# Impact of Missing Inter-Beat Intervals on Heart Rate Variability Features and Driver Drowsiness Detection


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

[1] National Technical University of Athens, 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 using physiological data, particularly heart rate variability (HRV), has emerged as a promising approach for assessing driver fatigue. Inter-beat intervals (IBIs) represent the time between consecutive heartbeats and are used to derive HRV, a measure of autonomic nervous system activity. HRV can be extracted from electrocardiogram (ECG), photoplethysmography (PPG) or other physiological signals, making it valuable for applications in health monitoring and driver state assessment. However, real-world conditions often lead to missing or corrupted data due to inconsistent sensor contact, motion artifacts, or signal interruptions. Such data loss can impact HRV feature extraction and affect downstream machine learning (ML) models. This study investigates the effects of missing data by systematically removing 15–30% of IBIs, either randomly or in sequential blocks. The distribution analysis indicated that IBI distributions largely retain their overall structure, with only minor deviations. Classification performance was robust to the investigated data losses, but when comparing 10-fold cross-validation to leave-one-participant-out validation, mean accuracy dropped by approximately 15%, and variability in accuracy across folds increased by around 10%. tSNE feature space visualization further revealed that class separability was much clearer at the participant level than at the group level. The findings underscore the need for personalized models tailored to each driver's physiological patterns, which has implications for drowsiness detection systems.**



***Index Terms*—IBI, HRV, driver drowsiness detection**


## I. Introduction

Drowsiness is a contributing factor in 20% of all crashes [1]. Driver drowsiness detection is a critical component of modern driver monitoring systems, aiming to enhance road safety by identifying drowsiness. Traditional approaches typically fall into three categories: vehicle-based, behavioral and physiological. Vehicle-based methods analyze driving patterns, including steering deviations and lane-keeping performance, offering non-intrusive monitoring but often struggling with variability across different driving contexts. Behavioral approaches, such as eye closure, head movements, and yawn detection through camera-based systems, have advanced significantly the past few years. However, challenges like occlusions, squinting, and individual variability still affect their robustness. Physiological-based methods, particularly those leveraging heart rate variability (HRV) [2], provide a direct and continuous measure of autonomic nervous system activity, but are also influenced by factors beyond drowsiness and typically require longer time windows, which can delay drowsiness detection. As no single modality offers a fully reliable solution, integrating multiple approaches holds promise for developing more accurate and robust systems. In this study, we aim to advance physiological-based drowsiness detection by improving the use of HRV.

With advancements in sensor technology, physiological data have become increasingly accessible through wearable devices [3], [4] and remote sensors embedded in steering wheels, seat belts, and seats [5], [6]. These innovations have renewed interest in HRV-based drowsiness detection, leveraging the relationship between cardiac function and drowsiness. HRV, derived from inter-beat intervals (IBIs) extracted from electrocardiogram (ECG), photoplethysmography (PPG) or even other measures of physiological signals [7], is a useful indicator of fluctuations in alertness and drowsiness.

However, real-time physiological monitoring is prone to data loss due to motion artifacts, sensor displacement, and inconsistent skin contact [8]. Wearable devices, such as ECG chest straps or PPG armbands, often suffer from poor electrode contact and temporary signal interruptions, leading to incomplete or unreliable data. Similarly, steering wheel-embedded sensors, such as CardioWheel [9], require both hands to be placed on the wheel for accurate signal acquisition. During real-world driving, drivers frequently remove their hands, resulting in intermittent signal loss and missing IBI data. These disruptions pose a significant challenge for HRV-based drowsiness detection, potentially affecting the reliability of extracted features and the performance of machine learning models.

In this study, we investigate the impact of missing IBI data on HRV-based drowsiness detection [10], [11], simulating real-world data loss scenarios. By systematically removing IBIs either randomly or in sequential blocks, we assess the robustness of HRV features and machine learning (ML) models under varying levels of missing data. The objective of this study is to determine the extent of data loss that can be tolerated without significantly compromising the performance of ML models, thereby informing future strategies for data imputation and model robustness.

## II. Methodology

The overall workflow of our method is illustrated in the following Fig. 1.

