

## Article

# Electrogastrogram-Derived Features for Automated Sickness Detection in Driving Simulator

Grega Jakus <sup>1</sup>, Jaka Sodnik <sup>1</sup> and Nadica Miljković <sup>2,1,\*</sup><sup>1</sup> Faculty of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia<sup>2</sup> School of Electrical Engineering, University of Belgrade, 11000 Belgrade, Serbia

\* Correspondence: nadica.miljkovic@etf.bg.ac.rs

**Abstract:** The rapid development of driving simulators for the evaluation of automated driving experience is constrained by the simulator sickness-related nausea. The electrogastrogram (EGG)-based approach may be promising for immediate, objective, and quantitative nausea assessment. Given the relatively high EGG sensitivity to noises associated with the relatively low amplitude and frequency spans, we introduce an automated procedure comprising statistical analysis and machine learning techniques for EGG-based nausea detection in relation to the noise contamination during automated driving simulation. We calculate the root mean square of EGG amplitude, median and dominant frequencies, magnitude of Power Spectral Density (PSD) at dominant frequency, crest factor of PSD, and spectral variation distribution along with newly introduced parameters: sample and spectral entropy, autocorrelation zero-crossing, and parameters derived from the Poincaré diagram of consecutive EGG samples. Results showed outstanding robustness of sample entropy with moderate robustness of autocorrelation zero-crossing, dominant frequency, and its median. Machine learning reached an accuracy of 88.2% and revealed sample entropy as one of the most relevant and robust parameters, while linear analysis highlighted spectral entropy, spectral variation distribution, and crest factor of PSD. This study clearly indicates the need for customized feature selection in noisy environments, as well as a complementary approach comprising machine learning and statistical analysis for efficient nausea detection.

**Keywords:** automated vehicle; electrogastrography; entropy; driving simulator; machine learning; motion sickness; nausea; noise reduction; random forest



**Citation:** Jakus, G.; Sodnik, J.; Miljković, N. Electrogastrogram-Derived Features for Automated Sickness Detection in Driving Simulator. *Sensors* **2022**, *22*, 8616. <https://doi.org/10.3390/s22228616>

Academic Editor: Juan Pablo Martínez

Received: 29 September 2022

Accepted: 4 November 2022

Published: 8 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Driving simulators are used in various industries to study human behavior, observe their performance and driving skills, and validate new human-machine interfaces in vehicles. They provide a robust, safe, and controllable testing environment, but sometimes cause simulation sickness and other unpleasant sensations [1,2]. The general approach to assessing sickness incorporates both subjective techniques mainly based on questionnaires (such as Motion Sickness Questionnaire or Simulator Sickness Questionnaire) and objective measures commonly comprising physiological recordings (heart rate, body temperature, electrodermal activity, electrogastrogram (EGG), etc.) [3–6]. The main difference between those two approaches is that subjective measures (driver's perception) are administered commonly after the recording session, while physiological measurements can be used for continuous assessment. Although different in nature, for complete practical assessment, a holistic approach comprising both objective and subjective measures is advised [7–9].

Real time or quasi real time continuous assessment with physiological measures is an attractive approach for the evaluation of nausea and simulator sickness phenomenon. However, these measures proved to be prone to a variety of factors either due to the environmental or noise contamination as it has been shown previously for skin temperature, EDA, and EGG [9,10]. Although the measurement of pupil diameter and pupillary rhythm

present an alluring non-contact method and promising approach to study not just sickness, but also emotional responses and cognitive load [11,12], these measures are not directly related to the stomach activity. Moreover, pupil diameter either measured by the eye tracker or infrared camera may have limited measurement agreements [13].

The symptoms of motion sickness include nausea, vomiting, sweating, eyestrain, difficulty focusing, headaches, oculomotor disturbances, disorientation, dizziness, vertigo, and others [14–18]. The most widely reported symptom is nausea [19,20]. Electrogastrogram is used for measuring gastric myoelectrical activity, and as such, it can be used to indirectly assess motion sickness, while it assesses nausea in a more direct manner as suggested by previous extensive research [14–18]. In this paper, we focused mainly on nausea occurrence as a result of simulator sickness being associated with gastric dysrhythmias that can be recorded via electrogastrogram [21]. An EGG-based approach for evaluating nausea occurrence presents a simple, efficient, continuous, and quantitative approach. Parameters derived from the EGG (e.g., mean amplitude, dominant frequency) and their effectiveness in nausea assessment can be observed, among others, in automated driving simulators. EGG amplitude and frequency content are altered as a consequence of nausea.

The most common changes in EGG signal caused by nausea occurrence are the frequency shift towards higher frequency ranges and amplitude increase. Specifically, the EGG signal loses its regularity (rhythm) and becomes more random in the course of nausea manifestation. EGG-derived features such as change of spectral power percentage in frequency ranges, dominant frequency, power of dominant frequency, and averaged amplitude with several less typical features such as crest factor of power spectrum and median frequency are already proposed to assess nausea-related EGG waveform alterations [4,10]. Typically, a combination of these features is used to study EGG changes caused by nausea elicitation. We calculate these customary parameters and at the same time propose novel features to reflect on overall nausea-related changes in EGG. Namely, we propose the adoption of the level of randomness (entropy), autocorrelation zero-crossing, and features derived from the Poincaré plot of consecutive EGG samples in time domain to quantify the self-similarity of the EGG time series and to detect nausea in noisy environments.

Our main goal is to examine in detail methodological aspects of EGG-based parameter usability. This comes with the unique terms of inspecting parameters' robustness to different noise levels as well as with the combined approach comprising both classical statistical analysis and the machine learning technique as proposed earlier for studying nausea with infra-red cameras [12]. Moreover, we introduce completely novel features for studying nausea occurrence (Poincaré plots and entropy). This research could possibly complement current efforts to standardize EGG technique and to make EGG-based evaluation widely adopted for nausea assessment, as well as for studying complex brain–gut interactions [22].

### *1.1. Rationale for Introduction of New EGG-Based Parameters*

The level of randomness is a particularly useful measure in many areas of data analysis. For example, in stock market analysis, approximate entropy as a measure of randomness suggests more data predictability in times of crisis indicating more pronounced repeated patterns [23]. Sample entropy is also a promising method for determining the regularity of signals based on the existence of patterns. The method is similar to approximate entropy, but it is independent of the signal length and has better relative consistency [24]. It was successfully used, for example, to separate uterine electromyogram (EMG) records of term and pre-term delivery groups [25]. On the other side, spectral entropy or the entropy of a signal normalized power distribution can be used to estimate the uniformity of signal power distribution and, as a result, discriminate between narrowband and wideband signals. The method was, for example, used to determine the depth of anesthesia from an electroencephalogram (EEG) [26,27]. For the EGG-based detection of nausea, we used both sample and spectral entropy as we hypothesize that these features would discriminate between less and more random changes in EGG signals corresponding to the baseline recording and nausea occurrence, respectively. In fact, this could fit perfectly into EGG-

based nausea assessment as repeated patterns are more pronounced in the baseline EGG waveform shape consisting mainly of normal gastric rhythm and revealing narrowband nature in comparison to the EGG signal during nausea occurrence.

The autocorrelation zero-crossing is another method for determining the randomness of signals. The method is based on calculating the first zero-crossing of the autocorrelation function: the closer it is to the maximum of the autocorrelation function, the more random the signal. This method has been used successfully in the analysis of the EMG and electrocardiogram (ECG) where it showed less proneness to the signal interferences [25,28]. For that reason, we hypothesize that it could be simply adopted for EGG-based analysis.

Previously, EGG frequency dynamics were expressed by utilizing the number of turning points as a test of randomness to study gastric coupling in animals [29,30]. Here, we adopt a common approach to examine heart rate variability based on standard deviations of a Poincaré plot to study EGG-related nausea alterations. A Poincaré plot allows for the evaluation of non-linear aspects of signal sequences, and it is commonly used in biomedical engineering to characterize heart rate variability [31]. Our hypothesis for the adoption of a Poincaré plot and its adaptation in EGG analysis is based on the premise that changes in EGG signal as a result of nausea occurrence would express more non-linear properties in comparison to the baseline EGG recordings.

### 1.2. Noise Effect on EGG-Based Parameters

Although promising, EGG-based assessment of nausea comes with a major drawback—exaggerated noise. Due to the relatively low amplitude and low frequency content, the EGG signal is easily affected by noises and artifacts, especially by the movement artifacts. This is clearly pronounced in motion-based driving simulators as they come with an interactive and dynamic environment causing EGG data quality to decrease due to the excessive noise [10].

To test the applicability of previously used and newly introduced EGG-based parameters and their robustness to noise levels, we added synthetic colored noise on datasets comprising EGG signals recorded in 20 healthy participants during driving simulation to create semi-synthetic datasets. To the best of our knowledge, this is the first study reporting effect of noise on EGG features for simulator sickness-related nausea assessment.

Moreover, we introduce and demonstrate a new machine learning approach to automatically evaluate nausea occurrence by EGG-based parameters in relation to noise. We report results of both statistical analysis and machine learning by reasoning for a middle ground approach between “data modeling” and “algorithmic modeling” cultures to yield a strong empirical foundation as recommended in [32].

### 1.3. Aims of the Study

Our objective is twofold. We firstly aim at the extensive exploration of known and novel EGG-based parameters for nausea assessment. Secondly, we seek features with proven higher levels of robustness to noises and artifacts that would be more appropriate for nausea detection in dynamic driving simulator environments. To achieve our aims, we employ automated techniques for the extraction of proposed EGG-based features with various Signal-to-Noise Ratios (SNRs) with added synthetic noise and contribute to the existing body of knowledge in the following ways:

1. We present an extended list of EGG-based features for nausea assessment following pertinent reasoning for their calculation.
2. We report on the sensitivity/robustness of the proposed EGG-based parameters to different levels of SNRs and the noise effect on nausea detection.

## 2. Materials and Methods

We introduce an extensive list of EGG-based parameters for nausea detection and for the evaluation of parameter robustness to different levels of SNRs. We use both traditional

statistical linear analysis and non-linear machine learning to test the usability of the selected EGG features for the detection of nausea incidence.

