Cognitive Computing in Sociotechnical Systems, Vol. 205, 2026, 12–20
https://doi.org/10.54941/ahfe1007357

Early Prediction of Physiological Strain Using Multivariate Time-Series Data

Yusuf Quadri[1], Julia Tschuden[1], Florian Haid[1], Anna Weber[1], Markus Bergen[1], Lucas Paletta[1], Michael Schneeberger[1], Gerald Bauer[2], and Thomas Hoelzl[3]

[1]JOANNEUM RESEARCH Forschungsgesellschaft mbH, Graz, Austria
[2]Austrian Federal Ministry of Defense, Science, Research and Development Division, Austria
[3]Austrian Armed Forces, Training Coordination Division, Austria

ABSTRACT

The increasing availability of wearable physiological sensors has enabled continuous monitoring of human performance in safety-critical and high-demand environments. Predicting physiological strain in advance provides valuable support for proactive decision-making and risk mitigation in sociotechnical systems. This study investigates short-term forecasting of physiological strain using multivariate wearable time-series data. The objective is to predict a bounded strain index several minutes in advance using heart rate (HR), respiratory rate (RR), and core body temperature (CBT) measurements. The dataset comprises continuous physiological recordings from 86 participants with approximately 734,000 time points, standardized to a sampling rate of 1 Hz. This work builds upon data collected in the RTVitalMonitor project, which develops predictive models for monitoring psychophysiological strain and performance in military task simulations. A multi-horizon forecasting framework predicts strain levels at 5, 10, and 15 minute intervals using sequential deep learning models. The methodological approach extends prior work conducted on a public graded exercise testing dataset, where exhaustion prediction was formulated as a binary classification task. In contrast, the present study reframes the problem as a regression-based forecasting task using a Gated Recurrent Unit (GRU) network, enabling continuous early warning of physiological strain. Models are evaluated using mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2). Results demonstrate that the GRU model achieves robust predictive performance across all forecast horizons and remains stable under parameter-reduction scenarios. The findings highlight the potential of cognitive computing approaches for privacy-aware, real-time physiological monitoring in sociotechnical systems. The proposed framework contributes to the development of predictive human performance monitoring solutions suitable for wearable and edge-based deployment.

Keywords: Physiological strain prediction, Multivariate time-series analysis, Wearable Vital-Sign-Sensors, Early warning systems, Human performance monitoring

INTRODUCTION

Advances in wearable sensing technologies have enabled continuous monitoring of physiological states in real-world environments. These

Received April 10, 2026; Revised May 17, 2026; Accepted June 4, 2026; Available online July 20, 2026

developments are particularly relevant for sociotechnical systems in which human performance and safety play a critical role. Continuous physiological monitoring allows for early identification of strain, fatigue, or performance decline, thereby supporting informed decision-making and proactive intervention strategies.

Traditional approaches to human performance assessment primarily focus on retrospective analysis or real-time detection of fatigue states. However, these reactive approaches provide limited opportunity for mitigation before performance degradation occurs. Predictive modeling of physiological strain offers a proactive alternative, enabling early warnings and supporting adaptive system responses.

Recent research within human-centered monitoring systems emphasizes the integration of wearable biosensors, machine learning, and cognitive computing frameworks to enhance decision support in complex operational environments (Weber et al., 2023). For example, semantic decision support systems have demonstrated the value of integrating physiological monitoring with risk stratification and real-time visualization tools (Haid et al., 2024).

This study contributes to this research direction by investigating shortterm forecasting of physiological strain using multivariate wearable sensor data. It frames physiological strain prediction as a multi-horizon regression problem with minimal wearable inputs, enabling continuous forecasting of strain evolution rather than discrete classification of stress states. This supports earlier and more actionable decision-making by estimating how strain is expected to evolve over the next 5, 10, and 15 minutes, rather than only indicating whether a fixed threshold has been crossed.

