City, University of London Institutional Repository

Technical note

Opening the envelope: Efficient envelope-based PPG denoising algorithm

George R.E. Bradley [*], Panayiotis A. Kyriacou

Research Centre for Biomedical Engineering, City, University of London, London, United Kingdom

ARTICLE INFO

ABSTRACT

Photoplethysmography (PPG) signals obtained from the skin's surface offer valuable insights into blood volume fluctuations. With the rising interest in continuous non-invasive physiological monitoring, PPG has garnered significant attention. However, PPG signals are often affected by various forms of noise, impeding reliable feature extraction. Robust data pre-processing approaches are vital for both retrospective and real-time analysis. Existing denoising methods, including recent machine learning techniques, often suffer from implementation challenges, computational inefficiency, and limited interpretability. Addressing this challenge, we propose a novel PPG denoising algorithm. The algorithm was evaluated using a dataset representing approximately 81,015.99 min or 1360.27 h of PPG data collected from 31 patients. The evaluation involved the calculation and analysis of five key metrics: Signal-to-Noise Ratio (SNR), Variance, Total Variation (TV), Shannon entropy, and Instances-per-second (IPS). Our results demonstrate a notable increase in SNR after denoising, indicating effective noise reduction while preserving signal content. Variance and TV values showed a reduction post-denoising, suggesting smoother and less variable signals, validating the noise suppression efficacy. Additionally, Shannon entropy exhibited a decrease after denoising, indicating successful noise reduction and enhanced signal regularity. The nonparametric Wilcoxon signed-rank test ($a = 0.05$) was employed to assess the statistical significance of the observed differences of these metrics before and after denoising. Furthermore, the computational speed analysis revealed the EPDA's potential for efficient processing of large datasets and real-time applications. This comprehensive evaluation approach allows for a thorough understanding of the EPDA's effectiveness in denoising PPG data, fostering advancements in non-invasive physiological monitoring and promoting the broader adoption of PPG-based healthcare technologies.

1. Introduction

Photoplethysmography (PPG) signals are non-invasive, cost-effective optical signals, easily obtained from the skin's surface using a light source and photodetector. PPG signals capture information about blood volume fluctuations by detecting variations in light absorption or reflection, induced by changes in blood volume. The growing interest in non-invasive physiological monitoring, coupled with the paradigm shift in healthcare towards continuous and ubiquitous patient monitoring rather than exclusive in-hospital care has triggered a significant surge in the adoption and exploration of PPG technology [1–3]. The magnitude of this interest is underscored by the significant rise in search results related to PPG over the past two decades which has seen an increase of 2392.72%, as evidenced by our search results of the terms "Photoplethysmography" or "Photoplethysmogram" in PubMed, Europe PubMed Central and Scopus from 2000 to 2022, as illustrated in Fig. 1.

This same trend has led to the proliferation of non-invasive medical devices as well as consumer-oriented wearable devices such as smartwatches and fitness trackers. Many of these devices use PPG technology

to enable non-invasive and continuous monitoring of various physiological markers, including heart rate, blood pressure, oxygen saturation, and sleep patterns [3].

