

Article

Reduction of Artifacts in Capacitive Electrocardiogram Signals of Driving Subjects

Tamara Škorić 

Faculty of Technical Sciences, University of Novi Sad, 21000 Novi Sad, Serbia; ceranic@uns.ac.rs or tamara.ceranic@gmail.com; Tel.: +381-65-667-3311

Abstract: The development of smart cars with e-health services allows monitoring of the health condition of the driver. Driver comfort is preserved by the use of capacitive electrodes, but the recorded signal is characterized by large artifacts. This paper proposes a method for reducing artifacts from the ECG signal recorded by capacitive electrodes (cECG) in moving subjects. Two dominant artifact types are coarse and slow-changing artifacts. Slow-changing artifacts removal by classical filtering is not feasible as the spectral bands of artifacts and cECG overlap, mostly in the band from 0.5 to 15 Hz. We developed a method for artifact removal, based on estimating the fluctuation around linear trend, for both artifact types, including a condition for determining the presence of coarse artifacts. The method was validated on cECG recorded while driving, with the artifacts predominantly due to the movements, as well as on cECG recorded while lying, where the movements were performed according to a predefined protocol. The proposed method eliminates 96% to 100% of the coarse artifacts, while the slow-changing artifacts are completely reduced for the recorded cECG signals larger than 0.3 V. The obtained results are in accordance with the opinion of medical experts. The method is intended for reliable extraction of cardiovascular parameters to monitor driver fatigue status.

Keywords: cECG filter; movement artefacts; binarized approximate entropy; KNN; DDNN



Citation: Škorić, T. Reduction of Artifacts in Capacitive Electrocardiogram Signals of Driving Subjects. *Entropy* **2022**, *24*, 13. <https://doi.org/10.3390/e24010013>

Academic Editor: Carlos M. Travieso-González

Received: 22 October 2021
Accepted: 6 December 2021
Published: 22 December 2021

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



Copyright: © 2021 by the author. 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

The automotive industry has been making efforts to develop smart health systems, as part of supporting smart cars that can communicate with each other, transmit data to the cloud, and use smart e-health service systems [1–5]. Measurement of electrocardiograms (ECG), electroencephalograms (EEG), and respiratory activities, as well as assessment of the parameters of these time series in real-time, would contribute to the insight into the health condition of the driver while driving. The driver's ECG, as well as the parameters derived from this signal, enable the assessment of alarming traffic situations such as driver fatigue [6], drowsiness [7], prediction of infarction development (of particular importance for the older group of drivers) [8], and EEG signal measurement contributes detection of driver fatigue [9], as well as prediction of emergency braking situations to activate the brake pedal when drivers are not able to react at the appropriate speed [10,11]. An additional motivation is the current demographic situation and the significant presence of older drivers in traffic. Ford has developed a car seat for heart rate monitoring (HR) [12]. Capacitive electrodes are installed in the car seat for non-contact (without direct contact with the skin) measurement of ECG signals through the driver's clothing [1]. The main goal of this system was to monitor HR as an important parameter for assessing the health status of drivers [13], or the presence of drowsiness in drivers [14]. In addition to the capacitive electrodes that are built into the car seat, in Reference [15], a steering wheel covered with a conductive fabric-based dry electrode material is proposed, while Reference [5] describes a solution that additionally uses sensors built into the steering wheel and belt. The signals are sent via Bluetooth® and processed on the built-in computer. A car seat equipped with Internet of Things (IoT) sensors for measuring ECG and EEG signals with a suitable transmitter

that sends physiological data to the database for further processing and prediction of the driver's health has been developed [3,4].

Limiting factors are specific conditions in which signal measurements are performed, i.e., recording while driving in a car, measurement without direct contact with the driver (recording over clothes made of different materials), and the dominant presence of movements, as well as the distractions while driving. The measured signals have very small amplitudes, with the dominant presence of artifacts that represent an obstacle to a reliable assessment of health parameters.

The presence of artifacts in the signals, as well as the estimation of parameters without a priori signal processing, can lead to the generation of false alarms or isolation of false outliers, which is becoming a relevant topic in database processing [16]. Recording by a capacitive electrode in clinical conditions, while the subjects were sitting still, was analyzed to determine the possibility of clinical application [17]. Detection of cardiac arrhythmia, however, is not reliable due to the impossibility of reliable identification of P and T waves, mostly due to the presence of movement artifacts [17]. Therefore, medical experts have pointed out that it is necessary to improve the algorithm for the reduction of artifacts in ECG signals recorded by capacitive electrodes [17]. For this reason, various efforts are being made to reduce artifacts. The initial proposal was to eliminate all beats that would have atypical values (below 30 beats/min or over 120 beats/min), i.e., these signal segments would be treated as artifacts [18]. The authors themselves stated that this very simple approach is not good enough for a realistic estimate of the signal parameters [18]. In addition, this approach removes the parameters that indicate the occurrence of unwanted pathologies. In Reference [1], an algorithm for artifact reduction based on principal component analysis (PCA) was proposed, with mandatory pre-processing of data to construct an appropriate training data set for PCA. The use of a training data set with a dominant presence of artifacts could lead to inaccurate classification by PCA, so pre-processing of signals recorded while driving is a mandatory step [1]. Rough signal preprocessing was performed based on the QRS complex quality assessment to reduce the number of artifacts in the data. QRS complexes were detected using open-source ECG analysis (OSEA) [19], and values that were many times higher than expected (so-called "outliers") were categorized based on a threshold representing the product of parameters describing the amplitude and standard deviation of QRS complexes [1]. After pre-processing of the data, the final elimination of the artifacts was performed based on the calculated eigenvalues, and, for that purpose, the threshold based on the Hotelling T square value, described in Reference [20], was used. The authors pointed out that the algorithm needs to be improved because it also eliminates QRS complexes of small amplitudes compared to noise [1].

One of the available algorithms for the assessment and detection of cECG signal artifacts is based on the injected signal and modeling of the capacitive system [21]. The model has been tested only in laboratory conditions and in simulations [21], so further confirmation in real systems is necessary [22]. Hardware upgrades have also been proposed [22], i.e., it is necessary to install an additional electrode that would provide motion information to form a reference signal for active noise cancellation (ANC), including an adaptive filter. There are many methods developed for the reduction of artifacts for the traditional measurement of ECG signals with direct contact with the patient's body and with the application of a gel. In the review paper of Reference [23], these techniques are categorized as techniques based on Empirical mode decomposition (EMD), wavelet transforms, hybrid models, deep-learning models, and various filters. However, for artifacts on cECG signals, EMD and wavelet-based techniques did not yield adequate results due to the dominant presence of large amplitudes and irregular motion noise characteristics, which was also observed in Reference [22]. Hopes have been pinned on a modified noise reduction method, Detrended Fluctuation Analysis (DFA)-EMD, based on the idea that, in addition to decomposing signals into intrinsic mode functions (IMF), the DFA parameters for each IMF should be estimated to characterize the IMF, as a carrier noise or information [24]. In Reference [25], however, it was stated that the DFA-EMD method failed to mitigate

the presence of artifacts in the cECG signal recorded while driving, due to the dominant presence of very large amplitudes of the artifacts.

The aim of this paper is to develop a method for reducing artifacts from ECG signals recorded by capacitive electrodes (cECG) in moving subjects. The basic idea of the proposed method is that quantified signal fluctuation around a linear trend of fixed-length segments can detect segments with artifacts. It is assumed that segments with a very high or very low value of fluctuations correspond to segments with coarse and slowly changing artifacts (frequency range from 0.5 to 15 Hz), respectively. We expect that the assessment of an acceptable level of fluctuation will indicate coarse and slowly changing artifacts and enable their elimination. Then, the reliability of cardiovascular parameters estimated from cECG signals without artifacts would increase, as well as the accuracy of classification techniques.

2. Materials and Methods

2.1. Materials

Acquisition of cECG time series while driving was performed using 6 electrodes built into the car seat. The electrodes are arranged on three levels of two electrodes in the upper part of the car seat (a detailed photograph is given in References [1,18]). Three electrodes with the most reliable measurements were manually selected [18]. The process of recording the cECG signal involved 6 volunteers (male, aged 39.8 ± 26.2 years) who drove in the city (about 2 h of recording), on the highway (about 8.8 h of recording), and the polygon in Belgium (about 2.5 h of recording) [18]. Thirty-one measurements were performed, comprising three cECG signals (designated as $cECG_1 = \text{Electrode}_1 - \text{Electrode}_2$, $cECG_2 = \text{Electrode}_2 - \text{Electrode}_3$, and $cECG_3 = \text{Electrode}_3 - \text{Electrode}_1$), as well as the reference ECG signal. Thus, each measurement while driving, whether driving on a highway, polygon, or city, provided four simultaneously recorded signals: $cECG_1$, $cECG_2$, $cECG_3$, and a reference signal, a total of 124 signals. The reference signal was measured using the equipment of the biosignal amplifier g. Bamp from g.tec medical engineering GmbH Schiedlberg, Austria (details in Reference [26]) [18]. An A/D converter (NI-USB6259) was used for capacitive electrodes ($cECG_1$, $cECG_2$, and $cECG_3$) and reference measurements, with an amplitude resolution of 16 bits per sample and sampling frequency of 1000 Hz [1], except for two signals with a sampling frequency of 200 Hz [18].

The cECG time series, recorded while lying on the bed, was performed using 12 built-in electrodes in the bed, from which three ones with the highest quality of measurements were automatically selected [18]. $cECG_1$, $cECG_2$, and $cECG_3$ were formed in the same way as measurements recorded during driving by three automatically selected measurements. Ten volunteers (aged 27.8 ± 4.3 years) participated in the recording procedure. They moved according to a certain protocol to simulate movements during sleep and generate coarse artifacts [18]. In the first half of the measurement, the volunteers moved every 60 s, while, in the second half of the measurement, they were asked to lie for 120 s, then to move for 60 s, and then lie down again for 120 s [18]. In this experiment, the sampling frequency was 400 Hz. The reference signal was measured by an MP70 (details in Reference [27]) of Philips (Eindhoven, the Netherlands) [18].

The database with all experimental recordings is publicly available [28], and the experiment is described in detail in Reference [18]. All volunteers gave written consent [18].

Duration of recording and amplitudes value of cECG time series, publicly available [28], are shown in Tables 1 and 2.

Table 1. Absolute amplitude of cECG [V], expressed as a mean \pm standard deviation SD.

	Number of Measurements in Each Group	$cECG_1$	$cECG_2$	$cECG_3$	Reference Signal
CAR	31	0.19 ± 0.67 [V]	0.29 ± 0.89 [V]	0.21 ± 0.69 [V]	0.81 ± 0.93 [V]
BED	20	1.18 ± 2.22 [V]	1.85 ± 2.64 [V]	1.66 ± 2.58 [V]	0.37 ± 0.52 [V]

Table 2. Duration of extracted ECG segments [s], expressed as a mean \pm standard deviation *SD*.

	Number of Measurements in Each Group	$cECG_1$	$cECG_2$	$cECG_3$	Reference Signal
CAR	31	1556.93 ± 2166.13 [s]	1556.93 ± 2166.13 [s]	1556.93 ± 2166.13 [s]	1556.93 ± 2166.13 [s]
BED	20	312.45 ± 10.54 [s]	312.45 ± 10.54 [s]	312.45 ± 10.54 [s]	312.45 ± 10.54 [s]

Table 1 shows the mean values of the absolute signal amplitudes \pm standard deviation (SD) recorded capacitive electrodes. It is noticeable that the intensity of recorded $cECG$ signals while driving is very low, while its duration is very large (Table 2). $cECG_1$, $cECG_2$, and $cECG_3$ have the same length because they are recorded simultaneously, while driving or moving on the bed.

