

ECG-Based Driver Drowsiness Detection Using a CNN-LSTM Framework

Aristotelis Styaniadis^{1,2}, André Lourenço^{2,3,4}, Carlos Carreiras², Christer Ahlström^{5,6},
Apostolos Ziakopoulos¹, George Yannis¹

¹ National Technical University of Athens, Athens, Greece

² CardioID Technologies, Lisbon, Portugal

³ Instituto Superior de Engenharia de Lisboa - Instituto Politécnico 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—Driver drowsiness is a major contributor to road crashes worldwide, motivating the development of driver monitoring systems based on behavioral, vehicular, and physiological signals. Among physiological modalities, electrocardiography (ECG) has attracted increasing attention due to its ability to capture autonomic nervous system activity associated with fatigue and reduced alertness, while remaining relatively easy to acquire in real-world driving settings through unobtrusive sensors embedded in the steering wheel or vehicle seat. Most existing ECG-based drowsiness detection approaches rely on handcrafted heart rate variability (HRV) features derived from domain knowledge. While these features are well established, it remains unclear whether additional information relevant to drowsiness can be learned directly from the raw ECG waveform. In this work, we explore whether an end-to-end deep learning framework can learn complementary representations directly from raw ECG signals and achieve competitive performance without explicit feature extraction. The proposed architecture combines convolutional neural networks (CNNs) to extract relevant features from the physiological data with long short-term memory (LSTM) networks to capture their temporal dependencies and dynamics. We evaluate the proposed framework on a driving dataset comprising approximately 120 hours of data acquired during simulated and naturalistic driving conditions and compare its performance against a conventional approach based on handcrafted HRV features. Experimental results demonstrate that the CNN-LSTM architecture achieves an accuracy of 60.55% under a leave-one-subject-out (LOSO) evaluation strategy, demonstrating that automatically learned ECG representations can provide competitive performance while avoiding explicit feature extraction. These findings suggest that deep learning offers a promising alternative for ECG-based driver drowsiness detection and motivate further investigation on larger and more diverse datasets.

Index Terms—driver drowsiness, deep learning, electrocardiography.

I. INTRODUCTION

Driver drowsiness constitutes one of the major contributing factors to road crashes worldwide, significantly increasing the risk of slower reaction time, impaired decision-making, and loss of vehicle control [1]. As modern transportation systems continue to evolve, the development of reliable driver monitoring systems has become increasingly important for improving road safety and reducing fatigue-related crashes.

Existing drowsiness detection approaches can generally be categorized into three major groups: behavioral-based, vehicular-based, and physiological-based systems. Behavioral-based approaches typically rely on visual indicators such as facial expressions, eye closure, blinking frequency, and head pose estimation, while vehicle-based approaches monitor driving patterns including steering wheel movements, lane deviation, vehicle speed, and pedal usage. In contrast, physiological-based methods analyze biosignals directly associated with the driver's internal state, including electroencephalography (EEG), electrocardiography (ECG), and electrooculography (EOG), offering the potential for earlier and more objective detection of fatigue and drowsiness [2].

Recent advances in wearable sensing technologies and artificial intelligence have significantly accelerated the development of non-invasive driver monitoring systems. In particular, these developments have enabled the acquisition of physiological signals in real-world environments using compact and unobtrusive sensing platforms [3], [4]. Modern vehicles are increasingly equipped with embedded sensing technologies capable of capturing physiological signals without interfering with the driver [5]. Examples include steering-wheel-integrated electrodes that enable ECG acquisition and subsequent estimation of heart rate variability (HRV).

In parallel, alternative contactless sensing modalities have emerged, including seat-integrated radar and ultra-wideband (UWB) sensors capable of monitoring vital signs through micro-movements associated with cardiac and respiratory activity [6]. Together, these unobtrusive sensing paradigms provide a promising pathway toward practical and scalable drowsiness detection systems for next-generation intelligent transportation environments.