It seems reasonable to suggest that PPG-based monitoring has the potential to disrupt healthcare. However, despite its promise, PPG signals are inherently vulnerable to various forms of noise that can distort the signal, posing challenges in obtaining reliable and accurate information. Among the sources of noise are power line interference, low and high-frequency noise, baseline drift, motion artifacts, and saturation of the photodiode by light [4]. The prevalence of such noise underscores the importance of data denoising approaches for both retrospective and real-time signal processing and analysis. Addressing these challenges, we present and evaluate a novel envelope-based denoising algorithm, hereafter referred to as the "Envelope PPG Denoising Algorithm" (EPDA), designed to accomplish two objectives: (i) the identification and removal of anomalous data while preserving the indices of the removed data, and (ii) the assurance of computational
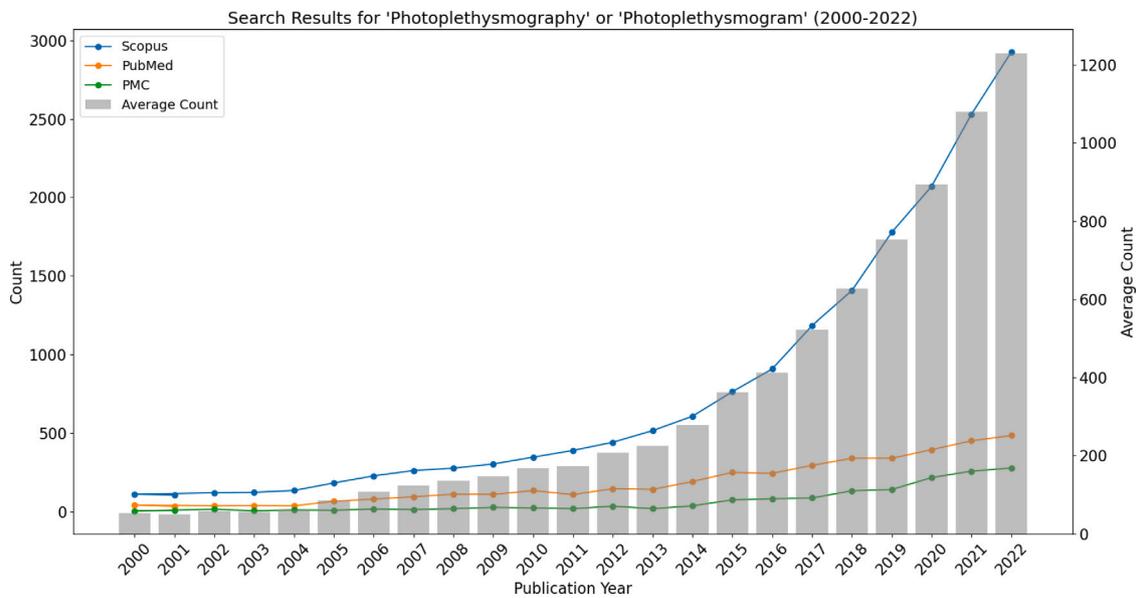
Fig. 1. Bar chart depicting the annual number of articles indexed in PubMed from 2000 to 2022, using the keywords "Photoplethysmography" or "Photoplethysmogram".

efficiency, recognising the practical constraints of real-time processing and resource utilisation.

We distinguish between two main approaches to denoising: (i) anomaly detection denoising and (ii) filter-based denoising. Anomaly detection denoising refers to methods that identify and eliminate unwanted anomalous data points or indexes within a dataset. On the other hand, filter-based denoising involves modifying the frequency content of a signal by selectively allowing or attenuating specific frequency components.

Within these definitions, the EPDA is classified as an anomaly detection denoising approach. Anomaly detection-based denoising proves particularly valuable when dealing with continuously labelled datasets, where maintaining synchronisation between the data and the label source is crucial. This approach involves storing the indexes of anomalous data, facilitating the removal of anomalous data points from both the data being denoised and the corresponding indexes from the label source, thus preserving data-label integrity.

There is a dearth in research focusing on non-filter based PPG denoising. Recently there has been a rise in the use of machine learning (ML) techniques for denoising. Lee et al. introduced a Bidirectional Recurrent Auto-Encoder (BRDAE) for PPG denoising, showcasing a 7.9 dB improvement in signal-to-noise ratio for noise-augmented data during validation [5]. Kwon et al. proposed a transformer-based deep generative model to eliminate noise within PPG signals [6]. Mohagheghian et al. developed a convolutional autoencoder approach for noise reduction, reducing the average detected heart rate and root mean square error (RMSE) by 45.74% and 23%, respectively, on arterial fibrillation and non-arterial fibrillation data [7]. Xu et al. created a motion artifact removal Time-Delay Neural Network (TDNN) that uses the signal envelope to normalise PPG signal amplitudes while preserving other waveform features such as dominant frequency and pulse width [8]. There have also been alternative statistical approaches to denoising. Lin et al. introduced a denoising method which characterises the signal through the calculation of features using fiducial points (onset, peak, and offset points) and employs adaptive thresholds to classify and remove anomalous segments of the signal [9]. Dao et al. presented a motion artifact detection algorithm that utilises time–frequency spectral (TFS) features [10]. These TFS features help distinguish between motion artifact-corrupted segments of data and clean data segments. Substantial distinctions arise when comparing the EPDA to other denoising algorithms. The EPDA, as a statistical approach, offers distinct advantages over ML methods, primarily in terms

