

,”

Article

Effects of Ballistocardiogram Peak Detection Jitters on the Quality of Heart Rate Variability Features: A Simulation-Based Case Study in the Context of Sleep Staging

Ahmad Suliman [†], Md Rakibul Mowla [‡], Alaleh Alivar [§], Charles Carlson ^{*}, Punit Prakash , Balasubramaniam Natarajan, Steve Warren  and David E. Thompson

Mike Wiegers Department of Electrical & Computer Engineering, Kansas State University, Manhattan, KS 66506 USA; ahmad.suliman@fda.hhs.gov (A.S.); mdrakibul-mowla@uiowa.edu (M.R.M.); ; alaleh.alivar@nyulangone.org (A.A.); prakashp@k-state.edu (P.P.); bala@k-state.edu (B.N.); swarren@k-state.edu (S.W.); davet@k-state.edu (D.E.T.)

* Correspondence: cwcral@k-state.edu

† Current affiliation: Office of Science and Engineering Laboratories, Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD 20993, USA.

‡ Current affiliation: Department of Neurosurgery, Roy J. and Lucille A. Carver College of Medicine, The University of Iowa, Iowa City, IA 52242, USA.

§ Current affiliation: NYU Langone Health, Department of Radiology, 660 First Avenue, 4th Floor, New York, NY 10016, USA.

Abstract: Heart rate variability (HRV) features support several clinical applications, including sleep staging, and ballistocardiograms (BCGs) can be used to unobtrusively estimate these features. Electrocardiography is the traditional clinical standard for HRV estimation, but BCGs and electrocardiograms (ECGs) yield different estimates for heartbeat intervals (HBIs), leading to differences in calculated HRV parameters. This study examines the viability of using BCG-based HRV features for sleep staging by quantifying the impact of these timing differences on the resulting parameters of interest. We introduced a range of synthetic time offsets to simulate the differences between BCG- and ECG-based heartbeat intervals, and the resulting HRV features are used to perform sleep staging. Subsequently, we draw a relationship between the mean absolute error in HBIs and the resulting sleep-staging performances. We also extend our previous work in heartbeat interval identification algorithms to demonstrate that our simulated timing jitters are close representatives of errors between heartbeat interval measurements. This work indicates that BCG-based sleep staging can produce accuracies comparable to ECG-based techniques such that at an HBI error range of up to 60 ms, the sleep-scoring error could increase from 17 % to 25% based on one of the scenarios we examined.

Keywords: ballistocardiogram; Bayes error; classifier; electrocardiogram; heartbeat interval; k-nearest neighbor; R-J interval; R-R interval; support vector machine



Citation: Suliman, A.; Mowla, M. R.; Alivar, A.; Carlson, C.; Prakash, P.; Natarajan, B.; Warren, S.; Thompson, D. E. Effects of Ballistocardiogram Peak Detection Jitters on the Quality of Heart Rate Variability Features: A Simulation-Based Case Study in the Context of Sleep Staging. *Sensors* **2023**, *1*, 0. <https://doi.org/>

Academic Editors: Riccardo Pernice, Raquel Bailón and Massimo W. Rivolta

Received: 7 November 2022

Revised: 23 February 2023

Accepted: 27 February 2023

Published:



Copyright: © 2023 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

Heartbeat intervals (HBIs) vary over time, and that variance can be quantified as a heart rate variability (HRV) feature, a quantity usually based on HBIs determined from R-R intervals in electrocardiograms (ECGs) [1–9]. Such HRV features support many health-related applications, including long-term health monitoring [10,11], sleep staging [1,12–15], and sleep quality assessment [16,17].

Electrocardiography is the gold-standard technique for HRV parameter measurement in clinical settings and, more recently, short-term, home-health monitoring environments, given the advent of wearable devices. For long-term health monitoring, the wires and electrodes required for electrocardiography introduce significant challenges.

A ballistocardiogram (BCG), which is a signal derived from tissue micromovements caused by circulatory dynamics in the body, is an alternative source for HRV parameter

estimation. A BCG-based measurement is preferable, in some settings, to an ECG-based measurement because a BCG does not require sensors to be attached to the body. However, the HBIs (and corresponding HRV features) obtained from a BCG are numerically different from those obtained from an ECG [2–9]. We refer to the differences between BCG- and ECG-based HBI and HRV values as HBI and HRV errors, respectively. This terminology reflects the fact that ECG-based HBI and HRV parameters are considered gold standards in this arena. Figure 1 illustrates several aspects of the ECG and BCG waveforms over a few cardiac cycles, demonstrating how even simultaneously recorded ECGs and BCGs can lead to different HBI estimates.

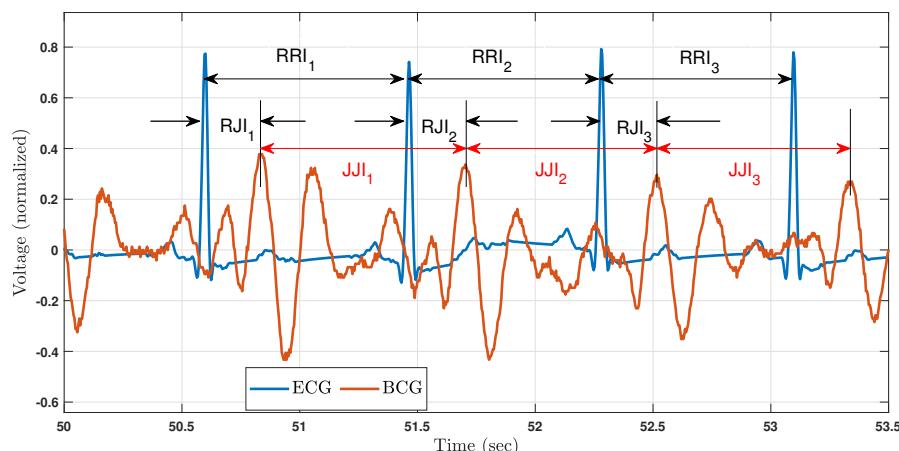


Figure 1. Heartbeat interval estimates from simultaneous ECG and BCG recordings. An R–R interval (RRI) represents the time between consecutive R peaks—the prominent peaks of an ECG that correspond to the onset of ventricular depolarization events. Similarly, a J–J interval (JJI) represents the time between consecutive BCG J peaks, where J peaks are the prominent peaks of a BCG, believed to be generated by blood ejection events at the aortic artery. Variations in RJIs due to physiology are also indicated in this figure, labeled as RJI_n . Although RJI_1 , RJI_2 , and RJI_3 are visually similar, they are numerically different. The signal amplitudes are normalized relative to their respective ranges across the entirety of each recording.