Manual notation of R peaks (maximum in the QRS complex) was performed by two medical experts independently. If there was a difference of opinion, the R peaks were marked as *NaN* [18]. Results of peaks detection by OSEA software (described in Reference [19]) were analyzed in Reference [18] and are available in Reference [28].

Examples of the raw $cECG$ signals from two sets are shown in Figure 1a,b. Red markers point to the annotated R peaks according to medical experts. The original purpose of the labeled R peaks was to assess the accuracy of algorithms for automatic detection of R peaks in $cECG$, such as OSEA software, and the possibility of reliable estimation of HR [1,18]. The results confirmed the possibility of a reliable assessment of HR while driving in the case of eliminating time intervals with artifacts [1]. Our aim is to eliminate the artifacts and prepare the signal for further analysis. The beat-to-beat HR time series is one of the targets.

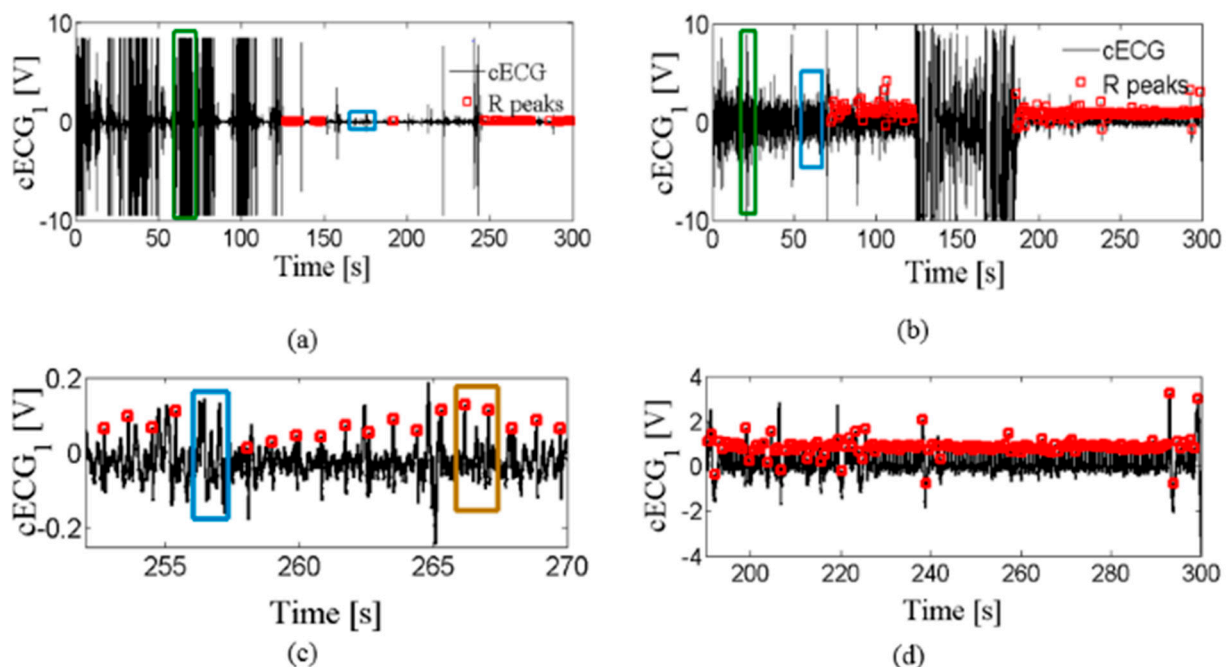


Figure 1. An example of a $cECG$ time series recorded (a) while driving in car, (b) while lying on the bed, (c) enlarged part of useful segments during driving, and (d) enlarged part of useful segments while lying. Red rectangles indicate R peaks marked by medical experts; an example of coarse artifacts is marked by the green rectangular border in (a,b); examples of slow-changing artifacts are marked by the blue rectangular border in (a–c). An example of the useful signal segment that is similar to the useless segment is marked by the brown rectangular border in (c). Slow-changing artifacts were determined by checking the overall accuracy in eliminating useless signal segments.

Two types of artifacts are distinguished: coarse artifacts, which occur as amplitude peaks but different amplitude values (examples are marked by green rectangles in Figure 1a,b), and slow-changing artifacts that are very similar to the useful part of a signal (examples are marked by blue rectangles in Figure 1a–c). Figure 1c,d show enlarged useful segments that medical experts have marked while driving and lying on the bed, respectively. The signal segment is treated as useful if R peaks can be detected by medical experts, or as useless if the R peaks, due to artifacts, cannot be detected. There are no manually marked artifact types in the publicly available database [28], neither coarse nor slow-changing artifacts.

After visual inspection, we noted that the cECG time series also differs in the number of coarse artifacts. Figure 2 shows the cECG time series (a) with a moderate amount of artifacts and (b) without coarse artifacts, while Figure 1a corresponds cECG time series with a very large amount of coarse artifacts. In accordance with the opinion of medical experts, useful parts of the cECG time series have been marked in red. Table 3 shows the number of recordings for each observed group during driving. The $cECG_3$ time series is not included in the analysis because there is no publicly available notation of the useful part of the signal. The total number of cECG time series per groups is small, but the long duration of recording (Table 2) enabled reliable signal analysis. Unfortunately, it is noted that a few of the cECG time series are into groups of cECG with a moderate amount of artifacts, and without artifacts, as a consequence of recordings conditions during driving. $cECG_1$, $cECG_2$, and $cECG_3$ time-series recorded while lying on the bed belong to a group of cECG with a large amount of artifacts due to moving subjects according to the protocol (total number of cECG is 60).

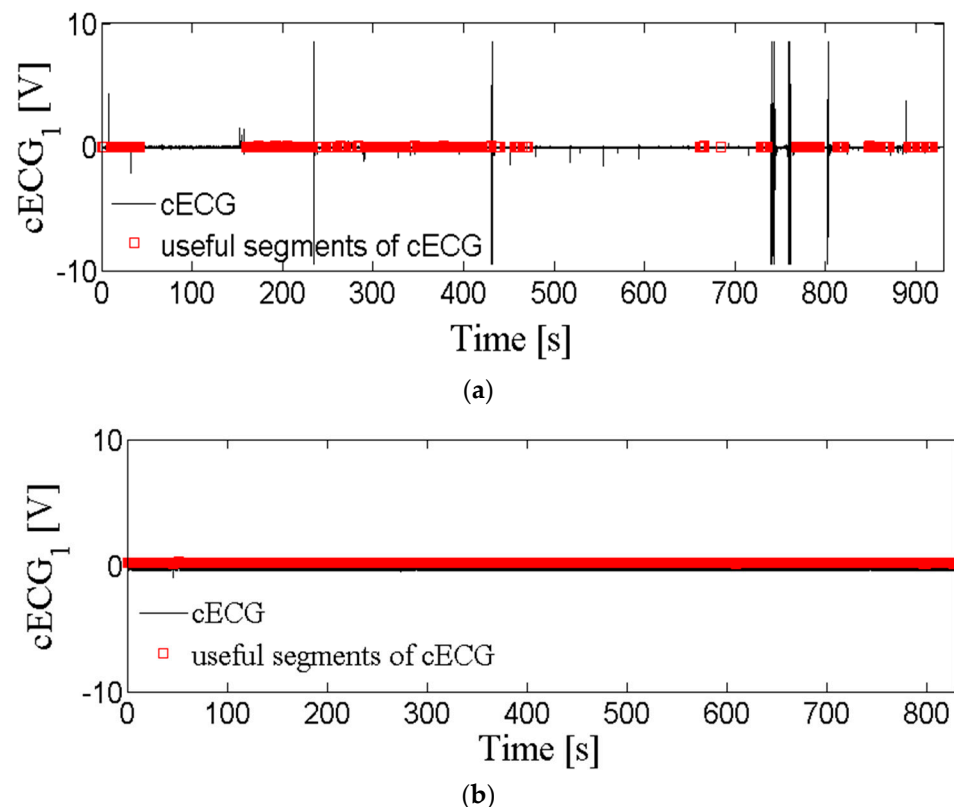


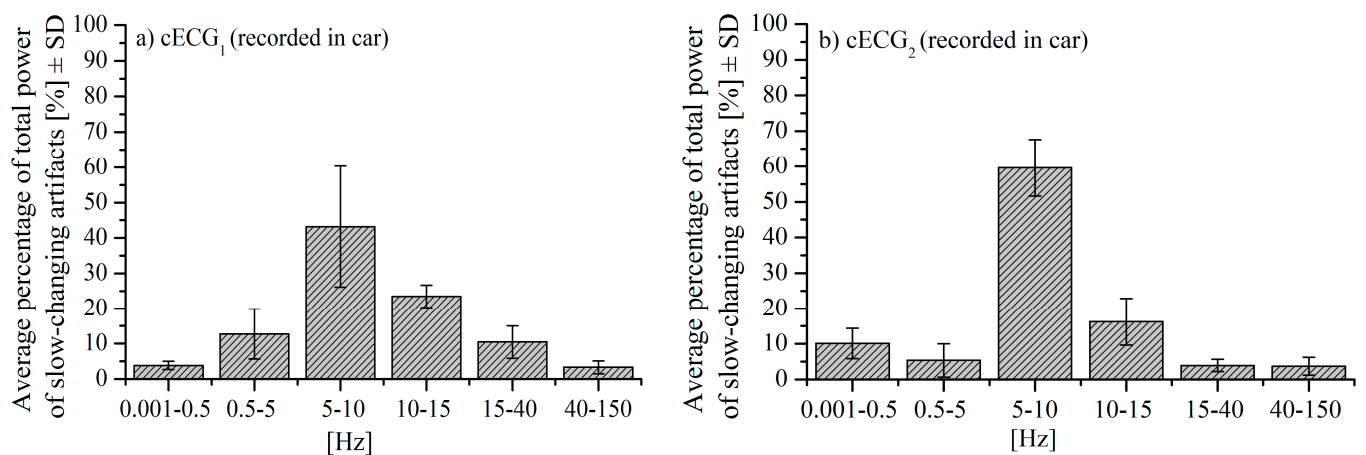
Figure 2. cECG time series recorded during driving: (a) $cECG_1$ with a moderate amount of coarse artifacts; (b) $cECG_1$ without coarse artifacts. Red marks useful segments according to medical experts.

Table 3. Number of cECG per observed groups during driving.

	$cECG_1$	$cECG_2$
cECG with large amount of artifacts	26	25
cECG with moderate amount of artifacts	2	3
cECG without artifacts	3	3

The signal-to-noise ratio is expressed as $SNR = 10 \cdot \log\left(\frac{P_s}{P_n}\right)$, where P_s corresponds to the power of cECG time series comprising only the useful segments according to the expert opinion, and P_n corresponds to the power of all segments that were declared as useless. The signal segment is treated as useful if R peaks are labeled, or as useless if the R peaks are not labeled by medical experts. Its mean value is equal -40.01 ± 33.64 dB for $cECG_1$, and -30.99 ± 23.39 dB for $cECG_2$, during driving. The power of time series is estimated as the sum of squared amplitude divided by length of time series. The negative value of SNR is a consequence of very high amplitude values of the coarse artifacts if compared to the useful parts of the signal. The high SD is a consequence of the different amounts of coarse artifacts in time series.