of enhanced transparency and interpretability, aspects inherently embedded within statistical methods but often lacking in ML approaches. ML techniques for denoising often require a large amount of labelled data for effective training, a resource that may not be readily accessible or easily generated. Furthermore, certain denoising approaches attempt to reconstruct noise-contaminated PPG data, potentially resulting in the loss of important features present in the uncorrupted PPG signal. Some algorithms within this domain exclusively target the capture of HR (Heart Rate) and SpO2 (Oxygen Saturation) information, sidelining the importance of preserving the signal's morphology and amplitudes, which may be of significance for other physiological markers [11,12].

While certain research utilises the signal envelope this usage typically occurs as a step not central to the denoising itself. For instance, it is has been employed as a preprocessing step for data normalisation or for identifying fiducial points within noisy data [8,13]. To the best of the authors' knowledge, there is no existing denoising approach that leverages the envelope difference (calculated as the absolute difference between the upper and lower envelope of the signal), as the foundational basis for denoising. Another attribute of the EPDA, not prevalent in many denoising approaches is the preservation of the indices of identified anomalous data. This preservation becomes particularly valuable when dealing with synchronised labelled data. Existing statistical approaches to denoising tend to incur large computational costs through the calculation of spectral features, transformation of the data or the calculation of morphological features for each pulse cycle. We hypothesise that the EPDA through the efficient calculation of the envelope difference and use of the envelope difference to characterise the signal incurs a lower computational expense with respect to other denoising approaches which becomes particularly important when considering the application of algorithms to large datasets retrospectively or for real time use.

It therefore seems reasonable to suggest that the EPDA helps to address a research gap in the development of easily implementable, computationally efficient, and interpretable denoising approaches which are data efficient and which can preserve the indexes of identified anomalous data for use in synchronised, labelled datasets.

This paper introduces and evaluates the efficacy and computational efficiency of the EPDA using a comprehensive dataset of PPG signals. The EPDA is tailored towards the denoising of PPG signals where the primary application is the extraction of temporal and morphological features from the pulsatile AC component. We aim to encourage new avenues in the domain of PPG data processing, helping to lead to the
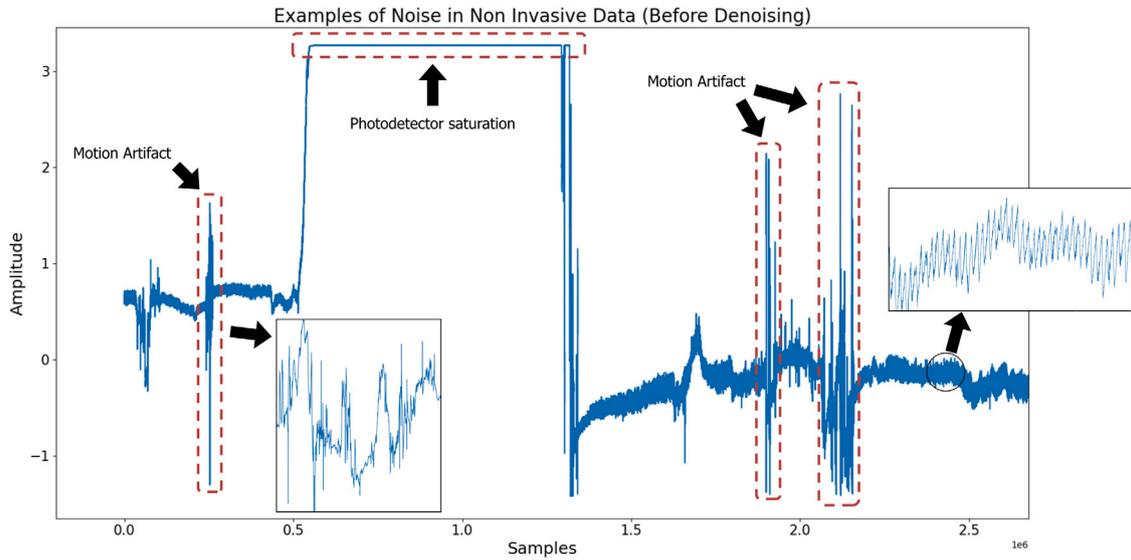
Fig. 2. PPG signal data with highlighted noise attributed to motion artifact and photodetector saturation. Visual aids emphasise the presence of noise within the window.

extraction of more accurate and reliable physiological measurements in the context of PPG based monitoring.