### 2.1. Available EGG Data and Recording Procedure

The data for the analysis are obtained from the study by Gruden et al. [4]. In the continuation, we summarize the most important information on the data collection required to explain the main goals of the present research.

The study was conducted at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia. Twenty individuals (two females), mostly students or staff from the Faculty of Electrical Engineering in Ljubljana participated in the study. The participants were between 19 and 40 years old, had a valid driving license, and had more than one year of driving experience. They were instructed to fast for at least 6 h and not to drink for at least 2 h before the study [33].

The study was performed in the Nervtech driving simulator (Nervtech d.o.o., Trzin, Slovenia) [34,35] with a motion platform with 4 degrees of freedom (yaw, pitch, roll, and heave) (Figure 1). The cockpit consisted of an adjustable car seat, a Fanatec ClubSport Wheel Base V2 steering wheel with dynamic feedback, and a Fanatec ClubSport Pedals V3 pedal set with three pedals (both from Fanatec, Endor AG, Landshut, Germany) [36,37]. The driving environment was simulated using the SCANeR simulation software (AV simulation, Boulogne, France) [38], a virtual reality headset (Oculus, Facebook Technologies LLC, Menlo Park, CA, USA) [39], and a stereo speaker set. Based on the definition of physical and functional fidelity as defined by Kinkade and Wheaton [40] and Hays [41], the simulator used in this study can be described as a high-fidelity driving simulator. After the participants were introduced to the experiment procedure, they signed informed consents in accordance with the Declaration of Helsinki and University Code of Ethics. The participants then completed a test trial to get acquainted with the experimental environment. They were instructed to raise their hand to stop the experiment at any time if sickness was too severe to be able to continue. The main part of the experiment was divided into three parts:

1. Baseline measurement before the driving simulation.
2. Driving simulation in autonomous vehicle.
3. EGG measurement while resting after driving simulation.



**Figure 1.** Illustration of the Nervtech driving simulator interior. Image is distributed under the Creative Commons Attribution License and taken from the study by Gruden et al. [4].

The driving simulation consisted of a less dynamic drive, which took part on a highway road and a more dynamic drive, which took part on a countryside road. The duration of each driving scene was approximately 7 min. The screenshots of the highway and countryside roads are given in Figure 2. The user study had a within-subject design as the participants' state was measured before, during, and after AV driving simulation.



**Figure 2.** Screenshots of the driving scenarios employed in this study. (Left) highway with low traffic; (Right) country road with many vehicles and icy roads. Image is distributed under the Creative Commons Attribution License and taken from the study by Gruden et al. [4].

In order to induce higher levels of sickness and at the same time to reduce the potential and exaggerated EGG artifacts originating from motion or steering, fully autonomous driving was used (SAE level 5) [42]. The participants' primary task was, therefore, only to observe the autonomous driving.

Five Ag/AgCl surface electrodes (H92SG, Kendall/Covidien, Dublin, Ireland) were placed on the participants' stomach to measure three-channel EGG following the recommendations [33,43,44]. The EGG signal was recorded using an amplification and filtering device [10] and digitized using Biopac UIM100C MP150 Analog-to-Digital Converter (ADC) (Biopac Systems, Goleta, CA, USA) [45]. The sampling frequency was set to 2 Hz and the resolution was 16 bits.

## 2.2. EGG Preprocessing and Creation of Semi-Synthetic EGG Dataset

Preprocessing and feature extraction are performed offline in Matlab version R2019b (The Mathworks, Natick, MA, USA). For feature extraction, we used four segments corresponding to the baseline recording, countryside driving, highway driving, and EGG measurement after driving simulation while resting. Furthermore, participants were instructed to continuously report any experience of nausea or sickness by pressing a button, which was recorded alongside EGG data. As the frequency of EGG slow waves is approximately 3 cycles per minute, and therefore, a large segment of data is required to reliably detect the parameters, the nausea incidence variable was set for each segment to 1 for a segment if at least one button press was detected in that segment; otherwise, it was set to 0.

In the previous study examining noise effect on EGG analysis [46], SNRs from  $-50$  dB to  $15$  dB with step  $5$  dB were used, but the majority of changes are seen from  $-20$  dB to  $15$  dB. Hence, we created a semi-synthetic dataset by adding pseudo-random noise matching SNR values of  $-20$  dB,  $-10$  dB,  $0$  dB,  $10$  dB, and  $20$  dB. We report on exact mean SNRs as we are not able to keep SNRs at a constant level due to the application of a pseudo-random generator and the fact that noise power was set before the filtering which removed the majority of the wideband noise. The actual obtained SNRs in this study are  $-23$  dB,  $-13$  dB,  $-3$  dB,  $7$  dB, and  $17$  dB.

Semi-synthetic EGG signals are digitally filtered with the 6th order zero-phase-distortion Butterworth bandpass filter with cutoff frequencies set at 1 cycle per minute (cpm) and 10 cpm ( $0.0167$  Hz and  $0.167$  Hz). This process colored the noise by filtering it in a range from 1 cpm to 10 cpm. Due to the relatively low EGG amplitude (in the range of  $\mu$ V), the raw EGG signal is very sensitive to noises and motion artifacts. The recorded EGG signals are therefore inspected by an experienced researcher who removed parts with identified

motion artifacts. Following an established procedure, these artifacts were identified as short, high-amplitude spikes in contrast to a slowly changing, low-amplitude signal representing gastric activity [4,10,22,47]. From the three measured EGG channels, the channel with the fewest artifacts is selected for further analysis. As 3 subjects are excluded as a consequence of low signal quality, the remaining analysis is performed in 17 subjects [4]. Succeeding feature extraction and nausea detection procedures are completely automated.

### 2.3. Automated Procedure for EGG-Based Features Extraction

All extracted EGG-based parameters are presented in Table 1. For the calculation of the level of randomness, we use Sample Entropy (SampEnt) of time series (SampEntT) and of Power Spectral Density (PSD) (SampEntP), as well as Spectral Entropy (SpectEnt). Additionally, we calculate the autocorrelation zero-crossing, as well as geometrical features corresponding to the Poincaré plots (SD1, SD2, and SDEGG in Table 1).

**Table 1.** 17 EGG-based features calculated from EGG segments recorded during a simulated drive in 17 subjects (SampEntT and SampEntP are separately evaluated for embedding dimensions  $m = 2, 3,$  and  $4$ ). The number of observations for all four segments is  $17 \times 4 = 68$  and the total number of features for 17 subjects is  $17 \times 68 = 1156$ . PSD stands for Power Spectral Density.

Feature	Explanation	Unit	References
RMS (Root Mean Square)	RMS of the amplitude of EGG selected segment	$\mu\text{V}$	[4,10]
median	Median frequency of PSD of selected EGG segment	cpm (cycles per minute)	[4,10]
DF (Dominant Frequency)	Dominant frequency of PSD of selected EGG segment	cpm	[4,10,48]
MagDF	Magnitude of DF in PSD of selected EGG segment	$\text{mV}^2/\text{Hz}$	[4,10,48]
CS (Crest Factor)	CS of PSD of selected EGG segments	/	[4,10]
SDV (Spectral Variation Distribution)	PSD with magnitude higher than 25% of DF	%	[4]
SampEntT (Sample Entropy of Time Series)	Embedding dimensions $m = 2, 3,$ and $4$	/	Introduced here and inspired by [49]
SampEntP (Sample Entropy of PSD)	Embedding dimensions $m = 2, 3,$ and $4$	/	
SpectEnt (Spectral Entropy)	/	/	
Autocorr (Autocorrelation zero-crossing)	The first lag of autocorrelation function of EGG at which autocorrelation equals 0	S	Introduced here and inspired by [46]
SD1	Transverse line of the Poincaré plot in the perpendicular direction. A Poincaré plot presents a scatter plot of the current EGG sample in relation to the prior EGG sample.	$\mu\text{V}$	Introduced here and inspired by [31,50]
SD2	Longitudinal line of the Poincaré plot in the perpendicular direction.	$\mu\text{V}$	
SDEGG	Standard deviation of EGG samples obtained from the SD1 and SD2.	$\mu\text{V}$	

By visual inspection of SampEntT for embedding dimensions  $m = 2, 3,$  and  $4$ , we conclude that there are two distinct groups with entropy higher and lower than 10. We

use this information as empirical reasoning for splitting the data into two groups and for transforming SampEntT into categorical data for further analysis.

#### 2.4. Statistical Analysis and Machine Learning Approach

For statistical analysis and machine learning, we use R programming language v4.1.2 [51] in R Studio environment (R Studio, Inc., Boston, MA, USA) with the following R packages from the Comprehensive R Archive Network (CRAN): dplyr [52], e1071 [53], caret [54], randomForest with randomForestExplainer [55], ggplot2 [56], effsize [57], and pROC [58]. For the sake of computational reproducibility, we shared R code and tables with EGG-based parameters on Zenodo repository [59].

To observe the impact of noise on the selected set of parameters, we statistically compare the values of calculated parameters in different SNR conditions. For normally distributed data, paired sample *t*-test is used to evaluate the existence of statistically significant differences between original and noisy EGG-based parameters for all SNRs, while Cohen's *d* (*Cd*) is used to estimate the effect size. For non-normally distributed data, paired Wilcoxon's Signed-Ranks test is used along with Cliff's *delta* (*Cdelta*) as effect size measure. To test normality of the data, we use Shapiro–Wilk's normality test with *p* set to 0.05.

Statistically significant differences among parameters corresponding to the nausea occurrence with those revealing the non-nausea occurrence are explored by *t*-test for normally distributed data, while non-parametric Wilcoxon–Mann–Whitney's U test is used to compare dependent variables for two independent groups for non-normally distributed data. Here, we also calculated *Cd* and *Cdelta* for normally and non-normally distributed data, respectively.

Summary statistics for categorical entropy parameters (SampEntT) in relation to nausea occurrence is reported in conjunction with Pearson's Chi square test with a simulated *p* value that is used to explore the significant correlation of nausea incidence within the categorical features. If not stated otherwise, the level for statistical significance is set at 0.05.