Among physiological modalities, EEG has demonstrated high performance in detecting drowsiness and transitions between alertness states, making it one of the most extensively studied biosignals in the literature [7]–[9]. However, real-world deployment remains challenging due to the need for multiple electrodes, sensitivity to noise, and substantial inter-subject variability. ECG represents a promising alternative,

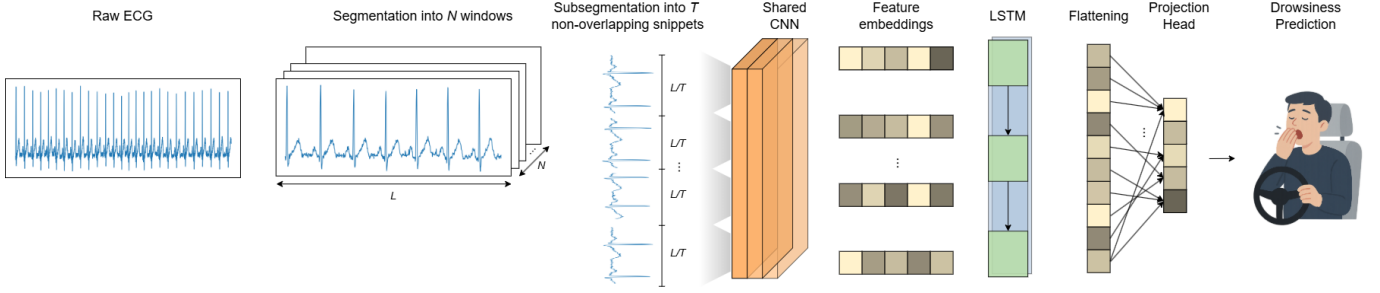


Fig. 1: Overview of the proposed CNN-LSTM architecture for ECG-based drowsiness detection. The raw ECG signals are first segmented into N windows, each of length L . Every window is further divided into T non-overlapping subsegments of equal size (L/T), forming a temporal sequence of raw ECG snippets. Each subsegment is processed by a shared convolutional neural network (CNN), producing a sequence of T feature embeddings per window. These embeddings are then passed to a Long Short-Term Memory (LSTM) network to capture temporal dependencies across subsegments. The final hidden state of the LSTM is flattened and fed into a projection head to generate the predicted drowsiness score for each window.

particularly with the emergence of wearable and vehicle-integrated sensing technologies. Although these modalities often provide lower-quality signals than clinical-grade ECG recordings, they offer a practical path toward continuous monitoring. Despite the strong relationship between cardiac autonomic activity and fatigue, most ECG-based drowsiness detection studies rely on handcrafted HRV features and conventional machine learning techniques [10]–[12]. In this work, we investigate whether representations relevant to drowsiness can be learned directly from ECG signals, providing a foundation for future studies in unobtrusive ECG monitoring systems.

II. MATERIALS AND METHODS

A. Dataset

The *SleepEye* dataset, a widely used benchmark for driver drowsiness detection research, consists of 20 participants (10 male, 10 female) aged 30–60 years, recruited from the Swedish national vehicle owner register [13]. Despite its widespread use, its limited number of participants presents a notable challenge for developing robust and generalizable drowsiness detection models. All participants were healthy, of normal weight, non-shift workers, non-professional drivers, and did not wear glasses. They adhered to regular sleep schedules, obtaining at least seven hours of sleep per night during the three nights prior to 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 signals, including ECG recorded at 256 Hz using a Vitaport 3 system, were collected together with vehicle data such as speed and lateral position. Subjective drowsiness was assessed every five minutes throughout the sessions using the Karolinska Sleepiness Scale (KSS) [14], which ranges from 1 (extremely alert) to 9 (very sleepy).

B. Data Preprocessing

The ECG signals were first resampled to 128 Hz to ensure a consistent temporal resolution while preserving the mor-

phological characteristics of the cardiac waveform. Baseline wander, primarily caused by respiration and electrode motion, was removed using a 4th-order high-pass Butterworth filter with a cutoff frequency of 0.5 Hz. Powerline interference was suppressed using a notch filter centered at 50 Hz, while high-frequency noise and electromyographic artifacts were attenuated using a low-pass Butterworth filter with a cutoff frequency of 40 Hz, resulting in an effective bandwidth of 0.5–40 Hz. The preprocessed ECG signals were then segmented into fixed-length, non-overlapping windows of 15,360 samples, corresponding to approximately 2 minutes of driving. Segments containing saturation or extreme amplitude deviations, typically associated with severe motion artifacts, were excluded from the analysis.