2. Materials and methods

2.1. Data

The findings presented in this paper are predicated upon data produced by an in-house, multi wavelength, near-infrared spectroscopy (NIRS) optical sensor [14]. The core dataset is comprised of reflectance PPG data obtained from 31 patients with traumatic brain injury (TBI). The data was collected with the aim to investigate the affect of intracranial pressure on the morphological features of PPG data. Within this context, the research focused upon the 810 nm wavelength which was selected because of the isobestic point where oxyhemoglobin and deoxyhemoglobin exhibit the same absorption characteristics [15]. This enabled the extraction of an optical signal independent of blood oxygenation and eliminating it as a confounding factor. Subsequently this study focuses on the data collected from the 810 nm wavelength.

The sensor was affixed to the patient's forehead below the hairline. During the collection of the data, the patients were in hospital beds, the majority of patients were sedated. The average age of the patients was 43.92 years, with a male-to-female ratio of 13:2. Data collection lasted, on average, for 40.55 h (SD 9.63). The total length of the dataset is 0.486 billion instances, representing approximately 81,015.99 min or 1350.27 h.

2.1.1. Manifestation of noise

Within the data, two main sources of noise are hypothesised to be present (i) motion artifacts, which are identified by irregular signal morphology or high amplitude variance, and (ii) photodetector saturation, characterised by areas with little or no amplitude variance referred to as "flat lines". Fig. 2 illustrates the presence of suspected motion artifacts and photodetector saturation within the dataset.

2.2. Envelope PPG denoising algorithm

The EPDA operates through a 6-step process: (i) the filtering and calculation of the upper and lower envelopes of the signal, (ii) the calculation of the *envelope difference*, (iii) the calculation of thresholds, (iv) detection of anomalous indexes, (v) the calculation of segmentation points, and (vi) the removal of identified anomalous data. In the initial step, a fourth-order bandpass Butterworth filter with cutoff frequencies

set at 0.5 and 12 is applied to the data removing undesired frequency components from the signal. Once the data has been filtered, the upper and lower envelopes of the signal are calculated. To calculate the envelopes, the peaks and troughs of the signal are detected using Scipy's "find peaks" function [16]. The "distance" input parameter of the "find peaks" function is set to 20% of the total length of the data window represented in seconds. This distance definition optimises computational efficiency by reducing the number of detected peaks and troughs. Consequently, fewer interpolation points are needed when computing the envelopes whilst preserving the ability to effectively characterise the anomalous regions of the signal. Once the peaks and troughs are identified, linear interpolation is used to calculate the upper and lower envelopes of the signal. Linear interpolation is preferred over other methods, such as spline interpolation, due to its computational efficiency. It estimates values by creating a straight line between two adjacent data points, capturing the relationships between points, and efficaciously representing anomalous segments of the data. The *envelope difference* is obtained by taking the absolute difference between the upper and lower envelopes. The *envelope difference* provides a simplified representation of the signal while preserving and emphasising significant changes in amplitude aiding in the characterisation of noise. Fig. 3 visually demonstrates the detected peaks and troughs, as well as the upper and lower envelopes, and the envelope difference. To demonstrate the step wise progression of the EPDA, the same window of data is used from Figs. 3 to 6.

The EPDA incorporates two distinct functions to effectively detect different types of noise. One function targets the identification of motion artifacts, while the other focuses on detecting flat line segments in the data. For motion artifact detection, the EPDA dynamically establishes upper and lower thresholds based on the interquartile range of the data. The thresholds are calculated using the Eqs. (1) and (2)

$$upper = 75\text{th } percentile + (threshold * interquartile\ range) \qquad (1)$$

$$lower = 25\text{th } percentile - (threshold * interquartile\ range) \qquad (2)$$