RELATED WORK AND METHODOLOGICAL EVOLUTION

Wearable Physiological Monitoring in Sociotechnical Systems

Wearable biosensors have become a cornerstone of modern human performance monitoring systems (Weber et al., 2023). Projects such as VitalMonitor have demonstrated the feasibility of real-time psychophysiological monitoring through heart rate, respiratory rate, and core body temperature measurements (Haid et al., 2024). Such systems support decision-making by providing continuous insight into physiological states while ensuring user safety and operational effectiveness.

Similarly, semantic decision support frameworks integrate physiological strain indicators with cognitive-emotional stress and situational awareness to enhance decision-making in safety-critical environments (Haid et al., 2024). Such systems highlight the importance of predictive analytics for proactive risk management.

Prior Seminar Work: Exhaustion Prediction as a Classification Problem

Previous work by the authors explored exhaustion prediction as a binary classification task using physiological signals from controlled exercise protocols based on a public graded exercise testing dataset (Mongin et al., 2021). The dataset included breath-by-breath physiological measurements such as heart rate, oxygen consumption (VO_2), carbon dioxide production (VCO_2), ventilation (VE), and respiratory rate. While effective for identifying imminent exhaustion events, this formulation did not capture the continuous evolution of physiological strain.

Classical machine learning models and deep sequential architectures, including Logistic Regression, Random Forests, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Temporal Convolutional Networks (TCN), were evaluated using accuracy, precision, recall, F1-score, and ROC-AUC (Hochreiter and Schmidhuber, 1997; Cho et al., 2014; Chung et al., 2014). Overall performance was strong across horizons, with TCN among the top-performing models. However, the classification setup still faced key limitations, especially class imbalance and reduced interpretability of binary exhaustion labels.

Importantly, the seminar work did not include the Physiological Strain Index (PSI) or regression-based forecasting. Instead, it focused on predicting imminent exhaustion as a discrete event, supported by an explicit exhaustion set point in the source dataset. In the current RT-VitalMonitor dataset, no equivalent exhaustion set point is available, making the same classification problem structure harder to define and less reliable. The current work therefore extends this direction by modeling strain as a continuous variable and predicting its trajectory over multiple future horizons.

From Classification to Forecasting of Physiological Strain

Classification-based formulations were also tested on the RT-VitalMonitor data; however, strong class imbalance led models to favor the majority non-strained class. For this binary setup, labels were defined manually using a future-threshold rule: a sample was marked as strained ("yes") for a given horizon (e.g., +5 min) when PSI crossed a predefined threshold for a minimum duration (in seconds) within that future interval. In this setting, overall accuracy remained deceptively high, but minority-class performance degraded, with low recall (sensitivity) for strained segments and unstable precision, resulting in weak F1-scores for the strained class. This behavior indicates a high false-negative rate for physiologically strained states and limits the operational usefulness of a binary classifier for early warning. Continuous strain forecasting therefore provides a more informative representation of physiological state without relying on a single discrete threshold.

The present study therefore adopts a GRU-based regression forecasting framework (Chung et al., 2014) to model continuous strain dynamics and support proactive decision-making in sociotechnical systems.

DATA AND EXPERIMENTAL SETUP

Dataset Overview

The dataset used in this study is a curated subset of the RT-VitalMonitor dataset, a large-scale physiological monitoring study conducted with soldiers of the Austrian Armed Forces (Haid et al., 2025). While the full study involved more than 170 soldiers, this work focuses on 86 soldiers selected for modeling and evaluation.

The RT-VitalMonitor project was designed to analyze physiological and cognitive strain during military task simulations, in which participants performed standardized infantry-specific activities under controlled yet realistic conditions (Haid et al., 2025). The experimental protocol included:

- loaded marching (8 km with approximately 43 kg load),
- overcoming an obstacle (1.9 m wall climb),
- fire-and-movement drills (15 x 6,6 m sprints into different firing positions),
- casualty evacuation (dragging a 123 kg dummy over 15 m), and
- repetitive load-handling tasks (lifting and carrying sandbags 20 x 30 m).

These tasks were designed by military subject-matter experts to reflect real-world operational demands, enabling the collection of ecologically valid physiological data (Haid et al., 2025). Physiological signals were recorded using a multi-sensor wearable setup, including heart rate (HR), respiratory rate (RR), and core body temperature (CBT), alongside motion and environmental measurements such as temperature and humidity (Haid et al., 2025).

