Mining, Metallurgy & Exploration (2026) 43:1951–1972
https://doi.org/10.1007/s42461-025-01415-w

Predicting Fatigue of Haul Truck Operators Using Reaction Time

Elaheh Talebi[1] · W. Pratt Rogers[1] · Frank A. Drews[2]

Received: 18 July 2025 / Accepted: 11 November 2025 / Published online: 3 February 2026

Abstract
Fatigue is a critical concern for haul truck operators in the mining industry, and early detection is essential for ensuring safe operations. Advances in wearable technology have made it possible to use smartwatches and applications to monitor an operator's reaction time continuously during the shift. Smartwatch applications can measure reaction time through simple tests such as a response time test or visual reaction time test. For this study, a custom application was used that displays different colors on the smartwatch's screen to measure users' reaction times. The data from these tests were analyzed to determine the operator's fatigue levels. Results showed that average reaction time slowed by approximately 30 milliseconds during the early morning hours (2–5 AM), corresponding to the circadian trough of alertness. This technology provides a non-intrusive, cost-effective, and continuous measurement of the operator's reaction time, enabling the detection of early signs of fatigue. This paper explores the feasibility, limitations, and benefits of using smartwatch applications to measure reaction time as a threshold indicator of fatigue in haul truck operators, building on recent advancements in fatigue modeling and monitoring in mining.

Keywords Fatigue · Wearable technology · Reaction time · Mining · Haul truck operators

1 Introduction

Fatigue is a significant issue in the mining industry due to the long and often irregular working hours, physically demanding work, and monotonous tasks. It can lead to reduced cognitive function, slower reaction times, impaired decision-making, and diminished alertness, all of which increase the risk of accidents and injuries. In a mining context, fatigue-related accidents can be particularly severe due to the size and complexity of the equipment involved and the hazardous nature of the work environment. Fatigue-related incidents can result in serious injuries, fatalities, damage to equipment, and production downtime. In addition to these direct impacts, fatigue also contributes to higher insurance costs, reduced productivity, and damage to the mining company's reputation. Therefore, mining companies should adopt comprehensive fatigue risk management strategies to effectively reduce the likelihood of fatigue-related incidents in their operations.

This study is part of a broader, multi-year research program on fatigue in mining. Drews et al. [1] conducted foundational focus group research with haulage operators, documenting the cognitive, psychological, and behavioral symptoms of fatigue and highlighting the limitations of current monitoring systems from the operator's perspective [1]. Their findings underscore the need for monitoring systems that are not only technologically effective but also socially acceptable and user-centered — a need this wearable-based study directly addresses.

Recognizing this, recent research efforts specific to mining operations have begun to systematically investigate and address operator fatigue. For example, Talebi et al. [2] developed a machine-learning-based model to identify environmental and work factors that drive fatigue in haul truck drivers [2]. Using large operational datasets, they modeled how variables like shift schedules, overtime hours, and haul

✉ Elaheh Talebi
elaheh.talebi@utah.edu

W. Pratt Rogers
pratt.rogers@utah.edu

Frank A. Drews
frank.drews@psych.utah.edu

1 Department of Mining Engineering, University of Utah, Salt Lake City, UT 84112, USA

2 Department of Psychology, University of Utah, Salt Lake City, UT 84112, USA

cycle details impact an individual operator's fatigue state. Building on the understanding that fatigue has multifactorial causes, Rogers et al. [3] introduced an IoT-enabled wearable fatigue-tracking system for mine operators. In their approach, a custom smartwatch application was used to collect both objective performance data and subjective fatigue ratings in real time [3].

This study continues that line of research by focusing on a wearable reaction-time-based fatigue monitoring approach for haul truck operators. We explore the feasibility of using a smartwatch reaction time test as an early indicator of fatigue, analyze how the measured reaction times vary with circadian factors (e.g., time of day) and compare to expected fatigue patterns, and discuss how the objective measurements might diverge from subjective perceptions of fatigue. By leveraging insights from prior fatigue prediction models e.g., Talebi et al., [2] and wearable monitoring implementations e.g., Rogers et al., [3], and the experiential data provided by Drews et al., [1], our aim is to evaluate a practical, real-time tool for fatigue detection that could be integrated into mining operations [1–3].

The objectives of this study are: (1) to evaluate the feasibility of using a smartwatch-based reaction time (RT) application as a practical and non-intrusive tool for monitoring operator fatigue during mining haulage operations. (2) to analyze how reaction time varies with shift timing and circadian patterns, with particular attention to fatigue-prone periods such as the early morning hours. (3) to compare objective RT-based fatigue indicators with subjective self-reported fatigue levels to assess their alignment and divergence. (4) to identify key operational and individual factors, including shift type, worksite, and operator variability, that contribute to differences in fatigue-related performance. (5) to establish a data-driven reaction time threshold capable of distinguishing between low and high fatigue states, thereby providing a foundation for future real-time fatigue detection and management systems in mining operations.

2 Literature Review

Fatigue is a well-documented contributor to workplace accidents and injuries across multiple industries. A systematic review and meta-analysis of 27 observational studies found that workers experiencing sleep problems had a 62% higher risk of occupational injury, with approximately 13% of all work-related injuries attributable to insufficient sleep [4]. Experimental research further shows that 17–19 h of sustained wakefulness impairs performance to a degree comparable to a blood alcohol concentration of about 0.05%, and 24 h to roughly 0.10% [5, 6]. In transportation contexts, fatigue has been implicated in 9–10% of all crashes, including a disproportionate share of the most severe cases—approximately 13% of serious-injury and up to 18% of fatal crashes [7–9]. Within mining operations, the National Institute for Occupational Safety and Health (NIOSH) identifies fatigue as a major human-factors hazard that diminishes attention and situational awareness, particularly in haul truck operations [10]. Collectively, these findings demonstrate that fatigue not only increases the likelihood of accidents and injuries but also amplifies their severity when they occur.

2.1 Reaction Time as a Fatigue Indicator

Reaction time (RT) is a widely used metric in psychology to assess alertness and responsiveness. There are two primary methods for measuring reaction time: simple reaction time tasks and choice reaction time tasks [11]. Simple reaction time tasks involve responding to a single stimulus (e.g., pressing a button as soon as a light appears), while choice reaction time tasks involve responding to one of several possible stimuli (for example, pressing one button if a green light appears and a different button if a red light appears). Each method has its advantages and disadvantages. Simple RT tasks are straightforward and primarily measure basic alertness and motor response speed. In contrast, choice RT tasks are more complex and can provide greater insights into cognitive processing (due to the need for stimulus discrimination and decision-making) at the cost of slightly longer baseline response times and more variability [11]. Choice tasks, by introducing a go/no-go or decision element, can also be useful for evaluating attention and inhibitory control, which are relevant in fatigue – fatigued individuals may have not only slower responses but also more difficulty in correctly executing or withholding responses as required.

RT is influenced by numerous factors, including age and gender. Studies have shown that reaction time generally slows with age, with older adults having slower RTs than younger adults [12]. This age-related decline is thought to be due to changes in neural processing speed and efficiency. Additionally, there are slight gender differences in reaction time, with some studies finding women having marginally faster RTs than men, though the magnitude of these differences is small and can be task-dependent [13]. Cognitive abilities such as working memory, attention, and processing speed are strongly related to reaction time performance; individuals who can sustain attention and process information quickly tend to have faster RTs [14]. Neuroimaging studies have linked reaction time performance to activity in brain regions like the prefrontal cortex, which is involved in attention and working memory, further indicating that RT is a proxy for the efficiency of cognitive processes [15].

In the context of fatigue, slower reaction time is a well-established indicator of reduced alertness. As an individual becomes fatigued (whether from sleep deprivation, circadian effects, or prolonged mental/physical work), their cognitive processing slows down and lapses in attention become more frequent, leading to longer response times on average. Importantly, the interpretation of RT measures in fatigue studies must account for various confounding factors. Task demands and stimulus characteristics can influence absolute RT values, so it is crucial to measure changes relative to a well-rested baseline for each operator when possible. Researchers recommend using appropriate statistical techniques, controlling for extraneous variables (like practice effects or time-of-day influences), and using multiple measures of RT (e.g., different tasks or repeated trials) to obtain a robust assessment of cognitive slowing [11]. By carefully designing RT measurements and analyses, one can draw accurate conclusions about an operator's level of fatigue from changes in their reaction times.

2.2 Employee Alertness Testing in the Workplace

Beyond laboratory-style reaction time tests, there is a growing interest in practical alertness testing tools that can be deployed in work settings. Employee alertness testing refers to brief, repeatable tests (often based on reaction time or other simple cognitive tasks) administered during work shifts to gauge workers' alertness and fatigue levels in real time. The idea is to have employees periodically complete a quick test on a smartphone, tablet, or computer, and compare their performance to their own baseline. A well-known example is the Psychomotor Vigilance Test (PVT) and its variants, which measure reaction times over several minutes to detect lapses in attention.

Dinges [16] first proposed on-duty alertness tests as a fatigue countermeasure, and in recent years there have been pilot implementations of such tests via mobile apps [16]. For instance, the AlertMeter® is a brief graphic-based alertness test that has been tested in occupational settings. Brossoit et al. [17] evaluated the perceptions and openness of employees and managers in safety-sensitive industries (including mining) to implementing an alertness testing program with the AlertMeter® app [17]. In that study, employees were encouraged to take a 60-second alertness test on a smart device at the start of each shift, and managers received immediate notifications if an employee's score fell outside their normal range [17]. The test involved responding rapidly to a sequence of symbols, and it produced an alertness score based on reaction time and accuracy. While the findings were generally positive about the potential of alertness testing (seeing it as a promising strategy for managing fatigue-related risk), there were concerns noted. Some challenges included questions of cost-effectiveness, the logistics of testing (the study only tested at the beginning of shifts, potentially missing fatigue buildup later in the shift), and user acceptance. Notably, employees and managers had to adjust to incorporating a new routine, and the study's small sample size limited the generalizability of results [17].

Overall, the literature suggests that alertness testing is a feasible approach to detect and prevent fatigue on the job [18, 19]. These tests provide objective measures (like reaction time) that correlate with alertness and can act as early warning indicators. However, most of what is known about their efficacy comes from controlled research settings rather than widespread industry adoption [17]. Occupational health and safety professionals currently have limited real-world guidance on how to effectively implement such testing programs [20]. Thus, while the concept is promising – potentially allowing intervention (e.g., a break or a shift change) before a fatigued worker has an accident – there remain practical hurdles in making alertness testing a standard safety practice. Key among these hurdles are gaining management buy-in, ensuring employee trust and participation (avoiding perceptions of surveillance or punitive use of results), and integrating the system smoothly into daily operations.

2.3 Fatigue Detection Technologies and Wearables

A variety of fatigue detection technologies have been developed to monitor workers for signs of impairment due to fatigue, with varying degrees of success in industrial applications. Traditional approaches in mining and transportation have included camera-based systems that monitor eye indicators such as PERCLOS (Percentage of Eyelid Closure) and wearable EEG headbands or caps that directly measure brainwave patterns associated with drowsiness. Each technology has its strengths and limitations. PERCLOS, for example, is effective in detecting eyelid droop and microsleeps and is relatively noninvasive, but it can be hampered by environmental conditions (poor lighting, driver position), and it provides only an indirect measure of cognitive state [2]. EEG-based fatigue monitors (like SmartCap systems) can measure microsleep events very accurately by detecting characteristic brainwave patterns, but they tend to be more intrusive (requiring the user to wear a headband or cap) and costly to implement broadly [21]. Other emerging solutions use oculomotor measures (pupil response and blink characteristics), head posture sensors, steering or equipment operation behavior, and even speech analysis to infer fatigue levels [22].

A recent report by the National Safety Council [22] surveyed impairment detection technology (IDT) vendors and highlighted the landscape of tools available for workplace fatigue and impairment monitoring [22]. It identified at least fifteen technologies, using methods ranging from psychomotor vigilance tests (reaction-time based) to eye tracking, motion sensors, and cognitive games, many of which target fatigue detection. Interestingly, only a minority of these systems (about 4 out of 15) utilize reaction time as a direct measure of impairment [22], despite the proven sensitivity of reaction time to fatigue. The most common approach among vendors was oculomotor (eye-movement) testing, used by 60% of companies, followed by general cognitive performance tests (40%). The report also noted that most IDTs are used pre-shift (as a screening) with some moving toward continuous monitoring throughout the work period [22]. Key barriers to adoption of these technologies include employee resistance to new monitoring, lack of management buy-in, union concerns, and privacy or comfort issues [22]. Cost can be a factor as well, although many solutions are moving to subscription models to reduce upfront expenses [22].