Binary Random Forest (RF) algorithm is constructed with the aim of classifying EGG-based parameters obtained from EGG data recorded with and without nausea occurrence. The choice of hyperparameters is based on the previous publication on a similar dataset classifying the physiological data [50] and on the Breiman and Cutler's Random Forests for Classification and Regression [60,61]. The seed is set to 100 for the sake of reproducibility. The rationale for RF selection is that previous studies focused on the exploration of noise effect on classification performance, which showed that RF is relatively resistant to noise probably due to the bagging ensemble procedure [62,63]. All parameters from Table 1 are used as input to the RF classifier. We perform leave-one-out cross-validation on the training data as this type of validation is more suitable for datasets with a smaller number of instances [64]. The data are split into training (75%) and test sets (25%) taking into account the distribution balance within the splits by createDataPartition Caret procedure [54] to ensure that both training and test sets are representative of the dataset. Furthermore, we use the automated Caret procedure for tuning the classifier parameters.

The Caret procedure for RF resampling across tuning parameters resulted in the determination of optimal parameters termed *mtry* that corresponds to the number of variables randomly sampled at each data split. Accuracy is used by the automatic procedure to determine the optimal *mtry* by using the largest value of accuracy. For parameters obtained from noisy data, we apply two types of RF classifiers: (1) RF trained and tested on noisy data and (2) RF trained on original and tested on noisy data. Together with reported parameters for machine learning (ML) evaluation (Kappa, confidence interval for 95%, accuracy, sensitivity, specificity, precision, AUC (Area Under the Receiver Operating Curve) for both training and tests sets, and recall), we present feature importance plots. Features are ranked according to the score obtained by the sum of the number of times the feature is selected by all trees in created binary RF.

### 3. Results

Table 2 presents the results of the statistical analysis for the EGG-based parameters obtained on noisy data with different SNR levels in comparison to the parameters without synthetically added colored noise.

**Table 2.** Results of paired *t*-tests and Cohen’s *d* (for normally distributed data), as well as paired Wilcoxon’s Signed-Ranks tests and Cliff’s *delta* (for not normally distributed data) for comparison of non-noisy and noisy EGG-based parameters are presented. For testing normality, *p* was set at 0.05. Values with *p* > 0.05 (no significant difference due to the noise is observed, i.e., EGG-based features are robust to the noise levels) are highlighted in bold. SNR stands for Signal-to-Noise Ratio, *V* stands for V-statistics, *p* for probability, *Cdelta* for Cliff’s *delta*, *Cd* for Cohen’s *d*, *t* for *t*-statistics, and *df* for degrees of freedom.

Features	SNR = 17 dB	SNR = 7 dB	SNR = −3 dB	SNR = −13 dB	SNR = −23 dB
RMS	<i>V</i> = 321, <i>p</i> < 0.001, <i>Cdelta</i> = 0.021	<i>V</i> = 77, <i>p</i> < 0.001, <i>Cdelta</i> = 0.092	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = 0.344	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = 0.733	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = 0.951
median	<b><i>V</i> = 188.5, <i>p</i> = 0.750, <i>Cdelta</i> = 0.002</b>	<b><i>t</i> = −1.761, <i>df</i> = 67, <i>p</i> = 0.083, <i>Cd</i> = −0.060</b>	<b><i>t</i> = −2.852, <i>df</i> = 67, <i>p</i> = 0.006, <i>Cd</i> = −0.265</b>	<b><i>t</i> = −4.5567, <i>df</i> = 67, <i>p</i> &lt; 0.001, <i>Cd</i> = −0.652</b>	<i>V</i> = 460, <i>p</i> < 0.001, <i>Cdelta</i> = −0.457
MagDF	<i>V</i> = 739, <i>p</i> = 0.008, <i>Cdelta</i> = −0.007	<i>V</i> = 812, <i>p</i> = 0.028, <i>Cdelta</i> = −0.031	<i>V</i> = 206, <i>p</i> < 0.001, <i>Cdelta</i> = −0.254	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.653	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.919
DF	<b><i>V</i> = 64, <i>p</i> = 0.476, <i>Cdelta</i> = 0.007</b>	<b><i>V</i> = 240, <i>p</i> = 0.461, <i>Cdelta</i> = −0.061</b>	<b><i>V</i> = 508.5, <i>p</i> = 0.730, <i>Cdelta</i> = −0.034</b>	<i>V</i> = 647, <i>p</i> = 0.009, <i>Cdelta</i> = −0.253	<i>V</i> = 845, <i>p</i> = 0.045, <i>Cdelta</i> = −0.176
CS	<i>V</i> = 1591, <i>p</i> = 0.011, <i>Cdelta</i> = 0.031	<i>V</i> = 1688, <i>p</i> = 0.002, <i>Cdelta</i> = 0.098	<i>V</i> = 1872, <i>p</i> < 0.001, <i>Cdelta</i> = 0.246	<i>V</i> = 2108, <i>p</i> < 0.001, <i>Cdelta</i> = 0.533	<i>V</i> = 2136, <i>p</i> < 0.001, <i>Cdelta</i> = 0.546
SDV	<b><i>V</i> = 495, <i>p</i> = 0.342, <i>Cdelta</i> = −0.012</b>	<i>V</i> = 335, <i>p</i> < 0.001, <i>Cdelta</i> = −0.135	<i>V</i> = 344.5, <i>p</i> < 0.001, <i>Cdelta</i> = −0.344	<i>V</i> = 173.5, <i>p</i> < 0.001, <i>Cdelta</i> = −0.674	<i>V</i> = 97.5, <i>p</i> < 0.001, <i>Cdelta</i> = −0.764
SampEntT_m2	<i>V</i> = 920, <i>p</i> = 0.46, <i>Cdelta</i> = 0.031	<b><i>V</i> = 610, <i>p</i> = 0.256, <i>Cdelta</i> = −0.052</b>	<b><i>V</i> = 1213, <i>p</i> = 0.249, <i>Cdelta</i> = 0.085</b>	<i>V</i> = 1918, <i>p</i> < 0.001, <i>Cdelta</i> = 0.413	<i>V</i> = 2330, <i>p</i> < 0.001, <i>Cdelta</i> = 0.754
SampEntT_m3	<b><i>V</i> = 626.5, <i>p</i> = 0.350, <i>Cdelta</i> = 0.067</b>	<b><i>V</i> = 444, <i>p</i> = 0.556, <i>Cdelta</i> = −0.018</b>	<b><i>V</i> = 787, <i>p</i> = 0.083, <i>Cdelta</i> = 0.136</b>	<i>V</i> = 1666, <i>p</i> < 0.001, <i>Cdelta</i> = 0.410	<i>V</i> = 2285, <i>p</i> < 0.001, <i>Cdelta</i> = 0.682
SampEntT_m4	<b><i>V</i> = 254, <i>p</i> = 0.914, <i>Cdelta</i> = 0.039</b>	<b><i>V</i> = 221, <i>p</i> = 0.948, <i>Cdelta</i> = 0.018</b>	<b><i>V</i> = 437, <i>p</i> = 0.338, <i>Cdelta</i> = 0.101</b>	<i>V</i> = 1598, <i>p</i> < 0.001, <i>Cdelta</i> = 0.419	<i>V</i> = 2209, <i>p</i> < 0.001, <i>Cdelta</i> = 0.671
SampEntP_m2	<b><i>V</i> = 1289, <i>p</i> = 0.480, <i>Cdelta</i> = 0.030</b>	<i>V</i> = 1349, <i>p</i> = 0.284, <i>Cdelta</i> = 0.040	<b><i>V</i> = 1278, <i>p</i> = 0.523, <i>Cdelta</i> = −0.006</b>	<b><i>V</i> = 1373, <i>p</i> = 0.223, <i>Cdelta</i> = 0.030</b>	<b><i>V</i> = 1221, <i>p</i> = 0.772, <i>Cdelta</i> = 0.008</b>
SampEntP_m3	<b><i>V</i> = 1306, <i>p</i> = 0.418, <i>Cdelta</i> = 0.015</b>	<b><i>V</i> = 1252, <i>p</i> = 0.631, <i>Cdelta</i> = 0.006</b>	<b><i>V</i> = 1275, <i>p</i> = 0.535, <i>Cdelta</i> = −0.013</b>	<b><i>V</i> = 1369, <i>p</i> = 0.232, <i>Cdelta</i> = 0.083</b>	<b><i>V</i> = 1086, <i>p</i> = 0.597, <i>Cdelta</i> = −0.042</b>
SampEntP_m4	<b><i>V</i> = 1293, <i>p</i> = 0.465, <i>Cdelta</i> = 0.008</b>	<b><i>V</i> = 1344, <i>p</i> = 0.297, <i>Cdelta</i> = −0.009</b>	<b><i>V</i> = 1417, <i>p</i> = 0.137, <i>Cdelta</i> = 0.020</b>	<b><i>V</i> = 1385, <i>p</i> = 0.196, <i>Cdelta</i> = 0.047</b>	<b><i>V</i> = 955, <i>p</i> = 0.184, <i>Cdelta</i> = −0.154</b>
SpectEnt	<i>V</i> = 727, <i>p</i> = 0.006, <i>Cdelta</i> = −0.019	<i>V</i> = 340, <i>p</i> < 0.001, <i>Cdelta</i> = −0.154	<i>V</i> = 180, <i>p</i> < 0.001, <i>Cdelta</i> = −0.416	<i>V</i> = 156, <i>p</i> < 0.001, <i>Cdelta</i> = −0.695	<i>V</i> = 161, <i>p</i> < 0.001, <i>Cdelta</i> = −0.724
Autocorr	<b><i>V</i> = 30, <i>p</i> = 0.351, <i>Cdelta</i> = 0.024</b>	<b><i>V</i> = 27, <i>p</i> = 0.608, <i>Cdelta</i> = 0.027</b>	<i>V</i> = 540, <i>p</i> < 0.001, <i>Cdelta</i> = 0.245	<i>V</i> = 1066, <i>p</i> < 0.001, <i>Cdelta</i> = 0.492	<i>V</i> = 1215, <i>p</i> < 0.001, <i>Cdelta</i> = 0.557
SD1	<i>V</i> = 205, <i>p</i> < 0.001, <i>Cdelta</i> = −0.021	<i>V</i> = 19, <i>p</i> < 0.001, <i>Cdelta</i> = −0.096	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.384	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.766	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.958
SD2	<i>V</i> = 327, <i>p</i> < 0.001, <i>Cdelta</i> = −0.022	<i>V</i> = 80, <i>p</i> < 0.001, <i>Cdelta</i> = −0.093	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.345	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.733	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.951
SDEGG	<i>V</i> = 321, <i>p</i> < 0.001, <i>Cdelta</i> = −0.021	<i>V</i> = 76, <i>p</i> < 0.001, <i>Cdelta</i> = −0.093	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.344	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.733	<i>V</i> = 0, <i>p</i> < 0.001, <i>Cdelta</i> = −0.951

Comparative results of binary RF classification for original and noisy data are shown in Table 3. Evaluation ML parameters for classifiers trained on the original dataset and tested in noisy data are presented in Table 4. Results for RF classifiers with categorical SampEntT parameters for  $m = 2, 3$ , and 4 did not influence the results, so we decided not to present them.