### A. Pre-processing

To analyze physiological signals for drowsiness detection, the raw data must first be segmented into meaningful windows. This segmentation step ensures that subsequent feature extraction and classification models operate on standardized time intervals. In this study, a windowing approach was applied, where each driving session was divided into two-minute segments with a 50% overlap. This overlap preserves temporal continuity, allowing each segment to incorporate data from the preceding two minutes while maintaining a one-minute update rate.

Peak detection is a crucial step in extracting IBIs from physiological signals such as ECG and PPG. In this study, R-peaks were detected in ECG signals using the Pan-Tompkins algorithm [12], a widely used method for identifying QRS complexes in noisy environments. From these detected peaks, IBIs were computed as the time intervals between consecutive heartbeats within each segment. To ensure the reliability of HRV features, segments containing fewer than 60 IBIs within the two-minute window were excluded from further analysis.

For this study, a supervised learning approach was adopted, leveraging subjective sleepiness ratings as ground truth labels. Driver drowsiness was assessed using the Karolinska Sleepiness Scale (KSS) [13], a widely used self-reported measure that ranges from 1 (extremely alert) to 9 (very sleepy, fighting sleep). Mean KSS scores were computed for each segment and rounded to the nearest integer to establish class labels. This labeling process ensured that each segment was associated with a representative sleepiness rating, enabling machine learning models to learn patterns associated with drowsiness.

### B. Missing Data Simulation

To systematically investigate the impact of missing data on HRV features, a controlled data removal approach was implemented. In more detail, each analysis window was modified by removing a fixed percentage of interbeat intervals (IBIs), set at either 15% or 30%, ensuring consistency across all windows. This method allows for a direct assessment of how varying levels of data loss influence the computed HRV features.

By applying a uniform deletion rate across all windows, the analysis isolates the effect of missing data, preventing confounding factors that could arise from inconsistencies in data availability. This approach enables a clearer evaluation of the robustness of HRV metrics under different levels of missing data. Additionally, it provides insights into the extent to which HRV features remain reliable as data loss increases, which is particularly relevant for real-world applications where missing IBIs are common due to signal artifacts or sensor limitations.

Two different approaches were implemented to assess the effects of missing IBIs. First, a random deletion method was applied, in which different percentages of IBIs were removed without any specific pattern. This served as a baseline approach, introducing non-patterned noise to simplify the problem. By analyzing changes in the histograms of inter-beat intervals and the distributions of HRV features, this method allowed for an initial assessment of how random data loss affects feature stability. Additionally, it helped determine whether a sequential removal approach was necessary and established a threshold for the level of random data loss that a given processing pipeline could tolerate before significant distortions occurred.

The second approach involved sequentially removing IBIs, simulating real-world scenarios where sensor contact is lost for short periods. Unlike the random deletion method, this approach introduced structured gaps in the data, simulating events such as a driver lifting one or both hands off the steering wheel for several seconds. This more challenging yet realistic scenario provided valuable insights into how extended periods of missing data influence HRV feature extraction. Understanding the impact of sequential data loss is critical for developing robust preprocessing methods capable of handling gaps in ECG recordings without compromising the reliability of downstream analyses.

### C. HRV Feature Extraction

HRV features can be broadly categorized into time-domain, frequency-domain, and nonlinear domain [14], each providing different insights into autonomic nervous system activity. Time-domain HRV features are particularly suitable for analysis in the presence of missing data, as they rely on statistical measures of IBIs rather than their exact timing or frequency distribution. These features capture overall heart rate fluctuations over a given period, making them more robust to signal interruptions and gaps in the data.

Frequency-domain features, such as low-frequency (LF) power, high-frequency (HF) power, and the LF/HF ratio, were not computed due to the presence of missing IBIs. The calculation of frequency-domain HRV features relies on a continuous and evenly sampled IBIs. However, gaps in the data introduce artifacts that distort spectral estimates. When IBIs are missing, standard spectral analysis methods, such as the Fast Fourier Transform (FFT) and Lomb-Scargle periodogram, may produce unreliable results or fail to compute power spectrum estimates altogether.