For the present study, the data were standardized to a 1 Hz sampling rate (downsampled from 10 Hz), resulting in approximately 734,000 time points. The target variable is a normalized Physiological Strain Index (0 – 10 scale) (Moran et al., 1998), and a user-wise 80–20 train–test split was applied.

Input–Output Design

Figure 1 summarizes the multi-horizon forecasting setup, including the 5-minute input window, model inference stage, and three forecast outputs (+5, +10, and +15 minutes). Each prediction is generated from synchronized physiological time-series features and evaluated at increasing forecast distance. Two sensor configurations are considered: a full setting with HR, RR, and CBT, and a reduced setting with HR and CBT only.

METHODOLOGY

Preprocessing and Windowing

The data preprocessing pipeline includes signal cleaning, temporal alignment, normalization, and downsampling to 1 Hz. Sliding windows are constructed to capture temporal dependencies and improve model generalization.

Modeling Approach

The following models were evaluated:

- Persistence baseline.
- Random Forest (Breiman, 2001).
- Long Short-Term Memory (LSTM) (Hochreiter and Schmidhuber, 1997).
- Gated Recurrent Unit (GRU) (Chung et al., 2014).

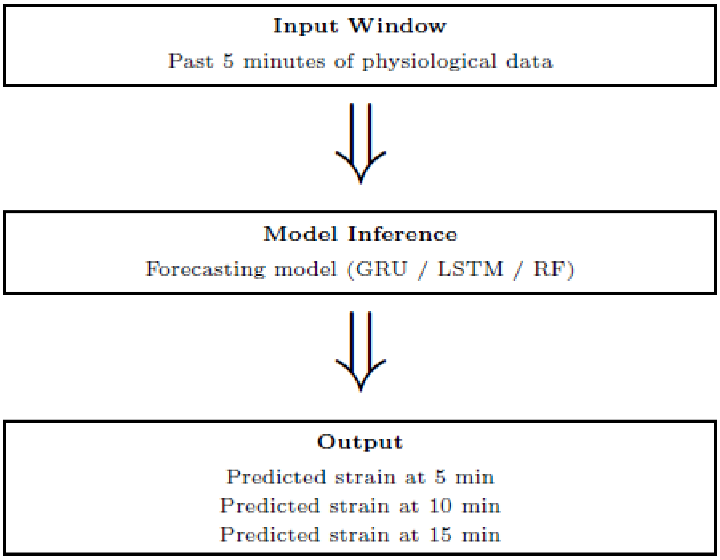

Figure 1: Input–output design for multi-horizon physiological strain forecasting.

The GRU model was selected as the final model because it balances predictive performance and real-time suitability in wearable systems, while capturing temporal dependencies more efficiently and with lower computational complexity than LSTM models.

Evaluation Metrics

Model performance is reported using standard regression metrics:

Mean Absolute Error (MAE)
Root Mean Squared Error (RMSE)
Coefficient of Determination (R^2)

RESULTS AND ANALYSIS

Forecasting Performance

Table 1 shows a clear ranking pattern across horizons: GRU performs best, with RF as the closest alternative. LSTM consistently improves on the persistence baseline but remains behind GRU. Across all models, forecast quality decreases as horizon length increases, which is expected in multi-step prediction.

Parameter Contribution Analysis

Ablation results in Table 2 provide additional insight into feature contribution. Because PSI is physiologically derived from heart rate and core body temperature (Moran et al., 1998), these two channels already encode most target-relevant information, and RR can become partly redundant. For GRU, removing RR changes MAE only marginally across horizons (approximately $\pm$ 0.004), confirming stable performance with HR+CBT alone. For LSTM, MAE consistently improves when RR is removed (largest gain at +15 min), suggesting that RR may introduce temporal variability that is less aligned with PSI dynamics and can act as noise in sequence learning. In contrast, RF shows performance degradation without RR, indicating that tree-based splits still extract useful complementary patterns from RR despite its weaker direct relevance to the PSI definition. Overall, the ablation indicates that HR and CBT form a strong minimal feature set for robust forecasting, while the benefit of RR is model-dependent and likely tied to how each architecture handles correlated and noisy inputs.