Within this landscape, wearable technology has emerged as a promising platform for fatigue monitoring. Smartwatches and other wearables offer a combination of sensing capabilities (accelerometers, heart rate sensors, etc.), user interface for simple tests, and connectivity, all in a device that workers may already find acceptable for daily use. Rogers et al. [3] specifically leveraged a wrist-worn device to measure both a performance metric (reaction time) and a subjective state (sleepiness via the Karolinska Sleepiness Scale (KSS)) [23] repeatedly during work shifts [3]. The rationale is that a wearable can integrate into the worker's routine with minimal intrusion and provide continuous or on-demand assessments. Early results from that study indicated that linking subjective and objective measures via a smartwatch application could provide a more complete picture of fatigue and potentially be used to determine fitness-for-duty in real time [3]. Moreover, because smartwatches have other everyday functionalities (timekeeping, communication, health tracking), operators may be more receptive to wearing them, as opposed to specialized fatigue devices that have no other use (hence improving user acceptance and engagement) [3].

Complementing the development of wearable monitors, data-driven modeling efforts like Talebi et al. [2] have shown that individualizing fatigue prediction is important. Their work demonstrated that factors driving fatigue can differ significantly from one operator to another, and that models benefit from accounting for each operator's unique profile (e.g., baseline alertness, sleep patterns, health conditions) [2]. In other words, a "one-size-fits-all" fatigue threshold may be suboptimal; personalized baselines and thresholds are needed for accurate fatigue detection. This insight is highly relevant for wearable systems: a smartwatch app could, for instance, establish each operator's typical reaction time when well-rested and then flag significant deviations from that baseline.

In summary, existing fatigue detection technologies provide a toolkit of options, but no single solution has emerged as universally superior for all situations [22]. Camera and EEG systems offer direct physiological measurements but can be invasive or limited in scope; simple alertness tests are easy to deploy but require user compliance and only capture moments in time. Wearables present a compelling middle ground, enabling continuous or frequent measurements that include both performance and physiological data, with relatively low burden on the user. The continued research and trials – including the present study – are aimed at validating these approaches in real-world mining environments. As fatigue remains a leading contributor to incidents and near-misses in mining, implementing effective monitoring coupled with proactive fatigue risk management could substantially improve safety outcomes.

3 Methodology

This study employed a wearable-based system using a smartwatch application to assess fatigue in haul truck operators in real-world mining operations. The system was developed to gather both objective performance data (via reaction time testing) and subjective fatigue reports (via self-ratings), enabling continuous and non-intrusive fatigue monitoring throughout a work shift. The goal was to create a tool that could seamlessly integrate into the daily workflow of operators without requiring external hardware or interfering with operational tasks.

3.1 Participants and Setting

The study included haul truck operators across three different mining sites with varying environmental, operational, and scheduling conditions. All participants were full-time employees engaged in day and night shift rotations. Participation was voluntary, and operators received brief training on how to use the smartwatch application before deployment. Data were collected during actual operating shifts to reflect natural work and fatigue conditions. A total of 46 haul truck operators participated in the study. Demographic information such as age, gender, and years of experience was not available, as these details were not provided by the participating companies due to confidentiality policies. Similarly, operators were not screened for sleep disorders or other health-related conditions because medical information was not accessible for privacy reasons. Consequently, the study sample represents a typical operational workforce without prior demographic or medical pre-screening.

3.2 Smartwatch Application and Testing Protocol

A custom application was deployed on commercially available smartwatches. The app featured a simple visual reaction time (RT) task, where the screen displayed a stimulus (e.g., a color change), and the operator was required to tap the screen as quickly as possible. Each reaction time test session required approximately 30–60 s to complete, making it feasible for operators to perform during natural pauses in haulage operations. Each RT test included several trials per session, and sessions were scheduled at fixed intervals throughout the shift (e.g., every 1–2 h), depending on the site's operational protocols. Figure 1 shows the user interface of the smartwatch application.

To capture subjective fatigue, the app also prompted operators to self-report their fatigue level using a 1–10 numeric scale, adapted from the Karolinska Sleepiness Scale (KSS). As Table 1 shows a score of 1 indicated high alertness ("very alert"), while a score of 10 indicated extreme drowsiness ("fighting sleep"). The fatigue rating prompt was triggered immediately after the RT test to ensure consistency in timing.

The application logged data automatically and transmitted it to a secure server for centralized analysis. Each record included:

Fig. 1 Smatwatch application user interface [3]

Table 1 Karolinska sleepiness scale [23]

Karolinska Sleepiness Scale (KSS)	
Extremely alert	1
Very alert	2
Alert	3
Rather alert	4
Neither alert nor sleepy	5
Some signs of sleepiness	6
Sleepy, but no effort to keep awake	7
Sleepy, but some effort to keep awake	8
Very sleepy, great effort to keep awake, fighting sleep	9
Extremely sleepy, cannot keep awake	10

- Timestamp.
- Reaction time (ms).
- Self-reported fatigue level (1–10).
- Shift type (day or night).
- Operator ID (anonymized).
- Site identifier.

3.3 Data Preprocessing and Normalization

For this study, a Fitbit smartwatch was used to collect the data. Fatigue levels were assessed using the Karolinska Sleepiness Scale (KSS). Following each KSS assessment, participants completed a reaction time (RT) test consisting of 20 visual stimuli. The smartwatch measured RTs in milliseconds and recorded whether each response was correct or incorrect. Data were collected from 46 haul-truck operators between October 2020 and March 2022, capturing both reaction time and subjective fatigue ratings.

A total of over 69,000 RT records and tens of thousands of fatigue ratings were collected. Prior to analysis, the data were cleaned to remove invalid or incomplete entries, and records with missing timestamps or inconsistent formatting were excluded. The initial dataset was screened to remove erroneous or physiologically implausible data. Trials with RT values below 250 ms (anticipatory responses) or above 1250 ms (delayed responses likely due to lapses) were excluded (~%34). Participants showing stereotypical or repetitive incorrect response patterns were also removed to ensure data reliability.

To minimize the influence of potential confounding factors such as caffeine intake, perceived stress, and varying workload, the analyses were designed using a within-person approach. Each participant served as their own control, allowing comparison of individual changes in reaction time (RT) across different shift conditions rather than between participants. Stable individual characteristics (e.g., habitual caffeine use or baseline alertness) were thereby held constant. Time of day and cumulative hours worked during a shift were incorporated into the models as covariates to capture diurnal variation and workload-related effects. This approach enabled the isolation of fatigue- or shift-related influences on RT from other individual or contextual factors.

Reaction time data were found to be positively skewed, as is typical in RT literature due to occasional lapses or microsleeps resulting in very slow responses. To address this, a Box-Cox power transformation was applied to normalize the distribution and reduce variance heterogeneity. A Box–Cox transformation was selected instead of a simple log-transform or non-parametric approach because it provides a flexible, data-driven method to correct for right-skewed reaction time distributions while preserving the assumptions required for parametric analyses such as ANOVA and regression. In this study, λ was estimated directly from the reaction time data using maximum likelihood estimation, which identifies the value that best approximates a normal distribution and stabilizes variance ($\lambda \approx -1.56$).

RT records were then categorized into key temporal groupings:

- Group 1: Early morning fatigue window (2:00–5:00 AM).
- Group 2: Afternoon window (2:00–5:00 PM).
- Group 3: All other hours.

The data were also segmented by shift type (day vs. night) and cumulative hours worked since shift start to evaluate time-on-task fatigue effects.

3.4 Statistical Analysis

We conducted one-way and two-way ANOVA tests to evaluate:

- Differences in reaction time by time-of-day groups.
- Effects of shift type and site on performance.
- Interaction effects between temporal and site-based variables.

Post-hoc analysis using Tukey's HSD was applied to compare group means and identify significant pairwise differences. To assess fatigue accumulation, regression analysis and trend plotting were used to track RT as a function of hours worked during the shift. While site, shift, and their interaction were modeled with controls for time of day and hours into shift, unmeasured site-specific scheduling practices (e.g., roster patterns, overtime policies) may still contribute to between-site differences and are noted as a limitation.

Subjective fatigue ratings were analyzed for distributional patterns and compared with objective RT results to evaluate alignment and divergence between perceived and actual performance.

3.5 Ethical Considerations

All data were anonymized and stored on secure servers. Participation in the study was voluntary, and operators could opt out at any time without consequence. The smartwatch application was non-invasive and used during non-critical moments, such as during scheduled pauses or vehicle idling.

4 Analysis and Results

After developing a smartwatch-based reaction time testing application and deploying it with haul truck operators, we collected a dataset of reaction times over the course of full work shifts. In addition, operator fatigue levels were recorded during the shift. Operators were prompted to select their fatigue level on a scale from 1 to 10. In this section, we present the analysis of these data, focusing on the overall distribution and variability of reaction times, differences in performance at various times of day (with particular attention to known fatigue-prone hours), and comparisons between objective reaction time measures and subjective fatigue assessments. All statistical analyses were performed using appropriate methods given the characteristics of the data, and key results are visualized in the figures and tables referenced below.

To reduce the influence of other factors like caffeine or stress, we set up the analysis so that each participant was compared to themselves over time. Models made based on individual differences (for example, regular caffeine habits) and included controls for time of day and hours worked in the shift, which reflect natural tiredness and workload. The results are also tested by removing unusually fast or slow reaction times. Shift hours are treated as a practical measure of workload, since this information is always available in mining operations.

4.1 Reaction Time Distribution and Variability

Prior to examining specific fatigue trends, we assessed the distribution of the reaction time data. As is common in cognitive performance measures, the reaction time data in this study were positively skewed – meaning that while most responses were relatively fast, there were a number of much slower responses that long-tailed the distribution [24]. This positive skew is typical for RT data, especially in contexts where fatigue or lapses of attention may produce occasional very slow responses (outliers). Because many statistical tests (e.g., t-tests, ANOVA) assume normally distributed data, this skew required attention. We applied a Box-Cox

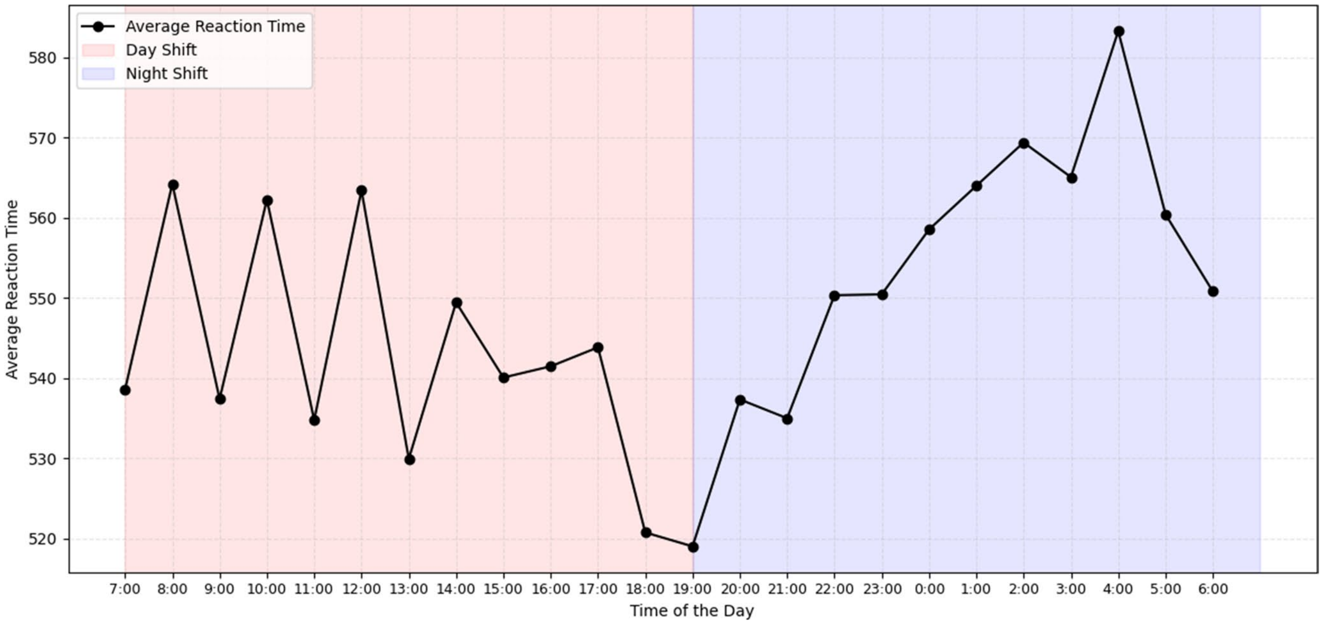

Fig. 2 Average reaction time for each hour of the day

power transformation to the reaction times to normalize the distribution (i.e., to reduce skewness and stabilize variance) before conducting inferential statistical tests. This step helps ensure that the results of significance tests are valid and not unduly influenced by non-normality.