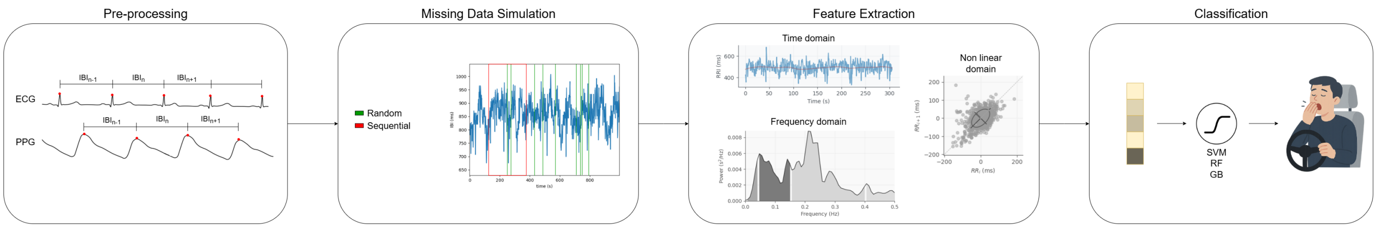


Fig. 1: Missing data handling workflow: (i) Pre-processing including signal segmentation, peak detection, and IBI extraction (ii) Missing data simulation by systematically removing 15% or 30% of IBIs, either randomly or in sequential blocks (iii) HRV feature extraction, limited in this study to time domain features (iv) Classification by feeding the extracted feature vector to classifiers for driver drowsiness detection.

Similarly, nonlinear HRV features, such as Poincaré plot parameters and entropy-based measures, were not extracted. These features are highly sensitive to data gaps and require long, uninterrupted time series to capture complex physiological patterns. Given the variability in IBIs availability in real-world conditions, the reliability of these features would be compromised.

By focusing on time-domain HRV features, the analysis minimizes the impact of data gaps and ensures robust feature extraction, aligning with the constraints of the available physiological recordings. Eight commonly used HRV time-domain features were extracted, as summarized in Table I. These features were standardized by subtracting the mean and dividing by the standard deviation to ensure comparability across participants and sessions.

TABLE I: HRV time-domain features.

| HRV feature | Description |
|---|---|
| meanHR | Mean heart rate |
| SDNN | Standard deviation of NN intervals |
| RMSSD | Root mean square of successive RR interval differences |
| SDSD | Standard deviation of successive RR interval differences |
| NN50 | Number of successive RR intervals that differ by more than 50 ms |
| pNN50 | Percentage of successive RR intervals that differ by more than 50 ms |
| NN20 | Number of successive RR intervals that differ by more than 20 ms |
| pNN20 | Percentage of successive RR intervals that differ by more than 20 ms |

## III. Evaluation

### A. Dataset

The dataset utilized in this study is the Valu3s dataset, provided by the Swedish National Road and Transport Research Institute (VTI) [15]. Data collection was conducted in Sweden and involved 20 participants (10 male, 10 female) aged between 20 and 60 years. Participants were required to hold a valid driver's license for passenger cars and to be regular drivers. Individuals with a body mass index (BMI) exceeding 35 were excluded to mitigate potential confounding factors, such as undiagnosed sleep disorders and motion sickness.

Data collection was conducted using two driving simulators. Both simulators were equipped with physiological sensors to facilitate the collection of drowsiness-related data. The dataset includes recordings from CardioWheel, a steering wheel-embedded ECG sensor, as well as additional physiological signals relevant to drowsiness detection, such as electrooculogram (EOG) and electrocardiogram (ECG). The EOG (vertical channel, right eye) and the ECG (lead-II) were recorded with a Vitaport 3-system (Temec Instruments BV, the Netherlands). The electrodes used for the EOG and ECG were of the disposable Ag/AgCl type. CardioWheel data were sampled at a frequency of 1000 Hz, while EOG and ECG signals were sampled at 256 Hz.

Each participant completed four driving sessions: two during daytime and two at night after having been awake since early morning. Each session lasted approximately 60 minutes. For this study, only the ECG data acquired from the chest strap were used, as it provided a continuous signal without interruptions. In contrast, CardioWheel ECG data were subject to signal loss whenever participants removed their hands from the steering wheel, making it unsuitable for this analysis.

