# A Real-time Unconstrained EEG-Classifier for Mental Workload Monitoring

**Ali Osia[1], Zeynab Tahamtan[1], Lin Zhao[2], Mahdi Davari[1], Mikael Nybacka[2]**

(1) InnoBrain AB, Enhagsvägen 18, 187 40 Täby, Sweden: {ali.osia, roja.tahamtan, mahdi.davari}@ innobraintech.com

(2) Vehicle Dynamics, Department of Engineering Mechanics, KTH Royal Institute of Technology, 10044, Stockholm, Sweden: {linzhao, mnybacka}@kth.se

***Abstract** – Real-time monitoring of mental workload (MWL) is critical for designing adaptive human-machine systems. This study introduces a subject-independent and task-independent EEG classifier trained on spectral power ratios (delta, theta, alpha, beta) from frontal, parietal, and occipital regions. Using a controlled arithmetic task with labeled difficulty levels (easy/hard), a Gaussian Naive Bayes model achieved 75.4% accuracy (LOSOCV) in distinguishing MWL states. Validated in a driving simulator, the model was highly sensitive to task difficulties and detected higher MWL in urban (overload) vs. rural (underload) scenarios ($p < 0.05$), aligning with NASA-TLX subjective ratings. Temporal analysis revealed declining MWL over time in both scenarios, reflecting cognitive adaptation, followed by a mental fatigue rise in the cumulative effect of prolonged cognitive effort during the overload scenario. The framework eliminates the need for individualized and task calibration, offering a scalable solution for real-world applications like automotive safety and virtual reality. By bridging controlled lab settings and naturalistic environments, this work advances EEG-based MWL monitoring for adaptive systems in high-stakes domains like driving and aviation.*

**Keywords:** *Mental Workload, EEG, Real-time Monitoring, Subject-Independent Classification, Mental Fatigue*

https://doi.org/10.82157/dsa/2025/3

## Introduction

Mental workload (MWL) refers to the mental effort or resources required to handle the cognitive demands of a task. As task difficulty increases, cognitive demands also increase, resulting in a higher MWL (Eggemeier et al., 1991). Evaluating MWL is crucial for designing human-machine interfaces that enhance comfort, satisfaction, efficiency, and safety in respective working and operating environments. Adjusting task demands to promote operator safety, health, and productivity is essential to avoid mismatches between cognitive resource capacity and task demands (Young et al., 2015; Rubio et al., 2004). In this study, we aim to assess the temporal patterns of mental workload in real-time, using an unconstrained Electroencephalogram (EEG) classifier, which provides insights into the mechanisms of learning and adaptation and the progression of mental fatigue over time.

Monitoring MWL in real-time is essential not only for advancing technology development, as it directly influences the design of systems such as driving simulators, gaming environments, and virtual reality platforms, but also for monitoring purposes, where maintaining safety and performance of individuals is essential. Assessing MWL allows developers to tailor user experiences that align with cognitive demands, ensuring systems are intuitive, engaging, and effective. This understanding is particularly valuable in applications where user performance and safety are critical. For instance, in driving simulators, insights into MWL can guide the development of adaptive training environments that respond to users' cognitive states. Additionally, monitoring MWL in real-time can enhance safety features in driving scenarios by alerting drivers. Many traffic accidents are associated with mental workload, as high MWL can slow reaction times and reduce accuracy, whereas low MWL can result in distractions and inattention (Brookhuis & de Waard, 2010; Makishita & Matsunaga, 2008; Paxion et al., 2014).

Furthermore, the temporal patterns of mental workload offer valuable insights. It is closely related to learning and mental fatigue. Mental workload decreases after a single practice session due to the effects of the learning process (Gevins et al., 1997;

Jaquess et al., 2018; Brookings et al., 1996; Haufler et al., 2000; Kerick et al., 2004). As learning progresses and a task becomes more familiar, mental workload gradually decreases (Radüntz, 2020). This refinement in how mental resources are engaged aligns with Fitts and Posner's model, which describes a transition from controlled processing to more automatic processing as skills develop, accompanied by a reduction in the mental effort required to perform the task (Fitts & Posner, 1967). In contrast, prolonged engagement in cognitively demanding tasks can lead to mental fatigue, which is accompanied by an increase in mental workload. In this state, individuals often compensate for reduced performance by exerting additional effort and reallocating cognitive resources, which can, in turn, increase mental workload (Mahdavi et al., 2024; Nakagawa et al., 2013).