We also analyzed the power distribution over the frequency bands of slow-changing artifacts in $cECG_1$ and $cECG_2$ during driving. The power distribution averaged over ten signals is shown in Figure 3. The signals comprise segments with slow-changing artifacts manually extracted from $cECG_1$ and $cECG_2$. The ECG spectral components (0.05–150 Hz), on the other hand, are mainly concentrated in the range of 0.05–35 Hz [29] so that the spectral overlap is observed, especially, for the frequency less than 15 Hz. For this reason, classical band-pass filters cannot be implemented for artifact removal [30,31].

**Figure 3.** Mean value of percentage of total power of slow-changing artifacts over frequency bands $\pm SD$: (a) $cECG_1$ during driving; (b) $cECG_2$ during driving.

2.2. Reduction of Artifacts Based on Fluctuation in cECG Time Series

The estimation of signal segment fluctuation was performed following the first part of DFA procedure [32]. The samples of the cECG time series x of length N are denoted by x_k , $k = 1, \dots, N$. In the first step, a vector of cumulative sums $Y(i)$, $i = 1, \dots, N$ is formed from the elements obtained by summing i successive centralized samples of the time series x [32]:

$$Y(i) = \sum_{k=1}^i [x_k - \langle x \rangle] \quad i = 1, \dots, N, \quad (1)$$

where the k -th centralized sample is formed by subtracting the mean value of the time series $\langle x \rangle$ from the k -th sample x_k .

The vector of cumulative sums Y is divided into non-overlapping segments of the same length [32]. Since the length of the vector of cumulative sums N need not be divisible

by the number of segments, the last segment is usually shorter and needs to be omitted. This is not a problem, as the time series are very long (Table 2), so removing one segment would not compromise the reliability of the results. In the case of short time series, the segmentation procedure is repeated twice, the first from the beginning, and then starting from the end of the time series. The total number of observed segments is doubled. In this way, the reliability of result would not be compromised. The short time series is defined as series with less than 10,000 samples [33]. In the case of driving, it is only 10 s of recording. So, the procedure of repeated segmentation from the other end of time series is not applied.

The vector of cumulative sums $Y(i), i = 1, \dots, N$, is divided into N/SL segments of length SL . The segments are denoted as $Y_j(k), j = 1, \dots, \frac{N}{SL}$, where k denotes the samples within a particular segment, $k = 1, \dots, SL$. Each segment is approximated by polynomial of the v -th order $p_{j,v}$ that represents the trend of segment number j . Subtracting a trend from a segment leads to a detrended segment [32]:

$$Y_{j,SL}(k) = Y_j(k) - p_{j,v}(k), j = 1, \dots, \frac{N}{SL}, k = 1, \dots, SL, \quad (2)$$

and

$$p_{j,v}(k) = a_v \cdot Y_j^v(k) + a_{v-1} \cdot Y_j^{v-1}(k) + \dots + a_0, \quad (3)$$

where a_v, a_{v-1}, a_0 polynomial coefficients on a segment; v polynomial order.

The most common polynomial that is used for this method is linear ($v = 1$) [34], so we implemented linear approximation.

The detrended fluctuation analysis function $F_D(j)$ of one segment of the time series is calculated as the sum of the square value of the difference between the original value of the time series and the trend of a given segment divided with SL [32]:

$$F_D(j) = \sqrt{\frac{1}{SL} \cdot \sum_{k=1}^{SL} \{ Y_{j,SL}^2(k) \}}, j = 1, \dots, \frac{N}{SL}. \quad (4)$$

The basic idea of the proposed method is to use the estimated value of fluctuation F_D for the artifacts reduction. The F_D values of extracted part of raw cECG signals recorded while driving is presented in the upper panel of Figure 4. Figure 4 shows an example of an extracted part of raw cECG from an available database, which is intended to show isolated characteristic segments for analysis. As expected, larger value of $F_D(j)$ was obtained for the segments with coarse artifacts, due to the expected larger deviation from the linear trend of segments. However, we should be careful about establishing criteria for an acceptable level of fluctuation, as the detrended fluctuation $F_D(j)$ of useless segments may be comparable to the correct segments. The range of F_D value depends on the time series, as described in detail in the next section.

Figure 4 shows that R peaks detected by OSEA software (marked by black rectangles) are in accordance with experts' opinions in useful segments. In addition, the amplitudes of the useful parts are very small compared to the coarse artifacts.

Unfortunately, the amplitudes of the useful parts are comparable to the parts of the signal with slow-changing artifacts in which the cECG was not detected. One example of such a case is isolated in Figure 1c. Figure 1c shows enlarged part of useful segments of the signal shown in Figure 1a, marked by brown a rectangle, and the useless segment by blue rectangle of comparable amplitude. So, artifact reduction cannot be performed on amplitude values alone.

Appropriate cECG pretreatment, shown in Figure 1, would contribute to greater accuracy of R peak detection algorithms, as well as parameters derived from cECG. So, it presents the first step in developing an auxiliary tool that, based on the parameters extracted (from other sources, not only cardiovascular), would detect possible fatigue in the driver and trigger an alarm that would warn him.

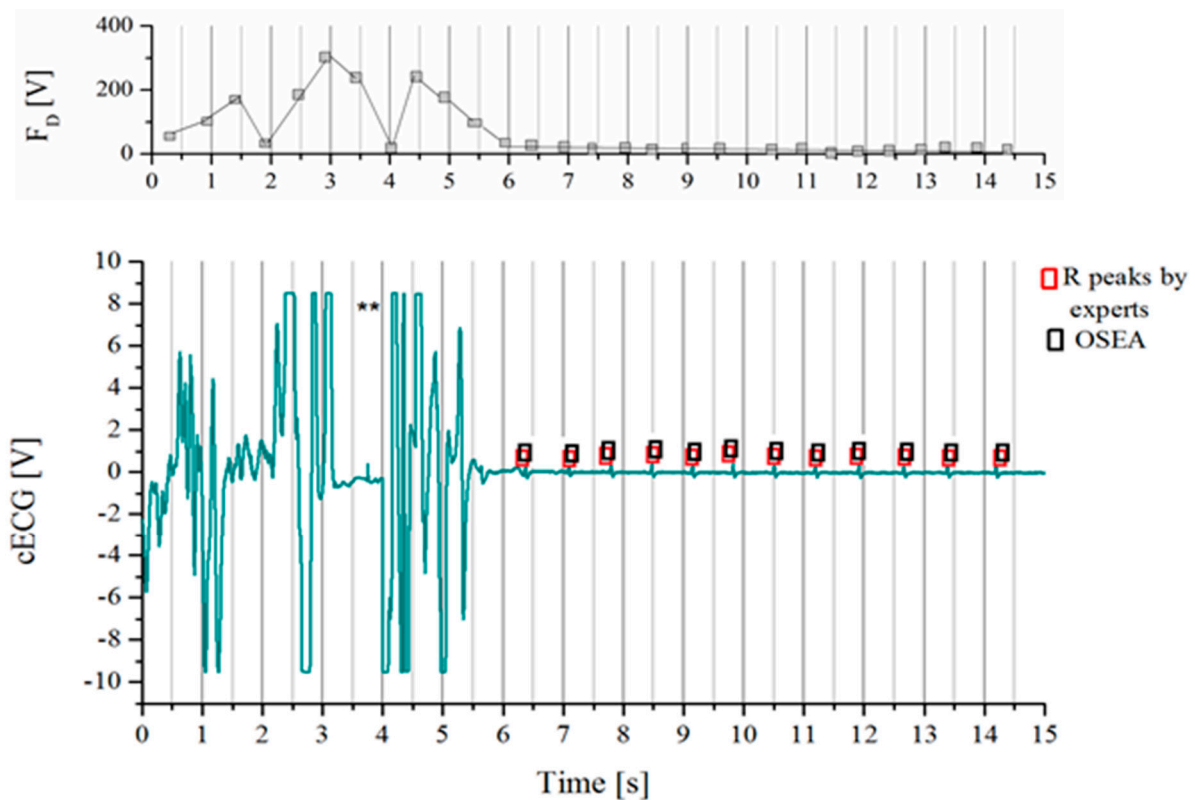


Figure 4. Example of extracted part of raw *cECG* time series recorded while driving. Red rectangles indicate *R* peaks of *cECG* signals marked by medical experts, black rectangles indicate peaks detected using OSEA software. Gray vertical lines indicate segments for which the value of the estimated fluctuation F_D is shown in the upper panel. Asterisks (**) indicate characteristic segments that appear next to segments with coarse artifacts.

Method for Artifacts Reduction

In the first step of the algorithm, the time series x should be divided into a non-overlapping segments of length SL . F_D is estimated according to Equation (4) for each segment.

Before artifact reduction, it is necessary to check the presence of coarse artifacts in the time series. Namely, the existence of the coarse artifacts is not known a priori. The parameters that influenced the formation of the criteria for checking the presence of coarse artifacts are the maximum and minimum value of time series x ($\max(x)$, $\min(x)$, respectively), and the square root of the second moment M . Comparing time series with a large amount and the moderate amount of coarse artifacts to time series without coarse artifacts (examples are given in Figure 1a, Figure 2a,c), we note that the difference between the maximum and minimum value of time series x , $\max(x) - \min(x)$, was larger for signals with coarse artifacts compared to signals without coarse artifacts. Subtracting the square root of the uncentralized second moment M , $M = E(x^2) = \frac{1}{N} \sum_{k=1}^N x_k^2$, from this difference made it possible to distinguish between signals with a large amount of coarse artifacts and a moderate amount of coarse artifacts, as the value of M is larger for time series with more coarse artifacts.

The final value of criterion for the presence of coarse artifacts was determined experimentally:

$$((\max(x) - \min(x))/2) - \sqrt{M} > 1, M = E(x^2) = \frac{1}{N} \sum_{k=1}^N x_k^2, \quad (5)$$

with $\max(x)$, $\min(x)$ -maximum, and minimum value of time series x , respectively, and M -the second moment of time series x .

Figure 5 shows the value of Equation (5) for all three observed groups of cECG (Table 3). The value of Equation (5) is lower than 1 for signals without the presence of coarse artifacts (below the gray line in Figure 5).

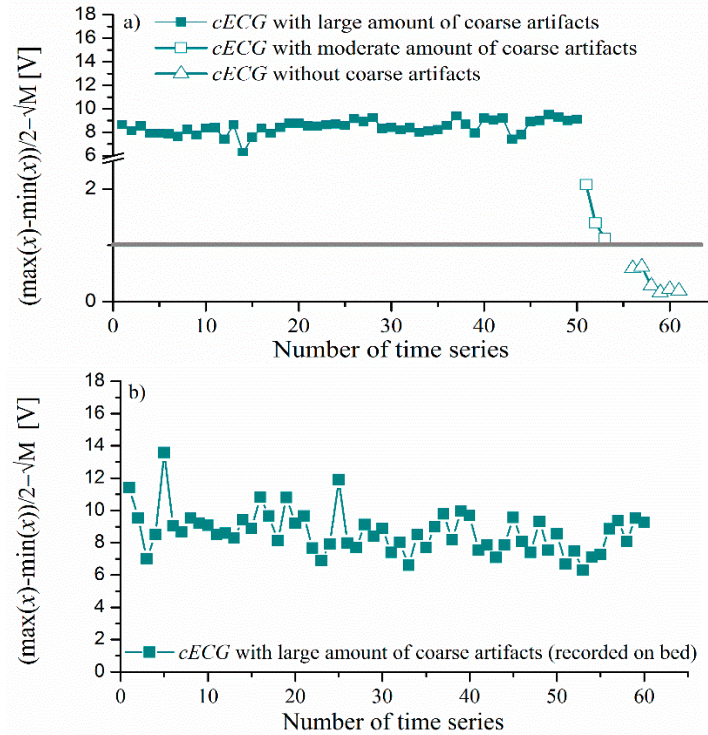


Figure 5. Difference between half of the difference maximal and minimum amplitude of the time series, and the square root of the second moment (a) $cECG_1$ and $cECG_2$ time series recorded while driving, (b) $cECG_1$, $cECG_2$, and $cECG_3$ time series recorded while lying in bed.