C. Problem Formulation and Notation

In this work, we consider N sliding windows extracted from the raw ECG signal, denoted as $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N) \in \mathbb{R}^{N \times L}$, where L represents the total number of samples per window. Each window is further segmented into T non-overlapping subsegments, each of length L/T , yielding a tensor representation of the input as $\tilde{\mathbf{X}} \in \mathbb{R}^{N \times T \times (L/T)}$. The corresponding target drowsiness labels are given by $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_N) \in \mathbb{R}^N$. We assume $N = 3072$ windows of fixed length $L = 15,360$, with each window further segmented into a fixed number of $T = 4$ subsegments of equal length. However, these choices are not restrictive and may be adapted depending on the task requirements or the temporal resolution of the data. A shared convolutional neural network (CNN) is applied to each subsegment to extract feature representations $\mathbf{E} = (\mathbf{e}_1, \dots, \mathbf{e}_N) \in \mathbb{R}^{N \times T \times D}$, where D denotes the dimensionality of the latent space. These embeddings are then fed into a Long Short-Term Memory (LSTM) network to model temporal dependencies across subsegments within each window. The final hidden state of the LSTM is used as a compact representation and passed through a projection head to predict the drowsiness labels $\tilde{\mathbf{y}} = (\tilde{\mathbf{y}}_1, \dots, \tilde{\mathbf{y}}_N) \in \mathbb{R}^N$.

D. Model Architecture

As illustrated in Fig. 1, the proposed architecture follows a CNN-LSTM design inspired by DeepSense [15]. The model consists of three stacked 1D CNN blocks with 32, 32, and 64 filters, respectively, all using a kernel size of 9. Each convolution is followed by a Leaky ReLU activation (slope 0.1), batch normalization, max pooling (pool size 4), and dropout (rate 0.2). The CNN feature extractor is followed by a global average pooling layer to summarize feature maps into compact embeddings. The resulting four representations are then fed into an LSTM layer with 64 hidden units to model temporal dependencies across subsegments. Finally, the last hidden state is passed through a fully connected layer with a sigmoid activation to perform binary classification.

E. Training Procedure

We employed a leave-one-subject-out (LOSO) cross-validation strategy. In each fold, one subject was held out as the test set, while two additional subjects from the remaining pool were used as a validation set. This ensured that training, validation, and test sets contained disjoint participants, preventing subject-specific leakage. Due to substantial inter-subject variability in ECG amplitude caused by factors such as electrode placement, skin conductivity, and body composition, subject-wise normalization was applied prior to model training. To mitigate class imbalance during training, samples were drawn using a weighted random sampler, which increased the probability of selecting instances from the minority class.

The model was trained using the Adam optimizer with a learning rate of 10^{-3} for up to 100 epochs. Early stopping was applied based on validation AUC with a patience of 15 epochs. Training was performed with a batch size of 16, using sequences composed of 4 temporal subsegments per input window.

Binary drowsiness labels were derived from the KSS by applying a threshold of 7, where ($KSS < 7$) is defined as the alert state and ($KSS \geq 7$) is defined as the drowsy state. When multiple KSS annotations were available within a window, the last recorded value was used as the ground truth before binarization. Windows for which no KSS annotation was available were excluded from the analysis to ensure reliable ground truth assignment.

F. Baseline and Performance Metrics

As a baseline, a conventional approach based on handcrafted HRV features extracted from the same ECG windows was considered. To ensure a fair comparison, HRV features were organized into the same temporal subsegments and provided as input to an LSTM network followed by a feed-forward classification layer. For each subsegment, 19 HRV features were extracted. Only time-domain features were considered due to the relatively short window length, as frequency-domain and nonlinear features are generally not physiologically meaningful at such durations. This baseline preserves the temporal modeling capability of the LSTM while enabling a direct

comparison between handcrafted features and representations learned by the proposed CNN feature extractor.

