm. *International Journal of Computer Applications in Technology*, 62(3):240–245.
- Chen, S., Webb, G. I., Liu, L., and Ma, X. (2020b). A novel selective naïve bayes algorithm. *Knowledge-Based Systems*, 192:105361.
- Cinar, A. C. (2020). Training feed-forward multi-layer perceptron artificial neural networks with a tree-seed algorithm. *Arabian Journal for Science and Engineering*, 45(12):10915–10938.
- Dang, Q., Kucukosmanoglu, M., Anoruo, M., Kargosha, G., Conklin, S., and Brooks, J. (2024). Auto detecting cognitive events using machine learning on pupillary data. *arXiv preprint arXiv:2410.14174*.
- Das, A. (2024). Logistic regression. In *Encyclopedia of quality of life and well-being research*, pages 3985–3986. Springer.
- Ding, L., Terwilliger, J., Parab, A., Wang, M., Fridman, L., Mehler, B., and Reimer, B. (2023). Clera: a unified model for joint cognitive load and eye region analysis in the wild. *ACM Transactions on Computer-Human Interaction*, 30(6):1–23.
- Duchowski, A. T. (2007). *Eye tracking methodology: Theory and practice*. Springer Science & Business Media.
- Ghosh, S., Dhall, A., Hayat, M., Knibbe, J., and Ji, Q. (2023). Automatic gaze analysis: A survey of deep learning based approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(1):61–84.
- Gorin, H., Patel, J., Qiu, Q., Merians, A., Adamovich, S., and Fluet, G. (2024). A review of the use of gaze and pupil metrics to assess mental workload in gamified and simulated sensorimotor tasks. *Sensors*, 24(6):1759.
- Holmqvist, K., Nyström, M., et al. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
- Khan, M. A., Asadi, H., Qazani, M. R. C., Lim, C. P., and Nahavandi, S. (2024). Functional near-infrared spectroscopy (fnirs) and eye tracking for cognitive load classification in a driving simulator using deep learning. *arXiv preprint arXiv:2408.06349*.
- Kosch, T., Karolus, J., Zagermann, J., Reiterer, H., Schmidt, A., and Woźniak, P. W. (2023). A survey on measuring cognitive workload in human-computer interaction. *ACM Computing Surveys*, 55(13s):1–39.
- Ktistakis, E., Skaramagkas, V., Manousos, D., Tachos, N. S., Tripoliti, E., Fotiadis, D. I., and Tsiknakis, M. (2022). Colet: A dataset for cognitive workload estimation based on eye-tracking. *Computer Methods and Programs in Biomedicine*, 224:106989.
- Mannaru, P., Balasingam, B., Pattipati, K., Sibley, C., and Coyne, J. T. (2017). Performance evaluation of the gaze-point gp3 eye tracking device based on pupil dilation. In *Augmented Cognition. Neurocognition and Machine Learning: 11th International Conference, AC 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9–14, 2017, Proceedings, Part I 11*, pages 166–175. Springer.
- Nasri, M., Kosa, M., Chukoskie, L., Moghaddam, M., and Hartevel, C. (2024). Exploring eye tracking to detect cognitive load in complex virtual reality training. In *2024 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 51–54. IEEE.
- Pisner, D. A. and Schnyer, D. M. (2020). Support vector machine. In *Machine learning*, pages 101–121. Elsevier.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372–422.
- Skaramagkas, V., Ktistakis, E., Manousos, D., Kazantzaki, E., Tachos, N. S., Tripoliti, E., Fotiadis, D. I., and Tsiknakis, M. (2023). esee-d: Emotional state estimation based on eye-tracking dataset. *Brain Sciences*, 13(4):589.
- Skaramagkas, V., Ktistakis, E., Manousos, D., Tachos, N. S., Kazantzaki, E., Tripoliti, E. E., Fotiadis, D. I., and Tsiknakis, M. (2021). Cognitive workload level estimation based on eye tracking: A machine learning approach. In *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*, pages 1–5. IEEE.
- Szczepaniak, D., Harvey, M., and Deligianni, F. (2024). Predictive modelling of cognitive workload in vr: An eye-tracking approach. In *Proceedings of the 2024 Symposium on Eye Tracking Research and Applications*, pages 1–3.