The IQR is robust to outliers and provides a measure of the spread of data that is less influenced by extreme values than measures like the standard deviation. By using the IQR, we ensure that our thresholds are based on the central 50% of the data, making them less sensitive to outliers and more representative of the typical data distribution. The choice of percentiles, specifically the 25th (lower quartile) and 75th (upper quartile) percentiles, is a common approach in statistics. These percentiles divide the data into quartiles, with the IQR representing the
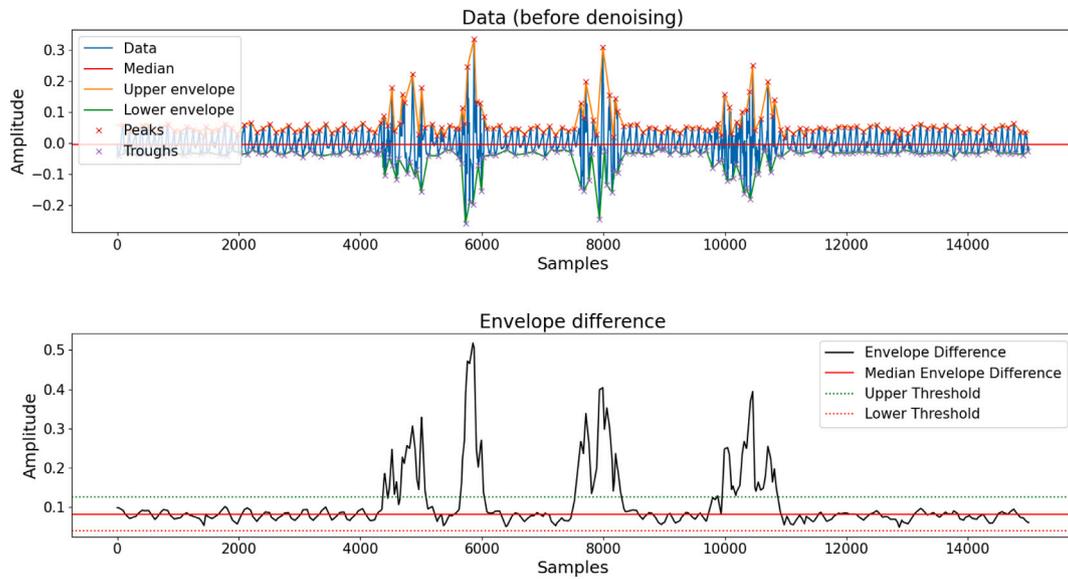
Fig. 3. Figure with two subplots: The top subplot displays the raw data, peaks, troughs, median of the raw data, and upper and lower envelope. The bottom subplot illustrates the envelope difference, median of the envelope difference, and upper and lower thresholds.
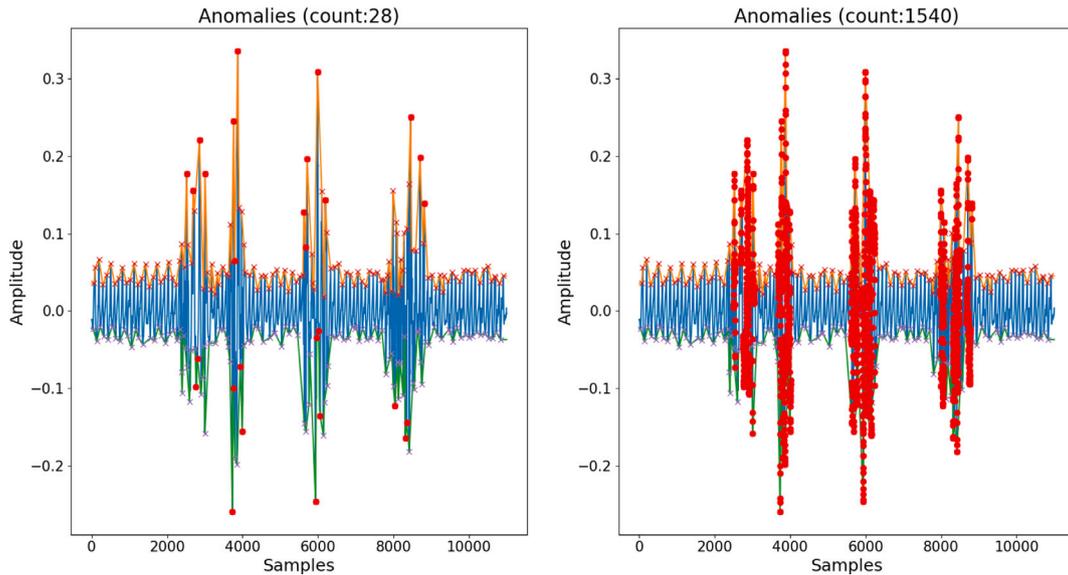
Fig. 4. Figure with two subplots: The left subplot shows the results of the optimised anomaly detection approach (count: 28). The right subplot displays the results of the unoptimised anomaly detection approach (count: 1540). A 5400% difference between the two approaches.