We then examined the reaction time data across different time windows of the shift to see how performance varied. A clear trend emerged: the average reaction time was substantially higher (slower responses) during the early morning hours roughly between 02:00 and 05:00 AM compared to the rest of the day. Figure 2 illustrates the mean reaction time for each hour of the day. The data points in the 2–5 AM window are consistently higher than those at other times, indicating significantly slower reactions during these hours. This pattern aligns with expectations, as the pre-dawn hours are associated with the circadian trough in human alertness. In fact, fatigue-related performance decrements are well-documented in this timeframe; for instance, accident statistics show a higher incidence of drowsiness-related crashes between about 2 AM and 6 AM [25]. The slowed reactions in our data during 2–5 AM likely reflect the combined effect of circadian low (biological night for most individuals) and potentially the accumulated fatigue of working a long shift into those hours.

Table 2 Details of the distribution of data

Category		Mean	Standard Deviation
Entire data	All data	548.8	139.4
Group 1	Data from 2 to 5 AM	572.6	140.7
Group 2	Data from 2 to 5 PM	544.4	141.5
Group 3	Rest of data	546.3	138.5

To quantify these differences, we categorized the data into groups corresponding to key periods: Group 1: 2–5 AM (peak fatigue hours), Group 2: 2–5 PM (afternoon period), and Group 3: all other times. Table 2 summarizes the distribution of reaction times in these groups and for the entire dataset. Group 1 (2–5 AM) had a mean reaction time of 572.6 ms with a standard deviation of 140.7 ms, noticeably slower than the overall dataset mean of about 548.8 ms. Group 3 (the rest of the day excluding the other groups) by contrast had a mean around 546.3 ms (with SD ≈ 138 ms). Group 2 (2–5 PM, a mid-afternoon window often associated with a mild "post-lunch" dip in alertness) interestingly showed the fastest mean reaction time at 544.4 ms, with an SD of 141.5 ms. This suggests that in our sample, the mid-afternoon did not produce a significant fatigue effect; in fact, operators might have been relatively alert during 2–5 PM (perhaps because this period could coincide with the start of a day shift or after a substantial break, depending on scheduling). In contrast, the early morning Group 1 not only had the slowest average reactions but also the highest variability. The standard deviation of ~ 140.7 ms in the 2–5 AM window was slightly higher than the variability during other times (~ 138.5 ms for Group 3). This greater spread implies that reaction times at night were more inconsistent – some responses were only moderately slower than baseline, while others were extremely slow (likely reflecting occasional lapses or microsleeps).

Fgure 3, 4, 5 and 6 show the detailed distribution histograms for All data, Group 1, Group 2, and the rest of the data (Group 3), respectively. All distributions are right-skewed to some degree, but the skew is most pronounced

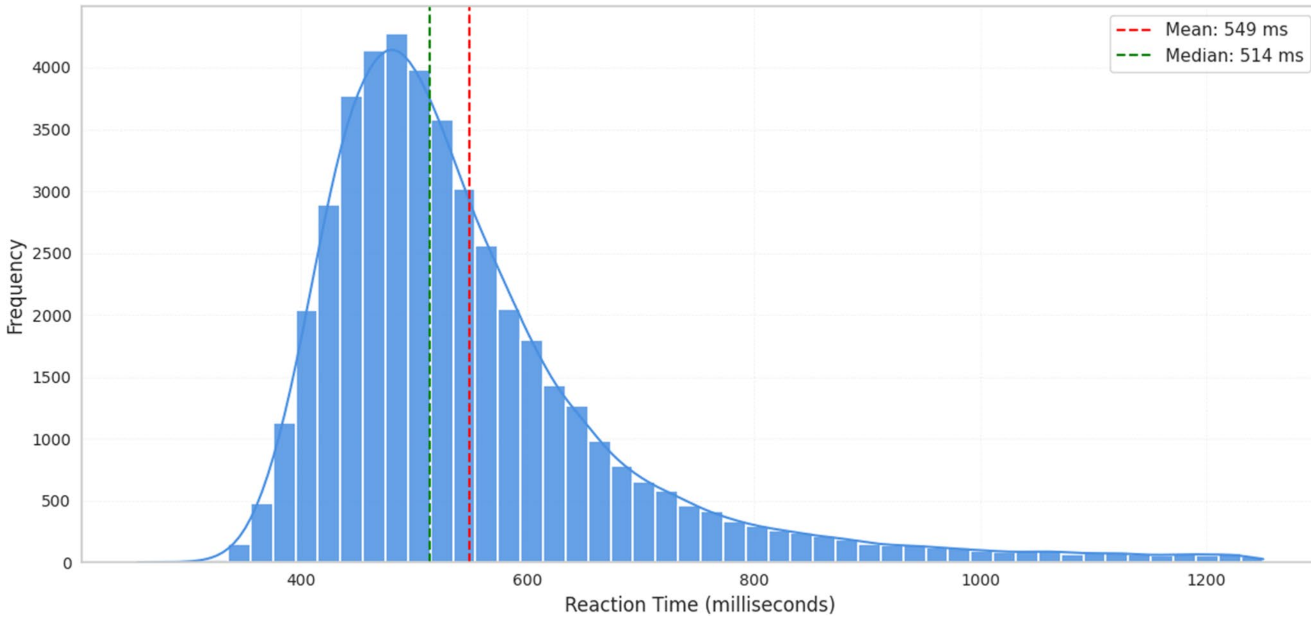

Fig. 3 Distribution of reaction time (All data)

in the 2–5 AM data (Fig. 1). In the distribution, a longer tail of very slow responses is evident, indicating that fatigue induced not just a uniform slowing, but sporadic reaction time outliers where an operator may have been momentarily inattentive. These could correspond to instances of micro-sleep or momentary lapse in concentration. The presence of such extreme slow responses is an important aspect of fatigue effects; rather than all reactions uniformly slowing by, say, 50 ms, what often happens is that most reactions remain near an individual's baseline but occasionally a very slow reaction occurs when the person momentarily loses alertness. This pattern is consistent with fatigue affecting vigilance: the person can perform nearly normally most of the time, but fails to respond promptly on some trials. The net effect is an increase in mean RT and an increase in variability. Indeed, our comparisons showed that the coefficient of variation (ratio of standard deviation to the mean) was highest during 2–5 AM

In practical terms, these findings mean that during the early morning hours, operators not only respond slower on average, but their performance is less reliable. One driver might have a reaction of 500 ms on one test (only slightly

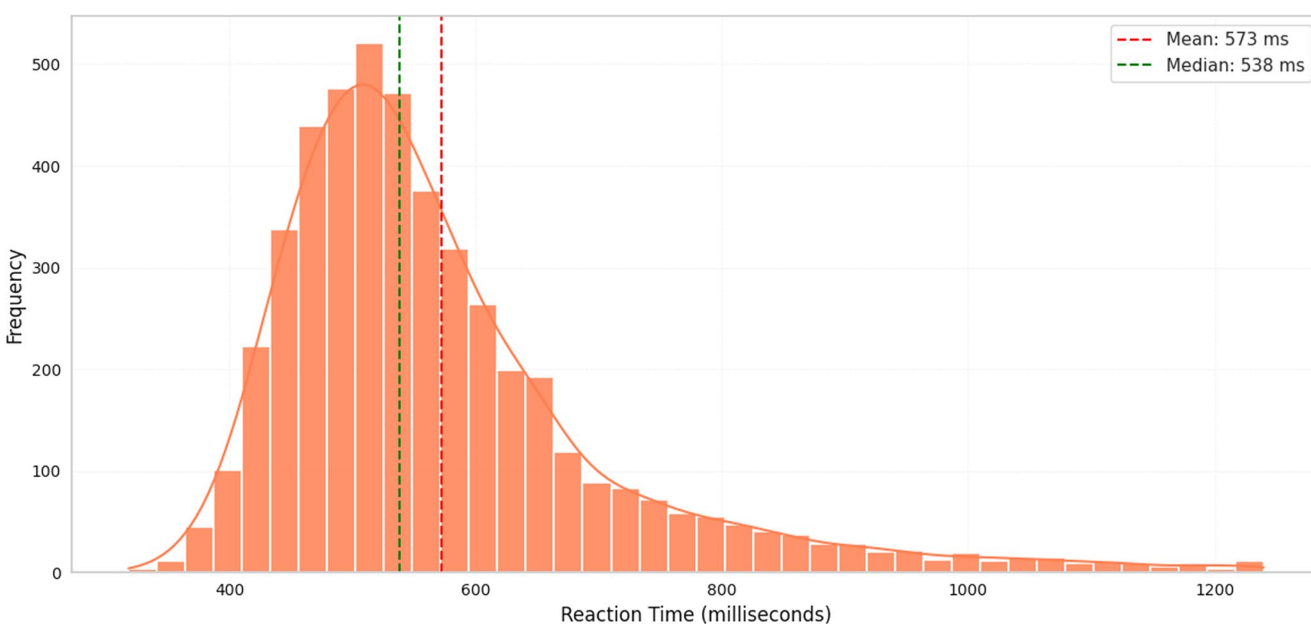

Fig. 4 Distribution of reaction time (2 to 5 AM)

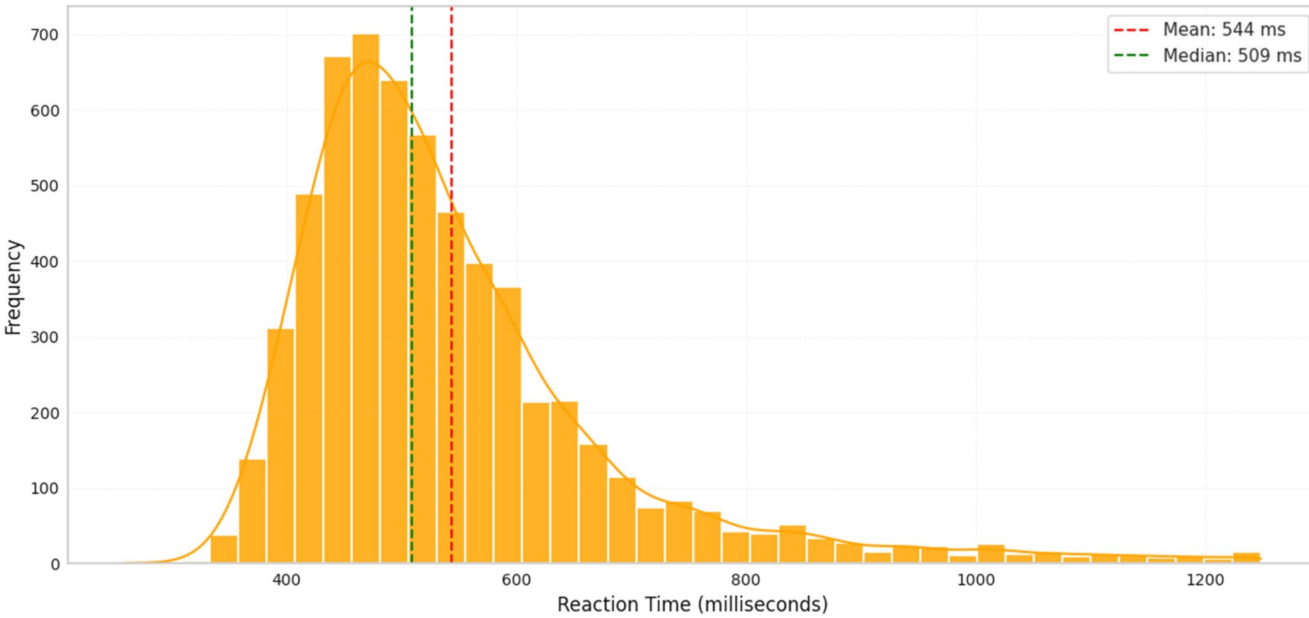

Fig. 5 Distribution of reaction time (2 to 5 PM)

worse than their typical 450 ms, for example) and then 800 ms on the next test, reflecting a lapse. Such variability can be dangerous in operational contexts, as it only takes one lapse at a critical moment to lead to an accident. The data reinforce the importance of fatigue management, especially for night shifts. The statistical significance of the time-of-day effect was confirmed by an ANOVA on the normalized reaction time data: differences among the three defined groups were significant ($p < 0.001$). Post-hoc comparisons using Tukey's HSD (Table 3) indicated that Group 1 (2–5 AM) was significantly different from both Group 2 and Group 3 ($p = 0.001$ for both comparisons), whereas the difference between Group 2 (2–5 PM) and Group 3 (rest of data) was not statistically significant ($p \approx 0.829$). These results support the conclusion that the early-morning fatigue window has a distinct impact on reaction time.

