IRTAD Conference 2026
Better Road Safety Data for Better Safety Performance
15-17 April 2026 Athens, Greece
International Transport Forum
irtad International Traffic Safety Data and Analysis Group
National Technical University of Athens

# Temporal Modeling of Heart Rate Variability for Driver Drowsiness Detection with LSTM Networks

**Aristotelis Styanidis[1,2], André Lourenço[2,3,4], Carlos Carreiras[2], Christer Ahlström[5,6], Apostolos Ziakopoulos[1], George Yannis[1]**

*[1]Department of Transportation Planning and Engineering, NTUA, Athens, Greece*
*[2]CardioID Technologies, Lisbon, Portugal*
*[3]Instituto Superior de Engenharia de Lisboa, Lisbon, Portugal*
*[4]NOVA Laboratory for Computer Science and Informatics, Lisbon, Portugal*
*[5]Department of Biomedical Engineering, Linköping University, Linköping, Sweden*
*[6]Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden*

## Abstract

Drowsiness detection plays a crucial role in enhancing safety and preventing accidents in domains such as transportation and industrial operations. This work presents a data-driven approach for detecting driver drowsiness using electrocardiogram (ECG) signals from 20 sleep-deprived participants driving on a motorway in real-world driving conditions. Heart rate variability (HRV) features were extracted from ECG signals and used as sequential input to Long Short-Term Memory (LSTM) networks, which are well-suited for modeling temporal physiological patterns associated with drowsiness. The model was trained and evaluated using 10-fold cross-validation and leave-one-subject-out (LOSO) cross-validation strategies to ensure generalization across individuals. Experimental results demonstrate the model’s ability to distinguish between alert and drowsy states (10-fold accuracy = 87.8%, LOSO accuracy = 61.0%), highlighting the potential of HRV-based deep learning approaches for robust, non-intrusive drowsiness detection. This work underscores the importance of physiological signal analysis and deep learning techniques in developing reliable real-time driver monitoring systems.

**Keywords**: Driver Monitoring, Drowsiness, Physiological Signals, ECG, LSTM

## Introduction

Driver drowsiness is a major contributing factor to road crashes worldwide [1], posing a serious threat to road safety. Detecting and mitigating drowsiness in real time can significantly reduce the likelihood of fatigue-related crashes and improve overall road safety. Drowsiness monitoring approaches can be broadly classified into three categories: vehicle-based, which rely on lane deviation or steering behavior; behavioral-based, which assess facial expressions or eye movements; and physiology-based, which analyze internal body signals such as heart rate, brain activity, or ocular responses [2]. Among these, physiological measures provide a more direct and objective representation of the driver’s state, complementing external behavioral indicators.

Physiology-based monitoring has received increasing attention in recent years. However, several research challenges remain open. Signal acquisition and processing often require careful calibration and noise handling, while inter-individual variability complicates model generalization [3]. Nevertheless, the main advantage of physiological measures lies in their ability to detect subtle autonomic changes—such as variations in sympathetic and parasympathetic nervous system activity—before any visible behavioral signs or vehicle performance deviations occur. This early detection capability makes physiological signals particularly valuable for proactive drowsiness assessment and prevention systems.

This study contributes to this research direction by exploring a data-driven approach for drowsiness detection based on heart rate variability (HRV) features derived from electrocardiogram (ECG) signals. The goal is to establish a solid baseline and methodological foundation for future development of more advanced drowsiness

---

[1] * Corresponding author. Tel.: +30-697-067-7742;
*E-mail address:* astyanidis@mail.ntua.gr

modeling techniques. By validating a simple yet effective deep learning model, this work aims to bridge the gap between traditional HRV analysis and modern temporal deep learning frameworks, paving the way for more robust and early detection of driver drowsiness.

## Methodology

### *Materials*

The dataset [4], provided by the VTI, was collected in Sweden from 20 participants (10 male, 10 female) aged 30–60 years, recruited from the national vehicle owner register. All participants were healthy, of normal weight, non–shift workers, non–professional drivers, and did not wear glasses. They maintained regular sleep schedules with at least seven hours of sleep per night during the three nights preceding the experiment and completed pre-study sleep and wakefulness logs. Each participant completed four 90-minute driving sessions—two in a simulator and two on real roads—under both alert (late afternoon) and sleep-deprived (after midnight) conditions. Physiological data, including ECG signals recorded at 256 Hz using a Vitaport 3 system, were collected alongside vehicle data such as speed and lateral position. Subjective drowsiness was assessed every five minutes throughout the sessions using the Karolinska Sleepiness Scale (KSS) [5], which ranges from 1 (extremely alert) to 9 (very sleepy).

### *Pre-processing and feature extraction*

To enable the analysis of physiological signals for drowsiness detection, the raw ECG data were first segmented into temporally meaningful windows. A sliding-window approach was employed, dividing each driving session into two-minute segments with a 50% overlap. This design preserved temporal continuity while providing one-minute update intervals, allowing the model to capture gradual physiological changes associated with variations in alertness.