For a large and moderate groups, the value of Equation (5) is larger than 1, with a notable distinction between these two groups. All the cECG time series recorded during lying on the bed fulfilled the condition for the presence of a large amount of artifacts (Equation (5)), as expected (Figure 5b).

If the condition of Equation (5) is fulfilled, artifacts from time series should be reduced. To reduce the presence of coarse and slow-changing artifacts, we have developed a set of formulae for automatically estimating the level of detrended fluctuation of time series segments as the criteria for the useful or useless segments.

The first threshold TH_1 is intended to reduce the coarse artifacts.

Threshold value TH_1 is equal to:

$$TH_1 = \left(\sqrt{((\max(x) - \min(x))/2)^2 - M} \right) \cdot C \cdot \text{median}(F_D) + \frac{(SD(F_D) + SD(x))}{SD(F_D) \cdot SD(x)} \cdot C_1, \quad (6)$$

where $\text{median}(F_D)$ is the median value of detrended fluctuation function of all segments in time series, $SD(F_D)$ and $SD(x)$ are the standard deviation of F_D and x , respectively, $\max(x)$, $\min(x)$ -maximum, and minimum value of time series x , respectively, and M - the second moment, while value C is constant value from the range, $C \in \{0.15-0.35\} \frac{1}{\sqrt{2}}$ and $C_1 = 1$ V.

Figure 6a shows $\text{median}(F_D)$ for cECG record during lying on the bed (gray) or driving in the car (dark green). The median value is higher for signals with a large presence of artifacts (marked by filled squares) compared to signals with moderate (marked by unfilled squares) or no artifacts (marked by unfilled triangle) in both signal groups. The obtained results are in accordance with the expectations motivated by Figure 4, where higher values

of $F_D(j)$ were noticed in the segments in which coarse artifacts are present. In Figure 6b,c, we note that the values of $SD(F_D)$ and $SD(x)$ for signals with a moderate amount of artifacts or without artifacts are smaller compared to the standard deviation of time series with a large amount of coarse artifacts, which is also in line with expectations. $SD(F_D)$ and $SD(x)$ have a larger impact on the final value of TH_1 value, while the influence of the $median(F_D)$ is negligible for the signal with a moderate amount of artifacts. In the case of signals with a large amount of coarse artifacts, the influence of $median(F_D)$ is larger compared to $SD(F_D)$ and $SD(x)$.

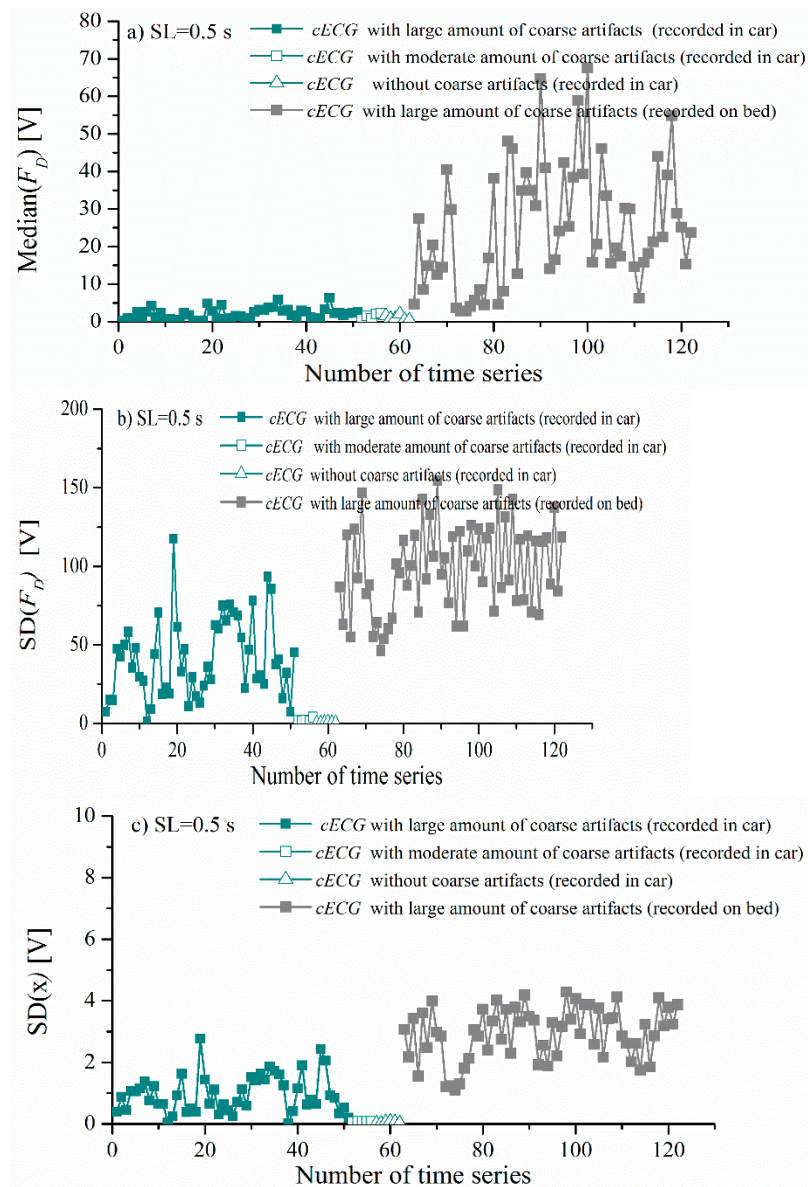


Figure 6. Statistical parameters of the recorded *cECG* time series for different recording conditions: driving car, lying in bed. (a) Median value of F_D , (b) standard deviation $SD(F_D)$, and (c) standard deviation of *cECG* time series.

The threshold evaluation includes an empirical parameter C . To find the most suitable value, we analyzed the percentage of preserved R peaks and the percentage of eliminated useless parts of the signal that might generate false. The results are presented in Figure 7, for the range of values $C \in \{0.05 \text{ to } 1\}$, and for the segment length $SL = 0.5$ s. The gray rectangle with C values from 0.15 to 0.35 indicates the range of values for which the best performances are achieved. Visual inspection of the *cECG* after the reduction of artifacts

can determine the presence of coarse artifacts. We used strict criteria to assess the presence of coarse artifacts, and a time series with at least one coarse artifact is treated as a time series in which coarse artifacts are not successfully reduced. Within this range, 90% to 97% of useful signal parts are preserved, and 96% to 100% coarse artifacts are eliminated. The recommended value is the median point, $C = 0.25$. R peak annotations were available for *cECG* groups recorded while driving and in bed.

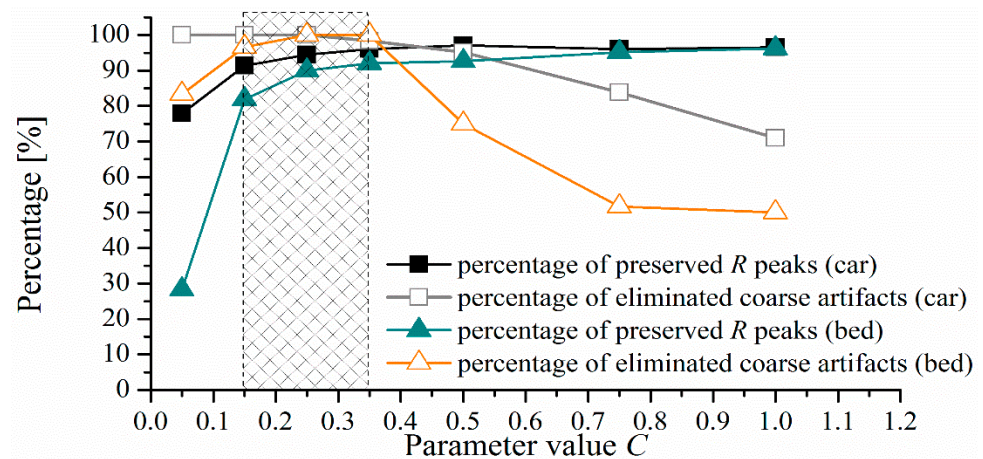


Figure 7. Influence of empirical parameter C on percentage of preserved R peaks marked by medical experts and percentage of eliminated coarse artifacts for $SL = 0.5$ s.

After elimination of all segments (i.e., excluding segments from further analysis) that fulfilled condition $F_D(j) > TH_1$, we check adjacent segments for the possibility of coarse artefacts partially spilling over the adjacent segments (an example of such segments is marked with asterisks in Figure 4). To be on the safe side, $F_D(j)$ values of adjacent segments are compared with the TH_2 .

$$TH_2 = \frac{TH_1}{2}. \quad (7)$$

If $F_D(j)$ values exceed half of TH_1 value, this segment is declared as useless, and it is eliminated from *cECG*.

The third threshold, TH_3 , has a role to eliminate segments with very small deviation from linear trend, i.e., slow-changing artifacts. If the square difference between the value of the sample, and the estimated trend of sample is equal to or less than 0.01, and if this condition is fulfilled for all samples in the segment, that segment is not treated as a carrier of useful information. In that case, (Equation (4)) is equal to $F_D(j) = \sqrt{\frac{1}{SL} \cdot SL \cdot 0.01} = 0.1$, so

$$TH_3 = 0.1. \quad (8)$$

The problem of elimination of this type of artifact is expressed in *cECG* recorded in cars, where the signal intensity is very low (Table 1), so, in this way, it is possible to eliminate significant parts of the useful signal (compare part of *cECG* marked by blue and red rectangles in Figure 1c). For these reasons, additional protection was introduced, and the comparison with the TH_3 threshold is made only if the difference between the mean value of F_D and $SD(F_D)$ is greater than TH_3 . In this way, the possibility of an incorrect elimination of the useful signal segments is reduced. In the database [28], there is no manual notation of slow-changing artifacts, and, since they are comparable to a useful part, it is difficult to be identifiable visually. For these reasons, the success in eliminating slow-changing artifacts is observed by the overall accuracy of the algorithm, i.e., by comparing the useless segments according to the assessment of the algorithm with the useless segments marked by the opinion of medical experts.

The pseudocode explaining Algorithm 1 is shown below.