To further understand the role of fatigue accumulation over the course of a shift, we examined how reaction times evolved as a function of time worked since shift start. Figure 7 plots the average reaction time (RT) against the

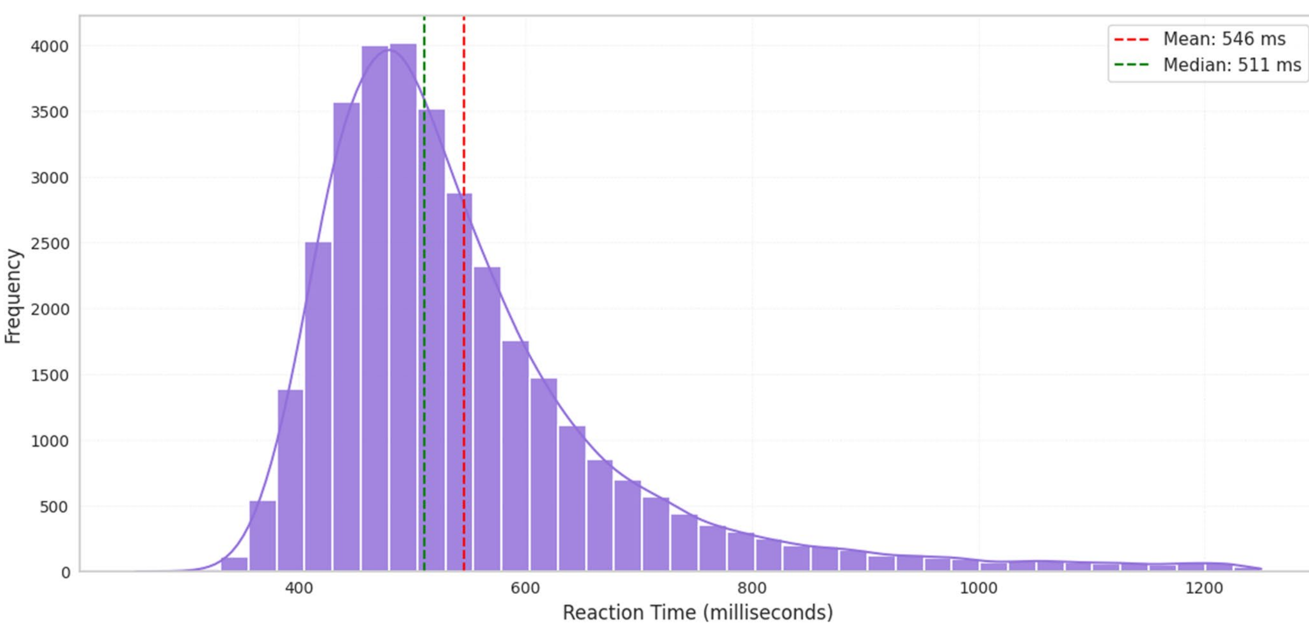

Fig. 6 Distribution of reaction time (Rest of data)

Table 3 Tukey HSD results: multiple comparison of means - Tukey HSD, FWER = 0.05

Dataset Group	Dataset Group	Mean Difference	*p*-value	Lower	Upper	Reject
Group 1	**Group 2**	−28.23	0.001	−34.50	−21.95	True
Group 2	**Group 3**	1.12	0.829	−3.38	5.63	False
Group 1	**Group 3**	−27.10	0.001	−32.15	−22.06	True

number of hours worked, separately for day shift (red) and night shift (blue). A distinct divergence is observed between the two shifts. For the day shift, average reaction time fluctuates within a relatively narrow range and shows no clear upward trend over time. In contrast, the night shift exhibits a consistent increase in RT, particularly after the fifth hour on duty, peaking around the 9th hour. This pattern suggests a cumulative fatigue effect during night work, where reaction time continues to deteriorate as the shift progresses. Notably, the initial RTs at the start of the night shift are faster than during the day shift, but this advantage diminishes rapidly, and by mid-shift, night shift RTs surpass those of the day shift and continue to rise. This finding complements the earlier time-of-day analysis (Fig. 2), indicating that both circadian timing and time-on-task fatigue jointly influence performance. The upward slope of night shift RTs supports the hypothesis that accumulated sleep pressure and biological night jointly impair vigilance and reaction consistency over time. Together, these results highlight the operational risks associated with prolonged night work, particularly during the latter half of the shift.

For each operator–shift, the pre-shift baseline was defined as the last RT recorded within 60 min before shift start. Absolute change ($\Delta RT = RT - \text{baseline}$) and percent change ($\%\Delta RT = 100 \times \Delta RT/\text{baseline}$) were computed for in-shift readings. To limit compositional bias from unequal sampling, operator-level shift means of %ΔRT were summarized and then aggregated by mine and shift. Welch's t-tests on operator-shift means compared Night vs. Day within each mine. (Interpretation: positive %ΔRT = slower than baseline; negative %ΔRT = faster than baseline.)

Across mines, Night shifts showed equal or smaller %ΔRT relative to baseline. Pooled across mines, the overall median %ΔRT was 5.89% on Day vs. 2.91% on Night (Night–Day = −2.98% points). Operator–shift sample sizes by mine were A: 28/28 (Day/Night), B: 60/80, and C: 51/58.

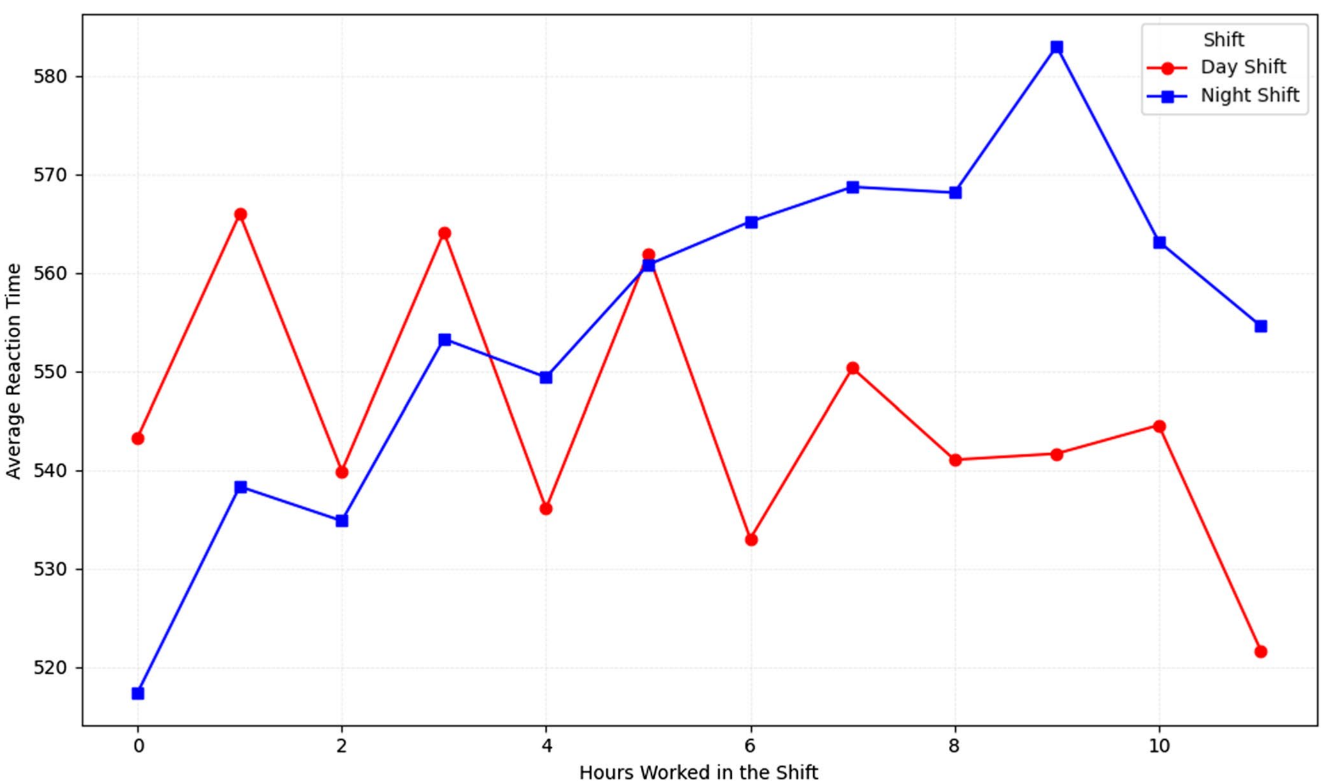

Fig. 7 Average reaction time vs. number of hours worked

4.2 Fatigue Level Distribution and Variability

Figure 8 illustrates the frequency distribution of self-reported fatigue levels based on the Karolinska Sleepiness Scale (KSS), which ranges from 1 ("Extremely alert") to 10 ("Extremely sleepy, cannot keep awake"). The distribution is strongly right-skewed, with the majority of responses concentrated at lower fatigue levels. The most frequently reported level is 3 ("Alert"), with more than 1,750 instances, followed by level 2 ("Very alert") and level 4 ("Rather alert"), each with several hundred responses. Moderate levels such as 5 ("Neither alert nor sleepy") and 6 ("Some signs of sleepiness") appear less frequently, while higher fatigue levels—7 through 10, which correspond to increasing degrees of sleepiness and reduced wakefulness—are rarely reported. Notably, extreme fatigue levels like 9 ("Very sleepy, great effort to keep awake, fighting sleep") and 10 ("Extremely sleepy, cannot keep awake") appear in very few cases. This distribution indicates that most individuals reported being in an alert or mildly fatigued state at the time of assessment.

Figure 9 displays the average fatigue level by month, separated by day and night shifts. Throughout the year, night shift workers consistently report higher average fatigue levels than those on the day shift. Night shift values range from approximately 2.7 in June to over 4.2 in September, with sustained elevations above 3.5 from August through December. In comparison, day shift fatigue levels are more stable and fall within a narrower range—from approximately 2.4 in February to 3.55 in August. The largest gap between the two shifts occurs in September, where night shift fatigue surpasses day shift levels by nearly a full point. A notable convergence is observed in August, where both shifts report similar fatigue levels around 3.5. While monthly fluctuations are present in both groups, the night shift consistently maintains higher average fatigue scores, particularly in the latter half of the year.

Figure 10 presents the variation in average fatigue levels across different hours of the day, distinguishing between the day shift (07:00–18:59, shaded red) and night shift (19:00–06:59, shaded blue). During the day shift, fatigue levels remain relatively stable, ranging between approximately 2.8 and 3.1. The lowest average fatigue is recorded between 17:00 and 18:00, with values just below 2.8. At the beginning of the night shift, average fatigue levels begin to increase, rising from approximately 3.1 at 19:00 to over 3.5 by 22:00.

From 23:00 onward, fatigue levels remain elevated. The highest levels are observed between 00:00 and 04:00, with a peak value nearing 3.9 around 01:00. This elevated pattern continues until around 04:00, after which a gradual decline is observed. By 06:00, average fatigue levels return closer to 3.0. The data indicates a distinct shift in fatigue patterns between day and night shifts, with a higher concentration of