From each ECG segment, inter-beat intervals (IBIs) were extracted as the basis for HRV analysis. R-peaks were detected using the Pan–Tompkins algorithm [6], a well-established method for identifying QRS complexes in noisy environments. IBIs were then computed as the time intervals between consecutive R-peaks within each segment, providing a reliable foundation for HRV feature extraction.

A comprehensive set of 25 HRV features was extracted from each segment, encompassing 19 time-domain metrics (e.g., mean NN, SDNN), 3 frequency-domain indices (LF, HF, LF/HF), and 3 nonlinear measures (SD1, SD2, SD1/SD2) [7]. To capture temporal dependencies in the physiological signals, consecutive feature vectors were grouped into sequences representing five minutes of driving.

Corresponding sleepiness labels were derived from the participants' self-reported KSS values. Mean KSS values were calculated for each segment and rounded to the nearest integer, ensuring that every window was assigned a representative level of subjective drowsiness for supervised learning. The KSS value corresponding to the final segment in each sequence was used as the ground truth label for model training.

### *Model Architecture*

The proposed drowsiness classification model is based on a Long Short-Term Memory (LSTM) neural network [8], designed to capture temporal dependencies inherent in physiological data. A single-layer LSTM architecture was employed to model the sequential relationships among the extracted HRV feature vectors. The output of the LSTM layer was subsequently passed through a fully connected linear layer, which transformed the learned temporal representations into the corresponding driver drowsiness states. This architecture provides an effective balance between model complexity and interpretability, enabling the network to learn temporal patterns relevant to fluctuations in alertness.

*Experimental Setup*

The drowsiness detection task was formulated as a binary classification problem, where each data instance was labelled as either drowsy or alert based on the corresponding KSS value. Specifically, samples with KSS values greater than or equal to 7 were categorized as drowsy, while all other samples were labelled as alert. This thresholding approach follows common practice in sleepiness-related research, distinguishing between normal alertness and drowsiness levels.

The model was evaluated using two distinct validation strategies: 10-fold cross-validation and leave-one-subject-out (LOSO) cross-validation. In the 10-fold cross-validation, the dataset was split into training and test sets with an 80–20% ratio, and the training process was repeated across ten different folds to reduce dependency on a specific data partition. The reported performance corresponds to the mean results across all folds.

In the LOSO cross-validation, data from a single participant were reserved for testing, while data from all remaining participants were used for training. This procedure was iteratively repeated so that each participant served as the test subject exactly once, ensuring a subject-independent evaluation. The final reported metrics correspond to the average performance across all iterations.

Model performance was evaluated using multiple metrics, including accuracy (Acc), sensitivity (Se), specificity (Sp), and F1-score, providing complementary insights into the model's ability to discriminate between drowsy and alert states.

## Results and Discussion

Table 1 summarizes the overall classification performance of the proposed model under the two evaluation schemes. As expected, the model achieved higher performance in the cross-validation setting. This is reasonable, given that in cross-validation the model is trained on data segments from all subjects, and therefore the testing partitions may contain sequences that are similar to those seen during training.

In contrast, the LOSO results indicate a notable drop in performance. This decline is expected because, in this setup, the test data come from an unseen subject whose physiological patterns may differ substantially from those of the training subjects. Consequently, the model faces greater challenges in generalizing to new individuals.

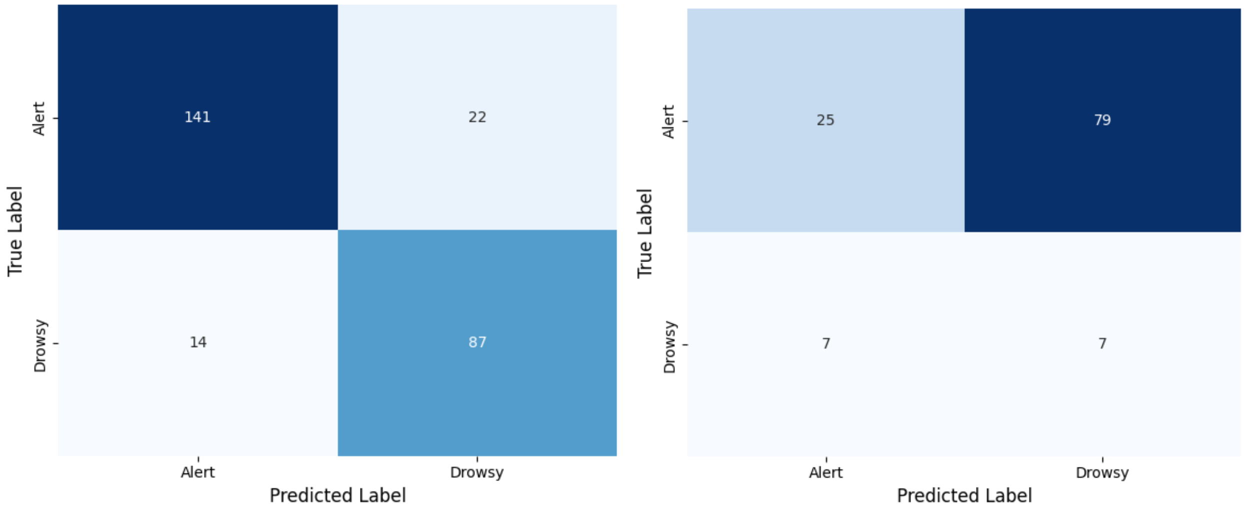
Another important observation concerns the class balance. The cross-validation setting benefits from a more balanced distribution between the drowsy and alert classes, while in the LOSO configuration the balance may vary significantly across subjects. Some participants exhibit only a few drowsy samples, which leads to increased variability in the results. This imbalance can be visually confirmed in Figure 1, which presents representative confusion matrices from both validation schemes. The LOSO example clearly illustrates the limited number of drowsy instances and the strong class imbalance for that specific participant.