Algorithm 1: Reduction of Artifacts in *cECG* Time Series

Input: *cECG* time series

1. Divide the signal into non-overlapping segments $SL = 0.5$ s.
2. Estimate the fluctuation for each segment
3. Form array F_D from estimated fluctuation for each segment
4. Determine: the minimum value $cECG_{MIN}$, maximum value $cECG_{MAX}$, and the second uncentered moment M
5. Calculate TH_1
6. Calculate TH_2
7. Calculate TH_3
8. **if** the difference between difference of half $cECG_{MAX}$ and $cECG_{MIN}$ and square root of M is less than 1
 - if** element in array F_D is larger than TH_1 **do**
 - eliminate observed segment in *cECG*
 - if** difference between current element in array F_D and next element larger than TH_2
 - eliminate next segment in *cECG*
 - end if**
 - if** difference between current element in array F_D and previous element larger than TH_2
 - eliminate previous segment in *cECG*
 - end if**
- end if**
- end if**
9. Calculate the mean value of F_D and standard deviation of F_D
10. **if** difference between the mean value of F_D and $SD(F_D)$ is larger than TH_3 **do**
 - if** element in array F_D is less than TH_3 **do**
 - eliminate observed segment in *cECG*
 - end if**
- end if**

Output: *cECG* time series with reduced artifacts

2.3. Binarized Entropy (BinEn)

We also analyzed a method that does not require artifacts removal. Such methods are rare, almost non-existent. Binarized entropy (BinEn) [35] is one of them, developed for another harsh environment—mobile crowdsensing systems—where the reduction of artifacts is not feasible. A brief recapitulation of BinEn adapted to a single data set is below.

In the first step, time series x are binary differentially encoded and split into m -sized binary vectors [35]:

$$c = \left\{ \begin{array}{l} 0 \ x_{i+1} - x_i \leq 0 \\ 1 \ x_{i+1} - x_i > 0 \end{array} \right\} \ i = 1, \dots, N$$

$$C_m^i = [c_i, c_{i+\tau}, \dots, c_{i+(m-1)\cdot\tau}], \ i = 1, \dots, N - (m - 1)\cdot\tau, \tag{9}$$

where the delay τ is distances the elements of the vector from each other, and m is size of vectors. In most applications $\tau = 1, m \in \{1, 2, 3, 4\}$ [35]. In the BinEn, the vectors are binary, so the number of different vectors is 2^m , and each vector can be assigned a decimal number k [35]:

$$k = \sum_{n=0}^{m-1} c_{i+n\cdot\tau} \cdot 2^n. \tag{10}$$

$N_C^{(m)}$ represents the number of occurrences of a certain vector series in the observed time series C [35]:

$$N_C^{(m)}(k) = \sum_{i=1}^{N-(m-1)\cdot\tau} I\left\{\sum_{l=0}^{m-1} c_{i+l\cdot\tau} \cdot 2^l = k\right\}, \quad k = 0, 1, \dots, 2^m - 1. \quad (11)$$

$I\{\}$ —indicator function equal to one if the condition is met, and zero otherwise. The estimation of probability mass function of observed vectors in C is equal to:

$$\hat{P}_C^{(m)} = \frac{N_C^{(m)}(k)}{N - (m - 1) \cdot \tau}. \quad (12)$$

In the following step, it is necessary to find the distance d between each pair of vectors. Distance d is calculated according to the Hamming distance [35]:

$$d(C_m^i, C_m^j) = \sum_{k=0}^{m-1} c_{i+k\cdot\tau} \oplus c_{j+k\cdot\tau} = \sum_{k=0}^{m-1} I\{c_{i+k\cdot\tau} \neq c_{j+k\cdot\tau}\} \\ i, j = 1, \dots, N - (m - 1) \cdot \tau, \quad (13)$$

where \oplus notes ex-or logic function, and $I\{\cdot\}$ indicators function. The distance $d(C_m^i, C_m^j)$ between the vectors is a discrete variable that can have one of the $m + 1$ values, that is, $d(C_m^i, C_m^j) \in \{0, 1, \dots, m\}$ [35].

The matrix of Hamming distance is denoted by $H(m)$ [35]. Elements of matrix H are the distance between the vector whose decimal represents k and the vector whose decimal represents n , $(h_{k,n})$ [35]. The probability that vector C_m^i occurs in C is estimated based on the value in matrix H , which gave information on which vectors are in distance less than r from C_m^i , and Equation (11) that gave information about a number of vectors that are at the same distance [35]:

$$\hat{p}_k^m(r) = Pr\left\{d(C_m^i, C_m) \leq r\right\} \\ = \frac{1}{N-(m-1)\cdot\tau} \cdot \sum_{n=0}^{2^m-1} N_C^{(m)}(n) \cdot I\{h_{k,n}^{(m)} \leq r\} \\ = \sum_{n=0}^{2^m-1} \hat{P}_C^{(m)}(n) \cdot I\{h_{k,n}^{(m)} \leq r\}. \quad (14)$$

In the next step, value of summand $\hat{\Phi}$ is calculated as average of logarithm \hat{p}_k^m [35]:

$$\hat{\Phi}^m(r, N, \tau) = \frac{1}{N-(m-1)\cdot\tau} \cdot \sum_{k=0}^{2^m-1} N_C^{(m)}(k) \cdot \ln(\hat{p}_k^m(r)) \\ = \sum_{k=0}^{2^m-1} \hat{P}_C^{(m)}(k) \cdot \ln(\hat{p}_k^m(r)). \quad (15)$$

The final value of *BinApEn* is estimated on model of approximate entropy (detail in Reference [36]) [35]:

$$BinApEn(m, r, N, \tau) = \hat{\Phi}^m(r, N, \tau) - \hat{\Phi}^{m+1}(r, N, \tau). \quad (16)$$

BinSampEn is a binarized version of sample entropy (proposed in Reference [37]), which excludes self-similarity (comparison of vectors with themselves) [35]:

$$BinSampEn(m, r, N, \tau) = -10 \cdot \log\left(\frac{\hat{\Phi}^{m+1}(r, N, \tau)}{\hat{\Phi}^m(r, N, \tau)}\right). \quad (17)$$

2.4. Classifiers

The K nearest neighbors (KNN) algorithm is a simple supervised machine learning algorithm that classifies data based on estimates of K the nearest neighbors [38]. The K

nearest neighbors are found by the distance between test and training objects in feature space [38]. The test object is classified into the appropriate class, in which the majority of K neighbors belong [38]. We used Euclidean distances to determine K nearest neighbors, the number of observed classes is two (driving in the city and driving on open roads). The number of K is selected by cross-validation of 10% of training data sets ($K = 5$).

The Deep Dense Neural Network (DDNN) is a deep learning technique that includes an input, output layer and a fully connected layer between those two layers [39]. We used model custom architecture (15 fully connected layers with 128 units, followed by drop layer output). The proposed architecture is very simple due to the existence of only two classes and a small database (details described in Section 2.5). We used an Adam optimizer [40] during DDNN training and cross-entropy as loss function [41]. The drop layer output is used for the regularization procedure to reduce overfitting to the training data set. The purpose of the experiment was to test the sensitivity of the classifier to the presence of artifacts in the signal.

2.5. Statistical Analysis

Some of the illustrative results are presented as graphs showing mean \pm standard deviation. Statistical significance between observed groups was checked by t -test for paired samples in MATLAB R2013a. We used significance level $p < 0.01$ for all compared groups.

To form a database of appropriate sizes for testing and training KNN and DDNN, we divided the $cECG_2$ time-series recorded while driving (large duration of time series, Table 2) into non-overlapped parts with a duration of 50 s. We opted for $cECG_2$ because it has a better recording quality compared to $cECG_1$ (higher amplitude, Table 1) and manual notation of useful segments available in Reference [28]. The total number of signals was 648, out of which 70% were used for the training set, and 30% for the test set. Data with a large amount of coarse artifacts, a moderate amount of coarse artifacts, and without coarse artifacts are uniformly arranged into a training set and a test set. For validation, we used only 10% of the training data because of the small size database. In the case of DDNN, the model has converged after several hundred epochs, so the number of observed epochs was set to 500. There are 2 classes in total, driving in the city and on the open road. We used a list of features: the value of the R peak, the HR (estimated as inversion of time intervals between adjacent R peaks), the BinEn of cECG time series, and pNN (percentage of successive normal cardiac interbeat intervals), that corresponds the percentage of RR intervals greater than 50 ms after reduction of artifacts (details are described in Reference [42]).

Classification performance is calculated according to the following expressions:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}, \quad (18)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \quad (19)$$

$$\text{Specifity} = \frac{TN}{FP + TN}, \quad (20)$$

$$\text{Positive prediction} = \frac{TP}{TP + FP}, \quad (21)$$

$$\text{Negative prediction} = \frac{TN}{TN + FN}. \quad (22)$$

The classification performance was tested in the context of quantifying the success of recognition of the driving location—open roads or city. In this context, TP denotes the number of cases correctly identified as driving in the city, FP denotes the number of cases incorrectly identified as driving in the city, TN denotes the number of cases correctly identified as driving in open roads (highway or polygon), and FN denotes the number of cases incorrectly identified as driving in open roads (highway or polygon).

In the context of quantifying the success of the artifact reduction by the proposed method, TP denotes the number of correctly identified useful segments (marked by medical experts), TN denotes the number correctly identified useless segments (segments without R peaks and categorized as artifacts according to experts), FP denotes the number of segments incorrectly identified as useless, and FN denotes the number of segments incorrectly identified as useful.

3. Results and Discussion

Figure 8a,b comparatively show the mean value of the percentage of eliminated time series by the proposed method, and the percentage of eliminated time series according to the notion of medical experts. These results are in excellent accordance for all three groups of $cECG_1$, $cECG_2$, and $cECG_3$ time series recorded while lying on the bed (Figure 8a). The high percentage of eliminated segments in time series while lying in bed is due to the movement of volunteers required by protocol. The high values of standard deviations shown in Figure 8 show a great variability of the amount of artifacts in the recorded signals. Figures 1a and 2 show that the $cECG$ time series contain different amounts of artifacts during driving. Besides, the controlled movement of the volunteers in an experiment driving car affects the electrodes in the upper part of the body more than it affects other electrodes [1].

The difference between the eliminated artifacts and medical expert's opinion is larger for $cECG$ time series recorded during driving, due to the slow-changing artifacts (Figure 8a). The elimination of slow-changing artifacts is complicated by the low intensity of the recorded signal while driving (Table 1), which makes the difference between a useful and a useless segment imperceptible (examples are marked in Figure 1c).

Figure 8b shows the high percentage of preserved R peaks marked by medical experts. Included in the analysis are all $cECG$ time series that are labeled by medical experts. Unfortunately, notation of R peaks for $cECG_3$ time series recorded during driving are not available in Reference [28]. As a consequence of the fact that useful parts of the signal are comparable to the parts of the signal with slow-changing artifacts, slow-changing artifacts are partly survived, especially, while driving a car. Fortunately, the presence of slow-changing artifacts does not significantly affect the estimation of the mean value of the absolute amplitude $cECG$, which is confirmed in Figure 8c.

To check the presence of a significant difference between mean value of absolute amplitude of $cECG$ after the elimination of the artifact by the proposed method and after the elimination of the artifact following the opinion of experts, we used a t -test for paired samples. The presence of statistical significance was observed between the group of raw recorded signals and signals after artifact elimination by the proposed method (Figure 8c, marked *), as well as to signals after artifact elimination, by the opinion of medical experts (Figure 8c, marked #). There is no statistical significance between the signals after the reduction of artifacts by the proposed method and following the opinion of experts.

Figure 9c,d show examples of the $cECG$ time series after artifacts reduction by the proposed method in the car and on the bed, respectively.

Figure 10a,b show the percentage of preserved useful segments (segments with R peaks according to medical experts), the percentage of eliminated useless segments (segments without R peaks and categorized as artifacts according to experts), and overall accuracy (Equation (18)), depending on length SL . A high level (95–100%) of preservation of the useful segment of time series is achieved, as well as reduction of coarse artifacts (98–100%), while the percentage of eliminated R peaks is around 10% (except for $cECG_1$ recorded during driving with the lowest amplitude intensity) for a length of 0.5 s (Figure 10c–d) for all signal groups. Overall, accuracy is slightly lower for time series recorded while driving (Figure 10a), due to the impossibility of eliminating slow-changing artifacts as a consequence of the very low intensity of recorded time series (Table 1). In addition, it has been shown that slow-changing artifacts, which are not eliminated from the time series, do not significantly affect changes in statistical parameters (Figure 8c). The same problem was observed in Reference [1], where the impossibility of detecting QRS complex in segments that are comparable to noise.