There are three primary ways to assess MWL: performance-based measures, subjective measures, and physiological measures. Performance-based measures quantify behavioral metrics, including task completion time and error rate. However, they have limitations, including a lack of sensitivity to changes in mental workload levels. Performance may remain consistent even when mental workload increases, leading to inaccurate assessments. Subjective measures rely on individuals reporting their perceived workload using questionnaires like the NASA Task Load Index. This method is easy and widely used, but does not provide real-time information. Among these, physiological measures uniquely offer objective, real-time assessments of MWL by tracking physiological responses, such as heart rate and brain activity, without interfering with task performance (Meshkati et al., 1995). Between these physiological measurements that assess MWL, only the EEG accurately reflects workload in real-time, allowing for second-by-second measurement (Berka et al. 2007). Furthermore, EEG's direct measurement of brain activity makes it the most direct indicator of different cognitive states among physiological measures (Debie et al., 2019).

## EEG & Mental Workload

There is a well-established body of literature on assessing mental workload using EEG. Generally, changes in task demands have been shown to correspond with alterations in EEG frequencies (Charles & Nixon, 2019). Specifically, a decrease in alpha band activity in the posterior region has been identified as a reliable indicator of higher mental workload. Additionally, theta power has been observed to increase significantly with greater task difficulty, particularly at anterior sites (Gevins et al., 1997; Smith et al., 1999; Jaquess et al., 2018; Di Flumeri et al., 2018; Borghini et al., 2014; Smith & Gevins, 2005; Brookings et al., 1996; Wilson, 2002). These findings are consistent across a variety of tasks, from controlled laboratory settings involving arithmetic tasks to real-world environments with aircraft pilots and car drivers. This aligns with the understanding that alpha activity is inversely related to general arousal and attentional processes, while frontal theta power is positively associated with working memory engagement and conscious control over attention (Smith et al., 1999; Jaquess et al., 2018). This suggests that, despite differences in task type, mental workload relies on a common neural infrastructure across diverse tasks.

In parallel with traditional spectral analysis, recent years have seen a growing use of machine learning (ML) techniques to estimate mental workload from EEG data. These approaches aim to move beyond fixed frequency-based heuristics by automatically learning patterns from data. A critical challenge in this domain is *subject dependency* and *task dependency*. Most ML models show significantly better performance when trained and tested on data from the same subject (subject-dependent models), as they can exploit individual-specific EEG features (Roy et al., 2016; Kingphai & Moshfeghi, 2024). However, this restricts the general applicability of such models, particularly in practical scenarios like driving, where pre-training on every new user is infeasible. On the other hand, subject-independent models—those trained across a group of individuals and tested on unseen subjects—often suffer a noticeable drop in accuracy due to inter-subject variability in anatomy, neural responses, and cognitive strategies (Zheng & Lu, 2015; Zhou et al., 2022). This variability is a well-documented limitation in EEG-based analysis. It presents a key difference from traditional methods, which often rely on population-level heuristics like increased frontal theta or decreased posterior alpha without being sensitive to individual-level variability.

To address this, researchers have explored domain adaptation, transfer learning, and personalized calibration techniques. For example, approaches that combine general features like spectral power ratios with data-driven features have shown promise in improving cross-subject generalization while preserving interpretability (Roy et al., 2016; Zhang et al., 2018). Nonetheless, ML-based models still face trade-offs between flexibility, transparency, and robustness. Unlike traditional EEG metrics that are relatively interpretable but less adaptive, ML models offer enhanced performance potential but often act as black boxes, making them harder to deploy in regulated or safety-critical environments like automotive settings.

## Contribution

While personalized models can offer highly accurate classifications of cognitive states, their deployment in real-world settings is often impractical. This is primarily due to the time and resources required to collect individualized training data and develop complex subject-specific models. In contrast, this study focuses on evaluating the reproducibility and sensitivity of the

MWL assessment using EEG signals within a subject-independent and task-independent model applied to uncontrolled environments.

To this end, we first designed a controlled scenario based on an arithmetic task specifically chosen for its ability to elicit distinct levels of mental workload. Task demands were systematically manipulated to induce variations in cognitive load, and participants provided subjective ratings of perceived difficulty and effort to assist with accurate labeling and model validation. The training phase included two defined difficulty levels—easy and hard—intended to reliably produce differing MWL states. While arithmetic tasks and driving differ in nature, both reliably engage the working memory and attention system. Prior studies (e.g., Gevins et al., 1997; Borghini et al., 2014) have shown consistent EEG workload patterns across cognitive and sensorimotor tasks, supporting the transferability of these findings. By training on arithmetic tasks that systematically vary in difficulty, we capture MWL-related neural signatures generalizable to driving scenarios, where urban/rural conditions similarly modulate cognitive demand.