It should also be noted that the way the binary labels were defined based on the KSS values directly influences the model's performance. In this study, samples with KSS $\geq 7$ were labelled as drowsy, and the remaining as alert. Choosing a different KSS threshold would alter the class distribution and consequently affect model performance.

Furthermore, the relatively high standard deviation observed in the LOSO results highlights substantial inter-individual variability, suggesting that physiological responses to drowsiness vary considerably between subjects. This variability, combined with the limited dataset size, imposes additional constraints on the model's generalization capability.

**Table 1** Classification Performance (mean±std)

| | Acc | Se | Sp | F1 |
|---|---|---|---|---|
| 10-fold cross-validation | 0.878 ± 0.017 | 0.821 ± 0.030 | 0.913 ± 0.026 | 0.837 ± 0.021 |
| LOSO cross-validation | 0.610 ± 0.169 | 0.427 ± 0.300 | 0.671 ± 0.298 | 0.379 ± 0.257 |



(a) (b)

*Figure 1: Confusion matrices illustrating model performance under different validation schemes. (a) Example confusion matrix obtained from 10-fold cross-validation. (b) Example confusion matrix obtained from the leave-one-subject-out (LOSO) cross-validation.*

## Conclusion and Future Work

Overall, the results emphasize the inherent difficulty of developing generalized models for drowsiness detection using HRV signals alone. These findings suggest that personalized or hybrid modeling approaches may provide a more effective solution for achieving robust detection performance in real-world scenarios. The comparison between cross-validation and leave-one-subject-out evaluations revealed the sensitivity of the model to inter-individual variability, as well as the importance of balanced data distributions across subjects. Despite these challenges, the study demonstrates that HRV-based features carry valuable information related to drowsiness levels and can serve as a promising basis for non-invasive driver monitoring systems.

In summary, this work contributes to the understanding of how HRV can be leveraged for drowsiness detection under different evaluation schemes. The findings underline the need for careful consideration of data balance, subject variability, and labelling strategies when developing and assessing physiological models of alertness. Moreover, the results highlight the limitations of relying solely on HRV data and motivate the integration of additional physiological and contextual signals to enhance detection robustness.

Future work will focus on extending the present framework in several directions. First, incorporating multimodal information—such as electroencephalography (EEG), electrooculography (EOG), or driving behavior metrics—could substantially improve the model's ability to generalize across individuals and driving conditions. Second, personalized modeling approaches, such as transfer learning or subject adaptation techniques, may help account for inter-individual variability. Finally, expanding the dataset to include a larger and more diverse population, along with more balanced drowsy and alert states, will also be essential for improving the reliability of the results.

## Acknowledgements

This research is based on work carried out within the IVORY project. The project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101119590. The authors are grateful to the data collection team at VTI for providing the dataset used in this study.

## References

[1] P. Philip and T. Åkerstedt, "Transport and industrial safety, how are they affected by sleepiness and sleep restriction?" Sleep Medicine Reviews, vol. 10, no. 5, pp. 347–356, 2006

[2] M. S. Al-Quraishi, S. S. A. Ali, M. Al-Qurishi, T. B. Tang, and S. Elferik, "Technologies for detecting and monitoring drivers' states: A systematic review," Heliyon, vol. 10, no. 20, 2024.

[3] J. Heikenfeld, A. Jajack, J. Rogers, P. Gutruf, L. Tian, T. Pan, R. Li, M. Khine, J. Kim, J. Wang, and J. Kim, "Wearable sensors: modalities, challenges, and prospects," Lab Chip, vol. 18, pp. 217–248, 2018.

[4] A. Anund, Camera-Based Sleepiness Detection: Final Report of the Project SleepEYE, Swedish National Road and Transport Research Institute (VTI), 2016.

[5] A. Shahid, STOP, THAT and One Hundred Other Sleep Scales. Springer, 2012.

[6] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," IEEE Transactions on Biomedical Engineering, vol. BME-32, no. 3, pp. 230–236, 1985

[7] A. Persson, H. Jonasson, I. Fredriksson, U. Wiklund, and C. Ahlström, "Heart rate variability for driver sleepiness classification in real road driving conditions," in Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC), 2019, pp. 6537–6540.

[8] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.