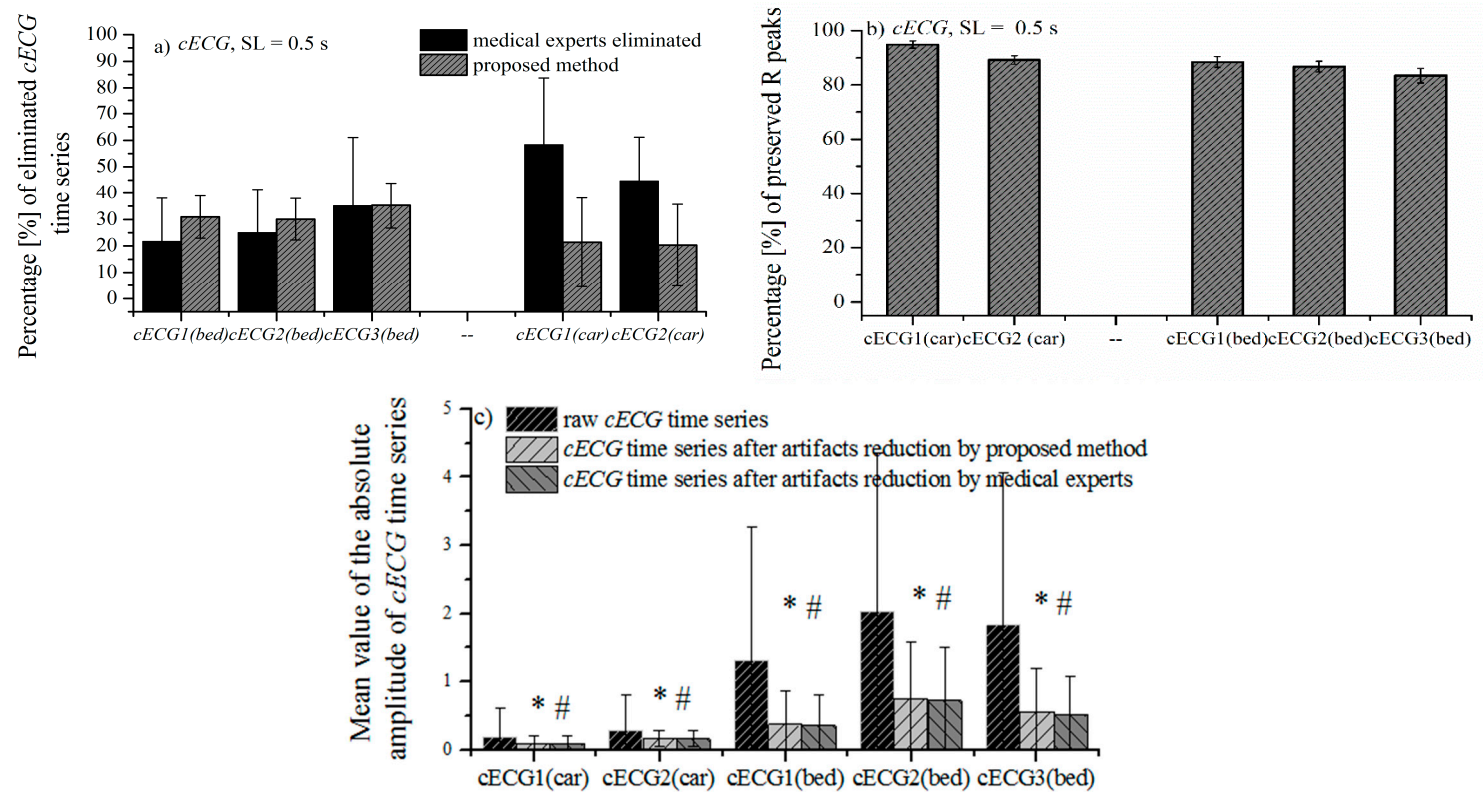


Figure 8. Comparative presentation of the mean value \pm SD percentage of eliminated time series by the proposed method and by the opinion of experts: (a) *cECG* time series recorded while lying on the bed, (b) *cECG* time series recorded while driving and (c) mean values of absolute amplitudes of *cECG* time series \pm SD. Statistical significance is observed between the raw *cECG* and *cECG* after elimination of the artifacts by the proposed method (marked *), as well as the *cECG* after elimination, according to the experts (marked #). Statistical significance does not exist between the group of signals after elimination of the artifact by the proposed method and in the opinion of experts. We used a *t*-test for paired samples, with significance levels $p < 0.01$ for all compared groups.

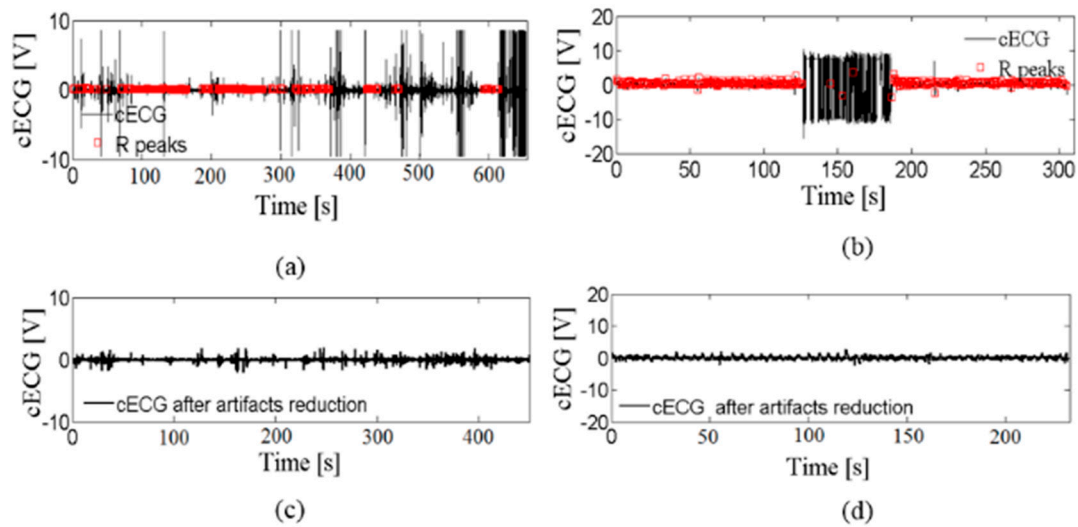


Figure 9. (a) Raw $cECG_1$ time series recorded while driving, (b) $cECG_3$ time series recorded while lying on the bed, (c) $cECG_1$ time series after elimination of artifacts by the proposed method, and (d) $cECG_3$ time series after elimination of artifacts by the proposed method.

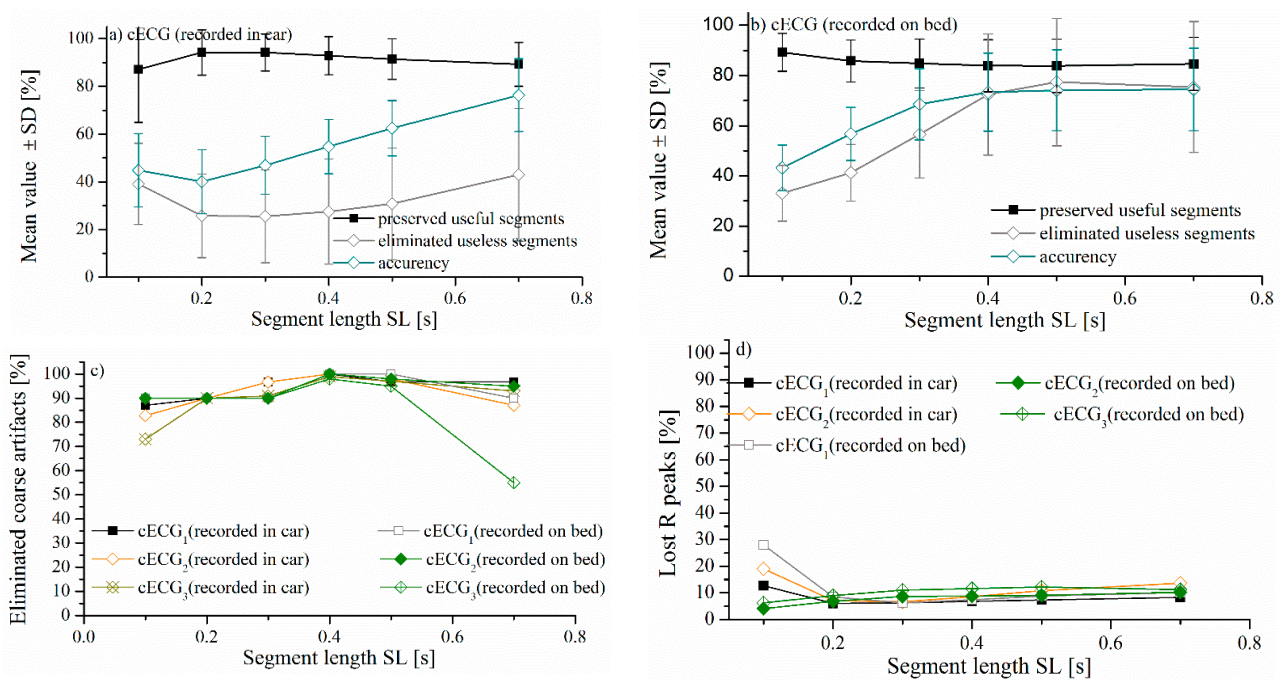


Figure 10. Demonstration of the success of eliminating artifacts and preserving the useful part of the signal by the proposed method depending on SL. (a) $cECG_1$, $cECG_2$ recorded in the car, (b) $cECG_1$, $cECG_2$, and $cECG_3$ while lying on the bed, (c) the percentage of time series in which the coarse artifacts are fully eliminated, depending on the SL, and (d) the percentage of lost R peaks after artifacts reduction, depending on the SL.

To the best of our knowledge, there are two methods for reducing cECG signaling artifacts of mobile subjects (described in detail in References [1,18]). These methods estimate the duration of the interval between R peaks (detected by OSEA) in raw cECG and reject atypical values as artifacts. In Reference [18], the HR value that is larger than 120 beats/min and lower than 30 beats/min was treated as artifacts. The authors pointed out that, for real application, the procedure should be improved [18]. In Reference [1], the possibility of reliable estimation of HR during driving is investigated. The proposed method is based

on QRS detection in raw cECG times series by OSEA software. It was noted that many false-positive QRS were detected by OSEA, as a consequence of the shaped pulse of cECG that is very similar to QRS complexes, so specific boundaries have been introduced [1].

A comparative analysis of the results of the proposed method the existing algorithms is not possible. Our method eliminates the artifacts before extracting the parameters, such as HR and QRS, and, more importantly, without predefined ranges. Thus, our method enables detection of potential cardiovascular pathology from the corrected signals, which is not possible in methods based on predefined ranges.