Previous research, typically using event-related potentials (ERPs) and event-related spectral perturbation (ERSP) (e.g., Zhou et al., 2022; Roy et al., 2016), faces limitations in generalizing to real-world, uncontrolled, and realistic environments. This is partly because ERPs and ERSP are often tied to specific events. In contrast, using spectral markers not assigned to specific events offers a significant advantage by enabling continuous monitoring. In this study, spectral power ratios were extracted from the frontal, parietal, and occipital regions after the training data were collected and pre-processed. The data was then normalized, and a Gaussian Naive Bayes classifier was used for training (Grimes et. al., 2008). The reason to use the Naive Bayes classifier is its simplicity and a small number of parameters that give us generalizability, interpretability, and easy deployment. Given the noise-prone nature of EEG signals and limited labeled data, Naive Bayes offers robustness through its probabilistic assumptions and low variance, making it well-suited for real-time, subject-independent EEG classification. The trained model was evaluated in a simulated driving environment to assess its potential for real-time MWL estimation in operational contexts such as automotive applications.

We will aim to answer the following questions: Can real-time, subject-independent EEG measurements in a driving simulator reliably detect changes in task difficulty and cognitive demands in naturalistic, real-world environments, given that most prior research in this area has focused on subject-dependent methods in controlled settings? Furthermore, what are the temporal patterns of MWL during tasks of varying difficulty, and how do these patterns reflect cognitive adaptation and skill automation?

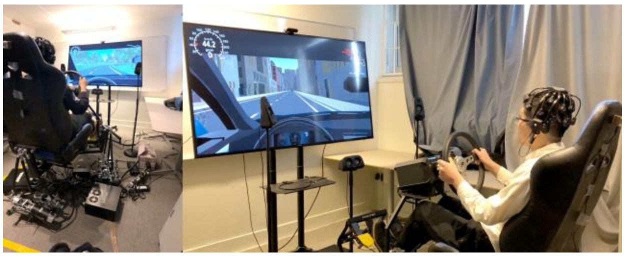

Figure 1: 6-DOF driving simulator with headset

While the effect of practice and task difficulty level on mental workload has been investigated, it remains unclear; 1) how the temporal dynamics of mental workload evolves during a single session under varying levels of difficulty, and 2) how these changes relate to learning and adaptation at the beginning of a task and mental fatigue toward the end of a task.

We hypothesize that: 1) higher mental workload will be observed in an overload scenario, which is more difficult, compared to an underload scenario; 2) a decreasing trend in mental workload will occur during both scenarios due to the learning process; and 3) effects of mental fatigue will be observed, with an increasing trend in mental workload at the end of the overload scenario, indicating that individuals need to allocate more cognitive resources and exert greater effort to perform the task.

# Methodology

To investigate mental workload using EEG, we required data with reliable ground truth labels. Therefore, we first designed a controlled arithmetic task experiment that allowed us to define two levels of task difficulty explicitly. This setup provided us with labeled EEG data suitable for training our model. To assess the model's generalizability in a more realistic context, we conducted a second experiment using a 6-DOF driving simulator, see Fig. 1). We designed a 30-minute scenario that included both rural and urban driving segments with respective speed limits and traffic signs and usual pedestrian density, see Fig. 2. While this dataset does not offer precise, moment-to-moment labels, the urban driving environment is generally associated with higher cognitive demands due to factors like traffic lights, pedestrians, and increased complexity. In contrast, rural driving is typically less mentally demanding. Thus, we used the arithmetic task data to train the model and then applied it to the driving dataset to evaluate its performance in estimating mental workload in a semi-naturalistic setting. This approach allowed us to explore how well the model could transfer from a well-labeled lab environment to a more complex, real-world scenario.

## Arithmetic Experiment

Twenty volunteers (12 females, 8 males; average age = 29.1 years, SD = 3.5) from the local community participated. All participants had normal or correctedto-normal vision and no history of neurological disorders. Informed consent was obtained from all participants, and they were instructed to get sufficient sleep and avoid alcohol for 24 hours before the study.

The EEG was recorded from Cognionics (CGX) Quick-20r v2 headset, a 21-channel, dry, wireless EEG device with sampling rate of 500 Hz at standard 10/20 locations (Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, P8, O1, O2). The electrode impedance was maintained below 300 kΩ. The signals were recorded using a left earlobe reference electrode and then re-referenced offline to the mean of the left and right earlobes and filtered with a bandpass of 0.5 – 40 Hz.

The experimental design consisted of six blocks with two levels of difficulty. The order of presenting levels in each block was randomized. Each level's difficulty increases based on the number of digits in the arithmetic problems presented. In the low mental workload condition, participants were required to sum up a one-digit number with a one-digit number or a two-digit number with a one-digit number. In contrast, the high mental workload condition involved summing a three-digit number with a three-digit number. Each participant had 66 seconds to complete each level, with trials separated by a visual cue - a dot displayed for two seconds. Following each level, participants were given a 15-second break to minimize fatigue and maintain focus. The order of difficulty levels was randomized for each participant and repeated six times throughout the experiment.