Throughout the driving sessions, participants were asked to verbally assess their sleepiness using the Karolinska Sleepiness Scale (KSS). Sleepiness ratings were recorded as an average value every fifth minute during the driving sessions. These KSS ratings serve as ground truth labels for training classification models in subsequent analyses.

### B. Drowsiness Classification and Metrics

The HRV 2-minute features were divided into two subsets: 80% for training and 20% for testing. The training set was used to establish the classifiers, and the test set was used to evaluate the performance of the classifiers by computing the accuracy, precision, recall, f1-score and matthews correlation coefficient (MCC).

To ensure that the results were not dependent on a specific data partitioning, this process was repeated ten times. The final test results represent the mean values obtained across these ten iterations (10-fold cross-validation). Additionally, all classifiers were evaluated using leave-one-participant-out (LOO) cross-validation, a more stringent validation method that assesses subject-independent performance. This approach is particularly challenging due to the inherent inter-individual variability in physiological responses.

For classification, two sleepiness classes were defined: alert (KSS $\leq 6$) and drowsy (KSS $> 6$). By framing the problem as a binary classification task, the analysis accounted for the subjectivity of KSS scores, which can vary across individuals. This redefinition was necessary to enhance the robustness of the classification model and mitigate inconsistencies arising from inter-individual differences in self-reported sleepiness [16]. The binary KSS distribution per session is illustrated on Fig. 2.

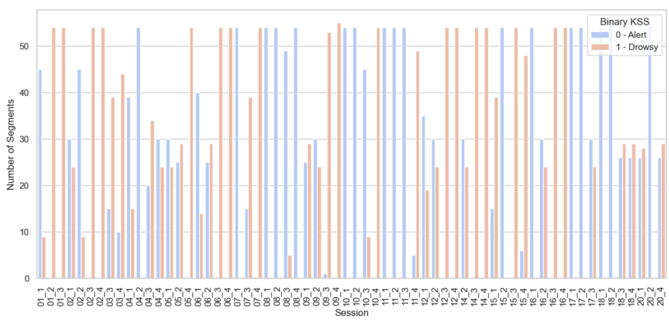


Fig. 2: Binary KSS distribution per session

Three machine learning classifiers were implemented for the binary classification of driver drowsiness: Support Vector Machine (SVM), Random Forest (RF), and Gradient Boosting (GB). The SVM model utilized a radial basis function (RBF) kernel with a regularization parameter 1.0 and an automatically scaled gamma value. The Random Forest classifier consisted of 100 decision trees with no maximum depth constraint, employing balanced class weights and parallel processing to optimize performance. The Gradient Boosting model was configured with 100 estimators, a learning rate of 0.1, and a maximum tree depth of 3, balancing predictive accuracy and model complexity. A fixed random seed ensured reproducibility across all models.

### *C. Distribution Analysis*

Since it is not possible to visually determine whether the removal of IBIs introduces a noticeable skew or flattens the distribution, a quantitative metric is required to assess these shifts. To evaluate how the inter-beat interval distribution is affected by missing IBIs, we employ the Kullback–Leibler (KL) divergence, also known as relative entropy.

KL divergence measures the difference between two probability distributions, quantifying how much one distribution deviates from another. Given two probability distributions, p (the reference distribution) and q (the altered distribution due to missing values), the KL divergence is defined as:

$$\mathrm{KL}(p \parallel q) = \sum_{i=1}^{N} p(x_i) \log \frac{p(x_i)}{q(x_i)}$$

where $p(x_i)$ and $q(x_i)$ represent the probability mass functions of the respective distributions. In this study, $p$ corresponds to the original IBI distribution, while $q$ represents the IBI distribution after the introduction of missing IBIs.

A higher KL divergence value indicates a greater shift in the distribution, meaning the removal of IBIs has significantly altered the statistical properties of the inter-beat intervals. Conversely, a lower KL divergence suggests that the distribution remains relatively unchanged despite missing data. This metric provides an objective means of assessing the extent to which missing IBIs affect the overall structure of the inter-beat interval distribution.