Statistical significance of entropy as a measurement of the complexity and unpredictability of time series [43] was observed between groups of ECG recorded with and without disturbing the driver while driving [44]. We test the possibility of applying binarized approximated entropy (BinEn), a method developed for entropy estimation on signals that do not require artifact elimination. We estimated BinEn for approximate entropy (BinApEn) and sample entropy (BinSampEn) for different groups of parameters. Statistical significance was observed in estimating BinApEn and BinSampEn between the raw signal and the signal after artifact removal by the proposed method only for parameters ($m = 3, r = 1$) for the $cECG_1$ group recorded in the car (time series with the smallest amplitude in Table 1). For other groups of signals with higher value of amplitudes, no statistical significance was noticed. The possibility for BinEn to distinguish between cECG recorded while driving in the city and on the open road (highway and polygon) after artifacts reduction is shown in Figure 11e.

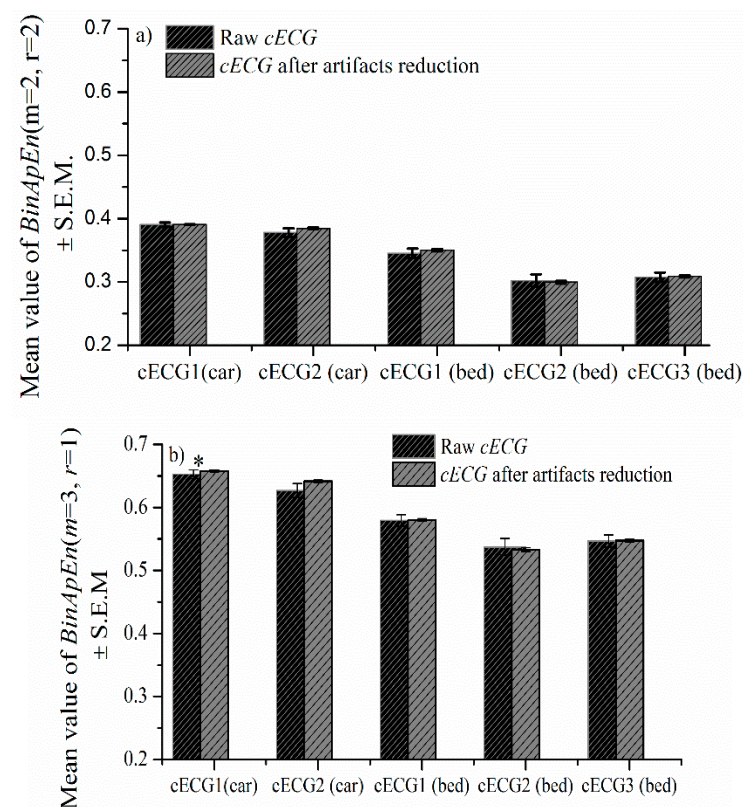


Figure 11. Cont.

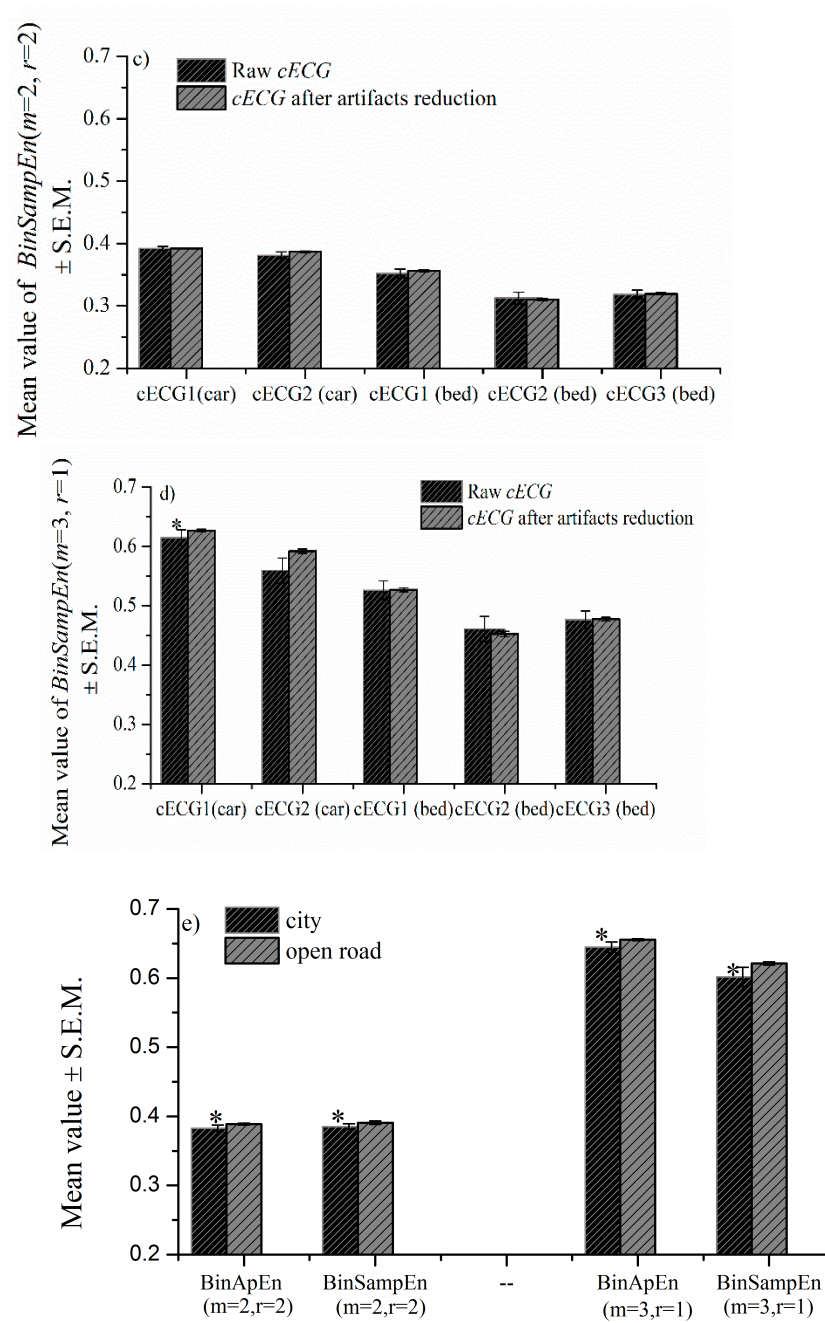


Figure 11. $BinApEn$ and $BinSampEn$ for $cECG$ of all observed groups. (a) Mean value of $BinApEn$ ($m = 2, r = 2$) \pm SD, (b) mean value of $BinApEn$ ($m = 3, r = 1$) \pm SD, (c) $BinSampEn$ ($m = 2, r = 2$) \pm SD, (d) $BinSampEn$ ($m = 3, r = 1$) \pm SD and (e) mean value of $BinApEn$ \pm SD and $BinSampEn$ \pm SD for ($m = 2, r = 2$) and ($m = 3, r = 1$) of the $cECG$ recorded during driving in the city and in the open road. Statistical significance between $BinEn$ of $cECG$ after elimination of the artifacts during driving in the city and the open road is marked by *.

The values of the threshold TH_1 for $cECG$ recorded during driving and while lying on the bed are shown in Figure 12. Introducing a constant threshold or range of values would not lead to an adequate result because the final value TH_1 depends on the statistical parameters of the time series. TH_1 values are slightly lower for time series of lower intensity, recorded in the car, but there are isolated cases whose threshold value is measurable with the time series recorded while lying on the bed (higher intensity; see Table 1). In addition, eliminating coarse artifacts based on the amplitude value (e.g., >4 V is a coarse artifact) would not be a good solution because it is a value that would eliminate partly coarse

artifacts in the case of a cECG recorded while driving, but, for cECG recorded during lying, it would eliminate R peaks (Figure 1a,b).

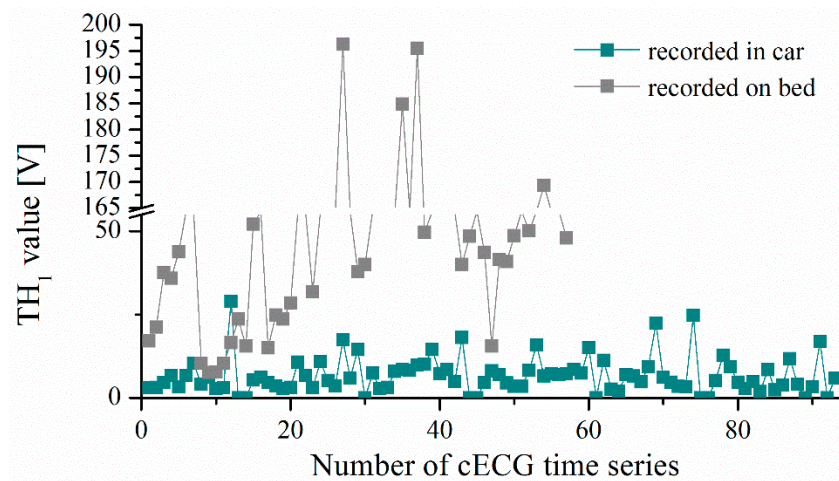


Figure 12. Threshold value of TH_1 for cECG of all observed groups.

We use the KNN technique [38] and DDNN technique [39] to classify driving in city conditions and in the open road (highway or on the proving ground). To investigate the sensitivity of KNN and DDNN on artifacts, we compare results on raw cECG and cECG after artifacts reduction. The list of features consists of the value of the R peak, the HR, and the BinEn of cECG time series. Table 4 shows the impact of artifacts on the accuracy of machine learning techniques. Results of the KNN technique [38] show that 23% accuracy has been improved for cECG time series after artifact elimination (in Table 4, noted as KNN²) and 39% for DDNN techniques [39] (in Table 4, noted as DDNN²). In addition, 55% growth was achieved for sensitivity and for positive prediction of 59%, while, for the DDNN, it was even growth for 73.47%, 63.18%, respectively. The accuracy of KNN classifications has increased after the expansion of the feature list with pNN50, a method which requires the reduction of artifacts (in Table 4, noted as KNN³), but, for DDNN, it was a slight increase of 0.51%. In addition, we can note the increase of all classification performance for KNN³ in comparison to KNN². DDNN is more sensitive to the presence of artifacts compared to KNN, but it is also slightly accurate compared to KNN.

Table 4. Classification performance [%].

ML Technique	Accuracy	Sensitivity	Specificity	Positive Prediction	Negative Prediction
KNN ¹	65.64	14.28	82.87	21.87	74.23
KNN ²	88.21	69.40	94.52	80.95	90.20
KNN ³	92.68	77.55	97.95	92.68	92.86
DDNN ¹	53.33	20.41	64.38	16.13	70.68
DDNN ²	92.31	93.88	91.78	79.31	97.81
DDNN ³	92.82	95.83	91.84	79.31	98.54

KNN¹ for raw cECG. KNN² for cECG after reduction artifacts. KNN³ for cECG after reduction artifacts with extended features with pNN50. DDNN¹ for raw cECG. DDNN² for cECG after reduction artifacts. DDNN³ for cECG after reduction artifacts with extended features with pNN50.