As the figure indicates, underlying timing differences exist. However, the importance of those differences is application-specific. We have chosen to focus on sleep staging. Quantifying the effects of HBI error on sleep staging accuracy is important, in part because it affects algorithm selection and development. While a few studies have used BCG-based HRV metrics for sleep staging [18–20], the effects of HBI error on the resulting outcomes seem to have been overlooked.

Note that HBI/HRV errors are due to several sources, e.g., physiology, sensing modality, and peak detection algorithm. We focus on those errors stemming from the BCG peak detection process. As observed in our previous works [21,22], peak detection methods introduce perturbations in the temporal locations of detected heartbeats. Missed and false peaks also contribute to HBI errors, but the focus of this research is the HBI error due to peak-detection-based timing jitter. Figure 2 illustrates a sample segment of BCG data where a detected heartbeat does not always align with the actual beat location. Importantly, such a mismatch is not always consistent; the error varies between heartbeats. This variability is due to the unpredictable nature of the error sources.

Another source of differentiation between ECG- and BCG-based HBIs is the variable time between ventricular depolarization and aortic blood ejection. The R peak of the ECG originates at the beginning of ventricular depolarization, whereas the J peak of the BCG is time aligned with blood ejection. Consequently, the R to J-peak interval (RJI) is not constant from beat to beat (see Figure 1). This RJI variability is a physiological component of the HBI error whenever a BCG is used, and it will be addressed in this work.

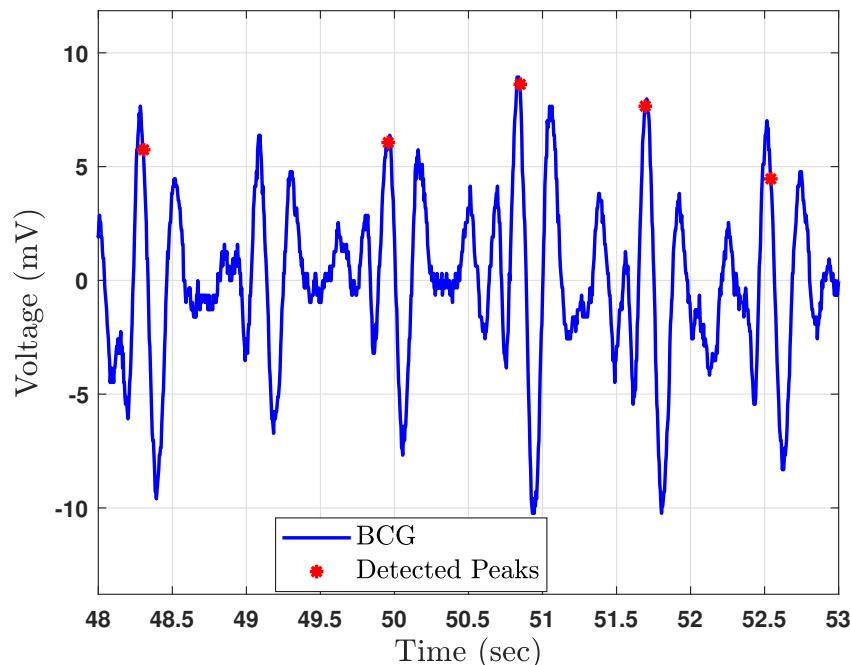


Figure 2. An example of perturbations in J–peak time estimates due to peak detection algorithm performance.

Based on our review, no published dataset incorporates simultaneous ECG and BCG recordings along with time-aligned expert sleep scores. However, the National Sleep Research Resource (NSRR) database [23–26] contains ECG and expert sleep scores, lacking only the BCG. Here, we simulate the missing BCG in order to allow an initial investigation of the feasibility of BCG-based sleep staging.

To that end, this research empirically investigates the effects of HBI errors on HRV feature quality and sleep staging accuracy. We use the results of sleep-stage classification and its sensitivity to HBI errors to demonstrate the effects of HBI error on sleep-scoring performance.

This work presents three primary contributions to the literature, (i) this is a novel, systematic investigation into the contribution of HBI error in sleep-scoring performance. The framework may be extended to other application-specific investigations. (ii) As part of this investigation, we have developed an estimated relationship between HBI error and sleep staging accuracy. This may serve as an assessment tool for other researchers developing sleep-scoring systems based on cardiac information before conducting human sleep studies. (iii) Finally, we believe this to be the first five-class automated sleep scoring using only HRV data. This study outlines an SVM-based sleep-scoring performance using HRV features only from a clean ECG dataset with well-labeled sleep scores. This establishes a reference for subsequent sleep-scoring studies.

2. Related Work

Both ECGs and BCGs have been used to extract HRV features for sleep staging. As indicated in Table 1, the performance metrics for the affiliated classification algorithms are roughly comparable. The column “Stages Classified” identifies the labeling schemes employed by the different studies. SWS stands for slow wave sleep, and N-SWS stands for non-SWS. Likewise, REM stands for rapid eye movement sleep, and N-REM stands for non-REM sleep.

Higher accuracies, of up to 89%, are also reported from other studies, as noted in [27]. However, those studies included other signals for sleep staging, e.g., respiratory effort data [28,29], in addition to ECGs.

Table 1. Classification performance metrics based on HRV features acquired from ECG and BCG signals.

Signal Used	Stages Classified	Acc. (%)	Sen. (%)	N	Cohen's Kappa	Ref.
ECG	SWS vs. N-SWS	90	69	45	0.56	[27]
BCG + Actigraphy	SWS vs. N-SWS	93	81	4	0.62	[19]
ECG	REM vs. N-REM	87	87	25	0.61	[30]
BCG + movement	REM vs. N-REM	80	N/A	18	0.43	[20]

While BCG-based HRV features appear to yield comparable sleep staging results when employed in lieu of ECG-based HRV features (e.g., as laid out in Table 1), BCG-based sleep staging is not yet widely used. Therefore, the effects of any differences in BCG versus ECG-based HRV estimates are not well understood, where discrepancies stem from HBI errors as determined from JJI versus RRI.

A few studies have addressed such errors. For example, [20,31,32] described errors in HRV features due to HBI errors, and [6,17] reported correlation coefficients between BCG- and ECG-based HRV features. However, these papers do not discuss the effects of HBI errors on sleep staging accuracy. One study [33] did compare sleep stages determined with a partially BCG-based system [34] to polysomnogram-based sleep scores as determined by a polysomnographic technician. However, the impacts of HBI errors on the sleep staging accuracy cannot be calculated for this study since BCGs were not the only signals used for sleep staging.

Based on the review of the available literature, it appears that no formal investigation of the effects of HBI differences (as determined from ECG versus BCG signals) has been performed in the context of sleep stage determination. Sleep-based studies that employ BCG-based HRV features seem to be limited to [14,17–20,32,35–37]. Therefore, it seems worthwhile to investigate the effects of ECG- versus BCG-based HRV features on sleep-scoring performance results, particularly in light of the promise that BCG-based systems hold for unobtrusive sleep monitoring. Such an analysis can help set a performance target for BCG-based heartbeat detection algorithms, which must deal with natural beat-to-beat changes in signal morphology that are not as prevalent in ECGs.