### *D. tSNE Analysis*

A tSNE analysis was conducted to explore the distribution of time-domain HRV features in the embedding space. Initially, we applied tSNE across all participants (Fig. 3), where no clear distinction between alert (blue dots) and drowsy (orange dots) states was observed, suggesting that HRV features do not form separable clusters in a global context. To further investigate potential individual differences, we performed tSNE per participant. In the case of participant 1 (Fig. 4), distinct clusters emerge with only minor outliers, indicating some degree of separability between alertness states. Conversely, participant 5 (Fig. 5) represents a scenario where alert and drowsy states are not clearly separated, with a substantial overlap in the feature space. These findings highlight both the subjective nature of KSS self-reports and the uniqueness of HRV time-domain features, reinforcing the necessity for personalized modeling approaches in driver state monitoring.

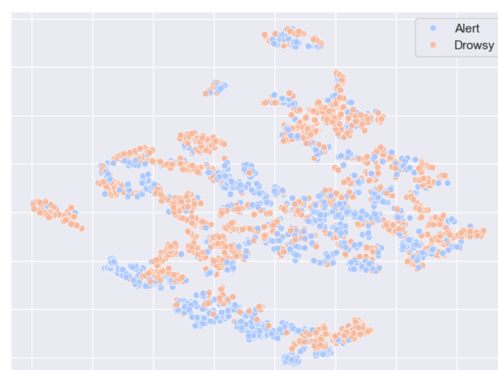


Fig. 3: tSNE plot across all Participants

## IV. Results

The classification performance is summarized in Tables II and III. As expected, classification performance decreases as the proportion of missing data increases, both in cases of randomly missing data and sequential blocks of data.

10-fold cross-validation (CV) demonstrates greater stability and higher overall performance compared to leave-one-participant-out (LOO) cross-validation. This superior performance is expected, as 10-fold CV model has access to data

TABLE II: Classification Performance (mean±std) with 10-Fold Cross-Validation Under Random and Sequential Missing Data

| Model | Metric | Baseline | Random Missing Data | | Sequential Missing Data | |
|---|---|---|---|---|---|---|
| | | | 15% | 30% | 15% | 30% |
| SVM | Accuracy | 71.20 ± 3.0 | 70.10 ± 2.8 | 69.55 ± 2.4 | 70.13 ± 2.5 | 70.71 ± 2.3 |
| | Precision | 71.41 ± 3.1 | 70.39 ± 2.9 | 69.90 ± 2.4 | 70.32 ± 2.6 | 70.95 ± 2.4 |
| | Recall | 71.20 ± 3.0 | 70.10 ± 2.8 | 69.56 ± 2.4 | 70.13 ± 2.5 | 70.71 ± 2.3 |
| | F1-score | 71.13 ± 3.0 | 69.98 ± 2.8 | 69.40 ± 2.4 | 70.05 ± 2.5 | 70.62 ± 2.3 |
| | MCC | 42.60 ± 6.2 | 40.48 ± 5.8 | 39.44 ± 4.9 | 40.44 ± 5.1 | 41.65 ± 4.7 |
| RF | Accuracy | 78.77 ± 2.6 | 77.38 ± 2.3 | 74.26 ± 1.6 | 77.38 ± 1.6 | 75.90 ± 2.0 |
| | Precision | 78.80 ± 2.5 | 77.47 ± 2.2 | 74.38 ± 1.4 | 77.44 ± 1.6 | 75.97 ± 2.0 |
| | Recall | 78.77 ± 2.6 | 77.38 ± 2.3 | 74.26 ± 1.6 | 77.38 ± 1.6 | 75.90 ± 2.0 |
| | F1-score | 78.76 ± 2.6 | 77.35 ± 2.3 | 74.22 ± 1.7 | 77.37 ± 1.6 | 75.89 ± 2.0 |
| | MCC | 57.57 ± 5.2 | 54.85 ± 4.5 | 48.64 ± 3.1 | 54.82 ± 3.2 | 51.88 ± 4.1 |
| GB | Accuracy | 75.82 ± 1.9 | 74.66 ± 2.5 | 72.76 ± 2.0 | 74.40 ± 2.5 | 72.76 ± 2.0 |
| | Precision | 75.93 ± 2.0 | 74.77 ± 2.4 | 72.86 ± 2.0 | 74.49 ± 2.5 | 72.86 ± 2.0 |
| | Recall | 75.82 ± 1.9 | 74.66 ± 2.5 | 72.76 ± 2.0 | 74.40 ± 2.5 | 72.76 ± 2.0 |
| | F1-score | 75.79 ± 1.9 | 74.63 ± 2.5 | 72.72 ± 2.0 | 74.37 ± 2.5 | 72.72 ± 2.0 |
| | MCC | 51.75 ± 3.9 | 49.44 ± 4.9 | 45.62 ± 4.0 | 48.89 ± 5.1 | 47.46 ± 4.1 |