4. Conclusions

The main contribution of this paper is the development of a method for the reduction of artifacts in cECG signals. The detailed analysis of cECG signals publicly available on [28] reveals that medical experts have concluded that about 30% of the signals represent noise, as well as that there are only a few cECG time series without coarse artifacts or with a moderate amount of artifacts. Such domineering presence of artifacts is not aligned

with the theoretical requirements according to which the signals should be stationary and without the artifacts. Besides, transferring the recorded cECG signals to the cloud, without prior reduction of the artifacts, would significantly increase the amount of transmitted traffic, while data processing in the cloud would not be reliable. The proposed method can be applied online. The first 50 s of cECG provide statistical parameters of F_D and the corresponding thresholds. Each subsequent recorded cECG segment affects the expansion of the F_D series of fluctuation values and updates the threshold value. To strengthen the motivation for our work, we have also shown that, even in a small size database, it is noticeable that the presence of artifacts weakens the performance of the classification of KNN and DDNN machine learning techniques, which differentiate urban and other driving conditions. Unfortunately, a small database can affect reliability in estimating the accuracy of classification, especially of DDNN techniques. An alternative to reducing artifacts is to develop methods that are resistant to the presence of artifacts. As an example, we have shown binarized entropy that operates on binary differential coded raw signals and yields good entropy estimates. The pseudocode is given for the easier implementation of the algorithm, and the code is available on request.

Funding: This research was funded by the Serbian Ministry of Education, Science and Technology Development, under Grant 451-03-68/2021-14/200156 (TR32040).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article. Datasets are available on <https://www.medit.hia.rwth-aachen.de/publikationen/unovis/> (accessed on 26 May 2011).

Acknowledgments: The author is grateful to Marian Walter, deputy chief for Medical Information Technology, Helmholtz Institute for Biomedical Engineering for pointing to the database and providing the necessary literature for understanding the procedure for measuring the ECG signal while driving.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wartzek, T.; Eilebrecht, B.; Lem, J.; Lindner, H.; Leonhardt, S.; Walter, M. EKG on the Road: Robust and Unobtrusive Estimation of Heart Rate. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 3112–3120. [[CrossRef](#)]
2. Walter, M.; Eilebrecht, B.; Wartzek, T.; Leonhardt, S. The smart car seat: Personalized monitoring of vital signs in automotive applications. *Pers. Ubiquitous Comput.* **2011**, *15*, 707–715. [[CrossRef](#)]
3. Park, S.J.; Hong, S.; Kim, D.; Seo, Y.; Hussain, I. Knowledge Based Health Monitoring During Driving. In Proceedings of the International Conference on Human-Computer Interaction, Las Vegas, NV, USA, 15–20 July 2018.
4. Park, S.J.; Subramaniam, M.; Hong, S.; Kim, D.; Yu, J. Conceptual Design of the Elderly Healthcare Services In-Vehicle using IoT. In Proceedings of the Conference: WCX™ 17: SAE World Congress Experience, Detroit, MI, USA, 4–6 April 2017.
5. Baek, H.J.; Lee, H.B.; Kim, J.S.; Choi, J.M.; Kim, K.K.; Park, K.S. Nonintrusive biological signal monitoring in a car to evaluate a driver's stress and health state. *Telemed. J. Health* **2009**, *15*, 182–189. [[CrossRef](#)] [[PubMed](#)]
6. Gromer, M.; Salb, D.; Walzer, T.; Madrid, N.M.; Seepold, R. ECG sensor for detection of driver's drowsiness. *Procedia Comput. Sci.* **2019**, *159*, 1938–1946. [[CrossRef](#)]
7. Jung, S.-J.; Shin, H.-S.; Chung, W.-Y. Driver fatigue and drowsiness monitoring system with embedded electrocardiogram sensor on steering wheel. *IET Intell. Transport. Syst.* **2014**, *8*, 43–50. [[CrossRef](#)]
8. Koh, D.-W.; Lee, S.-G. An Evaluation Method of Safe Driving for Senior Adults Using ECG Signals. *Sensors* **2019**, *19*, 2828. [[CrossRef](#)]
9. Jing, D.; Liu, D.; Zhang, S.; Guo, Z. Fatigue driving detection method based on EEG analysis in low-voltage and hypoxia plateau environment. *Int. J. Transp. Sci. Technol.* **2020**, *9*, 366–376. [[CrossRef](#)]
10. Nguyen, T.-H.; Chung, W.-Y. Detection of Driver Braking Intention Using EEG Signals During Simulated Driving. *Sensors* **2019**, *19*, 2863. [[CrossRef](#)] [[PubMed](#)]
11. Hernández, L.G.; Mozos, O.M.; Ferrández, J.M.; Antelis, J.M. EEG-Based Detection of Braking Intention Under Different Car Driving Conditions. *Front. Neuroinform.* **2018**, *12*, 00029. [[CrossRef](#)]
12. Ford Develops Heart Rate Monitoring Car Seat. Available online: <https://www.dicardiology.com/content/ford-develops-heart-rate-monitoring-car-seat> (accessed on 26 May 2011).

13. Perret-Guillaume, C.; Joly, L.; Benetos, A. Heart rate as a risk factor for cardiovascular disease. *Prog. Cardiovasc. Dis.* **2009**, *52*, 6–10. [[CrossRef](#)]
14. Jo, S.H.; Kim, J.M.; Kim, D.K. Heart Rate Change While Drowsy Driving. *J. Korean Med. Sci.* **2019**, *34*, e56. [[CrossRef](#)]
15. Choi, Y.-J.; Lee, J.-Y.; Kong, S. Driver EKG Measuring System With a Conductive Fabric-Based Dry Electrode. *IEEE Access* **2018**, *6*, 415–427. [[CrossRef](#)]
16. Nowak-Brzezińska, A.; Łazarz, W. Qualitative Data Clustering to Detect Outliers. *Entropy* **2021**, *23*, 869. [[CrossRef](#)]
17. Czaplik, M.; Eilebrecht, B.; Walocha, R.; Walter, M.; Schauerte, P.; Leonhardt, S.; Rossaint, R. The reliability and accuracy of a noncontact electrocardiograph system for screening purposes. *Anesth. Analg.* **2012**, *114*, 322–327. [[CrossRef](#)]
18. Wartzek, T.; Czaplik, M.; Antink, C.H.; Eilebrecht, B.; Walocha, R.; Leonhardt, S. UnoVis: The MedIT public unobtrusive vital signs database. *Health Inf. Sci. Syst.* **2015**, *3*, 1–9. [[CrossRef](#)] [[PubMed](#)]
19. Hamilton, P. Open source EKG analysis. In Proceedings of the Computers in Cardiology, Memphis, TN, USA, 22–25 September 2002.
20. Kiencke, U.; Schwarz, M.; Weickert, T. *Signalverarbeitung: Zeit-Frequenz-Analyse und Schätzverfahren*; Oldenbourg Verlag: Munich, Germany, 2008; pp. 1–430.
21. Serteyn, A.; Vullings, R.; Meftah, M.; Bergmans, J.W.M. Motion Artifacts in Capacitive EKG Measurements: Reducing the Combined Effect of DC Voltages and Capacitance Changes Using an Injection Signal. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 264–273. [[CrossRef](#)] [[PubMed](#)]
22. Choi, M.; Jeong, J.J.; Kim, S.H.; Kim, S.W. Reduction of Motion Artifacts and Improvement of R Peak Detecting Accuracy Using Adjacent Non-Intrusive EKG Sensors. *Sensors* **2016**, *16*, 715. [[CrossRef](#)]
23. Chatterjee, S.; Thakur, R.S.; Yadav, R.N.; Gupta, L.; Raghuvanshi, D.K. Review of noise removal techniques in EKG signals. *IET Signal. Process.* **2020**, *14*, 569–590. [[CrossRef](#)]
24. Mert, A.; Akan, A. Detrended fluctuation analysis for empirical mode decomposition based denoising. In Proceedings of the European Signal Processing Conference (EUSIPCO), Lisbon, Portugal, 1–5 September 2014.
25. Škorić, T.; Bajić, D. Redukcija artefakata EKG signala snimanog beskontaktnim elektrodama tokom vožnje. In Proceedings of the International Symposium on INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 18–20 March 2020.
26. Gtec Product. Available online: <https://www.gtec.at/product/> (accessed on 14 November 2021).
27. Philips Product. Available online: <https://www.philips.nl/healthcare/product/HC862451/iu22-xmatrix> (accessed on 14 November 2021).
28. UnoVis: The MedIT Public Unobtrusive Vital Sign Database. Available online: <https://www.medit.hia.rwth-aachen.de/publikationen/unovis/> (accessed on 3 March 2021).
29. Xie, L.; Li, Z.; Zhou, Y.; He, Y.; Zhu, J. Computational Diagnostic Techniques for Electrocardiogram Signal Analysis. *Sensors* **2020**, *20*, 6318. [[CrossRef](#)]
30. Buendía-Fuentes, F.; Arnau-Vives, M.A.; Arnau-Vives, A.; Jiménez-Jiménez, Y.; Rueda-Soriano, J.; Zorio-Grima, E.; Osa-Sáez, A.; Martínez-Dolz, L.V.; Almenar-Bonet, L.; Palencia-Pérez, M.A. High-Bandpass Filters in Electrocardiography: Source of Error in the Interpretation of the ST Segment. *ISRN Cardiol.* **2012**, *2012*, 706217. [[CrossRef](#)] [[PubMed](#)]
31. Lin, C.-H. Frequency-domain features for ECG beat discrimination using grey relational analysis-based classifier. *Comput. Math. Appl.* **2008**, *55*, 680–690. [[CrossRef](#)]
32. Kantelhardt, J.W.; Koscielny-Bunde, E.; Rego, H.H.A.; Havlin, S.; Bunde, A. Detecting long-range correlations with detrended fluctuation analysis. *Phys. A* **2001**, *295*, 441–454. [[CrossRef](#)]
33. Yuan, Q.; Gu, C.; Weng, T.; Yang, H. Unbiased detrended fluctuation analysis: Long-range correlations in very short time series. *Phys. A Stat. Mech. Its Appl.* **2018**, *505*, 179–189. [[CrossRef](#)]
34. Bryce, R.; Sprague, K. Revisiting detrended fluctuation analysis. *Sci. Rep.* **2012**, *2*, 315. [[CrossRef](#)] [[PubMed](#)]
35. Škorić, T.; Mohamoud, O.; Milovanović, B.; Japundžić-Žigon, N.; Bajić, D. Binarized cross-approximate entropy in crowdsensing environment. *Comput. Biol. Med.* **2017**, *80*, 137–147. [[CrossRef](#)]
36. Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)]
37. 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](#)]
38. Mitchell, T.M. Instance-Based Learning. In *Machine Learning*, 1st ed.; McGraw-Hill Science: New York, NY, USA, 1997; pp. 196–230.
39. Hvitfeldt, E.; Silge, J. Dense neural networks. In *Supervised Machine Learning for Text Analysis in R*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2021; pp. 250–282.
40. Kingma, D.P.; Ba, J. *Adam: A Method for Stochastic Optimization*; Proc. ICLR (Poster): San Diego, CA, USA, 2015.
41. Li, L.; Doroslovački, M.; Loew, M.H. Approximating the Gradient of Cross-Entropy Loss Function. *IEEE Access* **2000**, *8*, 111626–111635. [[CrossRef](#)]
42. Shaffer, F.; Ginsberg, J.P. An Overview of Heart Rate Variability Metrics and Norms. *Front. Public Health* **2017**, *5*, 258. [[CrossRef](#)]
43. Škorić, T.; Šarenac, O.; Milovanović, B.; Japundžić-Žigon, N.; Bajić, D. On Consistency of Cross-Approximate Entropy in Cardiovascular and Artificial Environments. *Complexity* **2017**, *2017*, 8365685. [[CrossRef](#)]
44. Yu, L.; Sun, X.; Zhang, K. Driving Distraction Analysis by EKG Signals: An Entropy Analysis. In Proceedings of the International Conference on Internationalization, Design and Global Development, Orlando, FL, USA, 9–14 July 2011.