**Table 3.** ML evaluation parameters for binary RF classification of nausea for original and noisy data. The best results are presented in bold. SNR stands for Signal-to-Noise Ratio, CI stands for Confidence Interval, and AUC stands for Area Under the Curve. Two AUC values are presented for the validation set and for the test set <sup>1</sup>.

Evaluation Classifier Metrics	Original Dataset	Noisy Data				
		SNR = 17 dB	SNR = 7 dB	SNR = −3 dB	SNR = −13 dB	SNR = −23 dB
Kappa	<b>0.452</b>	<b>0.452</b>	0.301	<b>0.452</b>	−0.214	−0.097
95% CI	<b>(0.636, 0.985)</b>	<b>(0.636, 0.985)</b>	(0.566, 0.962)	<b>(0.636, 0.985)</b>	(0.383, 0.858)	(0.501, 0.932)
Accuracy	<b>0.882</b>	<b>0.882</b>	0.823	<b>0.882</b>	0.647	0.765
Sensitivity	<b>1.000</b>	<b>1.000</b>	0.929	<b>1.000</b>	0.786	0.929
Specificity	<b>0.333</b>	<b>0.333</b>	<b>0.333</b>	<b>0.333</b>	0	0
Precision	<b>0.875</b>	<b>0.875</b>	0.867	<b>0.875</b>	0.786	0.812
Recall	<b>1.000</b>	<b>1.000</b>	0.929	<b>1.000</b>	0.786	0.929
AUC (training)	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>
AUC (test)	<b>0.667</b>	<b>0.667</b>	0.631	<b>0.667</b>	0.393	0.464

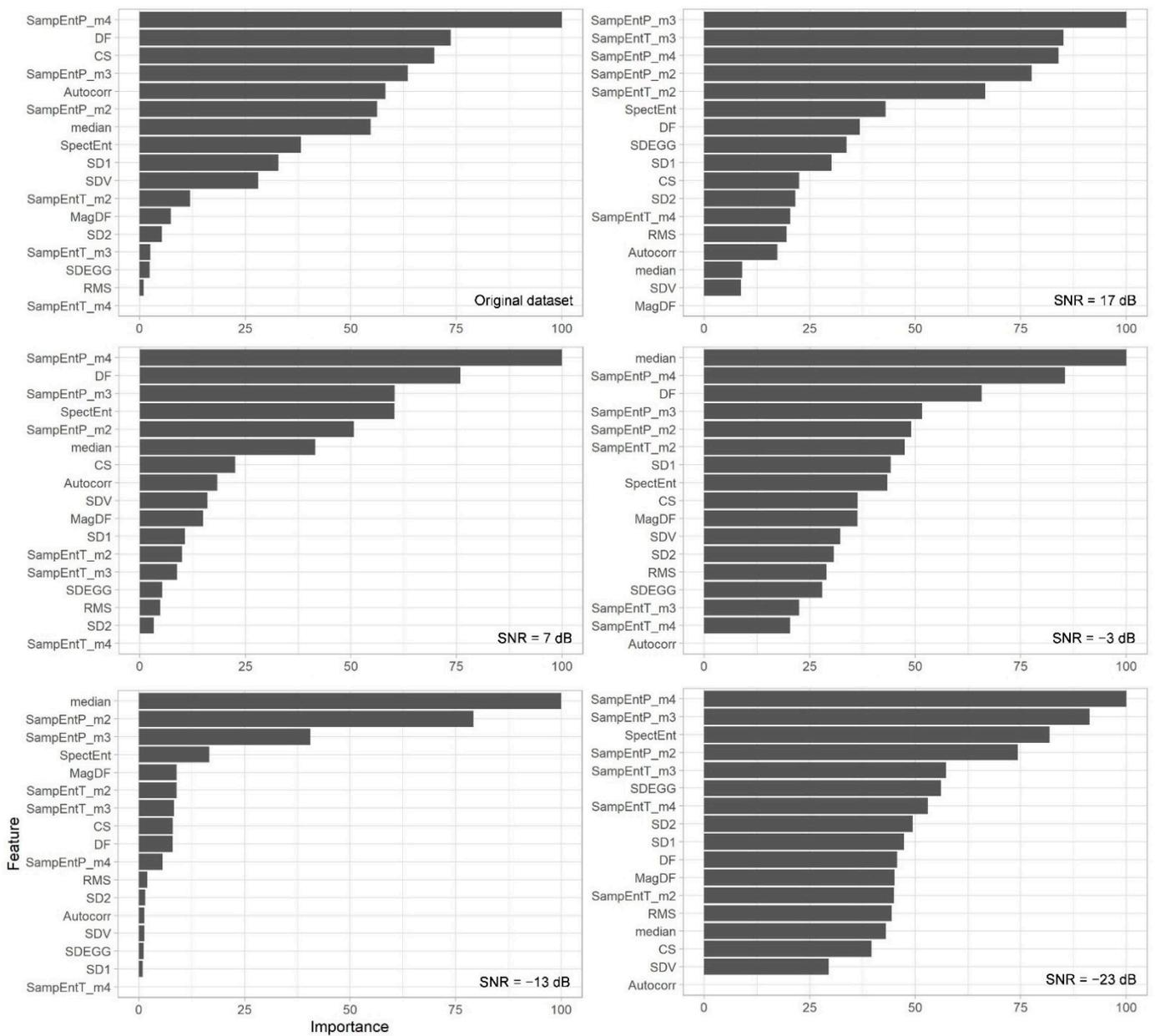
<sup>1</sup> As suggested by reviewer and literature [64], we applied leave-one-out cross validation. However, the leave-one-out validation performed only slightly better than 10-fold cross-validation (SNR = −13 dB in Table 3 and SNR = −23 dB in Table 4). This may be the result of similar instances in a dataset which did not have larger effect on separate models (in cases when similar instances are omitted). The discussion on selection of cross-validation procedures is out of scope of this paper and it has been discussed elsewhere [65].

**Table 4.** ML evaluation parameters for binary RF trained and validated on original dataset, and further tested on noisy data with different noise levels. The best results are presented in bold. SNR stands for Signal-to-Noise Ratio, CI stands for Confidence Interval, and AUC stands for Area Under the Curve. Two AUC values are presented for the validation set and for the test set.

Evaluation Classifier Metrics	Noisy Test Data				
	SNR = 17 dB	SNR = 7 dB	SNR = −3 dB	SNR = −13 dB	SNR = −23 dB
Kappa	<b>0.452</b>	<b>0.452</b>	<b>0.452</b>	0	0
95% CI	<b>(0.636, 0.985)</b>	<b>(0.636, 0.985)</b>	<b>(0.636, 0.985)</b>	(0.566, 0.962)	(0.566, 0.962)
Accuracy	<b>0.882</b>	<b>0.882</b>	<b>0.882</b>	0.823	0.823
Sensitivity	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	1.000	1.000
Specificity	<b>0.333</b>	<b>0.333</b>	<b>0.333</b>	0	0
Precision	<b>0.875</b>	<b>0.875</b>	<b>0.875</b>	0.823	0.823
Recall	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	1.000	1.000
AUC (training)	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>	<b>0.616</b>
AUC (test)	<b>0.667</b>	<b>0.667</b>	<b>0.667</b>	0.500	0.500

For the original dataset, the  $m_{try} = 17$  is selected by automatic tuning procedure with an accuracy of 0.790. For noisy datasets,  $m_{try}$  is set to 2 with an accuracy of 0.782, 17 with an accuracy of 0.827, 2 with an accuracy of 0.827, 17 with an accuracy of 0.827, and 2 with an accuracy of 0.810 for SNRs of −23 dB, −13 dB, −3 dB, 7 dB, and 17 dB, respectively.

Importance plots for RF applied on parameters obtained from noisy and original EGG data are presented in Figure 3.



**Figure 3.** Importance plots for original and noisy datasets with different SNRs as a result of RF binary classifier for nausea detection. SNR stands for Signal-to-Noise Ratio.

The results of the *t*-tests for normally distributed data and of the Wilcoxon–Mann–Whitney’s U tests for non-normally distributed data are given in Table 5. Table 6 presents summary statistics for categorical SampEntT EGG-based parameters in relation to nausea occurrence.

**Table 5.** Results of statistical tests for the comparison of parameters with and without nausea for non-noisy and noisy data with different noise levels. For testing normality,  $p$  is set at 0.05. Statistically significant differences are presented in bold ( $p < 0.05$ ). SNR stands for Signal-to-Noise Ratio,  $W$  stands for  $W$ -statistics,  $p$  for probability,  $Cdelta$  for Cliff's  $delta$ ,  $Cd$  for Cohen's  $d$ ,  $t$  for  $t$ -statistics, and  $df$  for degrees of freedom.