TABLE III: Classification Performance (mean±std) with Leave-One-Participant-Out (LOO) Cross-Validation Under Random and Sequential Missing Data

| Model | Metric | Baseline | Random Missing Data | | Sequential Missing Data | |
|---|---|---|---|---|---|---|
| | | | 15% | 30% | 15% | 30% |
| SVM | Accuracy | 55.61 ± 17.9 | 55.26 ± 17.9 | 54.32 ± 18.3 | 54.49 ± 18.1 | 55.85 ± 17.2 |
| | Precision | 71.14 ± 12.4 | 68.77 ± 18.8 | 66.22 ± 20.5 | 70.89 ± 12.2 | 71.22 ± 12.1 |
| | Recall | 55.61 ± 17.9 | 55.26 ± 17.9 | 54.32 ± 18.3 | 54.49 ± 18.1 | 55.85 ± 17.2 |
| | F1-score | 52.25 ± 20.5 | 51.80 ± 21.0 | 50.62 ± 21.3 | 51.27 ± 20.6 | 53.08 ± 19.3 |
| | MCC | 14.45 ± 20.1 | 14.44 ± 20.3 | 13.04 ± 20.3 | 12.91 ± 18.8 | 14.34 ± 18.9 |
| RF | Accuracy | 56.28 ± 17.7 | 56.05 ± 16.8 | 53.90 ± 16.5 | 56.81 ± 16.4 | 57.39 ± 16.3 |
| | Precision | 68.64 ± 12.6 | 69.48 ± 11.6 | 68.25 ± 11.7 | 68.75 ± 12.1 | 69.38 ± 11.6 |
| | Recall | 56.28 ± 17.7 | 56.05 ± 16.8 | 53.90 ± 16.5 | 56.81 ± 16.4 | 57.39 ± 16.3 |
| | F1-score | 54.63 ± 20.7 | 54.47 ± 19.4 | 52.10 ± 19.4 | 55.98 ± 18.3 | 56.54 ± 18.4 |
| | MCC | 15.06 ± 21.5 | 15.30 ± 19.7 | 11.05 ± 18.7 | 15.13 ± 20.1 | 15.32 ± 19.7 |
| GB | Accuracy | 54.88 ± 17.9 | 55.67 ± 16.8 | 53.87 ± 16.0 | 52.30 ± 16.2 | 55.01 ± 16.4 |
| | Precision | 70.67 ± 10.7 | 71.54 ± 11.8 | 68.88 ± 12.8 | 70.60 ± 12.4 | 70.21 ± 12.2 |
| | Recall | 54.88 ± 17.9 | 55.67 ± 16.8 | 53.87 ± 16.0 | 52.30 ± 16.2 | 55.01 ± 16.4 |
| | F1-score | 51.72 ± 21.9 | 53.49 ± 19.8 | 51.73 ± 18.6 | 49.39 ± 20.3 | 53.40 ± 19.3 |
| | MCC | 14.21 ± 20.1 | 15.22 ± 19.9 | 10.92 ± 19.4 | 12.79 ± 18.1 | 13.35 ± 20.0 |