Note that this effort parallels current research with photoplethysmograms (PPGs) as employed for sleep stage determination. Similar to a BCG, a PPG acquired from a finger or a wrist offers a remote indication of “pulse rate” that is different from “heart rate” as determined by an ECG, though we speak of each type of signal as providing HRV features. When acquiring a BCG or a PPG, the physical distance (and, therefore, the arterial layout) between the sensor and the heart affects the signal morphology, and beat-to-beat changes in signal shape are more pronounced in BCGs and PPGs when compared to ECGs. Additionally, the smoother, more oscillatory nature of each of these signals, when compared to ECGs with their distinct R waves, can make accurate heartbeat identification more of a challenge. Nonetheless, since light-based sensors used to acquire PPGs are less intrusive than electrodes used to acquire ECGs, recent efforts have investigated the benefits of PPGs to estimate HRV features in the context of sleep staging. For example, [38] compared PPG- and ECG-derived HR and HRV values, reporting good agreement based on the resulting correlation coefficients. However, other recent studies have indicated that using PPG-derived HR and HRV features can lead to a decrease in sleep staging performance [39].

3. Methods

3.1. Data Description

NSRR datasets from 50 participants were included in this study. Sleep scores identified for those 50 participants served as ground truth values for our sleep stage assessments. In order to investigate sleep stage scoring sensitivity to ECG-BCG timing differences, we introduced synthetic timing jitter to the ECG-based HBIs to mimic the effects of RJI variations.

Timing error can come from a combination of many sources, including BCG sensing modality, peak detection algorithm, and participant physiology. If these random sources

sum together, the resulting error should be Gaussian by the Central Limit Theorem. In addition, a Gaussian model is a reasonable fit for our measured timing jitter in another study (described in Section 3.5, below). Figure 3 depicts an overlay of the observed RJI distribution (blue line) and that of a theoretical Gaussian model (red line). For this histogram, we removed data from one participant because their BCG was collected from a different location than all other participants in the original study.

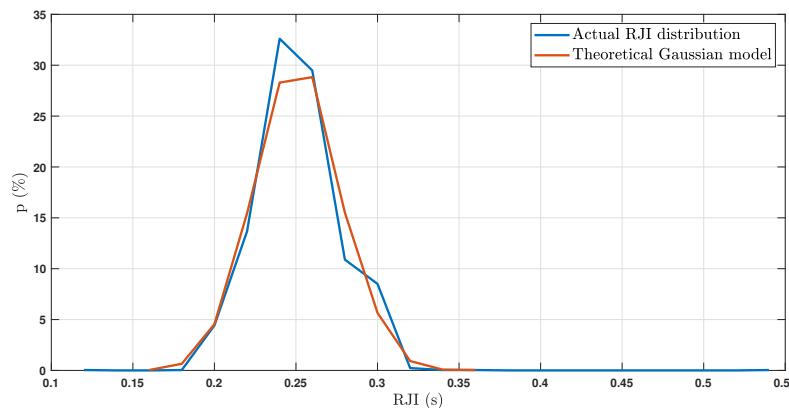
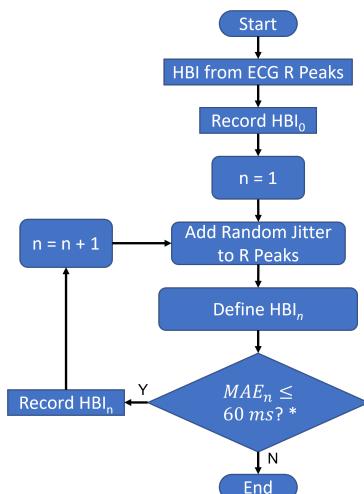


Figure 3. RJI data and a theoretical Gaussian fit based on mean and standard deviation of the RJI data.

For the jitters discussed in this paper, we used random time offsets drawn from a zero-mean Gaussian distribution because HBI calculations (and thus HRV calculations) are insensitive to a mean time offset.

ECG R peaks provided ground truth heartbeat times. We denote an HBI obtained from these R peaks as HBI_0 , where the zero subscript signifies zero added error. Then we obtained HBIs from the artificially perturbed R-peak locations. We denote HBIs obtained from these R peaks with timing jitter as HBI_n , where n corresponds to an n th-level of error contamination. Thus, HBI_n represents an HBI calculated from an R-peak stream with an n th-level error. We increased the standard deviation of these perturbations until we reached a mean absolute error (MAE) between HBI_n and HBI_0 of 60 ms. We specified 60 ms as a maximum based on the findings in [22], where HBI errors due to the worst-performing method were less than 60 ms for most of the participants. This process took 97 iterations, thus leading to $n = 97$ levels of synthetic HBI error. Figure 4 illustrates the process of obtaining HBIs with simulated errors starting with ECG R-peak locations provided in each dataset. The process was repeated 50 times to cover all of the datasets included in the study.



* $MAE_n = \text{mean}(|HBI_n - HBI_0|)$

Figure 4. The process of adding jitter to the R-peak locations and recording the resulting RRIs.

We refer to the MAE between each HBI_n and HBI_0 as MAE_n on an individual level, where HBI MAE is an overall metric. For example, MAE_{97} , the MAE between HBI_{97} and HBI_0 , is approximately 60 ms. MAE_n serves as the independent variable for this study.

3.2. HRV Features

We calculated frequency- and time-domain HRV features per 30-second epoch for HBI_0 and for each HBI_n time series, using five-minute-wide sliding windows with 90% overlap [40]. As described in Section 3.1, we had 97 levels of synthetic errors in addition to the original times, so 98 sets of HRV features were obtained.

3.2.1. Frequency-Domain Features

Low-frequency (LF) and high-frequency (HF) powers were calculated using Lomb–Scargle normalized periodograms [41,42] due to the unevenly sampled nature of all HBI time series. The LF and HF frequency ranges correspond to 0.04–0.15 Hz and 0.15–0.4 Hz, respectively [15]. In addition, the ratio of LF to HF powers (LFHF) and the one-dimensional median-filtered version of the ratio (MedFiltLFHF) were employed as frequency-domain features.

3.2.2. Time-Domain Features

The mean of the HBIs (denoted by HR as in [40]) and its standard deviation (SDNN) in each window were used as time domain features.