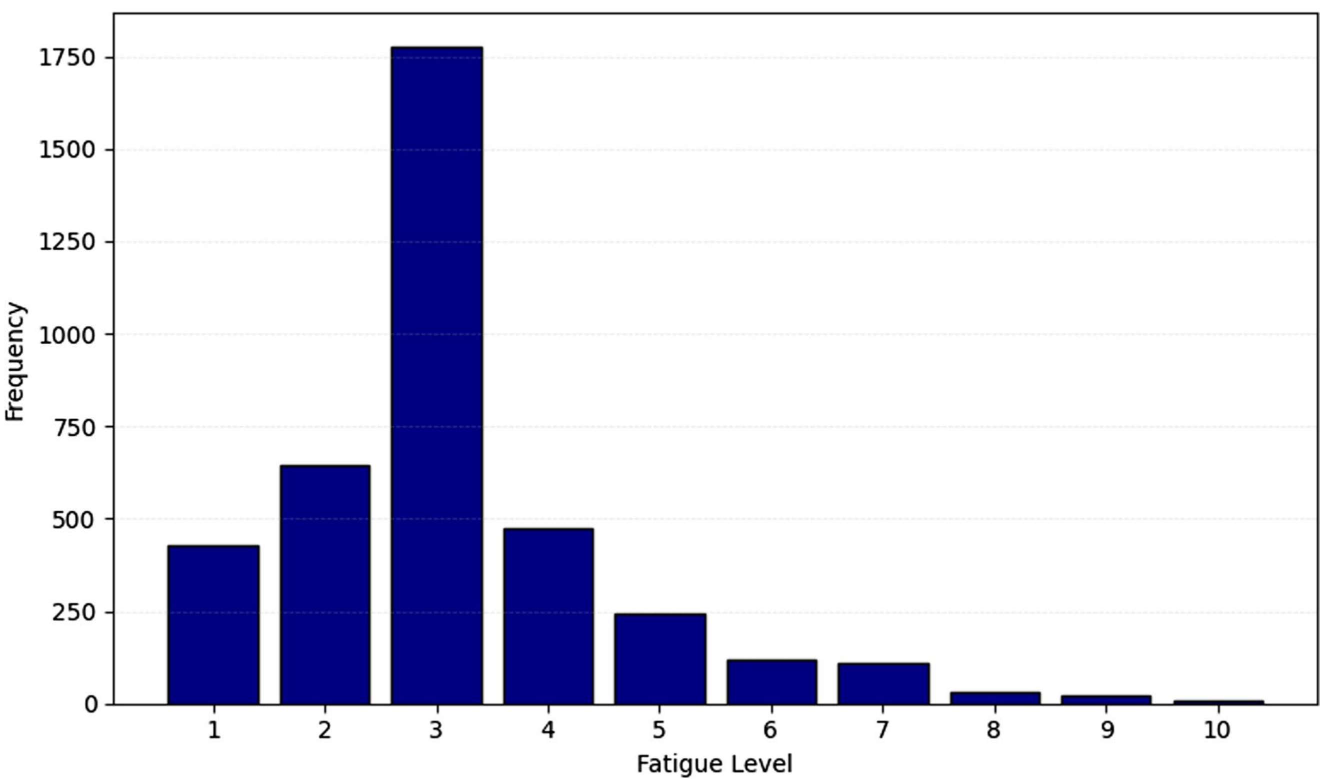

Fig. 8 Frequency of fatigue levels

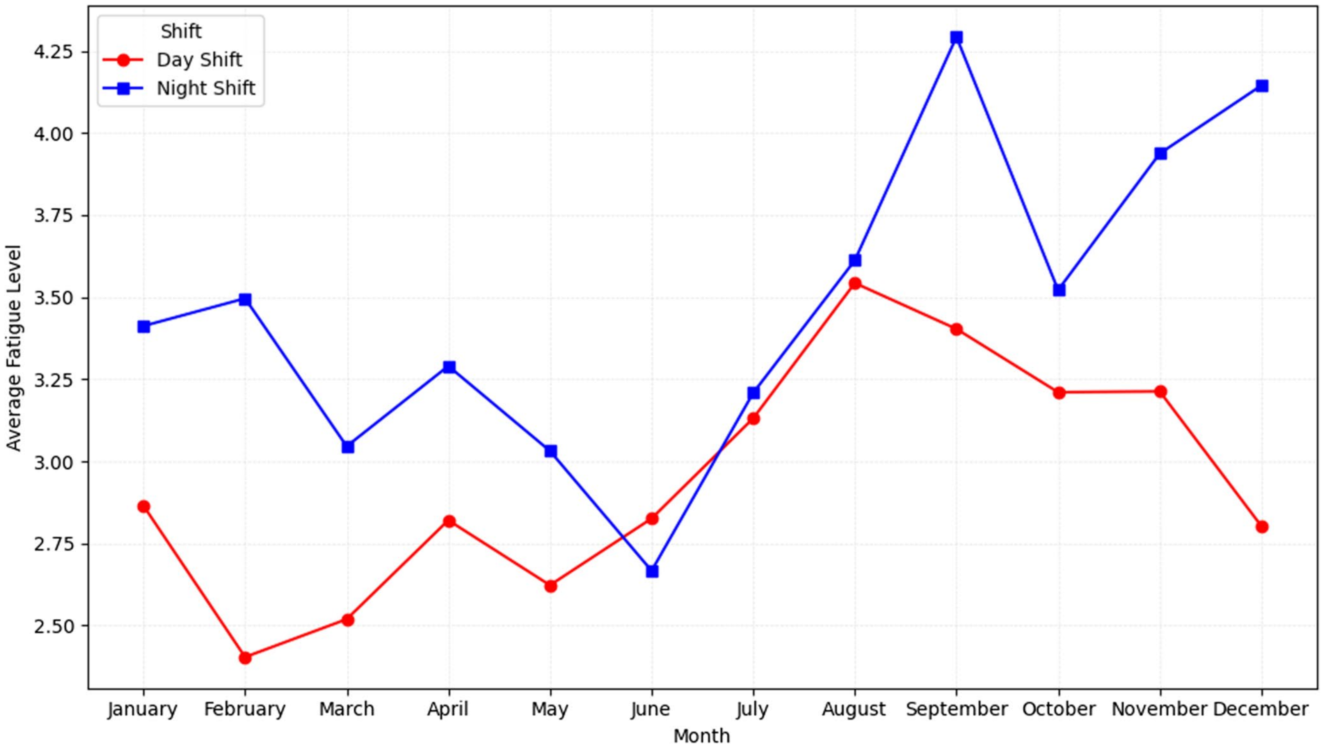

Fig. 9 Average fatigue level by month and shift

elevated fatigue levels during the late night and early morning hours.

The stacked bar chart (Fig. 11) presents the distribution of fatigue levels, rated from 1 to 10, reported during day and night shifts. Fatigue levels are grouped into three categories: low (levels 1–3), moderate (levels 4–6), and high (levels 7–10). During the day shift, the majority of reported fatigue levels fall within the low category, accounting for 52.2% of the total. Moderate fatigue levels represent 36.3%, while high fatigue levels make up only 11.5% of the responses. In contrast, the distribution during the night shift shifts notably. The proportion of low fatigue levels decreases to 29.9%,

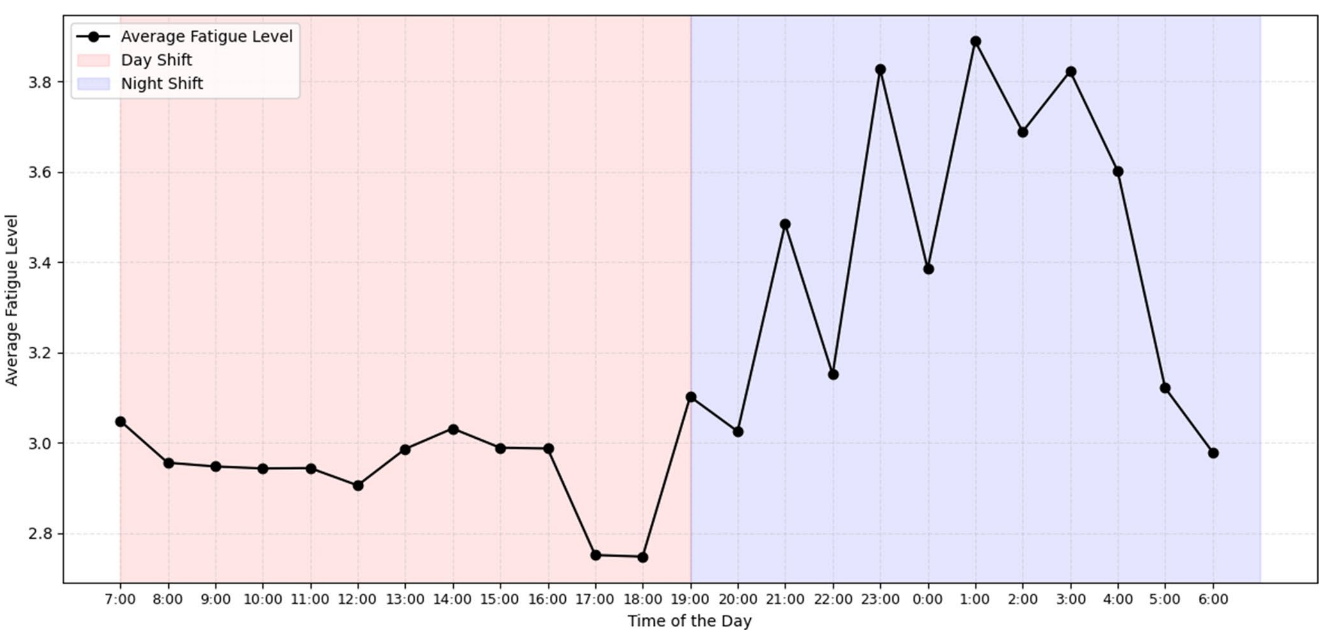

Fig. 10 Average fatigue level for each hour of the day

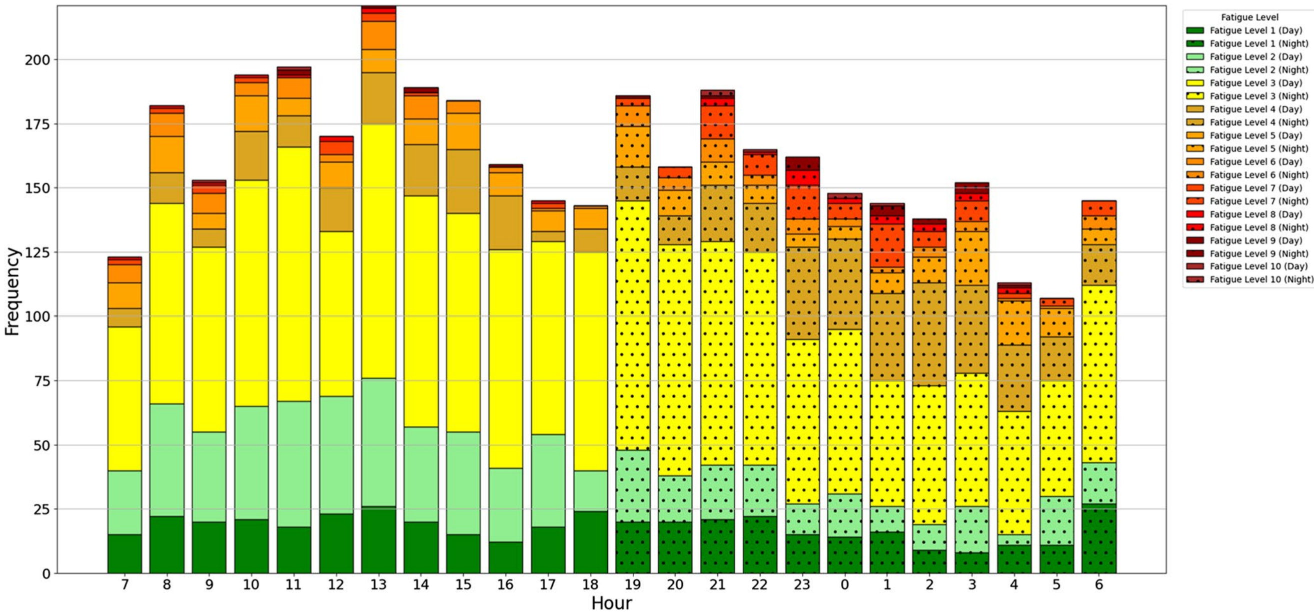

Fig. 11 Stacked frequency of fatigue level for each hour of the day

while moderate levels rise slightly to 40.8%. Most notably, high fatigue levels increase to 29.3%, indicating a broader spread across the fatigue spectrum during night hours. The data shows a clear shift in the distribution of fatigue levels between day and night shifts, with the night shift having a higher percentage of responses in the moderate to high fatigue categories and a lower percentage in the low fatigue category.

The box plots in Fig. 12 illustrate the distribution of reaction times across varying fatigue levels (0–9) for both the day shift and night shift. In both cases, a clear trend emerges: as fatigue level increases, reaction time also tends to increase, particularly at higher fatigue levels (levels 6–9), which are highlighted in blue. This trend is more pronounced in the night shift, where median reaction times and the interquartile range (IQR) increase substantially with fatigue. For example, night shift participants with fatigue levels of 8 and 9 show notably higher median reaction times and greater variability, indicating impaired cognitive performance under extreme fatigue.

In the day shift, reaction times remain relatively stable from fatigue levels 0 to 5, with modest increases beginning at level 6. However, the most significant increases in median reaction time and distribution spread occur beyond fatigue level 7, suggesting a potential fatigue threshold after which performance degradation becomes more severe.

To evaluate the effects of shift type (day vs. night) and site (Site A, B, and C) on participants' reaction time (RT), a two-way analysis of variance (ANOVA) was conducted (Table 4). The results revealed a statistically significant main effect of shift on reaction time ($F(1, 43645) = 52.30$, $p < 0.001$), indicating that average RTs differ meaningfully between day and night shifts. A highly significant main effect of site was also observed ($F(2, 43645) = 418.90$, $p < 0.001$), suggesting that RTs vary substantially across the different operational locations. Furthermore, there was a significant interaction between shift and site ($F(2, 43645) = 10.57$, $p < 0.001$), indicating that the impact of shift on reaction time was not consistent across all sites. For example, one site may experience a more pronounced degradation in reaction time during the night shift than others.

4.3 Multivariate Analyses of Reaction Time

To examine the relative contribution of operational and individual factors to variability in reaction time (RT), a multivariate regression analysis was conducted using a random forest model. Four predictor variables were included in the analysis: shift type, worksite, hour of the day, and operator ID. These features were selected based on their relevance to temporal patterns, task context, and inter-individual variability in fatigue and performance.

