Modeling Operational Fatigue in Emergency Medical Services Providers Under Real-World Conditions Using Wearable Sensors and Machine Learning

Yuxin Cai[1], Aditya Shanghavi[1], Ranjana Mehta[1]
[1]University of Wisconsin-Madison

Synopsis

This study applied a wearable sensor-based machine learning approach to model operational fatigue among EMS providers under real-world conditions. Using multimodal physiological, behavioral, and contextual data collected across five days, Logistic Regression achieved the best performance (AUROC = 0.87). Robust predictors shared across models, including time of day, cumulative work exposure, dispatched calls, and pulse rate dynamics, indicated the combined influence of circadian rhythm, workload accumulation, and physiological changes in fatigue. Findings support the feasibility of scalable wearable-based fatigue monitoring to enhance fatigue risk visibility in EMS operations.

Background

Fatigue is a well-recognized challenge among emergency medical services (EMS) providers and has been linked to degraded cognitive and operational performance, which may increase the risk of workplace injuries, medical errors, and other safety-compromising behaviors (Patterson et al., 2012). Monitoring fatigue among EMS professionals is essential as a critical component of fatigue risk management. However, most EMS studies still rely primarily on self-reported fatigue measures due to the difficulty of collecting continuous sensor data in dynamic field operations (Patterson et al., 2025). Recent advances in smartwatch-based wearable sensing, combined with machine learning (ML) algorithm, enable continuous monitoring of multimodal physiological and contextual data and have shown strong promise for fatigue modeling across safety-critical domains (Kakhi et al., 2025). Furthermore, identifying robust fatigue-related predictors through a data-driven ML approach can provide practical insights into the progression of ongoing fatigue. Previous studies in other safety-critical domains, such as aviation, construction, and healthcare, have demonstrated the feasibility of modeling fatigue using wearable-derived physiological and contextual data (Liu et al., 2023; Park et al., 2025). Despite these advances, several methodological limitations remain in existing literature. First, few studies systematically compare fatigue prediction performance across multiple ML architectures, making it difficult to determine whether identified predictors are robust or model specific. Second, many prior approaches rely on pre-post task comparisons or short pre-assessment windows (e.g., 10 minutes), which may not adequately capture the temporally evolving nature of fatigue. Fatigue in EMS is inherently cumulative, circadian-modulated, and often delayed in manifestation, requiring modeling approaches that account for extended temporal context (Karim et al., 2024). To address these gaps, this study presents a wearable sensor-based machine learning framework to model operational fatigue among EMS providers under real-world conditions. Specifically, we aim to (1) develop a scalable framework for fatigue modeling in real-world EMS operations and (2) identify robust fatigue-related predictors through a multi-model comparison approach.

Methods

We conducted a five-day longitudinal observational study with 11 active EMS professionals working a 3-on/2-off rotational schedule under real-world conditions. **Measurements and predictors.** Continuous physiological and behavioral data were collected using a wrist-worn wearable device (Embrace Plus, Empatica), while perceived fatigue (Samn-Perelli Fatigue Scale; Samn & Perelli, 1982), naps, and caffeine intake were self-reported every six hours. Predictor variables included multimodal physiological signals (pulse rate, pulse rate variability, respiratory rate, skin temperature, electrodermal activity, and metabolic equivalent of task), wearable-derived behavioral indicators (sedentary time and accelerometer magnitude), and

contextual operational features (e.g., work exposure time, time of day, dispatch calls, naps, and caffeine intake). For each fatigue report, the wearable-derived features were aggregated from the preceding six-hour period using multiple retrospective temporal windows to capture both signal magnitude and temporal dynamics. **ML modelling pipeline.** The fatigue prediction was formulated as a binary classification task, and four ML models representing linear and nonlinear approaches (Logistic Regression, Random Forest, XGBoost, and CatBoost) were applied. Prior to modeling, feature dimensionality was reduced by removing highly correlated variables (ρ>=0.8) followed by Minimum Redundancy Maximum Relevance feature selection (Peng et al., 2005) to retain the top 50 predictors. Model tuning and evaluation were performed using nested 5- fold cross-validation, with performance assessed using accuracy, precision, recall, F1-score, and AUROC metrics. To enhance interpretability, permutation importance was used to identify the top 10 fatigue-related predictors across models. (Breiman, 2001)

Results

Model Performance: As shown in Table 1, Logistic Regression achieved the highest predictive performance (AUROC = 0.87 ± 0.04) with well-balanced precision and recall.

Table 1. Performance of ML models for fatigue classification ($\bar{x}$ ± s) 5 fold cross-validation

Model	Accuracy ($\bar{x}$ ± s)	Precision ($\bar{x}$ ± s)	Recall ($\bar{x}$ ± s)	F1 ($\bar{x}$ ± s)	AUROC ($\bar{x}$ ± s)
Logistic Regression	**0.79 ± 0.05**	**0.79 ± 0.0568**	**0.79 ± 0.05**	**0.78 ± 0.06**	**0.87 ± 0.04**
Random Forest	0.75 ± 0.03	0.76 ± 0.05	0.75 ± 0.04	0.7175 ± 0.03	0.78 ± 0.06
CatBoost	0.73 ± 0.05	0.73 ± 0.07	0.73 ± 0.05	0.72 ± 0.05	0.79 ± 0.05
XGBoost	0.71 ± 0.03	0.75 ± 0.08	0.71 ± 0.03	0.64 ± 0.04	0.74 ± 0.08

Feature Importance: As seen in Table 2, feature importance analysis revealed a consistent set of robust fatigue-related predictors across models.