3.3. Sleep Labels

The NSRR database provides sleep stage labels of Wake, S1, S2, S3, S4 (rarely), and REM [23–26]. The label S4, which is not a standard level, was very rare among these participants. Therefore, those labels were merged with S3. In addition, sleep stages shorter than four epochs were merged into their previous stage. In most of the cases, such behavior was observed as a glitch where the stages before and after the glitch were the same. The average sleep duration among these participants was 8.43 hours with a standard deviation of $+/- 0.5$ h. The average time spent in each sleep stage for these participants is summarized in Table 2. Two scenarios were tested: (i) micro-labeling, where all sleep stage labels were provided as class labels, as available from the NSRR database, and (ii) macro-labeling, where non-REM sleep stages were merged into a single stage and assigned a single label resulting in the three stages of REM, N-REM, and Wake, similar to what was performed in [20,30] (see Table 1).

Table 2. Average percentage of sleep duration in each stage for the study cohort.

Wake	S1	S2	S3	REM
30.0%	2.6%	41.1%	11.6%	14.7%

3.4. Effects of HBI Error on HRV Feature Quality

We used Bayes error and classification error to quantify the effects of HBI error on HRV feature quality (see the sections below for the calculation methods used to determine these errors). Classification error was obtained using the support vector machine (SVM) classifier. SVM is a binary classifier that provides a class probability \hat{y} of a sample instance using the following equation [43]:

$$\hat{y}(\mathbf{x}) = \hat{w}_0 + \sum_{i=1}^N \alpha_i k(\mathbf{x}_i, \mathbf{x}) \text{ Where, } \alpha_i = \lambda_i y_i, \lambda \text{ is the } \ell_1 \text{ regularization term.} \quad (1)$$

SVM finds the optimal values for \hat{w}_0 and α by maximizing the minimum distance between the two separating hyperplanes. Given that SVM only finds the decision boundary between two discrete class instances, SVM can not be used directly to solve multi-class

problems. However, using one-vs.-all or one-vs.-one heuristic methods, we can solve multi-class classification tasks with SVMs. SVM is a well-established classifier with relatively compact representation of the training data and good test computation requirements. In addition, the SVM classifier is efficient in higher dimensional space and works well for a fewer number of samples than data dimensions. Therefore, we chose to use SVM here. In this study, we used the one-vs.-one strategy to implement SVM-based multi-class problems since the one-vs.-one approach creates more balanced binary datasets than one-vs.-all. To implement the strategy mentioned above, we used MATLAB's `templateSVM.m` function with the Gaussian kernel, and to train the model, we used `fitcecoc.m`. A 10-fold cross-validation approach was adopted for this purpose. All HRV feature sets obtained from HBI₀ to HBI₉₇ were examined individually to obtain Bayes and classification errors corresponding to each HBI MAE.

Two scenarios were evaluated based on the micro- and macro-labeling schemes identified in Section 3.3. In addition, two tests were performed in each scenario. In the first test, all HRV features were used to estimate Bayes and classification errors. In the second test, individual features were used to estimate Bayes and classification errors. The second test was performed to investigate the sensitivity of each HRV feature to HBI error.

3.4.1. Bayes Error Test

We used a non-parametric method to estimate the Bayes error. Parametric Bayes error estimation requires knowledge of class distribution parameters and prior probabilities, whereas non-parametric methods can be used in the absence of this information. A non-parametric method for an L-class classification problem is provided in [44], which relies on k-nearest neighbor (kNN) classification. The kNN classifier uses a predetermined distance metric to find the closest k training instances for a given test instance, and then it classifies the test instance based on the majority vote of the training labels for these “neighbors”. Given an L-class problem with a sufficiently large training dataset, the Bayes error bounds can be calculated using Equation (2) [45]:

$$\frac{L-1}{L} \left(1 - \sqrt{1 - \frac{L}{L-1} E_{kNN}} \right) \leq E_{bayes} \leq E_{kNN} \quad (2)$$

Here E_{kNN} is the kNN classifier error. Again, both micro- and macro-labeling test scenarios were addressed. Within each scenario, individual and multiple HRV features were considered.

3.4.2. Classification Error Test

For this test, classification error was obtained using the HRV feature sets. An overall classifier error pattern in response to HBI error is obtained as well. This error pattern is in the form of the slope of the classifier error curves due to each HRV feature within the n^{th} HRV feature set. This slope represents the change in classification error percentage per millisecond change in the HBI MAE of the underlying HBI _{n} . The slope was obtained from a linear regression of the classifier error against the HBI MAE _{n} , the independent variable.

3.5. Brain and Body Sensing Laboratory Data

While we have argued that simulated BCG data are better for our analyses, they do have one shortcoming. ECG-BCG timing differences include a physiological component—the variation between R-peak time and physical blood ejection—which cannot be estimated or subtracted from the BCG signal alone. This variability provides a lower bound for the timing error for BCG systems and must be estimated from a dataset that has both ECGs and BCGs. Therefore, we used simultaneous ECG-BCG data recorded in the Kansas State University Brain and Body Sensing Laboratory [21,22] to estimate this lower bound.

This archive included data from 30 healthy volunteers: 14 males (ages 30.9 ± 6.3 years) and 16 females (ages 46.0 ± 18.5 years). These data were initially collected for the purpose of algorithm comparison, as outlined in [22]. We excluded three volunteers for this analysis, as

their data lacked simultaneous ECG recordings. One hundred pairs of consecutive ECG- and BCG-based heartbeats were used for each subject, leading to a total of 2700 heartbeat pairs.

3.6. Comparing the Present Study with Previous Work: Simulated Versus Laboratory-Based ECG-BCG Timing Errors

We previously compared peak-detection algorithms for BCG signals [21,22], focusing on the time difference between each ground truth and detected J peak, an error we will denote here as e_p . The 'p' indicates that the source of the error is the peak detection algorithm. However, the error discussed in this present work is the total time difference between each ground-truth ECG R peak and the respective detected J peak. This total error then reflects not only the timing jitter from peak detection but also physiology and platform effects. The total error, which we denote here as e_T , is not directly comparable to e_p .

Therefore, we must augment our previously reported e_p values with platform and physiology effects, i.e., R to ground-truth J peak intervals (RGIs), before comparing them with e_T values obtained by adding synthetic timing jitter to the ECG peaks of the NSRR dataset employed in the current study. This may be performed by considering the following:

$$e_T = RGI + e_p \quad (3)$$

This equality stems from the fact that each peak is detected at a time offset, which is the sum of the platform offset, the physiology offset, and the peak detection offset. Taking the difference of adjacent detected peak times leads directly to the previous equation. The use of the triangle inequality allows the comparison of the two quantities:

$$|e_T| \leq |RGI| + |e_p| \quad (4)$$

Therefore, we will supplement the mean e_p values previously reported to account for platform and physiology effects, which are reflected in the R-to-J interval, or the difference between the times of arrival of the R and J peaks. While the mean RJI is unimportant because HRV features are insensitive to mean offsets, the difference between each RJI and the mean RJI is a source of error. Thus, we have calculated the total error by adding e_p and the mean absolute "jitter" between each RJI and the mean RJI. That is, the parameter $|RGI|$ in Equation (4), for each participant, i , is defined as

$$|RGI|_i = |RGI_i - \frac{\sum_{j=1}^{M_i} RGI_{j,i}}{M_i}| \quad (5)$$

where $M_i = 100$ is the number of R-J pairs from each participant, i , from our previous work [22] for whom we had collected time-aligned ECGs and BCGs.