## Driving Experiment

Thirty-two volunteers (6 females, 26 males; average age = 25.30 years, SD = 3.36) from the local community participated, distinct from the Arithmetic experiment's participants. All participants met the same rejection criteria as in the Arithmetic experiment and had at least 1–2 years of driving experience.

The EEG was recorded from a Cognionics (CGX) Quick-20r v2 headset, an 8-channel, dry, wireless EEG device with a sampling rate of 500 Hz at standard 10/20 locations (Fp1, Fp2, P3, P4, T3, T4, O1, and O2) which are a subset available on the CGX Quick-20r v2 headset. In this task, we used only the electrodes that were used for training our model. As in Experiment 1, the electrode impedance was kept below 300 kΩ. Signals were recorded using a left earlobe reference electrode, then re-referenced offline to the average of the left and right earlobes and filtered with a 0.5-40 Hz bandpass.

The experimental design includes two driving scenarios with a driving simulator categorized into two levels of difficulty: overload and underload scenarios. The driving simulator is a 6-DOF Stewart platform from DoF Reality. The software used for building up and simulating the environments and handling the vehicle dynamics is CarMaker 13 with the Cockpit Package.

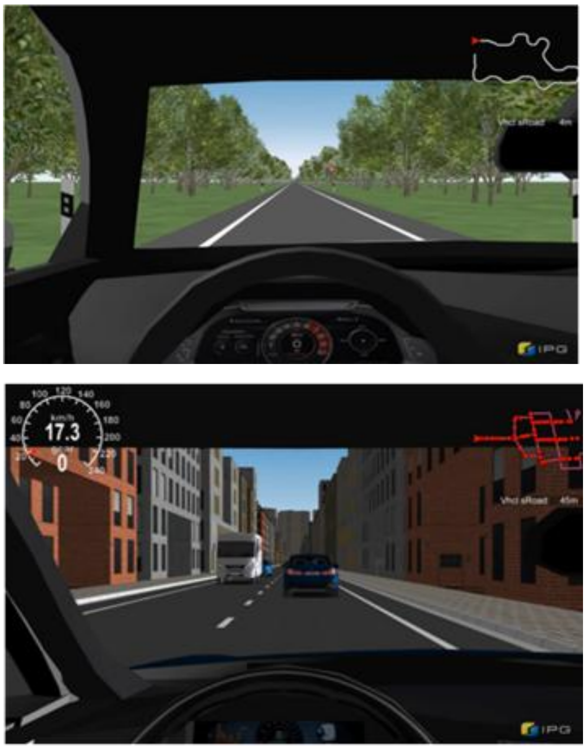

**Figure 2: The top and bottom images show the driving environment in CarMaker 13 for Underload and Overload Conditions, respectively**

To avoid misleading learning effects, overload and underload scenarios were experienced on separate days and presented in a randomized order. The Underload scenarios involved driving on a rural, one-way, straight road with minimal traffic and no pedestrians. The Overload scenarios involved driving on an urban, two-way road featuring distractions, traffic jams, traffic light changes, different road signs, and pedestrians crossing, as illustrated in Fig. 2. Each scenario lasted 30–45 minutes and began with a 5–10-minute familiarization period. After completing each experimental scenario, participants rated their perceived workload using the NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988). Also, they rated their overall drowsiness and the nausea level immediately following each experimental session using a numerical rating scale ranging from 0 to 20, where 0 indicated no drowsiness and 20 indicated maximal drowsiness.

## Preprocessing

For both experiments, after collecting the EEG data, a critical preprocessing step was undertaken to ensure data quality by systematically removing noisy epochs caused by artifacts such as muscle movements or eye blinks, which could compromise the model's accuracy. For this purpose, we epoched the data in 1-second intervals. After epoching, we applied several techniques to exclude signals containing significant artifacts, extreme voltage offsets, flat signals, or bad channels. If a channel within an epoch exhibited any of these issues, we labeled it as a 'bad channel' for that epoch and excluded it from further analysis without discarding the entire epoch. To identify extreme voltage offsets and flat signals, we flagged and removed channels with peak-to-peak amplitudes exceeding 90 µV or falling below 1 µV from subsequent analyses.

To eliminate noisy channels with significant muscle artifacts and high noise ratios, which exhibit distribution patterns differing from brain signals (Fitzgibbon et al., 2016; Barry & Blasio, 2021; Keil et al., 2022; Buzsáki & Mizuseki, 2014), a linear regression approach was employed to model the logarithmically transformed power spectral density (PSD) data. Using this approach, we identified channels that do not follow the lognormal pattern and labeled them as noisy (Fitzgibbon et al., 2016).