Table 2. Top 10 features identified by feature importance analysis in each model. * = Features common across two or more models.

Logistic Regression	Random Forest	CatBoost	XGBoost
Time of day*	Cumulative work exposure time*	Time of day*	Time of day*
Slope of pulse rate (5th hour)*	Time of day*	Slope of pulse rate (5th hour)*	Cumulative work exposure time*
Number of dispatched calls*	Number of dispatched calls*	Maximum respiratory rate (5th hour)	Sedentary time (1st hour)*
Mean skin temperature (3rd hour)	Slope of pulse rate (5th hour)*	Cumulative work exposure time*	Slope of pulse rate (5th hour)*
Cumulative work exposure time*	Standard deviation of skin temperature (5th hour)	Number of dispatched calls*	Number of dispatched calls*
Slope of pulse rate (past 1 hour)*	Minimum accelerometer variability (past 6 hours)	Slope of pulse rate (past 1 hour)*	Mean pulse rate variability (4th hour)
Maximum heart rate (1st hour)*	Slope of respiratory rate (4th hour)	Slope of respiratory rate (4th hour)	Minimum skin temperature (1st hour)*
Minimum accelerometer variability (past 2 hours)	Maximum respiratory rate (5th hour)	Standard deviation of skin temperature (5th hour)	Maximum respiratory rate (5th hour)
Minimum skin temperature (1st hour)*	Minimum skin temperature (1st hour)*	Relative caffeine intake compared to usual	Mean skin temperature (3rd hour)
Caffeine intake	Maximum heart rate (1st hour)*	Minimum accelerometer variability (past 6 hours)	Mean skin temperature (past 1 hour)

Discussion

This study demonstrates the feasibility of using wearable and contextual data to predict operational fatigue in EMS under real-world conditions, with Logistic Regression achieving the

best performance (AUROC = 0.87), indicating reliable discrimination between high- and low-fatigue states. The relatively high recall suggests that the model can detect a substantial proportion of fatigue cases, which is particularly important in safety-critical operational environments such as EMS. The low AUROC variability (SD) indicated stable generalization. Robust predictors shared across models, including time of day, cumulative work exposure, dispatched calls, sedentary time, and slope-based pulse rate features, indicating the roles of circadian rhythm, workload accumulation, and dynamic physiological responses in fatigue development. These findings support the potential of ML approaches for fatigue risk monitoring using wearable sensors in EMS. Future work with larger longitudinal datasets is planned to improve generalizability and enable richer model interpretability analyses.

References

Breiman, L. (2001). Random forests. Mach. Learn. 45, 5-32.

Kakhi, K., Jagatheesaperumal, S. K., Khosravi, A., Alizadehsani, R., and Acharya, U. R. (2025). Fatigue monitoring using wearables and AI: trends, challenges, and future opportunities. Comput. Biol. Med. 195:110461. doi: 10.1016/j.compbiomed.2025.110461

Karim, E., Pavel, H. R., Nikanfar, S., Hebri, A., Roy, A., Nambiappan, H. R., Jaiswal, A., Wylie, G. R., and Makedon, F. (2024). Examining the landscape of cognitive fatigue detection: a comprehensive survey. Technologies 12:38. doi: 10.3390/technologies12030038

Lawn, S., Roberts, L., Willis, E., Couzner, L., Mohammadi, L., and Goble, E. (2020). The effects of emergency medical service work on the psychological, physical, and social well-being of ambulance personnel: a systematic review of qualitative research. BMC Psychiatry 20:348. doi: 10.1186/s12888-020-02752-4

Liu, S. S.-H., Ma, C.-J., Chou, F.-Y., Cheng, M. Y.-C., Wang, C.-H., Tsai, C.-L., Duh, W.-J., Huang, C.-H., Lai, F., and Lu, T.-C. (2023). Applying a smartwatch to predict work-related fatigue for emergency healthcare professionals: machine learning method. West. J. Emerg. Med. 24, 693. doi: 10.5811/westjem.58139

Park, S., Cheon, B., Kim, H., and Kim, H. (2025). Investigated predictors of construction workers' ongoing fatigue: random forest approach. J. Manage. Eng. 41:04025021. doi: 10.1061/JMENEA.MEENG-6525

Patterson, P. D., Brailsford, J., Fishe, J., Roach, D. G. L., Sheffield, M. A., Lukacz, R. A., Thomas, R. N., Appio, M. C., Martin, S. E., and Martin-Gill, C. (2025). Barriers to implementing evidence-based guidance for fatigue risk mitigation in the prehospital setting. Prehosp. Emerg. Care 1-12. doi: 10.1080/10903127.2025.2527365

Patterson, P. D., Weaver, M. D., Frank, R. C., Warner, C. W., Martin-Gill, C., Guyette, F. X., Fairbanks, R. J., Hubble, M. W., Songer, T. J., Callaway, C. W., Kelsey, S. F., and Hostler, D. (2012). Association between poor sleep, fatigue, and safety outcomes in emergency medical services providers. Prehosp. Emerg. Care 16, 86-97. doi: 10.3109/10903127.2011.616261

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell. 27, 1226-1238. doi: 10.1109/TPAMI.2005.159

Samn, S. W., and Perelli, L. P. (1982). Estimating aircrew fatigue: a technique with application to airlift operations.