This method will lead to a conservative upper bound for the total system error. The true error might be lower since peak detection jitter can offset platform and physiology effects for at least some beats.

4. Results

4.1. Classifier Baseline Performance

Figure 5 is a confusion matrix based on classification performed on the HRV features when no timing jitters were introduced to the RRI data. For this confusion matrix, SVM was used. This confusion matrix is based on 33 participants out of 50 after discarding results that would contain zeros in any detection column.

4.2. HBI Error Effects

Table 3 summarizes the SVM classification errors, E_0 and E_{max} , for the extreme cases of HBI₀ and HBI₉₇, respectively (We did consult the kNN classifier-based results, and they showed overall worse performance than SVM, though it was slightly less sensitive to the

HRV errors added in the study. For brevity, we have included figures that relate only to the SVM classification method).

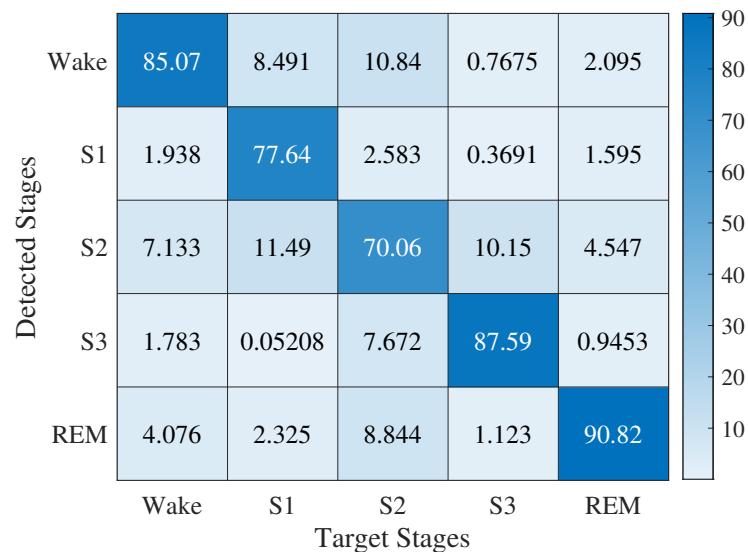


Figure 5. SVM classification baseline performance due to all HRV features for micro-labeled sleep stages.

Table 3. SVM classification errors averaged over 50 participants for HBI₀ and HBI₉₇.

Labeling	E_0 (%)	E_{max} (%)
Micro-Labeling	24	36
Macro-Labeling	17	27

Figure 6 illustrates the SVM classification error as a function of HBI MAE for the micro-labeled sleep stages, and Figure 7 illustrates the SVM classification error as a function of HBI MAE for the macro-labeled sleep stages. In both cases, all HRV features were provided to the classifier. Classification errors for each participant are graphed in light gray lines, whereas the dashed, bold black line in each figure is the ensemble average. As expected, increasing HBI MAE led to increasing sleep staging error. At the same time, the classification error seems to remain low ($\leq 20\%$ in the case of micro-labeled sleep stages) for roughly half of the simulated MAE_n values.

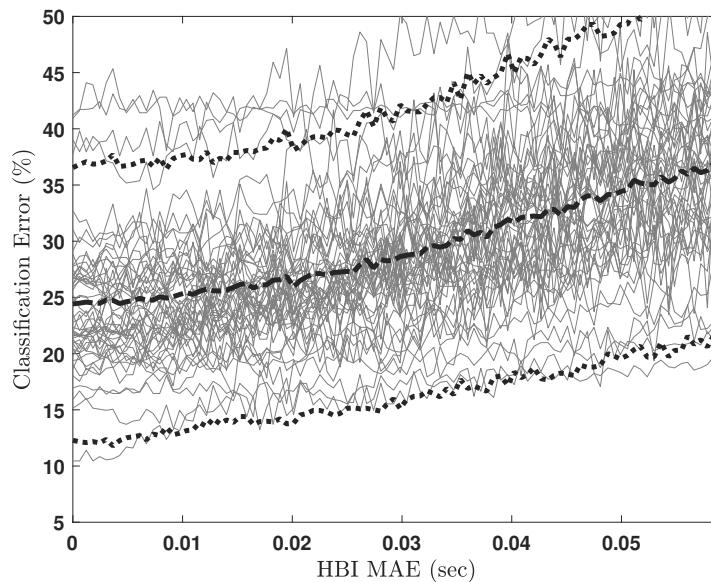


Figure 6. SVM classification error due to all HRV features for micro-labeled sleep stages. Light gray lines represent classification errors for each participant. The dashed, bold black line is the ensemble average, and the dotted bold lines show mean ± 1.96 standard deviations.

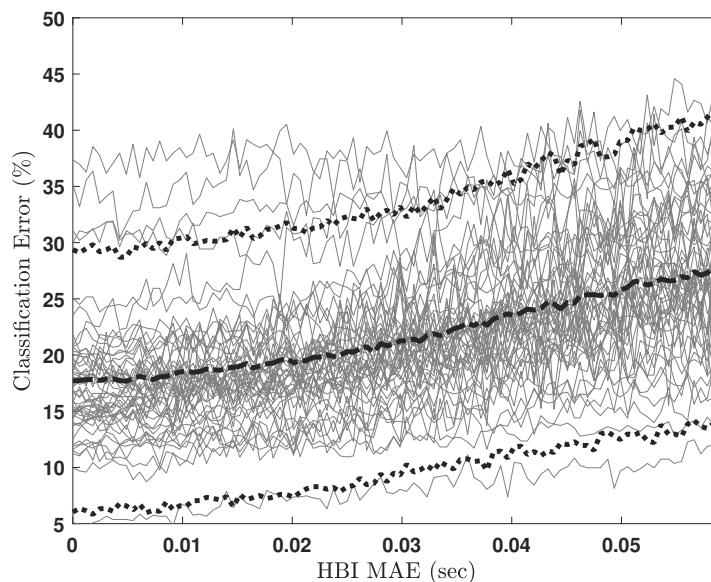


Figure 7. SVM classification error due to all HRV features for macro-labeled sleep stages. Light gray lines represent classification errors for each participant. The dashed, bold black line is the ensemble average, and the dotted bold lines show mean ± 1.96 standard deviations.