Model performance was evaluated using accuracy, sensitivity, specificity, macro F1-score, and AUC-ROC. All metrics were computed independently for each LOSO fold and subsequently averaged across folds.

III. RESULTS AND DISCUSSION

Table I reports the mean performance across LOSO folds for both the baseline approach based on handcrafted HRV features and the proposed CNN-based feature extraction framework. Figure 2 illustrates the per-subject performance of the CNN-LSTM model.

As shown in Table I, both approaches achieve moderate performance under the LOSO evaluation strategy. This outcome is expected, as LOSO represents a challenging and realistic scenario in which the model is required to generalize to entirely unseen individuals. Compared to conventional train-test splits, LOSO typically produces lower performance due to the substantial physiological variability observed across subjects.

The large standard deviations reported in Table I, together with the per-subject results shown in Fig. 2, further highlight the impact of inter-subject variability. While the proposed model performs effectively for certain participants, its performance deteriorates considerably for others. Such variability is common in physiological signal analysis due to differences in individual physiological characteristics, including baseline heart rate and autonomic nervous system activity. In addition, the KSS is inherently subjective, as it relies on each participant's self-assessment of sleepiness. Consequently, participants experiencing similar physiological states may assign different KSS scores, introducing additional variability that is difficult for the model to capture. Although subject-wise normalization was applied during preprocessing, these physiological and subjective differences remain, making it challenging to learn subject-independent representations.

This inter-subject variability may also be influenced by the limited amount of training data available in each LOSO fold. Because the validation set is composed of a small number of participants, its distribution may not always be representative of the held-out test subject. Consequently, model selection based on validation performance may not consistently reflect true generalization performance. These findings underline the difficulty of developing generalized drowsiness detection models from ECG signals alone and suggest that personalization or subject-adaptation strategies may be beneficial for improving real-world performance.

Finally, performance variability across folds may be further amplified by class imbalance at the subject level. For some participants, one class is considerably underrepresented, making it challenging to learn a robust decision boundary despite the use of a weighted sampling strategy during training. This issue highlights the need for larger and more balanced datasets to facilitate the development of more reliable subject-independent drowsiness detection systems.

TABLE I: Performance comparison of HRV-LSTM and CNN-LSTM models under leave-one-subject-out (LOSO) cross-validation (mean $\pm$ standard deviation)

	Accuracy (%)	Sensitivity (%)	Specificity (%)	F1 score (%)	AUC (%)
HRV-LSTM	62.49 $\pm$ 10.99	60.68 $\pm$ 15.97	66.83 $\pm$ 24.48	60.53 $\pm$ 11.15	69.17 $\pm$ 13.60
CNN-LSTM	60.55 $\pm$ 12.88	69.65 $\pm$ 34.78	58.08 $\pm$ 36.85	54.39 $\pm$ 16.37	75.91 $\pm$ 15.70



Fig. 2: Per-subject CNN-LSTM performance under leave-one-subject-out (LOSO) cross-validation. Radar plots depict (a) accuracy, and (b) F1 score for each test subject, illustrating the substantial inter-subject variability.

IV. CONCLUSION

In this study, we proposed an ECG-based approach for driver drowsiness detection using an end-to-end CNN-LSTM framework. The model leverages 1D convolutional layers to extract local temporal patterns from ECG subsegments, while LSTM layers capture longer-term temporal dependencies. We evaluated the proposed approach against a baseline using handcrafted HRV features and demonstrated the potential of learned representations for drowsiness detection.

Despite these encouraging results, the study is limited by the relatively small dataset size, which constrains both model complexity and generalization performance. In particular, more data-intensive architectures such as transformers were not explored due to data limitations. Moreover, substantial inter-subject variability was observed, with the model performing well for some individuals but failing to generalize consistently across all subjects. This highlights the importance of subject-specific adaptation strategies, such as domain adaptation, to improve robustness across subjects.