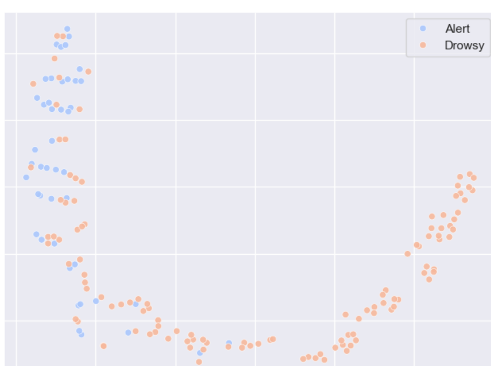


Fig. 4: tSNE plot for Participant 1

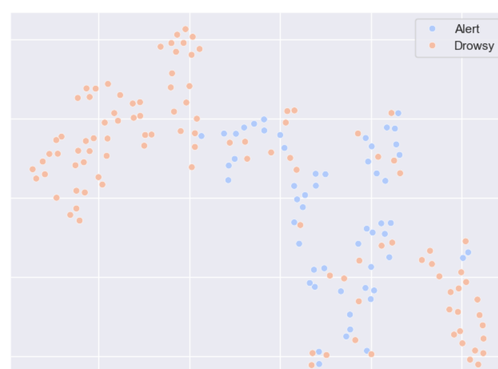


Fig. 5: tSNE plot for Participant 5

from all participants during training, including those it is evaluated on. In contrast, LOO cross-validation simulates a more realistic scenario where the model encounters entirely unseen individuals, resulting in lower and more variable performance. This discrepancy highlights that models benefit significantly from access to participant-specific data suggesting that physiological features like HRV are highly individual-dependent. The results underscore the challenge of building generalized models from HRV signals alone and point to the potential advantages of personalized approaches for improving detection performance in real-world applications.

Furthermore, the instability in LOO cross-validation can be attributed, in part, to class imbalance among certain participants, where one class may be significantly underrepresented. In such cases, the model faces difficulties in learning an effective decision boundary, further amplifying performance fluctuations across different validation folds.

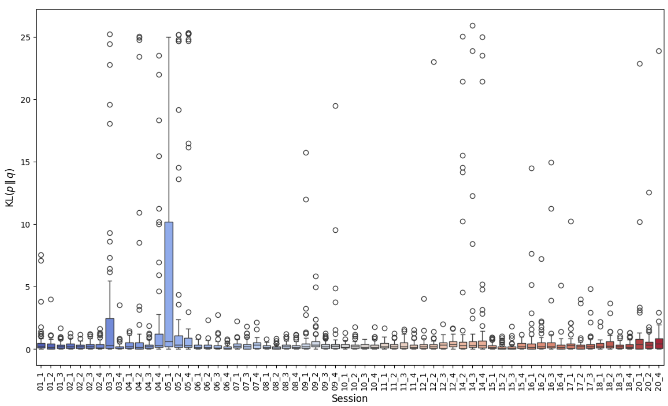


Fig. 6: KL divergence across sessions for 30% Random Missing Data

TABLE IV: Number of segments with KL divergence below 3 and 5 for different missing data conditions (total segments = 3462).

| | Random Missing Data | | Sequential Missing Data | |
|---|---|---|---|---|
| | 15% | 30% | 15% | 30% |
| $\mathrm{KL}(p\|\|q) < 3$ | 3412 | 3340 | 3419 | 3342 |
| $\mathrm{KL}(p\|\|q) < 5$ | 3423 | 3366 | 3427 | 3378 |

The KL divergence across all driving sessions, as shown in Fig. 6 and Table IV, demonstrates that IBI distributions remain relatively unchanged even at the highest levels of missing data considered in this study. This finding reinforces our earlier assumption that a reasonable degree of missing data does not significantly impact ML performance. However, Fig. 6 also highlights substantial variability across participants, with some exhibiting a greater number of outliers than others. This further supports the necessity of personalized models to account for individual differences in physiological responses.

## V. Conclusion

This study examined the impact of missing IBI data on HRV-based drowsiness detection models. The findings confirmed our initial expectations: a reasonable amount of missing data does not significantly degrade the performance of ML models. This could be translated into a real life scenario: suppose we have a system that requires data segments of 1 minute then even if we missed up to 20 seconds of the segment because the driver was hands off then a model would still be reliable. The analysis of KL divergence further reinforced this observation, showing that the overall distribution of HRV features remained relatively stable even with increasing data loss. However, the results also highlighted individual variability, underscoring the need for personalized models tailored to each driver's unique physiological patterns.

Several limitations should be acknowledged. First, the relatively small sample size may constrain the statistical power and the generalizability of the findings. Second, this study focused primarily on quantifying the impact of missing data on classification performance rather than systematically comparing different methods for handling missing data, such as interpolation or estimation techniques. Third, as with most experimental studies in this field, the use of a driving simulator may not fully capture the complex physiological responses that occur under real-world driving conditions. We acknowledge that these factors could affect how well the findings translate to real-world settings.