Table 4 summarizes the sensitivity of the HRV features to changes in HBI MAE. The numbers in this table are the slopes described in Section 3.4.2 averaged across participants. The sensitivities for HF, LFHF, and MedFiltLFHF are very close. Note that the units are in percentage points (pp) per second of change in MAE, so the change in performance is on the order of a few percentage points over the 60 ms range investigated in this study.

Table 4. Sensitivity of feature quality to HBI error in terms of classifier mean error slopes (units: percent error per second of HBI MAE).

Features	HR	SDNN	LF	HF	LFHF	MedFiltLFHF	Label
Slopes (%/s)	4.37	27.10	19.66	58.74	30.32	31.78	Micro
	4.43	27.26	19.44	59.02	30.37	32.01	Macro

4.3. BCG-Based HBI Error Limit

Table 5 lists some statistical metrics for $|RGI|$, $|e_p|$, and their sum. The numbers are based on 2700 heartbeats (100 beats from each participant, with a total of 27 participants), and the $|e_p|$'s are obtained from J peaks detected using an optimal BCG peak detection approach by Brüser [5] as identified in our previous papers [21,22].

Table 5. Statistical metrics for $|RGI|$, $|e_p|$, and their sum.

	Mean (ms)	Median (ms)	Min (ms)	Max (ms)
$ RGI $	7.16	6.09	0.94	44.44
$ e_p $	8.77	8.26	1.27	23.60
$ RGI + e_p $	15.93	14.35	2.21	68.04

As explained in Section 3.6, the sum of $|RGI|$ and $|e_p|$ is an upper bound for total mean error between BCG-based and ECG-based HBIs. Thus, this quantity is comparable to the HBI MAE parameter utilized for the horizontal axes in Figures 6 and 7. From Table 5, the mean for this quantity is approximately 16 ms. If we locate 16 ms on the curves in Figures 6 and 7, it corresponds to up to 26% and 18% sleep classification errors for micro- and macro-labeling schemes, respectively. Table 6 compares the macro-labeling errors projected from our analyses to state-of-the-art sleep-scoring errors compiled from Table 1.

Table 6. State-of-the-art macro-labeling performance with respect to projected performance based on the HBI error limit.

Signal Used	Stages Classified	Ref.	100-Acc. (%)	Projected Error (%)
ECG	SWS vs. N-SWS	[27]	10	
BCG + Actigraphy	SWS vs. N-SWS	[19]	7	
ECG	REM vs. N-REM	[30]	13	18
BCG + movement	REM vs. N-REM	[20]	20	

5. Discussion

We demonstrated that HBI errors due to timing jitters follow a Gaussian trend owing to the inherent variability across R to J peak intervals in the context of BCG. This trend is the basis of our simulation-based results with the following implications.

5.1. Timing Jitter in Heartbeat Detection

The best-case micro-labeling classification error on this dataset, with no added timing jitter, is 24 %. While any timing jitter decreases performance, all practical systems have some timing jitter. From the literature (see Table 1), a range of errors from 10% to 20% can be inferred, though all of those studies used macro-labeling, which makes it hard to put our micro-labeling performance in perspective. Macro-labeling performs better. The classification error for macro-labeling would be as low as 17% in the best case (see Table 3).

Ideally, the classification error would be lower. Yet detection of sleep staging from cardiac data is not trivial—fast heart rates could result from nightmares, waking events, and several other sources. Traditional polysomnography includes EEG electrodes as well as other non-cardiac data to achieve lower classification error. Here, we have investigated the use of BCG-only signals, which could allow sleep staging without wires or sensors on the participant, thus potentially allowing multi-month or multi-year studies, which are simply not practical with the existing technology.

5.2. HRV Feature Sensitivity to HBI MAE

The slopes of the error curves provide insight into each HRV feature's sensitivity to HBI MAE. From Table 4, we see that HR is the feature least sensitive to HBI MAE in

both the macro- and micro-labeling scenarios. The slope is noticeably smaller than it is for other features over the examined range of HBI MAEs, particularly for the macro-labeling scenario. This finding is consistent with our understanding of theoretical sensitivity. HR is calculated using windowed averages of HBIs, which smooths out the effects of introduced HBI perturbations. In fact, a perturbation to a heartbeat time inside of a window does not substantively change the calculated HR unless the perturbation affects one of the outside two heartbeats or if that perturbation moves a heartbeat into or out of the window. On the other hand, HF and LFHF are the features most sensitive to HBI MAE in macro- and micro-labeling scenarios, respectively. Again, this matches intuition—both of these features include frequency-domain content that is more sensitive to perturbations inside of the window. As expected, these sensitivities are lower in the case of macro-labeling since the number of labels is reduced, and thus, classification performance is improved.

5.3. Notes Regarding the Use of BCG-Based HBIs

It must be noted that when comparing an HBI estimation algorithm's performance with the HBI MAE values obtained here, the reference for both should be the R peaks of a simultaneous ECG. For other references, some calibration should first be considered. As mentioned in the Methods section, for instance, the HBI errors based on BCG ground truth peaks were calibrated to include the effects of RJI variability and the measurement platform.

5.4. Projected HBI-Based Sleep-Scoring Performance

As noted in Table 6, the projected sleep macro-labeling performance based on our proposed HBI error limit compares well with other recent results. Note that the 7% sleep labeling error reported in [19] is based on both BCG and actigraphy (acceleration) signals and is not, therefore, directly comparable.

5.5. Limitations and Future Work

This study relies on the results of simulated HBI errors. The BCG recording system introduced in [46–50] and the BCG peak detection method selected in [21,22] puts us in a position to estimate reliable HBIs and HRV features. This is because this work offers insight into the best- and worst-case HBI error scenarios, which will put the BCG-based HBIs and their related HRVs in perspective for us. As indicated earlier, there is no available dataset of simultaneous ECG, BCG, and sleep scores. Such a dataset would have enabled us to assess the other effects of peak detection algorithm artifacts, e.g., missed and false detection events. Nonetheless, this work can be a stepping stone toward further studies in that direction. Eventually, we hope to conduct sleep studies where we record polysomnograms and BCGs simultaneously on a larger population.

6. Conclusions

We examined the effects of HBI errors in the context of one BCG application: sleep staging. For that purpose, we introduced different levels of synthetic errors to ECG-based HBIs in order to simulate BCG-based HBIs. HRV features were then calculated from these altered HBIs and used for sleep staging. We used these HRV features with the SVM classification algorithm and one-vs.-one strategy to enable a multi-class classification. In addition, different sleep stage labeling and HRV feature combination scenarios were tested. Our results indicate a clear trend toward decreasing sleep staging accuracy with increasing timing jitter. The sensitivity varies substantially between HRV feature type. We cannot recommend a universal amount of acceptable jitter because the cost of staging errors will depend on the end application. Nevertheless, the results indicate that timing jitters as small as 20 ms demonstrate a measurable impact on the overall performance of HRV-based sleep labeling algorithms.