Features	Original	SNR = 17 dB	SNR = 7 dB	SNR = -3 dB	SNR = -13 dB	SNR = -23 dB
RMS	$W = 245$ , $p = 0.145$ , $Cdelta = -0.270$	$W = 247$ , $p = 0.154$ , $Cdelta = -0.265$	$W = 243$ , $p = 0.137$ , $Cdelta = -0.277$	$W = 256$ , $p = 0.201$ , $Cdelta = -0.238$	$W = 297$ , $p = 0.536$ , $Cdelta = -0.116$	$W = 299$ , $p = 0.557$ , $Cdelta = -0.110$
median	$t = -0.8408$ , $df = 19.075$ , $p = 0.411$ , $Cd = -0.232$	$W = 268$ , $p = 0.277$ , $Cdelta = -0.202$	$t = -1.7105$ , $df = 18.665$ , $p = 0.104$ , $Cd = -0.480$	$t = -1.7556$ , $df = 14.249$ , $p = 0.101$ , $Cd = -0.641$	$t = -2.97$ , $df = 18.279$ , $p = 0.008$ , $Cd = -0.846$	$W = 246$ , $p = 0.150$ , $Cdelta = -0.268$
MagDF	$W = 261$ , $p = 0.231$ , $Cdelta = -0.223$	$W = 263$ , $p = 0.243$ , $Cdelta = -0.217$	$W = 263$ , $p = 0.243$ , $Cdelta = -0.217$	$W = 263$ , $p = 0.243$ , $Cdelta = -0.217$	$W = 301$ , $p = 0.579$ , $Cdelta = -0.104$	$W = 317$ , $p = 0.766$ , $Cdelta = -0.056$
DF	$W = 319.5$ , $p = 0.796$ , $Cdelta = -0.049$	$W = 293.5$ , $p = 0.498$ , $Cdelta = -0.126$	$W = 245.5$ , $p = 0.147$ , $Cdelta = -0.269$	<b><math>W = 191</math></b> , <b><math>p = 0.020</math></b> , <b><math>Cdelta = -0.431</math></b>	$W = 296.5$ , $p = 0.530$ , $Cdelta = -0.117$	$W = 315.5$ , $p = 0.747$ , $Cdelta = -0.061$
CS	<b><math>W = 469</math></b> , <b><math>p = 0.033</math></b> , <b><math>Cdelta = 0.396</math></b>	<b><math>W = 468</math></b> , <b><math>p = 0.034</math></b> , <b><math>Cdelta = 0.393</math></b>	$W = 408$ , $p = 0.250$ , $Cdelta = 0.214$	$W = 377$ , $p = 0.515$ , $Cdelta = 0.122$	$W = 343$ , $p = 0.917$ , $Cdelta = 0.021$	$W = 385$ , $p = 0.435$ , $Cdelta = 0.146$
SDV	$t = -2.7527$ , $df = 36.441$ , $p = 0.009$ , $Cd = -0.559$	$W = 252$ , $p = 0.179$ , $Cdelta = -0.250$	$W = 266$ , $p = 0.263$ <sup>2</sup> , $Cdelta = -0.208$	$W = 353$ , $p = 0.791$ , $Cdelta = 0.050$	$W = 287.5$ , $p = 0.439$ , $Cdelta = -0.144$	$W = 308.5$ , $p = 0.663$ , $Cdelta = -0.082$
SampEntT_m2	$W = 421$ , $p = 0.174$ , $Cdelta = 0.253$	<b><math>W = 486.5</math></b> , <b><math>p = 0.016</math></b> , <b><math>Cdelta = 0.448</math></b>	$W = 443.5$ , $p = 0.085$ , $Cdelta = 0.320$	$W = 411.5$ , $p = 0.228$ , $Cdelta = 0.225$	$W = 393.5$ , $p = 0.359$ , $Cdelta = 0.171$	$W = 372$ , $p = 0.568$ , $Cdelta = 0.107$
SampEntT_m3	$W = 415.5$ , $p = 0.198$ , $Cdelta = 0.237$	<b><math>W = 475</math></b> , <b><math>p = 0.025</math></b> , <b><math>Cdelta = 0.414</math></b>	$W = 440$ , $p = 0.091$ , $Cdelta = 0.309$	$W = 406.5$ , $p = 0.259$ , $Cdelta = 0.210$	$W = 421$ , $p = 0.174$ , $Cdelta = 0.253$	$W = 388$ , $p = 0.407$ , $Cdelta = 0.155$
SampEntT_m4	$W = 420$ , $p = 0.136$ , $Cdelta = 0.250$	<b><math>W = 459.5</math></b> , <b><math>p = 0.034</math></b> , <b><math>Cdelta = 0.367</math></b>	$W = 446$ , $p = 0.054$ , $Cdelta = 0.327$	$W = 399.5$ , $p = 0.293$ , $Cdelta = 0.189$	$W = 382$ , $p = 0.464$ , $Cdelta = 0.137$	$W = 370$ , $p = 0.590$ , $Cdelta = 0.101$
SampEntP_m2	$W = 413$ , $p = 0.218$ , $Cdelta = 0.229$	$W = 438$ , $p = 0.102$ , $Cdelta = 0.303$	$W = 406$ , $p = 0.263$ , $Cdelta = 0.208$	$W = 420$ , $p = 0.179$ , $Cdelta = 0.250$	<b><math>W = 504</math></b> , <b><math>p = 0.007</math></b> , <b><math>Cdelta = 0.500</math></b>	$W = 256$ , $p = 0.201$ , $Cdelta = -0.238$
SampEntP_m3	$W = 408$ , $p = 0.250$ , $Cdelta = 0.214$	$W = 404$ , $p = 0.277$ , $Cdelta = 0.202$	$W = 401$ , $p = 0.299$ , $Cdelta = 0.193$	$W = 404$ , $p = 0.277$ , $Cdelta = 0.202$	<b><math>W = 525</math></b> , <b><math>p = 0.002</math></b> , <b><math>Cdelta = 0.562</math></b>	$W = 273$ , $p = 0.315$ , $Cdelta = -0.187$
SampEntP_m4	$W = 410$ , $p = 0.237$ , $Cdelta = 0.220$	$W = 416$ , $p = 0.201$ , $Cdelta = 0.238$	$W = 442$ , $p = 0.090$ , $Cdelta = 0.315$	<b><math>W = 466</math></b> , <b><math>p = 0.037</math></b> , <b><math>Cdelta = 0.387</math></b>	$W = 427$ , $p = 0.145$ , $Cdelta = 0.271$	$W = 339$ , $p = 0.968$ , $Cdelta = 0.009$
SpectEnt	<b><math>W = 172</math></b> , <b><math>p = 0.008</math></b> , <b><math>Cdelta = -0.488</math></b>	<b><math>W = 175</math></b> , <b><math>p = 0.010</math></b> , <b><math>Cdelta = -0.479</math></b>	<b><math>W = 160</math></b> , <b><math>p = 0.005</math></b> , <b><math>Cdelta = -0.524</math></b>	$W = 226$ , $p = 0.078$ , $Cdelta = -0.327$	$t = -2.032$ , $df = 17.237$ , $p = 0.058$ , $Cd = -0.606$	$t = -2.055$ , $df = 17.409$ , $p = 0.055$ , $Cd = -0.608$
Autocorr	$W = 439$ , $p = 0.084$ , $Cdelta = 0.306$	$W = 438.5$ , $p = 0.082$ , $Cdelta = 0.305$	$W = 447.5$ , $p = 0.060$ , $Cdelta = 0.332$	<b><math>W = 446.5</math></b> , <b><math>p = 0.047</math></b> , <b><math>Cdelta = 0.329</math></b>	$W = 399$ , $p = 0.261$ , $Cdelta = 0.187$	$W = 369$ , $p = 0.548$ , $Cdelta = 0.098$
SD1	$W = 232$ , $p = 0.096$ , $Cdelta = -0.309$	$W = 232$ , $p = 0.096$ , $Cdelta = -0.309$	$W = 231$ , $p = 0.093$ , $Cdelta = -0.312$	$W = 249$ , $p = 0.164$ , $Cdelta = -0.259$	$W = 285$ , $p = 0.417$ , $Cdelta = -0.152$	$W = 294$ , $p = 0.504$ , $Cdelta = -0.125$
SD2	$W = 245$ , $p = 0.145$ , $Cdelta = -0.271$	$W = 246$ , $p = 0.150$ , $Cdelta = -0.268$	$W = 243$ , $p = 0.137$ , $Cdelta = -0.277$	$W = 257$ , $p = 0.207$ , $Cdelta = -0.235$	$W = 298$ , $p = 0.546$ , $Cdelta = -0.113$	$W = 299$ , $p = 0.557$ , $Cdelta = -0.110$
SDEGG	$W = 245$ , $p = 0.145$ , $Cdelta = -0.271$	$W = 247$ , $p = 0.154$ , $Cdelta = -0.265$	$W = 243$ , $p = 0.137$ , $Cdelta = -0.277$	$W = 256$ , $p = 0.201$ , $Cdelta = -0.238$	$W = 297$ , $p = 0.536$ , $Cdelta = -0.116$	$W = 299$ , $p = 0.557$ , $Cdelta = -0.110$

<sup>2</sup> SDV for SNR = 7 dB and 17 dB appears statistically significant if  $p$  is set to 0.05 for testing normal distribution. All other parameters throughout the manuscript remained the same for  $p$  set to 0.001. We decided to report more rigor results.

**Table 6.** Summary statistics for categorical data. For more information, please see Materials and Methods section. SNR stands for Signal-to-Noise Ratio.

Feature	Proportions of Reported Nausea Corresponding to the Selected Feature (Low/High)					Proportions of Regular EGG Corresponding to the Selected Features (Low/High)						
	Original	SNR = 17 dB	SNR = 7 dB	SNR = -3 dB	SNR = -13 dB	SNR = -23 dB	Original	SNR = 17 dB	SNR = 7 dB	SNR = -3 dB	SNR = -13 dB	SNR = -23 dB
SampEntT_m2	0.58/0.25	0.58/0.17	0.50/0.25	0.42/0.25	0.67/0.08	0.92/0.00	0.27/0.30	0.21/0.25	0.23/0.32	0.25/0.12	0.70/0.07	0.93/0.00
SampEntT_m3	0.50/0.42	0.75/0.25	0.67/0.33	0.67/0.33	0.75/0.17	0.92/0.00	0.37/0.61	0.39/0.55	0.41/0.59	0.48/0.43	0.84/0.09	0.98/0.02
SampEntT_m4	0.33/0.58	0.50/0.50	0.42/0.58	0.50/0.50	0.67/0.33	0.92/0.08	0.30/0.70	0.25/0.71	0.29/0.70	0.36/0.62	0.70/0.25	0.95/0.05

Pearson's Chi square test is used to explore the correlation of nausea occurrence with categorical SampEntT parameters for  $m = 2, 3$ , and  $4$ , and for SNR = 17 dB, 7 dB,  $-3$  dB,  $-13$  dB, and  $-23$  dB. However, no statistically significant relationships are found.