spread of the middle 50% of the data. By setting the lower threshold at the 25th percentile and the upper threshold at the 75th percentile, we are effectively capturing the central range of data where most observations are expected to fall. This ensures that our thresholds are meaningful and relevant to the majority of data points. By introducing a user-defined threshold multiplier we aim to enable flexibility and adaptability in the anomaly detection process. By multiplying the IQR with the user-defined threshold value, we enable users to adjust the strictness of the thresholds according to their specific needs. A higher multiplier makes the thresholds more stringent, while a lower multiplier makes them more permissive. This parameter allows users to fine-tune the sensitivity of the algorithm to anomalies.

The user can define the *threshold value*, with a default value of 2 (within this study, the default value of 2 was maintained across patient data). When detecting flat line segments, both the *line height threshold* and *line temporal threshold* are user-defined. The *line height threshold* determines the minimum required amplitude to be considered a flat

line, while the *line temporal threshold* specifies the duration for which data must remain below the amplitude threshold to be identified as a flat line. Thus, if a segment of data remains below the amplitude threshold for a duration equal to or longer than the temporal threshold, it is identified as a flat line.

Within the function for motion artifact detection, all instances where a slope change occurs within the envelope difference is calculated. Any instance of data with a slope change that exceeds the upper threshold or falls below the lower threshold is identified as an anomaly. Pseudocode 1 outlines the motion artifact anomaly detection algorithm.

In order to increase the computational efficiency of the denoising algorithm the motion artifact detection ensures that not all instances of data between the first and last threshold crossing points are defined as an anomaly. Instead, only instances of data where a slope change occurs within this range are considered anomalies. This logic minimises unnecessary iterations when calculating the segmentation points for each detected anomaly and substantially reduces the computational

Pseudocode 1 Motion artifact detection

1: slope_change_indexes ← empty list ▷ Initialise an empty list to store slope change indexes
2: **for** $i \leftarrow 1$ **to** length(envelope_difference) − 1 **do**
3:     diff_prev ← envelope_difference[i] − envelope_difference[$i − 1$]
4:     diff_next ← envelope_difference[$i + 1$] − envelope_difference[i]
5:     **if** (diff_prev ≥ 0 **and** diff_next < 0) **or** (diff_prev < 0 **and** diff_next ≥ 0) **then**
6:         slope_change_indexes.append(i) ▷ Add current index to the list
7:     **end if**
8: **end for**
9: anomalies ← empty list ▷ Initialise an empty list to store anomalies
10: **for** index **in** slope_change_indexes **do**
11:     **if** envelope_difference[index] > Upper Threshold **or** envelope_difference[index] < Lower Threshold **then**
12:         anomalies.append(index)
13:     **end if**
14: **end for**

expense of the algorithm. Fig. 4 depicts the difference between the two approaches of detecting anomalies which amounts to a 5400% difference between approaches for the example data.

The logic for the detection of flat line segments of the data can be divided into two steps. The first step involves applying the user-defined *line height threshold* and representing the data as 0 s and 1 s. Values in the envelope difference that exceed the *line height threshold* are set to 1. All values less than the *line height threshold* are set to 0. This representation allows for the simple characterisation of the start and end points of potential flat line segments in the data. Once the data is represented as 0 s and 1 s, flat line segments are identified using the algorithm described in Pseudocode 2.