The model was trained on 36,475 out of 45,594 records and used feature importance scores to evaluate the relative predictive power of each variable. Feature importance quantifies how much each variable contributes to reducing model error. As summarized in Table 5, shift type emerged as the most influential predictor (29%), followed by worksite (26%), hour (24%), and operator ID (21%).

This approach allowed for quantifying and comparing the relative influence of each factor on RT while accounting for nonlinear interactions and interdependencies between

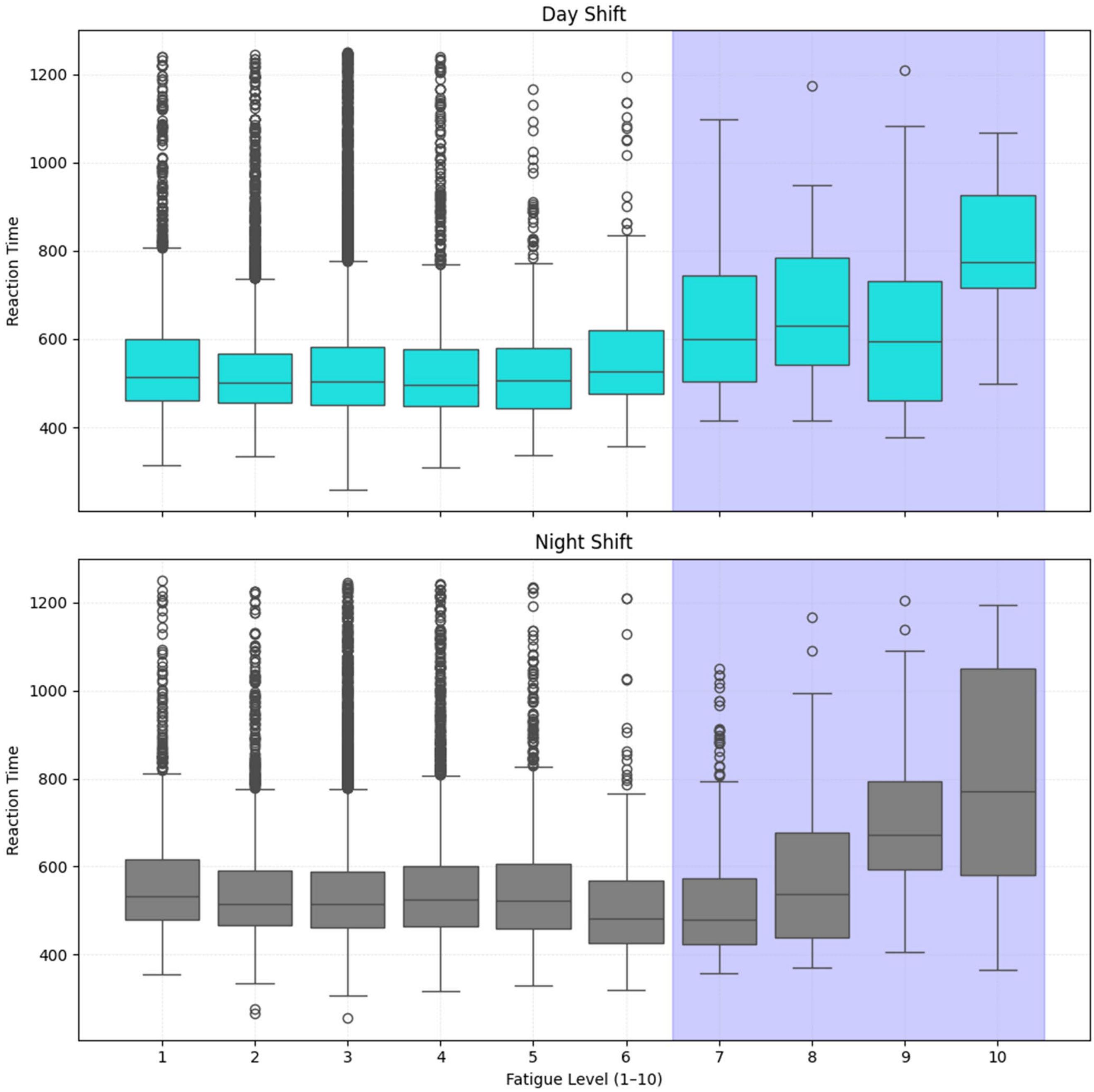

Fig. 12 Reaction time vs. fatigue level

Table 4 Statistical analysis of reaction time by shift and site

Source	Sum of Squares	Degree of Freedom	F-value	*p*-value	Significance
Shift	979,565	1	52.30	4.83e-13	Significant
Site	15,690,910	2	418.90	6.29e-181	Very significant
Shift and Site	396,023	2	10.57	2.57e-05	Significant interaction

Table 5 Feature importances in predicting RT

Feature	Importance
Shift	0.29
Worksite	0.26
Hour	0.24
Operator ID	0.21

variables. The close grouping of importance values suggests that RT is not driven by a single dominant factor, but rather a combination of temporal, contextual, and individual components.

To provide interpretable effect estimates alongside the exploratory random-forest results, a multiple linear regression (MLR) was applied to baseline-normalized reaction time, expressed as the percent change relative to each operator's own pre-shift baseline (%ΔRT; positive = slower than baseline). The primary unit of analysis was the operator–shift mean of %ΔRT from repeated observations within a shift. The model included shift (Night vs. Day), hours since shift start, and mine site indicators. Coefficients are reported as unstandardized β (percentage points, pp) and standardized β* (SD units), with 95% confidence intervals (CI) and p-values.

Results (operator–shift means). Night shifts were not associated with a meaningful change in %ΔRT relative to Day (β = −2.03 pp, 95% CI [−8.96, 4.89], $p = 0.565$; β* = −0.079). The linear effect of hours since shift start was negative but not statistically reliable (β = −9.85 pp per hour, 95% CI [−25.20, 5.49], $p = 0.208$; β* = −0.845), and the quadratic term suggested a modest curvature that did not reach conventional significance (β = 1.05, 95% CI [−0.21, 2.32], $p = 0.103$; β* = 0.965). Mine indicators were not statistically different from the reference site (e.g., Mine Site B: β = −2.49, 95% CI [−14.27, 9.30], $p = 0.679$; Mine Site C: β = 2.37, 95% CI [−7.86, 12.61], $p = 0.649$).

Taken together with the random-forest feature importance (which highlights several comparably influential factors), the MLR indicates no evidence that Night shifts worsen baseline-normalized RT compared with Day after adjusting for within-shift timing and mine site. Observed differences are small and statistically uncertain, reinforcing the view that RT variation reflects a combination of temporal, contextual, and individual components rather than a single dominant driver.

4.4 Case Study: Operator-Specific Threshold Determination

To determine an effective RT threshold for identifying high fatigue periods, it is focused on Operator ID = 71 due to their broad range of reaction times and a well-distributed set of fatigue scores. This made them an ideal candidate for threshold calibration. By comparing RT distributions across low and high fatigue states, it is aimed to identify a cut-off point that reliably separates these conditions. The motivation for this analysis was to establish an operator-specific threshold as a proof of concept, which can later be adapted to generate individualized fatigue thresholds for multiple operators in real-time monitoring systems. The RT = 590 ms value was selected based on its ability to visually and practically distinguish the majority of high fatigue responses from low fatigue ones with minimal overlap.

The 590 ms fatigue threshold presented here is based solely on data from a single operator (Operator ID = 71), who had the most continuous and well-distributed dataset suitable for this analysis. Therefore, this threshold should be interpreted as a case study example demonstrating the feasibility of determining operator-specific fatigue thresholds rather than as a universal cutoff applicable to all workers. The differences observed among individuals indicate that each operator should have a personalized reaction time threshold rather than a universal value. A data-driven model can be developed to dynamically estimate this individualized threshold in real time for each operator, accounting for their baseline performance and variability.

Figure 13 presents the distribution of RTs for both low and high fatigue states, along with the threshold indicated by a red dashed line. Fatigue levels were categorized as "low" when the self-reported fatigue score was less than 6, and "high" when the score was 6 or greater. Low fatigue responses are shown in blue, while high fatigue responses are shown in orange. The vertical red line represents the proposed threshold of 590 ms.

While the example shown in Fig. 13 demonstrates a clear separation between low- and high-fatigue responses for one operator, it is important to note that this threshold (RT = 590 ms) is not intended as a universal cutoff applicable to all individuals. Reaction time varies naturally across operators due to factors such as age, baseline alertness, cognitive processing speed, and habitual caffeine use. Therefore, defining a single reaction time threshold for an entire workforce would be inappropriate and could lead to misclassification.

In this study, the 590 ms value serves as an illustrative example to demonstrate the feasibility of establishing individualized fatigue thresholds. In practical applications, thresholds should be personalized and task-specific, determined relative to each operator's well-rested baseline and the cognitive demands of the job. For instance, a reaction time that represents a 15–20% decline from a worker's own baseline could serve as a trigger for a fatigue alert in future implementations. This individualized approach ensures that fatigue detection accounts for both personal variability and the specific operational context of haul truck driving.

To reduce the influence of other factors like caffeine or stress, the analysis is set up in a way that each participant was compared to themselves over time. Models that account for individual differences (for example, regular caffeine habits) were applied and controls for time of day and hours worked in the shift were included to reflect natural tiredness and workload. Also, results are tested by (i) looking only at the early part of shifts and (ii) removing unusually fast or

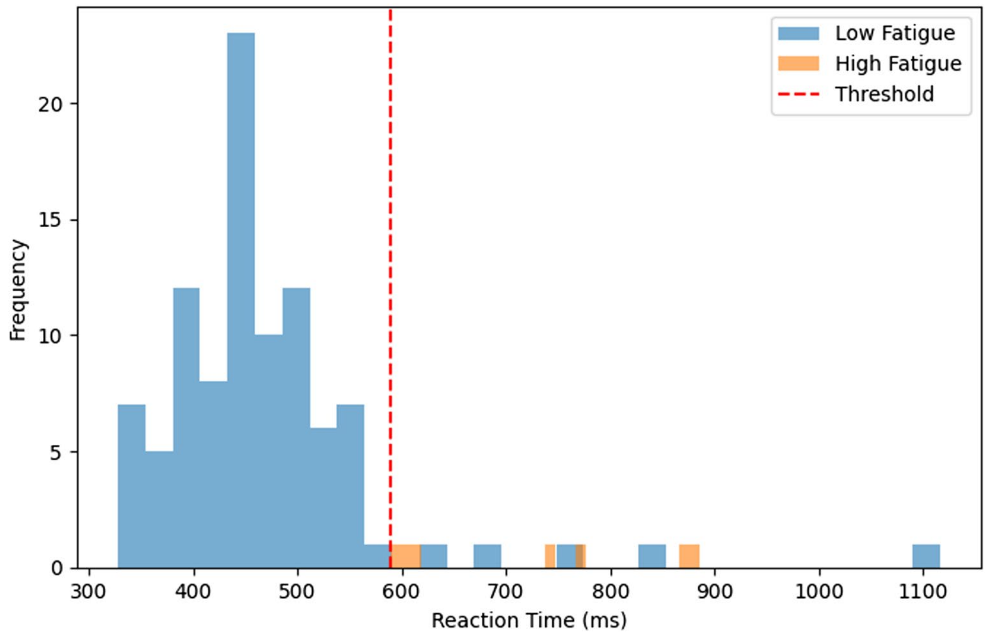

Fig. 13 Distribution of reaction time for the operator with well-distributed data (Operator ID = 71)

slow reaction times. Shift hours were treated as a practical measure of workload, since this information is always available in mining operations.

5 Discussion

5.1 Signal Detection Performance Under Fatigue

The reaction time test used in this study involved the appearance of different colored stimuli on the smartwatch screen, to which the operator had to respond as quickly as possible when the target color appeared. This paradigm can be viewed through the lens of signal detection theory (SDT). In our case, one can consider the presentation of a stimulus (e.g., a specific color that requires a button press) as a "signal" that the operator must detect and respond to, while other colors or the absence of a stimulus could be considered the "noise" or non-target condition. Under conditions of fatigue, not only do reaction times lengthen, but errors of omission (missed signals) or commission (false alarms) can increase. In other words, fatigue might cause an operator to sometimes fail to respond at all when they should (a miss), or respond when they shouldn't (a false alarm), though the latter is less likely in a simple vigilance task.

In our data, nearly all scheduled tests were completed, but it was observed anecdotally that during the night shift hours some operators missed responding to a stimulus entirely on a few test trials, which was recorded as either a very large reaction time or a null response. These missed responses are essentially extreme cases of slowed reaction (a lapse where the response was so late it was not registered as valid). From an SDT perspective, this can be seen as a failure to hit on a signal. Prior research indicates that as fatigue increases, the rate of such lapses (misses) tends to rise [3]. Hits (correct detections) become less frequent and the operator's sensitivity to the signal diminishes, even if they are trying to remain alert. Meanwhile, false alarms (responding when no target is present) could potentially increase if the individual becomes confused or overly trigger-happy to compensate, though in our simple task false alarms were not a major issue (since the task was straightforward and the operators were instructed clearly on when to respond).