## Learning

We utilized data from the Arithmetic experiment, excluding signals with artifacts. After data cleaning, the training phase consisted of 8,962 one-second epochs (4,872 labeled as overload and 4,090 as underload). Each epoch included 19 EEG channels, along with a 'bad channel' label indicating whether each channel was considered clean (True) or noisy (False).

To reduce noise and capture more stable cognitive patterns, we applied a 30-second window across each trial. This window length was selected based on prior studies indicating that cognitive states such as mental workload are more reliably detected over moderately long time windows (e.g., 20–60 seconds), as they allow for temporal smoothing while maintaining responsiveness to changes in cognitive state (Grimes et. al., 2008). To expand the dataset and balance the classes, we applied a window-level sampling strategy that generates an additional 30-second window while preserving the original signal characteristics. This resulted in a balanced dataset of 11,709 samples (5,944 overload and 5,765 underload).

The data from the Arithmetic experiment was originally composed of 19 channels. For each channel, Power Spectrum Density (PSD) is calculated, and then the average PSD is calculated for frontal, parietal, central, and occipital regions. Then, spectral power ratios were extracted from the delta, theta, alpha, and beta frequency bands. This resulted in a 16-dimensional feature vector for each sample. Following data normalization, a Gaussian Naive Bayes classifier was employed to maintain a low number of parameters while ensuring both generalizability and interpretability. We employed Leave-One-Subject-Out Cross-Validation to evaluate our subject-independent model.

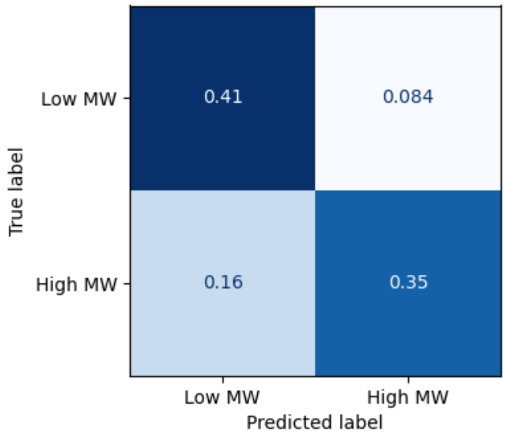

**Figure 3: Confusion matrix of the Arithmetic task using LOSOCV**

For the testing phase, we used data from the Driving experiment. After systematically excluding segments containing artifacts, we ran InnoBrain’s MWL prediction model on the dataset for both scenarios (overload and underload). The average mental workload score was calculated for each 5-minute time interval segment in both scenarios.

At the end, a paired t-test was conducted for each segment to evaluate whether InnoBrain’s MWL score is sensitive enough to detect differences in workload between high-demand and low-demand tasks. Linear Regression Analysis was used to assess the temporal trend in MWL, addressing patterns of cognitive adaptation and skill automation over time.

# Results

Accuracy, precision, and recall scores were used to evaluate the trained model in the Arithmetic experiment. Leave-One-Subject-Out Cross-Validation (LOSOCV) was used for evaluation to maintain subject independence. In this evaluation method, one subject is completely removed from training, and then the metrics are evaluated for that subject. This process is repeated for all subjects, and the average scores are then considered as subject-independent scores. We have achieved an average accuracy of 75.4% (SD = 10.8%) for our binary classifier of high vs. low mental workload using LOSOCV. Also, the normalized confusion matrix shows 80% precision and 68% recall. The confusion matrix is shown in Fig. 3.

For the Driving experiment, the results of the Paired T-Test on the NASA-TLX questionnaire indicate that the overload scenario was significantly more difficult than the underload scenario (t(31) = -3.34, p = 0.002, N = 32), and participants perceived greater mental workload and exerted more effort to perform in the overload scenario.

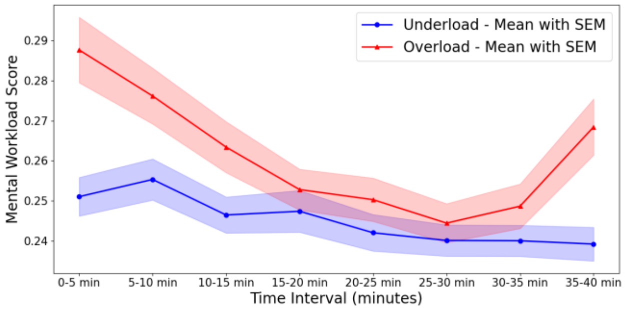

**Figure 4: Average Mental Workload Score with SEM by 5-minute time interval for both scenarios**

Due to the non-normal distribution of drowsiness and nausea ratings, two Wilcoxon Signed Rank tests were conducted to compare participants' drowsiness and nausea levels across scenarios. Results indicated that participants reported significantly higher drowsiness following the underload scenario compared to the overload scenario (Z = -2.20, p = .031, r = -0.40, N = 33). We did not observe any significant difference in nausea levels between the scenarios (Z = 0.4, p = 0.665, r = 0.1, N = 33).