We examined classification performance sensitivity to HBI errors while providing individual features to each classifier. As expected, classification based on heart rate was the least sensitive to this type of error.

Based on the HBI error range proposed in this work, the feasibility of BCG-based HBI estimation for sleep stage classification was assessed.

Author Contributions: Conceptualization, A.S., P.P., and D.E.T.; methodology, A.S. and D.E.T.; validation, A.S.; formal analysis, A.S.; investigation, A.S. and D.E.T.; resources, A.S., M.R.M., A.A., and C.C.; data curation, A.S. and C.C.; writing—original draft preparation, A.S.; writing—review and editing, A.S., M.R.M., and D.E.T.; visualization, A.S., M.R.M., and D.E.T.; supervision, D.E.T. and S.W.; project administration, D.E.T., S.W., and B.N.; funding acquisition, D.E.T. and S.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported in part by Kansas State University faculty startup funds, in part by the National Science Foundation General and Age-Related Disabilities Engineering (GARDE) Program under grants CBET-1067740 and UNS-1512564, and by the National Science Foundation Human-Centered Computing (HCC) and Integrative Strategies for understanding Neural and Cognitive Systems programs under grant CHS-1910526. Opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Kansas State University (protocol code 8344, approved 29 June 2016 and 9386, approved 5 July 2019).

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

Data Availability Statement: Not applicable.

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

References

1. Hayano, J.; Yuda, E.; Yoshida, Y. Sleep stage classification by combination of actigraphic and heart rate signals. In Proceedings of the 2017 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Taipei, Taiwan, 12–14 June 2017; pp. 387–388.
2. Zink, M.D.; Brüser, C.; Stüben, B.O.; Napp, A.; Stöhr, R.; Leonhardt, S.; Marx, N.; Mischke, K.; Schulz, J.B.; Schiefer, J. Unobtrusive Nocturnal Heartbeat Monitoring by a Ballistocardiographic Sensor in Patients with Sleep Disordered Breathing. *Sci. Rep.* **2017**, *7*, 13175.
3. Lee, W.K.; Yoon, H.; Han, C.; Joo, K.M.; Park, K.S. Physiological Signal Monitoring Bed for Infants Based on Load-Cell Sensors. *Sensors* **2016**, *16*, 409.
4. Paalasmaa, J.; Toivonen, H.; Partinen, M. Adaptive heartbeat modeling for beat-to-beat heart rate measurement in ballistocardiograms. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 1945–1952.
5. Brüser, C.; Winter, S.; Leonhardt, S. Robust inter-beat interval estimation in cardiac vibration signals. *Physiol. Meas.* **2013**, *34*, 123.
6. Brueser, C.; Winter, S.; Leonhardt, S. Unsupervised heart rate variability estimation from ballistocardiograms. In Proceedings of the 7th International Workshop on Biosignal Interpretation, Como, Italy, 2–4 July 2012.
7. Brüser, C.; Stadlthanner, K.; de Waele, S.; Leonhardt, S. Adaptive Beat-to-Beat Heart Rate Estimation in Ballistocardiograms. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 778–786.
8. Shin, J.H.; Hwang, S.H.; Chang, M.H.; Park, K.S. Heart rate variability analysis using a ballistocardiogram during Valsalva manoeuvre and post exercise. *Physiol. Meas.* **2011**, *32*, 1239.
9. Kortelainen, J.M.; Virkkala, J. FFT averaging of multichannel BCG signals from bed mattress sensor to improve estimation of heart beat interval. In Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007; pp. 6685–6688.
10. Arab, C.; Vanderlei, L.C.M.; da Silva Paiva, L.; Fulghum, K.L.; Fristachi, C.E.; Nazario, A.C.P.; Elias, S.; Gebrim, L.H.; Ferreira Filho, C.; Gidron, Y.; et al. Cardiac autonomic modulation impairments in advanced breast cancer patients. *Clin. Res. Cardiol.* **2018**, *107*, 924–936.
11. Boman, K. Heart rate variability: A possible measure of subjective wellbeing? 2018.
12. Yuda, E.; Yoshida, Y.; Sasanabe, R.; Tanaka, H.; Shiomi, T.; Hayano, J. Sleep Stage Classification by a Combination of Actigraphic and Heart Rate Signals. *J. Low Power Electron. Appl.* **2017**, *7*, 28.
13. Nurmi, S.; Saaresranta, T.; Koivisto, T.; Meriheinä, U.; Palva, L. *Validation of an Accelerometer Based BCG Method for Sleep Analysis*; Aalto University: Espoo, Finland, 2016.
14. Ni, H.; Zhao, T.; Zhou, X.; Wang, Z.; Chen, L.; Yang, J. Analyzing Sleep Stages in Home Environment Based on Ballistocardiography. In *Health Information Science*; Yin, X., Ho, K., Zeng, D., Aickelin, U., Zhou, R., Wang, H., Eds.; Number 9085 in Lecture Notes in Computer Science; Springer International Publishing: New York, NY, USA, 2015; pp. 56–68.
15. Xiao, M.; Yan, H.; Song, J.; Yang, Y.; Yang, X. Sleep stages classification based on heart rate variability and random forest. *Biomed. Signal Process. Control* **2013**, *8*, 624–633.

16. Dong, Z.; Li, X.; Chen, W. Frequency-network analysis of heart rate variability for obstructive apnea patient detection. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 1895–1905.

17. Jung, D.W.; Hwang, S.H.; Yoon, H.N.; Lee, Y.J.; Jeong, D.U.; Park, K.S. Nocturnal Awakening and Sleep Efficiency Estimation Using Unobtrusively Measured Ballistocardiogram. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 131–138.

18. Watanabe, T.; Watanabe, K. Noncontact method for sleep stage estimation. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1735–1748.

19. Choi, B.H.; Chung, G.S.; Lee, J.S.; Jeong, D.U.; Park, K.S. Slow-wave sleep estimation on a load-cell-installed bed: a non-constrained method. *Physiol. Meas.* **2009**, *30*, 1163.

20. Kortelainen, J.M.; Mendez, M.O.; Bianchi, A.M.; Matteucci, M.; Cerutti, S. Sleep staging based on signals acquired through bed sensor. *IEEE Trans. Inform. Technol. Biomed.* **2010**, *14*, 776–785.

21. Suliman, A.; Carlson, C.; Warren, S.; Thompson, D. Performance Evaluation of Processing Methods for Ballistocardiogram Peak Detection. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 502–505.