#### 4. Discussion

As expected, noise influences EGG signals and consequently EGG-based parameters [66]. Presented results suggest that EGG-based parameters have divergent robustness to the additive colored noise. Here, we provide a quantitative approach to evaluate the noise effect on the results of statistical analysis and RF performance for sickness-related nausea detection during automated driving simulation. The discussion on different levels of SNR should be taken with precaution as the original dataset already contains a certain level of noise in the studied frequency band (from 1 cpm to 10 cpm). Hence, we stress that reported SNRs of semi-synthetic EGG data are lower than the actual SNRs. Taking into account that we applied channel selection and manual deletion of segments with exaggerated noise following the procedure applied in [4], we may assume that the true SNRs are non-significant in comparison to the reported SNRs.

Our results provide some important and intriguing insights into the behavior of selected EGG parameters in noisy conditions demonstrating their usability to detect EGG signals affected by nausea and simulator sickness. At the same time, we are aware that a larger and more diverse sample should be used to confirm or contradict the following insights obtained from the presented results.

##### 4.1. Effect of Noise on EGG-Based Parameters

Expectedly, changing SNRs have different impacts on EGG-based parameters as their robustness to noise often degraded. Luckily, some features showed independent relationships with different SNRs. We therefore discuss each observed feature independently and indicate its potential and limitations in different noisy conditions.

Table 2 reveals that SampEntP is the least affected and does not change significantly across SNRs ranging from 17 dB to  $-23$  dB, whereas SampEntT shows sensitivity to noise at higher SNRs ( $-13$  dB,  $-23$  dB). This may be due to the fact that the PSD of colored noise should be relatively flat on a studied segment of EGG spectrum (and thus relatively deterministic), whereas the noise signal in time domain is random. The sensitivity of SampEntT to noise at higher SNR values is to be expected as the substantial noise compromises the calculations of sample entropy [24].

Both SampEntT and SampEntP remain unchanged for all three embedding dimensions  $m$  indicating that the choice of  $m$  (if kept small) has no effect on SampEntP sensitivity to the noise in EGG signal. Moreover, SampEntP is the only parameter that had effect size  $C_{delta}$  under 0.2 indicating negligible changes for all noise levels (Table 2). In all parameters,  $C_{delta}$  and  $C_d$  where applicable degraded, i.e., revealed increased difference for increased noise levels.

Autocorrelation zero-crossing and median frequency did not change significantly for positive SNRs. This is expected as the calculation of autocorrelation and median frequency depends on the signal power which is in cases of SNR  $> 0$  dB larger than the power of noise. On the other hand, DF remained stable in noisy conditions down to SNR =  $-3$  dB, and

unlike autocorrelation zero-crossing and median frequency, DF depends on just one peak location which could remain the same even in cases when noise power is larger than the signal power as long as it is not larger than the DF peak. SDV is not affected only at SNR of 17 dB, and this is expected as the variability of the signal is not the same as the variability of colored noise. Hence, SDV is easily affected, especially for lower SNRs.

Other parameters (RMS, MagDF, CS, SD1, SD2, and SDEGG) are influenced significantly for all SNR levels. Partly, this is expected as RMS and MagDF depend on the amplitude of the EGG signal which changes with added noise. SD1, SD2, and SDEGG depend on the relation between consecutive EGG samples, which is obviously highly sensitive to the colored pseudo-random noise. What is somewhat surprising is the fact that even relatively low noise levels (SNRs of 17 dB and 7 dB) had such a significant influence on these parameters.

#### 4.2. Effect of Noise on Random Forest Classifier for Nausea Detection

Accuracy for the binary RF classifier is satisfactory (>88%). This is almost twice as better than a result presented by Dennison et al. [67] where the unimodal classifier trained only on two EGG features (percentages of band power of slow and fast stomach activity) reached an accuracy of 48.52% for four-scale sickness severity assessment. On the other hand, Dennison et al. [67] reported an accuracy higher than 95% when heterogeneous sensor data or solely EEG features are fed to the classifier for nausea severity classification. Future work should definitely be focused on multi-modal data fusion and properly selected EGG-based features for even better classification accuracies. AUC showed a poor classification result (0.616 for training with the best result of 0.667 for test set) which is expected as classifiers have poor specificity of  $\leq 0.333$  (Tables 3 and 4). AUC remained constant for the training set throughout all results, probably indicating less training confidence. Although the AUC for the test set reached a maximum of 0.667 and slightly outperformed on unseen instances in comparison to AUC on the training set (0.616), it still performed poorly. It could be argued that without higher specificity, the data cannot reveal a strong machine learning pattern. The classifier performance remained stable even for noisy datasets (Table 3). The highest degradation happened at the lowest SNRs of  $-17$  dB and  $-7$  dB as expected. This is in line with previously reported results indicating that with higher noise levels, RF performance degrades while it is reasonably resistant to the noise procedure [62,63]. The reason for this may be in the fact that the majority of EGG-based features changed statistically significantly for these noise levels (Table 2) and that non-linear relations among parameters probably changed. A similar result is seen when a classifier previously trained on original data is tested with a noisy dataset (Table 4), indicating that RF may be considered a good candidate for nausea detection in simulated automated vehicles by EGG-based parameters.

Class imbalance may be the problem with the available dataset. Although all applied methods should compensate for imbalance, they cannot eliminate it. For nausea occurrence, overall, 12 out of 68 (~17.6%) EGG segments had positive nausea incidence. We hypothesize that RF deals well with imbalanced data in comparison with other classifiers as it uses data bootstrapping by random sampling with replacement [60]. The problem with nausea occurrence is that although sensitivity/recall is high (100%)—meaning that all subjects with nausea are correctly classified—the specificity is low (33.3%) so the classifier is not good at discerning those without reported nausea. This may not be caused solely by the RF, as it may also be the consequence of subjective reporting of nausea occurrence. We do not exclude the case that some subjects probably failed to report sickness when it actually happened.

Importance plots should be taken with precaution due to the existence of cross-correlations among introduced parameters that can influence importance. Feature cross-correlations cannot influence the RF accuracy [68,69], indicating that differences in the importance plot originate from the SNR influence. Importance plots reveal that SampEntP rose to the top five with the highest importance, indicating that it may be one of the most relevant features (Figure 3). This is rather important as SampEntP could not differentiate

between nausea and non-nausea data with a classical statistical approach (Table 5). The reason for this may be in the fact that the statistical test failed to detect non-linear relations in comparison to the ML algorithm.

Autocorrelation zero-crossing and CS appear only in the first five features in the importance plot for the original dataset, while SampEntT, DF, MagDF, and median also appear within the five most relevant features. Interestingly, RMS, SDV, and features derived from the Poincaré plot (SD1, SD2, and SDEGG) do not appear within the top five features in all importance plots incorporating original and semi-synthetic noisy EGG-based parameters (Figure 3). This may be the consequence of RF which fails to detect their influence or of the fact that the influence of parameters derived from the Poincaré plot is minor in comparison to other EGG-based parameters.

#### 4.3. Effect of Noise on Detection of Nausea through Statistical Tests

SpectEnt shows the statistically significant difference between those with and without reported nausea (Table 5) for all positive SNR values. This is not in line with the results reported in Table 2 where this feature remained statistically changed for all SNR values. We can argue that the changes in the signals introduced by nausea occurrence are more dominant in comparison to changes introduced by noise, which makes these two parameters mildly robust to the added colored noise. Similarly, CS does not show any tolerance to noise, but its ability to discern among EGG with and without nausea occurrence is stable for a relatively low noise level (SNR = 17 dB). However, this is only true if  $p$  is set to 0.001 for testing normal distribution, but not if  $p$  is set to 0.05. Median frequency, DF, autocorrelation zero-crossing, as well as SampEntT and SampEntP for all embedding dimensions  $m$  show no statistically significant difference between those features with and without nausea occurrence for the original dataset. The difference that arose with added noise is probably merely coincidental or falsely produced as a result of additive synthetic noise.

DF is not affected by the nausea occurrence and  $p$  value is much lower than in the previous study that reported results on the same dataset [4]. However, these results cannot be directly compared as different independent variables and different statistical tests were used in the current study.

Transformation of SampEntT for three embedding dimensions  $m$  into categorical variables for original data and all noise levels did not produce any significant result yielding to a conclusion that SampEntT should be treated as a numerical variable. Table 6 speaks in favor of such a finding as reported proportions are indecisive for the original data, and what is more convincing is that categorical SampEntT appear very sensitive to SNR levels, which is contrasting to the SampEntT robustness for all embedding dimensions  $m$  from Table 2.

Effect size reported by  $Cd$  and  $Cdelta$  tended to decrease with higher SNRs (Table 5) meaning that the higher noise contamination influences parameter sensitivity to nausea occurrence. In all cases where statistically significant differences are reported in Table 5, absolute effect size parameters ranged from small ( $>0.2$ ) to large differences ( $>0.8$ ). In all other cases, differences were rather small ( $<0.4$ ) or negligible ( $<0.2$ ), except for the SpectEnt which, despite the large effect size ( $>0.6$ ), did not reveal statistical significance for higher noise levels ( $-13$  dB and  $-23$  dB).

#### 4.4. Limitations of the Study

Although we present a detailed analysis of EGG-based parameters for simulator nausea evaluation, we recognize the following limitations:

1. We use a discrete set of predefined SNRs, and one should note that the actual SNRs were much higher, as our data were already contaminated with noises and artifacts. Despite the linear Butterworth filtering applied in the preprocessing stage, the noise with overlapping frequency content probably remains present in the semi-synthetic EGG dataset. Future efforts towards the generation of synthetic noises would provide a firm basis for exact SNR contamination and more reliable analysis.