Due to the fluctuating mental workload inherent in the unsupervised Driving experiment, accuracy cannot be precisely measured. However, it is expected that the overload scenario would generally yield a higher mental workload score, aligning with questionnaire results. To investigate this, the model was run on the entire time series using a 30-second window (consistent with training) and then compared between the two scenarios at 5-minute intervals. A Paired T-Test was employed to compare the mental workload scores between the two scenarios, and Linear Regression Analysis was used to examine the trend of mental workload changes over time.

Fig. 4 illustrates the average predicted mental workload in both scenarios across successive 5-minute time intervals, with the shaded regions representing the Standard Error of the Mean (SEM). The SEM indicates the variability of mental workload within each time interval, providing insights into the precision of the mean estimates. Each point corresponds to a time segment, starting from 0–5 minutes and progressing linearly. Consistent with the NASA-TLX questionnaire results, average mental workload values in the overload scenario were consistently higher across all time segments than in the underload scenario, particularly in the first and last segments. These differences were significantly greater in the overload scenario compared to the underload scenario during the first segment (t(28) = -2.22, p = 0.03, N = 29) and the last segment (t(22) = -1.98, p = 0.05, N = 23).

As shown in Fig. 4, a downward trend in average mental workload is observed in both scenarios. To examine the trend, linear regression analysis was conducted on the average MWL scores for each scenario. The results of the underload scenario indicate a statistically significant decreasing trend, with a slope of -0.00044 (p = 0.001). For the overload scenario, the decreasing trend has been seen just with marginal significance (slope of -0.00078 and p = 0.08) due to the increase after 30 minutes.

# Discussion

Mental workload exhibits a dynamic pattern, increasing with task difficulty and mental fatigue while decreasing with practice due to the learning effect (e.g., Gevins et al., 1997; Jaquess et al., 2018; Nakagawa et al., 2013). Our research aimed to detect real-time changes in MWL reliably in a subject-independent and task-independent manner using EEG measurements and to determine whether MWL trends and temporal patterns could reflect learning and mental fatigue in a single session. To achieve this, we first trained our model in a controlled laboratory environment using an arithmetic task with two levels of difficulty. For testing, we employed an uncontrolled, more realistic setting involving a driving simulator task, also with two levels of difficulty. We hypothesized that: 1) mental workload would be higher in a difficult (overload) scenario than in an easier (underload) one; 2) workload would decrease in both scenarios due to learning; and 3) Mental fatigue tends to increase workload in the later stages of an overload scenario due to the sustained effort required over time. Confirming these hypotheses, the proposed model demonstrates a notable capacity to effectively measure MWL. It clearly responds to task difficulty and captures temporal changes in mental workload during learning and mental fatigue in realistic settings.

Despite the known inter-subject variability in EEG signals, our use of spectral power ratios, less affected by individual anatomy, and LOSOCV evaluation in the Arithmetic experiment highlights a balance between generalizability and performance in realistic conditions.

Temporal analysis in the Driving experiment showed that the average of MWL in the overload scenario is higher than the average of MWL in the underload scenario, especially in the first and last segments, aligning with NASA-TLX questionnaire responses indicating higher perceived workload in the overload condition. This supports our hypothesis that MWL is greater in more difficult scenarios and confirms our model's sensitivity to task difficulty differences.

Linear regression analysis indicated a general decline in mental workload over time in both the underload and overload scenarios. In the underload scenario, this decline was fairly consistent across average values. Similarly, the overload scenario showed a steady reduction in average values during the first 30 minutes. This trend suggests that the reduction in the task's cognitive demands over time is due to a learning or adaptation effect. Initially, the workload was higher due to the cognitive demands of learning and adaptation, but it declined as proficiency improved, reflecting cognitive adaptation and skill automation (Jaquess et al., 2018).

Furthermore, the subsequent increasing trend observed beyond 30 minutes for the overload scenario suggests a potential shift in cognitive states due to accumulating mental fatigue, emphasizing the need for further investigation into the impact of prolonged activity on mental workload. This pattern highlights the dynamic nature of mental workload over time and suggests that prolonged activity leads to increased cognitive strain. Mental fatigue negatively impacts performance and increases the effort required for subsequent activities. To maintain performance standards, individuals experiencing mental fatigue must reallocate cognitive resources, often requiring additional support to achieve their goals effectively (Nakagawa et al., 2013).