Table 1: Forecasting performance across models and prediction horizon.

Model	Horizon (min)	MAE ↓	RMSE ↓	R^2 ↓
Baseline	5	0.479	0.659	0.912
Baseline	10	0.672	0.902	0.827
Baseline	15	0.872	1.162	0.701
RF	5	0.425	0.579	0.932
RF	10	0.558	0.736	0.885
RF	15	0.671	0.863	0.835
GRU	5	0.419	0.562	0.936
GRU	10	0.533	0.696	0.897
GRU	15	0.647	0.837	0.845
LSTM	5	0.446	0.597	0.928
LSTM	10	0.578	0.743	0.882
LSTM	15	0.710	0.906	0.818

Table 2: Impact of removing respiratory rate (RR) on MAE across models and horizons.

Model	Horizon	With RR	No RR	Δ
GRU	+5 min	0.414	0.413	-0.001
GRU	+10 min	0.530	0.534	+0.004
GRU	+15 min	0.642	0.643	+0.001
LSTM	+5 min	0.438	0.412	-0.026
LSTM	+10 min	0.573	0.531	-0.042
LSTM	+15 min	0.699	0.645	-0.054
RF	+5 min	0.442	0.452	+0.010
RF	+10 min	0.569	0.593	+0.024
RF	+15 min	0.689	0.726	+0.037

Visualization Insights

Figure 2 presents residual distributions for GRU forecasts at +5, +10, and +15 minutes. The residuals are centered close to zero across horizons, indicating low systematic bias in predictions. As the forecast horizon increases, the distributions become slightly wider, reflecting the expected growth in uncertainty for longer-term predictions. Overall, the plots support the quantitative results by showing stable short-horizon behavior and only moderate error spread at longer horizons.

CONCLUSION AND OUTLOOK

This study demonstrates that short-term forecasting of physiological strain is feasible at +5, +10, and +15 minutes using wearable HR, RR, and CBT signals. Across all horizons, GRU shows the most consistent performance, with RF as a competitive baseline and LSTM in between. As expected for multi-step forecasting, error increases with forecast distance, but the degradation remains gradual and operationally useful for short-horizon early warning.