Rogers et al. [3] analyzed wearable test data using a signal detection framework and observed clear performance differences between day and night shifts, particularly in accuracy and response reliability [3]. Our findings align with this: during the 2–5 AM window, in addition to slower reaction times, the few instances of missed responses indicate lapses in detection. While our primary metric was reaction time (for successful responses), incorporating an analysis of missed signals would provide an even clearer picture of fatigue impact. An operator who is so fatigued that they occasionally fail to respond at all within the test interval is an operator who might also fail to notice a warning signal or hazard in the real work environment. This emphasizes that reaction time monitoring can serve a dual purpose – measuring how fast responses are and also keeping track of whether responses are made at all when expected.

Going forward, framing the results in terms of signal detection theory could yield additional useful metrics: for example, calculating a d' (d-prime) sensitivity score for each operator for day vs. night conditions, or counting lapses (responses above a certain threshold, say 1000 ms, could be considered " lapses"). These metrics can complement average reaction time as indicators of fatigue. The key takeaway from the perspective of signal detection is that fatigue degrades an operator's ability to consistently detect and respond to critical signals. This degradation was evident in our data through both slower responses and the occasional failure to respond, mirroring findings in controlled experiments on fatigued participants [3]. Ensuring that such lapses are minimized is crucial in safety-critical operations; thus, if an increase in miss rate is detected via a wearable test, it should trigger an immediate fatigue mitigation action (like alerting the operator or supervisor for a break or other intervention).

5.2 Subjective vs. Objective Fatigue Indicators

In parallel with objective reaction time measurements, it is important to consider subjective fatigue indicators – essentially, how tired the operators feel. A common self-report metric is the Karolinska Sleepiness Scale (KSS), which asks individuals to rate their level of sleepiness on a numeric scale (typically 1 = extremely alert, up to 9 = extremely sleepy). In this study, subjective fatigue data were collected periodically by prompting operators to report their perceived fatigue or sleepiness level (using a KSS-style prompt on the smartwatch) at key times (e.g., start, middle, and end of shifts). This allows us to compare subjective self-assessments with the objective performance data from the reaction time tests.

One might expect that as reaction times slow (indicating increased fatigue), the subjective sleepiness ratings would also increase (indicating the person feels more fatigued or drowsy). To some extent, our data did show the expected trend: operators generally reported higher sleepiness (higher KSS scores) during night shifts and especially in the early morning hours, compared to the day shift. For example, KSS ratings collected toward the end of night shifts were often 7 or 8 ("sleepy, some effort to keep alert"), whereas during day shifts they were typically around 3 or 4 ("alert" to "rather alert"). This aligns with the reaction time results that showed degradation at night. An ANOVA on the subjective sleepiness scores by shift (day vs. night) confirmed a significant effect of shift, paralleling the objective data (in line with findings by Rogers et al. 2023, who also found KSS scores on night shifts to be significantly higher than on day shifts) [3].

However, when the correspondence between subjective and objective measures on a finer level are examined, a more complex picture emerges. In our analysis, the correlation between an individual's reaction time and their concurrent KSS self-rating was weak – in fact, it was statistically non-significant. This means that knowing someone's reaction time at a given moment did not reliably predict what they would report as their sleepiness level, and vice versa. Similarly, Rogers et al. [3] reported almost no correlation (Pearson $r \approx 0.13$, $p > 0.2$) between the KSS scores and reaction times in their wearable fatigue monitoring study [3]. In other words, subjective sleepiness and objective performance can diverge markedly [3].

There are several potential causes for this divergence. First, individuals have varying awareness of their own fatigue. Some operators may under-report their sleepiness either due to a belief that they are expected to stay alert (a kind of bias or even fear of admitting fatigue) or due to lack of self-awareness (they don't realize how impaired they are until a lapse happens). Conversely, some might report feeling very tired even if their reaction times haven't yet slowed dramatically – perhaps because they are cautious or have a low tolerance for fatigue feelings. Second, motivation and adrenaline can temporarily mask subjective fatigue. An operator might feel very drowsy (high KSS), but knowing that they are being tested or that they must perform might push them to concentrate harder during the reaction time test, leading to an okay performance despite feeling tired. This would produce a case of high subjective fatigue but relatively decent objective performance. The reverse can happen as well: an operator might report low sleepiness (perhaps due to caffeine or simply optimism about their state), yet still exhibit slowed responses and lapses because physiologically their body is fatigued. Essentially, the subjective state can lag or misrepresent the underlying neurocognitive state.

Another factor is that subjective and objective measures are capturing different aspects of fatigue. The KSS is essentially asking "how sleepy/drowsy do you feel right now?" which is a conscious perception of one's state. Reaction time, on the other hand, measures one aspect of performance (speed of responding). It is possible to feel sleepy but still marshal enough focus to respond quickly in a short test (especially if standing up or if one knows the test is underway). But sustaining that performance over time or in a non-test scenario might be much harder. On the flip side, one could feel alert but still have slow reactions due to accumulated fatigue effects on the brain that haven't registered as sleepiness. The lack of a strong correlation suggests each measure has unique variance – they are not redundant. Rogers et al. noted that each measure "appears to be tracking something different" and recommended further research to understand the lack of alignment [3].

For practical purposes, the divergence between KSS and reaction time means that neither measure alone gives a complete picture. An operator might be in a danger zone for fatigue-related impairment even if they report feeling only mildly tired. Thus, relying solely on self-reports can be misleading. Likewise, an objective test might flag someone as fatigued (slow RT) even if they say they feel fine; in such cases, the test might be revealing underlying fatigue that could soon manifest more obviously. The dual measurement approach (as taken in this study and by Rogers et al., 2023) is therefore valuable. In this study, it is seen that both measures independently indicated greater fatigue at night, which strengthens confidence in identifying those high-risk periods. But on an instance-by-instance basis, discrepancies remind us that triggering alerts or interventions should probably consider both data streams. For example, an operator who reports extreme sleepiness (KSS 9) even if their RT hasn't yet crashed should probably be attended to, and an operator who has very sluggish RT responses even if they claim to feel okay should likewise be considered at risk. The combination of measures helps cover the gaps each one has when used in isolation.

Although the smartwatch-based reaction time test proved effective in identifying fatigue patterns, several limitations should be considered when interpreting the findings. Reaction time captures primarily the cognitive and psychomotor components of fatigue and does not account for its emotional, motivational, or physiological dimensions. For instance, a worker may exhibit normal reaction times while still experiencing mental fatigue, decreased motivation, or microsleeps not reflected in the brief test. Moreover, external factors such as caffeine intake, stress, ambient temperature, lighting, or even operator engagement during testing could influence reaction times independently of true fatigue. These variables, along with demographic factors such as age and experience, were not included as predictors in the random forest model due to data availability constraints, which may have limited the model's ability to fully account for inter-individual variability. Thus, while slower responses are a useful and objective sign of reduced alertness, reaction time should be viewed as an indicator rather than a definitive measure of fatigue. Integrating it with complementary physiological or behavioral data—such as heart rate variability, eye movement patterns, or sleep duration—would yield a more comprehensive understanding of fatigue states in operational environments.

Beyond the conceptual limitations of the reaction time metric, several methodological constraints should also be acknowledged. Demographic and health information, including age, sleep history, or pre-existing medical conditions, were not available due to confidentiality restrictions. These factors may contribute to individual variability in baseline alertness and fatigue susceptibility. Additionally, differences in worksite conditions—such as scheduling practices, equipment type, terrain, or noise levels—could have influenced reaction time results, even though shift and site effects were statistically controlled. The self-reported fatigue ratings may also be subject to bias, as some participants might under- or overestimate their fatigue due to social or motivational factors. Together, these limitations suggest that while smartwatch-based reaction time testing is a promising and practical tool for monitoring operator alertness, it should be applied with caution and ideally complemented by other data sources to improve reliability and interpretability.

6 Next Steps

The 590 ms reaction time threshold presented in this study was derived from data collected for a single operator (Operator ID = 71), who had the most continuous and complete dataset among all participants. This value serves as a case-specific example illustrating how individualized reaction time data can be used to determine operator-specific fatigue thresholds in real time. In practice, each operator would have their own unique threshold value based on their baseline performance and variability. Future work should focus on developing and validating models capable of dynamically estimating these personalized thresholds across multiple operators and operational settings.

The promising findings from this preliminary study open several avenues for future work to enhance the fatigue monitoring system and its integration into mining operations. The following next steps are recommended to build upon and improve the current approach:

- Integrate Sleep and Circadian Data: Future versions of the monitoring system should incorporate data on operators' sleep duration and quality (e.g., from wearable sleep trackers or self-reported sleep logs). Fatigue levels are heavily dependent on prior sleep history and circadian rhythms; adding this context could improve the accuracy of fatigue predictions. For instance, knowing that an operator slept only 4 h in the past 24 h or is on their third consecutive night shift would help interpret their reaction time results. Integrating sleep data is crucial given that chronic sleep deprivation and circadian misalignment are primary contributors to fatigue in mining shift work [2]. By combining performance data with sleep metrics, the system could potentially forecast fatigue levels and issue preemptive alerts before a shift begins.
- Enhance Real-Time Feedback and Alerts: Providing immediate, actionable feedback to operators is critical for an effective fatigue management tool. Many current fatigue detection systems lack direct feedback to the user [3]. In the next iteration, when an operator's reaction

time test indicates significant fatigue (or if a lapse is detected), the smartwatch could instantly alert the operator – for example, through a vibration and a message suggesting a short break, a caffeine intake, or some alertness exercise. Simultaneously, an alert could be sent to a supervisor if a certain threshold is crossed. Real-time feedback not only helps the individual correct course (mitigating fatigue in the moment), but also keeps them engaged in the process, knowing that the system actively responds to their state. The feedback loops should be designed carefully to support the operator (as a buddy system or safety net) rather than punish them, thereby encouraging honest participation.

- Personalize Fatigue Thresholds: As highlighted by prior research, individual differences in fatigue susceptibility are significant [2]. One size does not fit all when determining what reaction time constitutes a dangerous level of fatigue. Therefore, the system should learn each operator's baseline performance and tailor alert thresholds accordingly. Personalization could involve calibrating the reaction time test for each user during a well-rested state, and establishing a custom threshold (e.g., a certain percentage slowdown from their baseline or a certain frequency of lapses). Additionally, personal factors such as age, experience, and even medical conditions (like sleep disorders) could be taken into account if available. The goal is to reduce false alarms and ensure that when the system flags fatigue, it truly reflects a significant deviation for that specific individual. This individualized approach aligns with the recommendation that fatigue risk management be flexible to each person rather than applying a uniform standard [2].
- Compare and Combine with Other Monitoring Systems: To validate and strengthen the wearable reaction time approach, it would be beneficial to conduct comparative studies with other fatigue monitoring solutions. For example, future tests could have operators concurrently monitored by a PERCLOS camera system or a SmartCap EEG-based system while using the smartwatch test. Comparing the alerts and metrics from these systems would shed light on the strengths and weaknesses of each. Perhaps the wearable detects performance impairment slightly earlier, while the EEG cap might catch micro-sleeps more directly – such insights could lead to a hybrid approach. At the very least, benchmarking against established systems will build confidence in the smartwatch method. Additionally, engaging with industry partners to pilot the system in real operations and gather feedback will be important. Given that no single fatigue detection technology fits all scenarios [22], the ultimate solution in practice might involve a combination of tools. The next steps should explore how the reaction-time wearable can complement other measures (such as scheduling software, fitness-for-duty tests, or operator wellness programs) as part of a comprehensive Fatigue Risk Management System (FRMS).
- User Training and Engagement Programs: Simply deploying a fatigue monitoring tool is not enough; its effectiveness will depend on user acceptance and proper usage. Future work should focus on training programs for operators and management. Educating users on the purpose of the wearable test, how to interpret their scores, and how to respond to alerts (e.g., take a break when prompted) will be vital. Gathering user feedback and involving them in the refinement of the system will improve adoption. Moreover, integrating the tool into the organizational safety culture (for example, incorporating fatigue scores into toolbox talks or shift handovers in a non-punitive way) can normalize its use. Since organizational culture and support were noted as important factors in the success of such interventions [3], this aspect should be a key part of future implementation strategies.

Each of these improvements will move the research closer to a deployable solution that can proactively manage fatigue in mining operations. The next steps will also involve larger-scale validation: increasing the sample size of operators, collecting data over longer periods (weeks or months) to capture more fatigue cycles, and testing the system's predictive power for actual safety incidents or performance metrics. By iteratively refining the technology and its integration into the human and organizational context, it is aimed to develop a robust tool that not only detects fatigue but also helps reduce it.