Notably, this trend was absent in the underload condition, further supporting the interpretation that the late-stage workload rise in the overload condition was driven by accumulated mental fatigue. Also, the observation of lower mental workload in our model, alongside higher self-reported drowsiness in the underload scenario, aligns with previous findings that mental workload tends to decrease during states of fatigue or drowsiness (Brookhuis & de Waard, 2010).

Our approach demonstrates subject independence, as it was evaluated on participants distinct from those used in training, confirming its ability to generalize across individuals. Similarly, task independence was achieved by testing the model on activities different from those used for training. Its strong performance under these conditions underscores its adaptability to diverse mental workload demands. Higher scores observed for the more challenging task (e.g, Gevins et al., 1997) during evaluation align with existing literature on learning effects (e.g., Radüntz, 2020) and mental fatigue (e.g., Nakagawa et al., 2013), suggesting the model not only classifies workload but may also capture subtle cognitive states consistent with established psychological phenomena.

# Conclusions

This study confirms the sensitivity of InnoBrain's MWL metric in detecting differences in task difficulty, with a higher workload observed in high-demand tasks during the initial phase. The downward trend in MWL over time reflects cognitive adaptation and skill automation, supporting its value for dynamic monitoring in real-time scenarios.

The development and validation of this subject-independent and task-independent EEG-based model represent a significant advancement in MWL research, as it eliminates the need for off-line labor-intensive personalized calibration and expands the applicability of cognitive state monitoring across diverse populations and environments. By successfully applying this model to a simulated driving environment, we have demonstrated its robustness and scalability, making it a viable tool for industrial applications.

For the scientific community, this work bridges the gap between controlled laboratory studies and dynamic, real-world settings. It provides a reproducible framework for EEG-based workload assessment and opens new avenues for understanding temporal cognitive patterns, including adaptation and mental fatigue. The high sensitivity and practicality of the metric position it as a foundational tool for future research in human-machine interaction, cognitive ergonomics, and adaptive system design.

For the industrial sector, this research offers actionable insights for optimizing user experiences in safety-critical applications. Incorporating real-time MWL monitoring into system design can improve operational safety, reduce cognitive overload, and enhance productivity.

# Limitations & Future Works

Precise labels in naturalistic driving would require intrusive secondary tasks; instead, we adopted segment-level difficulty labels. Road-scene coding is planned to create frame-wise labels. Subject-independent models typically show a 10–20 pp accuracy gap. However, the benefit is zero-calibration deployment, and we were expecting this as a trade-off, but will further develop the algorithm in future works. In the future, we will explore instance-based transfer (e.g., Riemannian alignment) to reduce residual variability. Because MWL is derived from global spectral ratios rather than oculomotor artefacts, we expect limited sensitivity to head rotation compared to a more immersive setup. Future work will replicate the protocol in a 270° wrap-around simulator and in-vehicle tests.

# References

Barry, R. J. and De Blasio, F. M., 2021. Characterizing pink and white noise in the human electroencephalogram. *Journal of Neural Engineering, 18*(3), 034001.

Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., ... and Craven, P. L., 2007. EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, space, and environmental medicine, 78*(5), B231-B244.

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., and Babiloni, F., 2014. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience & Biobehavioral Reviews, 44*, 58-75.

Brookhuis, K. A. and De Waard, D., 2010. Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis & Prevention, 42*(3), 898-903.

Brookings, J. B., Wilson, G. F., and Swain, C. R., 1996. Psychophysiological responses to changes in workload during simulated air traffic control. *Biological psychology, 42*(3), 361-377.

Buzsáki, G. and Mizuseki, K., 2014. The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience, 15*(4), 264-278.

Charles, R. L. and Nixon, J., 2019. Measuring mental workload using physiological measures: A systematic review. *Applied ergonomics, 74*, 221-232.

Debie, E., Rojas, R. F., Fidock, J., Barlow, M., Kasmarik, K., Anavatti, S., ... and Abbass, H. A., 2019. Multimodal fusion for objective assessment of cognitive workload: A review. *IEEE transactions on cybernetics, 51*(3), 1542-1555.

Di Flumeri, G., Borghini, G., Aricò, P., Sciaraffa, N., Lanzi, P., Pozzi, S., ... and Babiloni, F., 2018. EEG-based mental workload neurometric to evaluate the impact of different traffic and road conditions in real driving settings. *Frontiers in human neuroscience, 12*, 509.

Eggemeier, F. T., Wilson, G. F., Kramer, A. F., and Damos, D. L., 1991. Workload assessment in multi-task environments. In: D. L. Damos, ed. *Multiple Task Performance*. London, UK: Taylor & Francis, pp. 207-216.

Fitts, P.M. and Posner, M.I., 1967. Human performance.

Fitzgibbon, S. P., DeLosAngeles, D., Lewis, T. W., Powers, D. M. W., Grummett, T. S., Whitham, E. M., ... and Pope, K. J., 2016. Automatic determination of EMG-contaminated components and validation of independent component analysis using EEG during pharmacologic paralysis. *Clinical neurophysiology, 127*(3), 1781-1793.