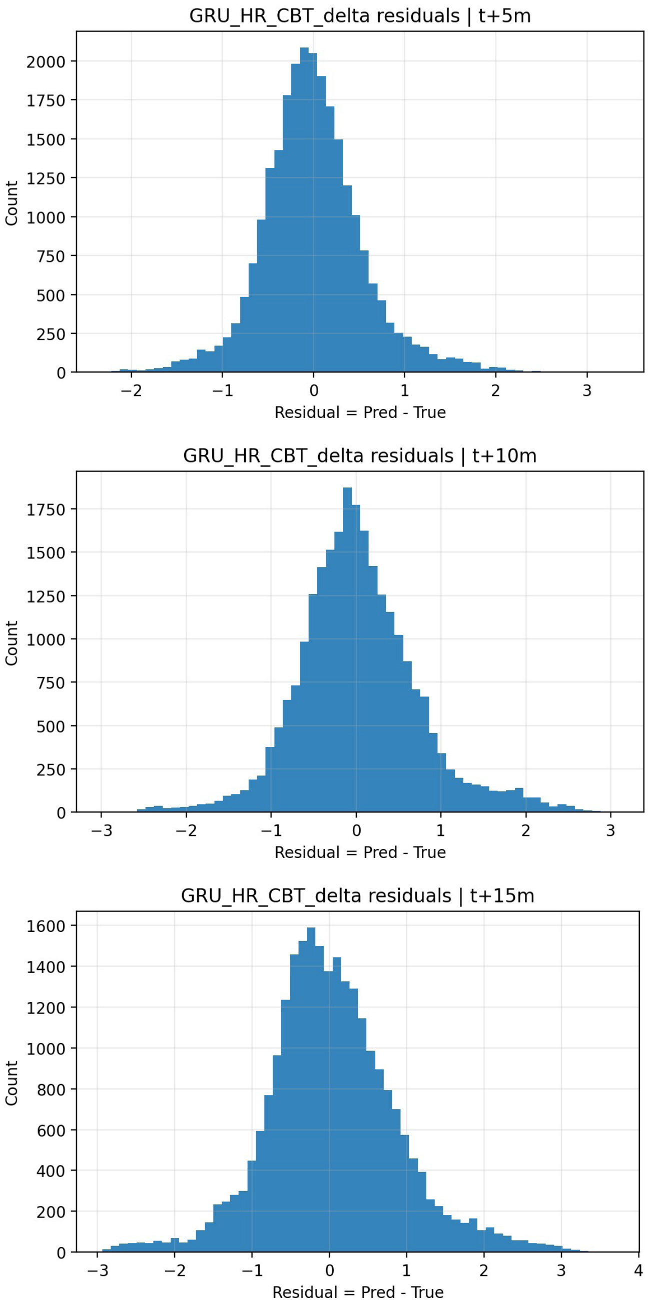

Figure 2: Residual distributions for GRU predictions at 5, 10, and 15 minute horizons.

The ablation results further clarify feature contribution. HR and CBT form a strong minimal feature set for PSI forecasting, while RR has modeldependent value: it is largely neutral for GRU, beneficial for RF, and can introduce noise for LSTM. This pattern indicates that compact sensor configurations can retain strong predictive quality when paired with suitable sequence models, which is important for robust wearable deployment under real-world constraints.

Overall, the proposed framework supports earlier risk awareness and more adaptive decision support in high-demand, safety-critical environments. Early prediction of physiological strain enables proactive intervention, allowing operators to adjust workload, hydration, or environmental exposure before critical thresholds are reached. Future work will focus on sequence-to-sequence forecasting, multimodal fusion, transfer learning across cohorts, and external validation across broader operational scenarios.

ACKNOWLEDGMENT

This work was conducted within the RT-VitalMonitor project and was funded and supported in part by the Austrian Defense Research Programme FORTE, the Austrian Research Promotion Agency (FFG), the Austrian Federal Ministry of Finance, the Austrian Federal Ministry of Defense, and the Austrian Armed Forces. The authors gratefully acknowledge the project partners involved in data collection, system development, and evaluation.

REFERENCES

Breiman, L. (2001). Random forests. *Machine Learning,* 45(1), 5–32.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of EMNLP.*

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555.*

Haid, F., Schneeberger, M., Carballo-Leyenda, B., Rodr´ıguez-Marroyo, J. A., Ladst ätter, S., Weber, A., Almer, A., Mosbacher, J. A., & Paletta, L. (2024). Semantic decision support for action forces with risk stratification from estimated physiological strain, cognitive-emotional stress and situation awareness. In *Cognitive Computing and Internet of Things* (pp. 21–32).

Haid, F., Tschuden, J., Schneeberger, M., Paletta, L., Weber, A., Almer, A., Bergen, M., Rausch, W., Bauer, G., & H¨olzl, T. (2025). Real-time monitoring in military task simulations: Insights from the RT-VitalMonitor Project. In L. Paletta (Ed.), *Cognitive Computing and Internet of Things*. AHFE (2025) International Conference, AHFE Open Access, 165. AHFE International, USA. https://doi.org/10.54941/ahfe1006096

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

Mongin, D., García Romero, J., & Alvero Cruz, J. R. (2021). Treadmill maximal exercise tests from the Exercise Physiology and Human Performance Lab of the University of Malaga (version 1.0.1). *PhysioNet.* RRID:SCR 007345. https://doi.org/10.13026/7ezk-j442

Moran, D. S., Shitzer, A., & Pandolf, K. B. (1998). A physiological strain index to evaluate heat stress. *American Journal of Physiology*, 275(1), R129–R134.

Weber, A., Almer, A., Haid, F., Paletta, L., Schneeberger, *M., Klöckl,* P., Eder, D., Wallner, D., Glanz, P., & Kuehr, O. (2023). Stress modelling based on specific stress characteristics of various military and civil forces for a real-time stress monitoring. *UOET Proceedings*.