7 Conclusion

This study demonstrated the feasibility and utility of using a wearable device – specifically a smartwatch with a custom reaction time testing app – to monitor fatigue in haul truck operators during their shifts. The key findings show that reaction time is a sensitive indicator of fatigue in the field: operators experienced significantly slower reaction times during the early morning hours (around 2–5 AM), which aligns with the expected circadian trough of alertness. This objective measure of performance degradation correlates with the times when fatigue-related risk is known to be high, underscoring that the smartwatch-based test can successfully capture meaningful fluctuations in alertness. For example, the average reaction time during the peak fatigue window was on the order of 30 ms slower than baseline, a difference that was statistically significant and practically relevant (in real driving tasks, such a delay could translate to many extra feet traveled before braking in response to a hazard). It is also found that the distribution of reaction times widens under fatigue (indicating more frequent lapses), reinforcing the idea that fatigue affects not just the average speed of response but also the reliability of performance.

By incorporating a subjective sleepiness scale (KSS) alongside the reaction time test, the study provided insights into how operators perceive their fatigue versus how it manifests in their performance. The lack of a strong correlation between subjective and objective measures is an important result: it suggests that relying on self-reported alertness alone would be insufficient for managing fatigue risk. An operator might feel "okay" even as their response times lengthen, or conversely feel very tired yet still perform adequately for a short period. The wearable system, by capturing both perspectives, offers a more comprehensive assessment. In practice, this could enable a more nuanced fatigue management strategy – one that can validate or challenge a person's self-assessment and ensure that hidden fatigue does not go unnoticed. These findings echo those of Rogers et al. [3], who noted the complementary nature of subjective and objective data for a fuller understanding of operator fatigue [3].

The value of the wearable approach was evident throughout the study. The smartwatch-based test proved to be non-intrusive and convenient: operators could complete a reaction time trial in under a minute during natural pauses in their work, and the device collected data passively otherwise. Unlike camera systems that require in-cab installation or specialized devices that workers may resist, the smartwatch was a familiar form factor. This suggests that uptake and compliance can be high if the system is introduced and managed well. Moreover, the continuous stream of data opens the door to real-time monitoring in a way that periodic fitness-for-duty tests cannot match. We showed that data can be transmitted and analyzed in near real time, allowing for the possibility of immediate interventions. In sum, this wearable solution represents a cost-effective and flexible alternative to traditional fatigue monitoring methods, aligning with recent recommendations that off-the-shelf technology can be leveraged to improve safety [3]. It directly measures performance (which is what ultimately matters for safety) rather than just proxies of fatigue, and it does so in the operator's actual work environment.

This study advances previous fatigue detection research in several specific and substantive ways. (1) From qualitative insights to validated field metrics. Drews et al. [1] characterized fatigue among mine operators through qualitative analyses of experience and perception. Our work extends theirs by operationalizing those insights into a validated, smartwatch-based reaction-time (RT) measure collected during live haulage operations. In doing so, the study transforms qualitative context into a continuous, objective field metric of alertness. (2) From predictive correlations to in-situ measurement. Talebi et al. [2] modeled environmental and work factors associated with fatigue using operational datasets. We complement and extend that line by directly measuring fatigue in situ through RT testing and quantifying how circadian windows, shift timing, and time-on-task manifest in individual performance. This provides direct empirical validation for factors previously inferred statistically. (3) From feasibility to inferential evidence. Where Rogers et al. [3] demonstrated the feasibility of a wearable approach, the present study contributes inferential rigor. We deployed the method across multiple sites and shifts, applied Box–Cox normalization to address RT skewness, and used mixed-effects and ANOVA/Tukey analyses to control for individual and temporal variability. These enhancements move the research from proof-of-concept to statistically validated application. (4) Objective–subjective integration and divergence. We jointly analyzed objective RT and subjective fatigue ratings (KSS-style), identifying systematic circadian trends and clear divergences between perceived and measured fatigue. This dual-stream evaluation provides novel evidence that single-channel monitoring may under- or over-estimate risk. (5) Toward actionable practice. Finally, we introduce a data-driven RT thresholding framework—including an operator-specific calibration example—to support real-time fatigue flagging. By quantifying individualized thresholds, the study takes a step toward operational implementation within Fatigue Risk Management Systems (FRMS). Collectively, these contributions shift the field from the question "can fatigue be measured in real operations?" to "how reliably, when, and how can it inform safety management?" They provide both methodological validation and practical guidance for integrating wearable-based monitoring into mining operations.

In terms of this work's contribution relative to past efforts: earlier studies like Talebi et al. [2] advanced our understanding of which factors contribute to fatigue by analyzing large operational datasets [2]. Our approach complements that by providing a tool to measure fatigue in real time on an individual level, potentially supplying the kind of data that those models need (for example, our reaction time measurements could be used as inputs or validation for fatigue prediction models). Meanwhile, Rogers et al. [3] established the initial viability of a wearable fatigue monitoring concept; our study adds to their work by applying a similar concept in a practical scenario and examining the results in detail, including distributional aspects and the interplay of subjective/objective measures [3]. Together, these efforts are converging on a new generation of fatigue management strategies for mining: ones that are data-driven, individualized, and embedded in daily operations.

In conclusion, continuous reaction time monitoring via wearable devices shows great promise as a cornerstone of fatigue risk management in mining and other industries. It offers a proactive means to detect early signs of fatigue and thus prevent accidents before they happen. By providing quantifiable, real-time indicators of diminished alertness, it enables timely interventions – whether that is as simple as advising a short break or as significant as stopping a task altogether. The findings from this study contribute to the growing body of evidence that managing fatigue can be significantly improved through technology that is both scientifically grounded and practically acceptable on the work floor. As we move forward, refining this approach and integrating it with other safety systems will be key. Ultimately, the goal is a safer work environment where operators are supported to maintain alertness, and where the perilous effects of fatigue are kept in check through smart monitoring and management – before a microsleep or lapse leads to a serious incident.

It is acknowledged that examining the relationship between heart rate variability (HRV) and reaction time (RT) would provide valuable insight into the physiological correlates of fatigue. HRV is recognized as a sensitive indicator of autonomic nervous system activity and may capture early signs of fatigue-related stress and reduced alertness. In future phases of this research, HRV monitoring could be incorporated—either through the existing smartwatch platform or through complementary wearable sensors—to enable a multi-modal assessment of fatigue. Through such integration, a more comprehensive understanding of the relationship between physiological and cognitive indicators of fatigue could be achieved, thereby strengthening the robustness and practical applicability of the monitoring approach within fatigue risk management systems. The present study takes an important step in that direction, demonstrating how modern wearable technology can be harnessed to protect workers in one of the world's most demanding industries.

Acknowledgements We would like to express our gratitude to the National Institute for Occupational Safety and Health (NIOSH) for funding and the mining companies, their management, and all of the people supporting this research.

Author Contributions For this research article, contributions of the authors are as follows: Conceptualization, E.T., W.P.R. and F.A.D.; methodology, E.T., W.P.R., and F.A.D.; software, E.T.; validation, W.P.R. and F.A.D.; formal analysis, E.T. and F.A.D.; investigation, E.T.; resources, W.P.R.; data curation, E.T. and F.A.D.; writing—original draft preparation, E.T., and W.P.R.; writing—review and editing, E.T. and W.P.R.; visualization, E.T.; supervision, W.P.R.; project administration, W.P.R.; funding acquisition, W.P.R. All authors have read and agreed to the published version of the manuscript.

Funding Funding for testing the application was partially provided by the National Institute for Occupational Safety and Health (NIOSH) with grant number 75D30119C05500.

Data Availability The data presented in this study are not publicly available due to confidentiality.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there are no conflicts of interest.

References

1. Drews FA, Rogers WP, Talebi E, Lee S (2020) The experience and management of fatigue: a study of mine haulage operators. Min Metall Explor 37:1837–1846. https://doi.org/10.1007/s42461-020-00259-w
2. Talebi E, Rogers WP, Drews FA (2022) Environmental and work factors that drive fatigue of individual haul truck drivers. Mining 2(3):542–565. https://doi.org/10.3390/mining2030029
3. Rogers WP, Marques J, Talebi E, Drews FA (2023) IoT-enabled wearable fatigue-tracking system for mine operators. Minerals 13(2):287. https://doi.org/10.3390/min13020287
4. Uehli K, Mehta AJ, Miedinger D, Hug K, Schindler C, Holsboer-Trachsler E, Leuppi JD, Künzli N (2014) Sleep problems and work injuries: a systematic review and meta-analysis. Sleep Med Rev 18(1):61–73. https://doi.org/10.1016/j.smrv.2013.01.004
5. Dawson D, Reid K (1997) Fatigue, alcohol and performance impairment. Nature 388(6639):235. https://doi.org/10.1038/40775
6. Williamson AM, Feyer A-M (2000) Moderate sleep deprivation produces impairments in cognitive and motor performance equivalent to legally prescribed levels of alcohol intoxication. Occup Environ Med 57(10):649–655. https://doi.org/10.1136/oem.57.10.649
7. Tefft BC (2012) Prevalence of motor vehicle crashes involving drowsy drivers, united States, 2009–2013. AAA Foundation for Traffic Safety, Washington, DC
8. Owens JM, Dingus TA, Guo F, Fang Y, Perez MA, McClafferty J, Tefft BC (2018) The prevalence of drowsy driving crashes: estimates from a large-scale naturalistic driving study. AAA Foundation for Traffic Safety, Washington, DC
9. AAA Foundation for Traffic Safety (2024) Drowsy driving in fatal crashes, United States, 2017–2021 (Research Brief). AAA Foundation for Traffic Safety, Washington, DC. https://aaafoundation.org/drowsy-driving-in-fatal-crashes-united-states-2017-2021/. Accessed 8 Oct 2025
10. Bauerle T, Dugdale Z, Poplin G (2018) Mineworker fatigue: a review of what we know and future decisions. Min Eng 70(3):33–40
11. Kosinski RJ (2013) A literature review on reaction time. Clemson University. Retrieved from https://www.clemson.edu/psych/research/cognitive/reactiontime.html
12. Salthouse TA (2000) Aging and measures of processing speed. Biol Psychol 54(1–3):35–54. https://doi.org/10.1037/a0019096
13. Der G, Deary IJ (2006) Age and sex differences in reaction time in adulthood: results from the UK health and lifestyle survey. Psychol Aging 21(1):62–73. https://doi.org/10.1037/0882-7974.21.1.62
14. Salthouse TA (1996) The processing-speed theory of adult age differences in cognition. Psychol Rev 103(3):403–428. https://doi.org/10.1037/0033-295X.103.3.403

15. D'Esposito M, Detre JA, Aguirre GK, Stallcup M, Alsop DC, Tippet LJ, Farah MJ (1995) A functional MRI study of mental image generation. Neuropsychologia 33(12):1563–1575. https://doi.org/10.1016/S0028-3932(96)00121-2
16. Dinges DF (1995) An overview of sleepiness and accidents. J Sleep Res 4(S2):4–14. https://doi.org/10.1111/j.1365-2869.1995.tb00220.x
17. Brossoit RM, Crain TL, Hammer LB, Bodner T, Olson R (2022) Alert at work? Perceptions of alertness testing and recommendations for practitioners. Occup Health Sci 6(1):1–24. https://doi.org/10.1007/s41542-022-00139-3
18. Guastaferro WP, Avena JS, Gorman CM (2018) Validity and reliability of a brief, computer-administered neurocognitive test battery for use in workplace fatigue research. J Occup Environ Med 60(6):560–568
19. Balkin TJ, Rupp TL, Picchioni D, Wesensten NJ (2019) Alertness management in safety-critical operations: enhancing performance and sleep health. J Occup Environ Med 61(Suppl 1):S49–S59
20. Barnes CM, Drake CL (2015) Prioritizing sleep health: public health policy recommendations. Perspect Psychol Sci 10(6):733–737. https://doi.org/10.1177/1745691615598509
21. National Safety Council (2021) Wearables for fatigue monitoring. https://www.nsc.org/getmedia/a4f3a4ac-0e45-4b1f-a8ce-bccb13e4f7d2/report-wearables.pdf. Accessed 30 Jun 2025
22. National Safety Council (2022) Impairment detection technology and workplace safety: report. National Safety Council
23. Åkerstedt T, Gillberg M (1990) Subjective and objective sleepiness in the active individual. Int J Neurosci 52(1–2):29–37. https://doi.org/10.3109/00207459008994241
24. Whelan R (2008) Effective analysis of reaction time data. Psychol Rec 58(3):475–482. https://doi.org/10.1007/BF03395630
25. Federal Aviation Administration (2010) Advisory Circular 120–100: basics of aviation fatigue. U.S. department of transportation. https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC%20120-100.pdf. Accessed 30 Jun 2025