Gevins, A., Smith, M. E., McEvoy, L., and Yu, D., 1997. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cerebral cortex* (New York, NY: 1991), 7(4), 374-385.

Grimes, D., Tan, D. S., Hudson, S. E., Shenoy, P., and Rao, R. P., 2008. Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 835-844).

Hart, S. G. and Staveland, L. E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology* (Vol. 52, pp. 139-183). North-Holland.

Haufler, A. J., Spalding, T. W., Santa Maria, D. L., and Hatfield, B. D., 2000. Neuro-cognitive activity during a self-paced visuospatial task: comparative EEG profiles in marksmen and novice shooters. *Biological psychology*, 53(2-3), 131-160.

Jaquess, K. J., Lo, L. C., Oh, H., Lu, C., Ginsberg, A., Tan, Y. Y., ... and Gentili, R. J., 2018. Changes in mental workload and motor performance throughout multiple practice sessions under various levels of task difficulty. *Neuroscience, 393*, 305-318.

Keil, A., Bernat, E. M., Cohen, M. X., Ding, M., Fabiani, M., Gratton, G., ... and Weisz, N., 2022. Recommendations and publication guidelines for studies using frequency domain and time-frequency domain analyses of neural time series. *Psychophysiology, 59*(5), e14052.

Kerick, S. E., Douglass, L. W., and Hatfield, B. D., 2004. Cerebral cortical adaptations associated with visuomotor practice. *Medicine & Science in Sports & Exercise, 36*(1), 118-129.

Kingphai, K. and Moshfeghi, Y., 2024. Mental workload assessment using deep learning models from EEG Signals: a systematic review. *IEEE Transactions on Cognitive and Developmental Systems.*

Mahdavi, N., Tapak, L., Darvishi, E., Doosti-Irani, A., and Shafiee Motlagh, M., 2024. Unraveling the interplay between mental workload, occupational fatigue, physiological responses and cognitive performance in office workers. *Scientific Reports, 14*(1), 17866.

Makishita, H. and Matsunaga, K., 2008. Differences of drivers' reaction times according to age and mental workload. *Accident Analysis & Prevention, 40*(2), 567-575.

Meshkati, N., Hancock, P. A., Rahimi, M., and Dawes, S. M., 1995. Techniques in mental workload assessment.

Nakagawa, S., Sugiura, M., Akitsuki, Y., Hosseini, S. H., Kotozaki, Y., Miyauchi, C. M., ... and Kawashima, R., 2013. Compensatory effort parallels midbrain deactivation during mental fatigue: an fMRI study. *PLoS One, 8*(2), e56606.

Paxion, J., Galy, E., and Berthelon, C., 2014. Mental workload and driving. *Frontiers in psychology, 5*, 1344.

Radüntz, T. (2020). The effect of planning, strategy learning, and working memory capacity on mental workload. *Scientific reports, 10*(1), 7096.

Roy, R. N., Charbonnier, S., Campagne, A., and Bonnet, S., 2016. Efficient mental workload estimation using task-independent EEG features. *Journal of neural engineering, 13*(2), 026019.

Rubio, S., Díaz, E., Martín, J., and Puente, J. M., 2004. Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. A*pplied psychology, 53*(1), 61-86.

Smith, M. E. and Gevins, A., 2005. Neurophysiologic monitoring of mental workload and fatigue during operation of a flight simulator. In *Biomonitoring for Physiological and Cognitive Performance during Military Operations* (Vol. 5797, pp. 116-126). SPIE.

Smith, M. E., McEvoy, L. K., and Gevins, A., 1999. Neurophysiological indices of strategy development and skill acquisition. *Cognitive Brain Research, 7*(3), 389-404.

Wilson, G. F., 2002. An analysis of mental workload in pilots during flight using multiple psychophysiological measures. *The International Journal of Aviation Psychology, 12*(1), 3-18.

Young, M. S., Brookhuis, K. A., Wickens, C. D., and Hancock, P. A., 2015. State of science: mental workload in ergonomics*. Ergonomics, 58*(1), 1-17.

Zhang, P., Wang, X., Zhang, W., and Chen, J., 2018. Learning spatial–spectral–temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment*. IEEE Transactions on neural systems and rehabilitation engineering, 27*(1), 31-42.

Zheng, W. L. and Lu, B. L., 2015. Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development, 7*(3), 162-175.

Zhou, Y., Xu, Z., Niu, Y., Wang, P., Wen, X., Wu, X., and Zhang, D., 2022. Cross-task cognitive workload recognition based on EEG and domain adaptation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 30*, 50-60.