Future work will focus on refining approaches to handle missing data. Specifically, we plan to investigate interpolation techniques for estimating the missing IBIs [17] and compare their impact against deep learning-based methods such as generative AI [18] and autoencoders [19]. This will allow us to assess how frequency-domain and nonlinear HRV features are affected when using generated values, and to determine their influence on overall model performance. We also aim to enhance classifier performance through hyperparameter optimization and conduct feature importance analysis to identify the most informative HRV features. To improve classification accuracy, we also intend to incorporate temporal dynamics of HRV features over multiple time intervals, rather than relying solely on single-segment data. Furthermore, because the subjectivity of the KSS can be influenced by individual rating tendencies, fatigue accumulation, and temporal variability, we plan on exploring ground truth alternatives such as self-supervised learning to leverage unlabeled data or multimodal approaches that integrate additional physiological signals. These advancements will contribute to more robust and accurate physiological-based 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. The authors are grateful to the driving simulator group and 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] A. Persson, H. Jonasson, I. Fredriksson, U. Wiklund, and C. Ahlström, "Heart Rate Variability for Driver Sleepiness Classification in Real Road Driving Conditions," vol. 2019, 07 2019, pp. 6537–6540.

[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] S. Leonhardt, L. Leicht, and D. Teichmann, "Unobtrusive Vital Sign Monitoring in Automotive Environments—A Review," *Sensors*, vol. 18, no. 9, 2018.

[6] 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.

[7] C. Gouveia, D. F. Albuquerque, P. Pinho, and J. Vieira, "Bio-Radar Cardiac Signal Model Used for HRV Assessment and Evaluation Using Adaptive Filtering," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–10, 2022.

[8] 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.

[9] A. Lourenço, A. P. Alves, C. Carreiras, R. P. Duarte, and A. Fred, "CardioWheel: ECG Biometrics on the Steering Wheel," in *Machine Learning and Knowledge Discovery in Databases*, A. Bifet, M. May, B. Zadrozny, R. Gavalda, D. Pedreschi, F. Bonchi, J. Cardoso, and M. Spiliopoulou, Eds. Cham: Springer International Publishing, 2015, pp. 267–270.

[10] D. Cajal, D. Hernando, J. Lázaro, P. Laguna, E. Gil, and R. Bailón, "Effects of missing data on heart rate variability metrics," *Sensors*, vol. 22, no. 15, 2022.

[11] R. Buendia, F. Forcolin, J. Karlsson, B. A. Sjöqvist, A. Anund, and S. C. and, "Deriving heart rate variability indices from cardiac monitoring—an indicator of driver sleepiness," *Traffic Injury Prevention*, vol. 20, no. 3, pp. 249–254, 2019.

[12] 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.

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

[14] J. Vicente, P. Laguna, A. Bartra, and R. Bailón, "Drowsiness detection using heart rate variability," *Medical & biological engineering & computing*, vol. 54, pp. 927–937, 2016.

[15] A. Meyerson, J. Eklind, F. Fischer, M. Aramrattana, I. Fredriksson, and C. Ahlström, "Effects of daylight and darkness at daytime versus night-time on driver sleepiness: A driving simulator study," *Transportation Research Interdisciplinary Perspectives*, vol. 24, p. 101087, 2024.

[16] R. Zemblys, C. Ahlström, A. Anund, and S. Fine;r, "Towards better subjective sleepiness ground truth data quality," in *9th International Conference on Driver Distraction and Inattention (DDI2024), University of Michigan, Ann Arbor, US, October 22-24, 2024*, 2024.

[17] M. Benchekroun, B. Chevallier, V. Zalc, D. Istrate, D. Lenne, and N. Vera, "The Impact of Missing Data on Heart Rate Variability Features: A Comparative Study of Interpolation Methods for Ambulatory Health Monitoring," *IRBM*, vol. 44, no. 4, p. 100776, 2023.

[18] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Networks," 2014.

[19] G. E. Hinton and R. R. Salakhutdinov, "Reducing the Dimensionality of Data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.