Future work will focus on improving generalization by expanding the dataset through the integration of multiple experimental sources, enabling more comprehensive cross-dataset evaluation and analysis of sample size effects. We also plan to investigate approaches for mitigating inter-subject variability, including adversarial learning strategies that encourage the extraction of subject-invariant representations while preserving information relevant to drowsiness detection. These directions aim to support the development of more robust and generalizable driver monitoring systems.

ACKNOWLEDGMENT

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.

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] G. Sikander and S. Anwar, "Driver fatigue detection systems: A review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 6, pp. 2339–2352, 2019.
- [3] H. Lee, J. Lee, and M. Shin, "Using Wearable ECG/PPG Sensors for Driver Drowsiness Detection Based on Distinguishable Pattern of Recurrence Plots," *Electronics*, vol. 8, no. 2, 2019.
- [4] S. Seneviratne, Y. Hu, T. Nguyen, G. Lan, S. Khalifa, K. Thilakarathna, M. Hassan, and A. Seneviratne, "A Survey of Wearable Devices and Challenges," *IEEE Communications Surveys Tutorials*, vol. PP, pp. 1–1, 07 2017.
- [5] J. R. Pinto, J. S. Cardoso, A. Lourenço, and C. Carreiras, "Towards a Continuous Biometric System Based on ECG Signals Acquired on the Steering Wheel," *Sensors*, vol. 17, no. 10, 2017.
- [6] S. Leonhardt, L. Leicht, and D. Teichmann, "Unobtrusive Vital Sign Monitoring in Automotive Environments - A Review," *Sensors*, vol. 18, no. 9, 2018.
- [7] S. Barua, M. U. Ahmed, C. Ahlström, and S. Begum, "Automatic Driver Sleepiness Detection Using EEG, EOG and Contextual Information," *Expert Systems with Applications*, vol. 115, pp. 121–135, 2019.
- [8] G. Yang, Y. Lin, and P. Bhattacharya, "A driver fatigue recognition model based on information fusion and dynamic bayesian network," *Inf. Sci.*, vol. 180, no. 10, p. 1942–1954, May 2010. [Online]. Available: <https://doi.org/10.1016/j.ins.2010.01.011>
- [9] M. Hultman, I. Johansson, F. Lindqvist, and C. Ahlström, "Driver Sleepiness Detection with Deep Neural Networks Using Electrophysiological Data," *Physiological Measurement*, vol. 42, no. 3, 2021.
- [10] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness Detection Using Heart Rate Variability," *Medical & Biological Engineering & Computing*, vol. 54, no. 6, pp. 927–937, 2016.
- [11] A. Persson, H. Jonasson, I. Fredriksson, U. Wiklund, and C. Ahlström, "Heart Rate Variability for Classification of Alert Versus Sleep Deprived Drivers in Real Road Driving Conditions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3316–3325, 2021.
- [12] M. Patel, S. Lal, D. Kavanagh, and P. Rossiter, "Applying neural network analysis on heart rate variability data to assess driver fatigue," *Expert Syst. Appl.*, vol. 38, no. 6, p. 7235–7242, Jun. 2011. [Online]. Available: <https://doi.org/10.1016/j.eswa.2010.12.028>
- [13] A. Anund, C. Fors, C. Ahlström, P. Sörner, J. Kovaceva, E. Hasselberg, M. Krantz, J.-F. Grönvall, and K. Kircher, "Camera-based sleepiness detection: Final report of the project sleep-eye," Swedish National Road and Transport Research Institute (VTI), Linköping, Sweden, Tech. Rep. 2011-6, 2011. [Online]. Available: <https://urn.kb.se/resolve?urn=urn:nbn:se:vti:diva-5270>
- [14] T. Åkerstedt and M. Gillberg, "Subjective and objective sleepiness in the active individual," *International Journal of Neuroscience*, vol. 52, no. 1–2, pp. 29–37, 1990.
- [15] S. Yao, S. Hu, Y. Zhao, A. Zhang, and T. F. Abdelzaher, "DeepSense: A Unified Deep Learning Framework for Time-Series Mobile Sensing Data Processing," *CoRR*, vol. abs/1611.01942, 2016. [Online]. Available: <http://arxiv.org/abs/1611.01942>