2. It should also be noted that sample entropy scaling parameter  $r$  is kept constant at the 0.15 of the noiseless data standard deviation. This value was determined empirically based on the recommendations [24]. Adjusting this value for different SNRs may have a further effect on the results and should be investigated in the future.
3. We apply procedures for automatic feature calculation. However, a guided visual observation and manual corrections are still considered a gold standard for the evaluation of EGG-based parameters especially in cases of excessive noises [10,70,71]. We use visual inspection only for channel selection. Despite this drawback, we obtained promising results in nausea assessment by both statistical and ML approaches.
4. We select the embedded dimension  $m$  for sample entropy calculation empirically. For future selection and discussion on embedding dimension selection, one may look at outstanding reasoning by Matilla-García et al. [72].
5. We did not apply unimodal or multi-modal machine learning algorithms, and we do not provide comparison of existing machine learning techniques as in [67].
6. Our method is applied only for nausea occurrence. Further customization of presented EGG-based parameters and complementary approach by RF and statistical analysis should yield at assessment of sickness levels similarly as in [67].
7. The dataset used for the analysis contains more male than female participants. However, we do not consider this to be a major drawback of our study, as we were not interested in the differences between the genders but focused on the relationships between the occurrence of nausea, the EGG parameters, and noise. Moreover, a systematic review performed by Grassini and Laumann [73] showed conflicting results in published studies focused on determining sex differences in experiencing simulator sickness.
8. We did not use multi biomarkers for the assessment of sickness occurrence as our focus was solely on the direct assessment of gastric activity. However, future studies should be focused on a promising heterogeneous approach as, for example, suggested by Dennison et al. [67].

## 5. Conclusions

The presented results highlighted the importance of appropriate EGG parameters selection when the higher levels of noise are anticipated during driving simulation for nausea detection. Although some EGG-based features are sensitive to the nausea occurrence, they may at the same time be sensitive to the higher noise levels. This is important for the study design of EGG-based nausea detection within driving simulators encompassing haptic frameworks.

Feature engineering and decision making by both machine learning and statistical tests may be fully automated for the future adoption of EGG-based nausea detection. These two approaches are complementary, as ML algorithms benefit from non-linear relations that cannot be revealed by statistical tests such as in the case of sample entropy parameters.

Sample entropy of EGG signals stands out among all other parameters due to its exceptional robustness to the colored noise and due to its ability to differentiate between EGG segments with and without nausea occurrence for signals recorded in the driving simulator. The potential of sample entropy to detect nausea in noisy EGG signals should be further explored in other EGG-related dynamic studies. The assessment of EGG signals with the sample entropy feature may open a door to the scientific experiments that were never conducted before as exaggerated EGG sensitivity to noises and artifacts may not present an obstacle anymore.

Sickness-related nausea detection in driving simulators by EGG-based parameters is an important aspect that could assist in the overall comfort improvement in both simulators and automated vehicles. This study emphasized the importance of proper EGG-based feature selection when dynamic and noisy EGG recording is anticipated (e.g., different levels of driving automation resulting in subject's maneuvers interfacing vehicle commands

or in-vehicle infotainment). Additionally, our results revealed the superiority of sample entropy in relation to other parameters and in combination with the RF ML algorithm.

**Author Contributions:** Conceptualization, N.M., J.S. and G.J.; methodology, G.J.; software, G.J. and N.M.; validation, N.M. and J.S.; formal analysis, N.M., J.S. and G.J.; investigation, G.J. and N.M.; resources, J.S.; data curation, G.J. and N.M.; writing—original draft preparation, N.M.; writing—review and editing, G.J. and J.S.; visualization, N.M.; supervision, N.M.; project administration, J.S.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by HADRIAN (Holistic Approach for Driver Role Integration and Automation Allocation for European Mobility Needs) EU Horizon 2020 project, grant number 875597. It was also partly supported by the Slovenian Research Agency within the research program ICT4QoL—Information and Communications Technologies for Quality of Life, grant number P2-0246. N.M. was partly supported by the Ministry of Education, Science, and Technological Development, Republic of Serbia, grant number 451-03-68/2022-14/200103.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Authors' Institutional Review Board of Department of ICT (7 September 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The raw data presented in this study and Matlab code for feature extraction are available on request from the corresponding author. R code and CSV tables with relevant EGG-based parameters are available on Zenodo repository and shared under GNU General Public License [59].

**Acknowledgments:** The authors' gratitude goes to Nenad B. Popović for his long collaboration in studies related to EGG research and for the fruitful discussions on feature extraction techniques followed by his scientific contributions published elsewhere. The authors also acknowledge Timotej Gruden, a student, for his exceptional work in experiment design and measurement conduction.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References

1. Brooks, J.O.; Goodenough, R.R.; Crisler, M.C.; Klein, N.D.; Alley, R.L.; Koon, B.L.; Logan, W.C., Jr.; Ogle, J.H.; Tyrrell, R.A.; Wills, R.F. Simulator sickness during driving simulation studies. *Accid. Anal. Prev.* **2010**, *42*, 788–796. [[CrossRef](#)]
2. Lucas, G.; Kemeny, A.; Paillot, D.; Colombet, F. A simulation sickness study on a driving simulator equipped with a vibration platform. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *68*, 15–22.
3. Wang, J.; Liang, H.N.; Monteiro, D.; Xu, W.; Xiao, J. Real-time Prediction of Simulator Sickness in Virtual Reality Games. *IEEE Trans. Games* **2022**. [[CrossRef](#)]
4. Gruden, T.; Popović, N.B.; Stojmenova, K.; Jakus, G.; Miljković, N.; Tomažič, S.; Sodnik, J. Electrogastrography in Autonomous Vehicles—An Objective Method for Assessment of Motion Sickness in Simulated Driving Environments. *Sensors* **2021**, *21*, 550. [[CrossRef](#)]
5. Classen, S.; Bewernitz, M.; Shechtman, O. Driving simulator sickness: An evidence-based review of the literature. *Am. J. Occup. Ther.* **2011**, *65*, 179–188. [[CrossRef](#)]
6. Dużmańska, N.; Strojny, P.; Strojny, A. Can simulator sickness be avoided? A review on temporal aspects of simulator sickness. *Front. Psychol.* **2018**, *9*, 2132.
7. Aykent, B.; Merienne, F.; Guillet, C.; Paillot, D.; Kemeny, A. Motion sickness evaluation and comparison for a static driving simulator and a dynamic driving simulator. *Proc. Inst. Mech. Eng. Part D J. Automob. Eng.* **2014**, *228*, 818–829.
8. Dennison, M.S.; Krum, D.M. Unifying research to address motion sickness. In Proceedings of the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 1858–1859.
9. Smyth, J.; Birrell, S.; Woodman, R.; Jennings, P. Exploring the utility of EDA and skin temperature as individual physiological correlates of motion sickness. *Appl. Ergon.* **2021**, *92*, 103315. [[CrossRef](#)]
10. Popović, N.B.; Miljković, N.; Stojmenova, K.; Jakus, G.; Prodanov, M.; Sodnik, J. Lessons Learned: Gastric Motility Assessment During Driving Simulation. *Sensors* **2019**, *19*, 3175. [[CrossRef](#)]
11. John, B. Pupil diameter as a measure of emotion and sickness in VR. In Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications, Denver, CO, USA, 25–28 June 2019; pp. 1–3.