22. Suliman, A.; Carlson, C.; Ade, C.; Warren, S.; Thompson, D.E. Performance Comparison for Ballistocardiogram Peak Detection Methods. *IEEE Access* **2019**. <https://doi.org/10.1109/ACCESS.2019.2912650>

23. Dean, D.A.; Goldberger, A.L.; Mueller, R.; Kim, M.; Rueschman, M.; Mobley, D.; Sahoo, S.S.; Jayapandian, C.P.; Cui, L.; Morrical, M.G.; et al. Scaling up scientific discovery in sleep medicine: The National Sleep Research Resource. *Sleep* **2016**, *39*, 1151–1164.

24. Zhang, G.Q.; Cui, L.; Mueller, R.; Tao, S.; Kim, M.; Rueschman, M.; Mariani, S.; Mobley, D.; Redline, S. The National Sleep Research Resource: Towards a sleep data commons. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1351–1358.

25. Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O'Connor, G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M.; et al. The sleep heart health study: Design, rationale, and methods. *Sleep* **1997**, *20*, 1077–1085.

26. Redline, S.; Sanders, M.H.; Lind, B.K.; Quan, S.F.; Iber, C.; Gottlieb, D.J.; Bonekat, W.H.; Rapoport, D.M.; Smith, P.L.; Kiley, J.P. Methods for obtaining and analyzing unattended polysomnography data for a multicenter study. *Sleep* **1998**, *21*, 759–767.

27. Yoon, H.; Hwang, S.H.; Choi, J.W.; Lee, Y.J.; Jeong, D.U.; Park, K.S. Slow-wave sleep estimation for healthy subjects and OSA patients using R–R intervals. *IEEE J. Biomed. Health Inform.* **2018**, *22*, 119–128.

28. Long, X.; Fonseca, P.; Aarts, R.; Haakma, R.; Rolink, J.; Leonhardt, S. Detection of Nocturnal Slow Wave Sleep Based on Cardiorespiratory Activity in Healthy Adults. *IEEE J. Biomed. Health Inform.* **2016**, *21*, 123–133.

29. Fonseca, P.; Long, X.; Radha, M.; Haakma, R.; Aarts, R.M.; Rolink, J. Sleep stage classification with ECG and respiratory effort. *Physiol. Meas.* **2015**, *36*, 2027.

30. Yoon, H.; Hwang, S.H.; Choi, J.W.; Lee, Y.J.; Jeong, D.U.; Park, K.S. REM sleep estimation based on autonomic dynamics using R–R intervals. *Physiol. Meas.* **2017**, *38*, 631.

31. Vehkaoja, A.; Rajala, S.; Kumpulainen, P.; Lekkala, J. Correlation approach for the detection of the heartbeat intervals using force sensors placed under the bed posts. *J. Med. Eng. Technol.* **2013**, *37*, 327–333.

32. Waltisberg, D. Sleep Assessment Using Bed-Mounted Sensors. Ph.D. Thesis, ETH Zurich, Zürich, Switzerland, 2017.

33. Mack, D.C.; Patrie, J.T.; Felder, R.A.; Suratt, P.M.; Alwan, M. Sleep assessment using a passive ballistocardiography-based system: Preliminary validation. In Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 2–6 September 2009; pp. 4319–4322.

34. Mack, D.C.; Patrie, J.T.; Suratt, P.M.; Felder, R.A.; Alwan, M. Development and preliminary validation of heart rate and breathing rate detection using a passive, ballistocardiography-based sleep monitoring system. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 111–120.

35. Park, K.S.; Yoon, H. Ballistocardiography. In *Seamless Healthcare Monitoring*; Springer: Berlin, Germany, 2018; pp. 127–155.

36. Samy, L.; Huang, M.C.; Liu, J.J.; Xu, W.; Sarrafzadeh, M. Unobtrusive sleep stage identification using a pressure-sensitive bed sheet. *IEEE Sens. J.* **2014**, *14*, 2092–2101.

37. Yi, R.; Enayati, M.; M Keller, J.; Popescu, M.; Skubic, M. Non-Invasive In-Home Sleep Stage Classification Using a Ballistocardiography Bed Sensor. In Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA, 19–12 May 2019; pp. 1–4.

38. Kinnunen, H.; Rantanen, A.; Kenttä, T.; Koskimäki, H. Feasible assessment of recovery and cardiovascular health: Accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. *Physiol. Meas.* **2020**, *41*, 04NT01.

39. van Gilst, M.; Wulterkens, B.; Fonseca, P.; Radha, M.; Ross, M.; Moreau, A.; Cerny, A.; Anderer, P.; Long, X.; van Dijk, J.; et al. Direct application of an ECG-based sleep staging algorithm on reflective photoplethysmography data decreases performance. *BMC Res. Notes* **2020**, *13*, 1–5.

40. Clifford, G. ECGtools/ecgBag. 2003. Available online: <http://www.robots.ox.ac.uk/~gari/CODE/ECGtools/ecgBag/> (accessed on 16 September 2019).

41. Lomb, N.R. Least-squares frequency analysis of unequally spaced data. *Astrophys. Space Sci.* **1976**, *39*, 447–462.

42. Scargle, J.D. Studies in astronomical time series analysis. II-Statistical aspects of spectral analysis of unevenly spaced data. *Astrophys. J.* **1982**, *263*, 835–853.

43. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.

44. Cover, T.M.; Hart, P.E. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27.

45. Tumer, K.; Ghosh, J. Bayes error rate estimation using classifier ensembles. *Int. J. Smart Eng. Syst. Des.* **2003**, *5*, 95–109.

46. Prakash, P.; Kuehl, P.; McWilliams, B.; Rubenthaler, S.; Schnell, E.; Singleton, G.; Warren, S. Sensors and instrumentation for unobtrusive sleep quality assessment in autistic children. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 800–803.
47. Warren, S.; Prakash, P.; Brokesh, E.; Singleton, G.; Fowler, K. Design projects to quantify the health and development of autistic children. In Proceedings of the 2014 Annual Conference & Exposition, Indianapolis, IN, USA, 15–18 June 2014; pp. 15–18.
48. Carlson, C.; Suliman, A.; Prakash, P.; Thompson, D.; Wang, S.; Natarajan, B.; Warren, S. Bed-based instrumentation for unobtrusive sleep quality assessment in severely disabled autistic children. In Proceedings of the 2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 4909–4912.
49. Carlson, C.; Suliman, A.; Alivar, A.; Prakash, P.; Thompson, D.; Natarajan, B.; Warren, S. A Pilot Study of an Unobtrusive Bed-Based Sleep Quality Monitor for Severely Disabled Autistic Children. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 4343–4346.
50. Carlson, C. Development of a Bed-Based Nighttime Monitoring Toolset. Ph.D. Thesis, Kennesaw State University in Georgia, Kennesaw, GA, USA, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.