12. Park, S.; Mun, S.; Ha, J.; Kim, L. Non-contact measurement of motion sickness using pupillary rhythms from an infrared camera. *Sensors* **2021**, *21*, 4642. [CrossRef]
13. Wachler, B.S.B.; Krueger, R.R. Agreement and repeatability of pupillometry using videokeratography and infrared devices. *J. Cataract Refract. Surg.* **2000**, *26*, 35–40.
14. Koohestani, A.; Nahavandi, D.; Asadi, H.; Kebria, P.M.; Khosravi, A.; Alizadehsani, R.; Nahavandi, S. A Knowledge Discovery in Motion Sickness: A Comprehensive Literature Review. *IEEE Access* **2019**, *7*, 85755–85770. [CrossRef]
15. Laviola, J.J. A discussion of cybersickness in virtual environments. *ACM Sigchi Bull.* **2000**, *32*, 47–56. [CrossRef]
16. Crampton, G.H. *Motion and Space Sickness*; CRC Press: Boca Raton, FL, USA, 1990; ISBN 978-0-8493-4703-0.
17. Golding, J.F. Motion sickness. In *Handbook of Clinical Neurology*; Furman, J.M., Lempert, T., Eds.; Neuro-Otology; Elsevier: Amsterdam, The Netherlands, 2016; Volume 137, pp. 371–390.
18. Davis, S.; Nesbitt, K.; Nalivaiko, E. Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters. In Proceedings of the 11th Australasian Conference on Interactive Entertainment (IE 2015), Sydney, Australia, 27–30 January 2015.
19. Iskander, J.; Attia, M.; Saleh, K.; Nahavandi, D.; Abobakr, A.; Mohamed, S.; Asadi, H.; Hossny, M.; Lim, C.P.; Hossny, M. From car sickness to autonomous car sickness: A review. *Transp. Res. Part F Traffic Psychol. Behav.* **2019**, *62*, 716–726.
20. Mühlbacher, D.; Tomzig, M.; Reinmueller, K.; Rittger, L. Methodological Considerations Concerning Motion Sickness Investigations during Automated Driving. *Information* **2020**, *11*, 265.
21. Koch, K.L. Gastric dysrhythmias: A potential objective measure of nausea. *Exp. Brain Res.* **2014**, *232*, 2553–2561. [CrossRef]
22. Wolpert, N.; Rebollo, I.; Tallon-Baudry, C. Electrogastrigraphy for psychophysiological research: Practical considerations, analysis pipeline, and normative data in a large sample. *Psychophysiology* **2020**, *57*, e13599.
23. Delgado-Bonal, A. Quantifying the randomness of the stock markets. *Sci. Rep.* **2019**, *9*, 1–11.
24. Delgado-Bonal, A.; Marshak, A. Approximate entropy and sample entropy: A comprehensive tutorial. *Entropy* **2019**, *21*, 541. [CrossRef]
25. Fele-Žorž, G.; Kavšek, G.; Novak-Antolič, Ž.; Jager, F. A comparison of various linear and non-linear signal processing techniques to separate uterine EMG records of term and pre-term delivery groups. *Med. Biol. Eng. Comput.* **2008**, *46*, 911–922. [CrossRef]
26. Li, X.; Li, D.; Liang, Z.; Voss, L.J.; Sleigh, J.W. Analysis of depth of anesthesia with Hilbert–Huang spectral entropy. *Clin. Neurophysiol.* **2008**, *119*, 2465–2475. [CrossRef] [PubMed]
27. Chhabra, A.; Subramaniam, R.; Srivastava, A.; Prabhakar, H.; Kalaivani, M.; Paranjape, S. Spectral entropy monitoring for adults and children undergoing general anaesthesia. *Cochrane Database Syst. Rev.* **2016**, *3*. [CrossRef]
28. Schruppf, F.; Sturm, M.; Bausch, G.; Fuchs, M. Derivation of the respiratory rate from directly and indirectly measured respiratory signals using autocorrelation. *Curr. Dir. Biomed. Eng.* **2016**, *2*, 241–245. [CrossRef]
29. Rangayyan, R.M. Filtering for removal of artifacts. In *Biomedical Signal Analysis*; IEEE: Piscataway, NJ, USA, 2015; pp. 91–231. [CrossRef]
30. Mintchev, M.P.; Stickel, A.; Bowes, K.L. Dynamics of the level of randomness in gastric electrical activity. *Dig. Dis. Sci.* **1998**, *43*, 953–956. [CrossRef] [PubMed]
31. Brennan, M.; Palaniswami, M.; Kamen, P. Do existing measures of Poincare plot geometry reflect nonlinear features of heart rate variability? *IEEE Trans. Biomed. Eng.* **2001**, *48*, 1342–1347.
32. Yarkoni, T.; Westfall, J. Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspect. Psychol. Sci.* **2017**, *12*, 1100–1122. [CrossRef]
33. Yin, J.; Chen, J.D. Electrogastrigraphy: Methodology, validation and applications. *J. Neurogastroenterol. Motil.* **2013**, *19*, 5.
34. Nervtech Simulation Technologies. Available online: <https://www.nervtech.com> (accessed on 20 September 2022).
35. Vengust, M.; Možina, D.; Pušenjak, N.; Zevnik, L.; Sodnik, J.; Kaluža, B.; Tavčar, A. NERVteh 4DOF motion car driving simulator. In Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Seattle, WA, USA, 17–19 September 2014; pp. 1–6.
36. Fanatec ClubSport Pedals V3. Available online: <https://fanatec.com/eu-en/pedals/clubsport-pedals-v3> (accessed on 20 September 2022).
37. Fanatec ClubSport Wheel Base V2.5. Available online: <https://fanatec.com/eu-en/racing-wheels-wheel-bases/wheel-bases/clubsport-wheel-base-v2.5> (accessed on 20 September 2022).
38. AV Simulation SCANer Studio. Available online: <https://www.avsimulation.com/scanerstudio/> (accessed on 20 September 2022).
39. Oculus Rift Oculus. Available online: <https://www.oculus.com/rift/> (accessed on 20 September 2022).
40. Kinkade, R.G.; Wheaton, G.R. Training device design. In *Human Engineering Guide to Equipment Design*; Van Cott, H.P., Kinkade, R.G., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 1972; pp. 668–699.
41. Hays, R.T. *Simulator Fidelity: A Concept Paper*; Army Research Inst for the Behavioral and Social Sciences: Alexandria, VA, USA, 1980.
42. SAE Levels of Driving Automation™ Refined for Clarity and International Audience. Available online: <https://www.sae.org/blog/sae-j3016-update> (accessed on 20 September 2022).
43. Riezzo, G.; Russo, F.; Indrio, F. Electrogastrigraphy in adults and children: The strength, pitfalls, and clinical significance of the cutaneous recording of the gastric electrical activity. *BioMed Res. Int.* **2013**, *2013*, 282757. [CrossRef]
44. Popović, N.; Miljković, N.; Popović, M.B. Simple gastric motility assessment method with a single-channel electrogastrigram. *Biomed. Tech. Eng.* **2018**, *64*, 177–185. [CrossRef]

45. Universal Interface Module UIM100C. Available online: <https://www.biopac.com/product/universal-interface-module/> (accessed on 20 September 2022).
46. Jovanović, N.; Popović, N.B.; Miljković, N. Combined approach for automatic and robust calculation of dominant frequency of electrogastragram. *arXiv* **2020**, arXiv:2009.09023.
47. Chang, F.Y. Electrogastrography: Basic knowledge, recording, processing and its clinical applications. *J. Gastroenterol. Hepatol.* **2005**, *20*, 502–516. [[PubMed](#)]
48. Koch, K.L.; Stern, R.M. *Handbook of Electrogastrography*; Oxford University Press: Oxford, UK, 2003.
49. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol.-Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)] [[PubMed](#)]
50. Boljanić, T.; Miljković, N.; Lazarević, L.B.; Knežević, G.; Milašinović, G. Relationship between electrocardiogram-based features and personality traits: Machine learning approach. *Ann. Noninvasive Electrocardiol.* **2022**, *27*, e12919. [[CrossRef](#)] [[PubMed](#)]
51. Team, R.C.R. A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. 2021. Available online: <https://www.r-project.org> (accessed on 20 September 2022).
52. Wickham, H.; Francois, R.; Henry, L.; Müller, K.; Dplyr, A. Grammar of Data Manipulation. R Found. Stat. Comput., Vienna. R Package Version 0.4.3. 2015. Available online: <https://CRAN.R-project.org/package=dplyr> (accessed on 20 September 2022).
53. Meyer, D.; Dimitriadou, E.; Hornik, K. Misc Functions of the Department of Statistics, Probability Theory Group (e1071), TU Wien. R Package Version 1–7. 2019. Available online: <https://CRAN.R-project.org/package=e1071> (accessed on 20 September 2022).
54. Kuhn, M.; Wing, J.; Weston, S.; Williams, A.; Keefer, C.; Engelhardt, A.; Cooper, T.; Mayer, Z.; Kenkel, B.; Team, R.C. Package ‘Caret’. *R J.* **2020**, *223*, 7.
55. Paluszynska, A.; Biecek, P.; Jiang, Y.; Jiang, M.Y. Package ‘randomForestExplainer’. Explaining and visualizing random forests in terms of variable importance. 2020. Available online: <https://CRAN.R-project.org/package=randomForestExplainer> (accessed on 20 September 2022).
56. Hadley, W. *Ggplot2: Elegant Graphics for Data Analysis*; Springer: Berlin/Heidelberg, Germany, 2016.
57. Torchiano, M. Effsize: Efficient Effect Size Computation. R Package Version 0.8.1. 2020. Available online: <https://CRAN.R-project.org/package=effsize> (accessed on 20 September 2022). [[CrossRef](#)]
58. Robin, X.; Turck, N.; Hainard, A.; Tiberti, N.; Lisacek, F.; Sanchez, J.; Müller, M. pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinform.* **2011**, *12*, 77. [[CrossRef](#)]
59. Jakus, G.; Sodnik, J.; Miljković, N. NadicaSm/Statistical-Analysis-and-Machine-Learning-for-EGG-based-Nausea-Detection: V1 (Version v1). *Version V1. Zenodo* **2022**. [[CrossRef](#)]
60. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
61. Liaw, A.; Wiener, M. Classification and Regression by Random Forest. *R News* **2002**, *2*, 18–22.
62. Schwalbe Lehtihet, O.; Åryd, V.A. *Comparison of Performance and Noise Resistance of Different Machine Learning Classifiers on Gaussian Clusters*; KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science: Stockholm, Sweden, 2021.
63. Ishii, S.; Ljunggren, D.A. *Comparative Analysis of Robustness to Noise in Machine Learning Classifiers*; KTH Royal Institute of Technology, School of Electrical Engineering and Computer Science: Stockholm, Sweden, 2021.
64. Wong, T.T. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit.* **2015**, *48*, 2839–2846.
65. Arlot, S.; Celisse, A. A survey of cross-validation procedures for model selection. *Stat. Surv.* **2010**, *4*, 40–79. [[CrossRef](#)]
66. Verhagen, M.A.; Van Schelven, L.J.; Samsom, M.; Smout, A.J. Pitfalls in the analysis of electrogastrographic recordings. *Gastroenterology* **1999**, *117*, 453–460. [[PubMed](#)]
67. Dennison, M., Jr.; D’Zmura, M.; Harrison, A.; Lee, M.; Raglin, A. Improving motion sickness severity classification through multi-modal data fusion. In *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications*; SPIE: Bellingham, WA, USA, 2019; Volume 11006, pp. 277–286.
68. Strobl, C.; Boulesteix, A.-L.; Kneib, T.; Augustin, T.; Zeileis, A. Conditional variable importance for random forests. *BMC Bioinform.* **2008**, *9*, 307. [[CrossRef](#)]
69. Tološi, L.; Lengauer, T. Classification with correlated features: Unreliability of feature ranking and solutions. *Bioinformatics* **2011**, *27*, 1986–1994. [[CrossRef](#)]
70. Komorowski, D. EGG DWPack: System for multi-channel electrogastrographic signals recording and analysis. *J. Med. Syst.* **2018**, *42*, 1–17. [[CrossRef](#)] [[PubMed](#)]
71. Du, P.; O’Grady, G.; Paskaranandavadivel, N.; Tang, S.J.; Abell, T.; Cheng, L.K. High-resolution mapping of hyperglycemia-induced gastric slow wave dysrhythmias. *J. Neurogastroenterol. Motil.* **2019**, *25*, 276. [[CrossRef](#)]
72. Matilla-García, M.; Morales, I.; Rodríguez, J.M.; Ruiz Marín, M. Selection of embedding dimension and delay time in phase space reconstruction via symbolic dynamics. *Entropy* **2021**, *23*, 221. [[CrossRef](#)]
73. Grassini, S.; Laumann, K. Are modern head-mounted displays sexist? A systematic review on gender differences in HMD-mediated virtual reality. *Front. Psychol.* **2020**, *11*, 1604. [[CrossRef](#)]