Pseudocode 2 Flat line detection

1: flat_line_segments ← empty list ▷ Defining a list to store the flat line segments of the signal
2: diffs ← list ▷ A list containing the differences between adjacent elements in envelope_difference
3: line_height_threshold ← float ▷ A user defined height threshold
4: line_temporal_threshold ← float ▷ A user defined temporal threshold
5: height_detection ← where(abs(data_chunk_filt) < line_height_threshold, 0, 1) ▷ Representing the data as either 0 or 1
6: segment_indices ← find_non_zero_segments(diffs) ▷ The indices in diffs where the difference between adjacent elements is non-zero
7: segment_indices ← [0] + segment_indices + [length(height_detection)] ▷ Concatenate the segment indices with the start and end indices
8: **for** $i \leftarrow 0$ **to** length(segment_indices) − 1 **step** 2 **do**
9:     start_index ← segment_indices[i]
10:     end_index ← segment_indices[$i + 1$] − 1
11:     **if not** any(height_detection[start_index : end_index]) **then** ▷ If all height_detection values between start_index and end_index are zeros
12:         **if** (end_index − start_index + 1) ≥ ($fs × line\_temporal\_threshold$) **then** ▷ Check if the candidate segment meets the temporal threshold
13:             **if** start_index < 0 **then**
14:                 start_index ← 0
15:             **end if**
16:             **if** end_index > length(envelope_difference) **then**
17:                 end_index ← length(envelope_difference) − 1
18:             **end if**
19:             flat_line_segments.append((start_index, end_index))
20:         **end if**
21:     **end if**
22: **end for**

The flat line detection code searches for 1 s within the augmented data which indicate the possible start and end points of flat line segments. If the distance between a start and end point is ≥ to the user defined *line temporal threshold*, the segment is identified as a flat line.

The output of the flat line detection code is a list of indexes representing the start and end indexes of the flat line segments in the data.

Once any motion artifacts and flat line segments are detected and their indexes recorded, suitable segmentation points are calculated for each identified anomaly. These segmentation points are determined based on the indexes where the envelope difference and the median of the envelope difference intersect or closely align. Once these indexes have been calculated the segmentation points for each anomaly is derived. The result of the segmentation code is a list of indexes that represent the start and end points of anomalous data. In the motion artifact detection algorithm, the crucial step of only considering instances of data that surpass a threshold and are indexes where slope change occurs as anomalies plays a significant role in the calculation of segmentation points. By focusing only on the instances where the slope change exceeds the threshold, the algorithm can highlight potential areas of interest, which are most likely to contain anomalies caused by motion artifacts. This approach reduces the number of iterations required and eliminates the computation of redundant segmentation indexes for the same identified anomaly. As a result, the computational complexity of the algorithm is significantly reduced. The algorithm can process the data more efficiently, leading to faster anomaly detection. This optimisation is important when dealing with large datasets, real-time processing, or resource-constrained environments, as it helps improve the overall performance of the motion artifact detection process.

Before removing the index ranges between the calculated segmentation points for each anomaly, the algorithm combines adjacent "segmentation sections" that are in close proximity to create larger segmentation sections. This process serves two purposes, it: (i) reduces the computational complexity of the algorithm by minimising the number of segmentation sections that need to be iterated, and (ii) incorporates potential data instances between segmentation sections that may not have been explicitly identified as anomalies but exhibit poor morphological quality. Once the anomalous regions of the data have been removed, the data is filtered using a lowpass 2nd order butterworth filter with a cutoff frequency of 10 hz. Fig. 6 present the data before and after denoising.

2.3. Algorithm evaluation

To evaluate the performance of the EPDA, an assessment was conducted using a PPG dataset comprised of approximately 81,015.99 min or 1360.27 h of data collected from 31 patients. In the absence of a "clean" reference dataset, the evaluation was carried out through the calculation and analysis of 5 metrics. These included: (i) Signal-to-noise Ratio (SNR), (ii) Variance, (iii) Total Variation (TV), (iv) Shannon entropy, and Instances-per-second (IPS)

SNR is used to quantify the strength of the signal relative to noise through the ratio of the power of the signal to the power of the noise. Throughout the data collection process, the patients were situated in hospital beds, the majority of patients were sedated. Given this, the patients remained predominantly motionless during data collection, occasionally being repositioned by nursing staff. As a result of this environment, it was inferred that the primary sources of noise within the data were likely to be of higher frequency outside of the typical frequency bands of the PPG. Given this context and the absence of a standardised method to calculate the SNR a cutoff frequency of 10 Hz was deemed appropriate in order to isolate the morphology of the PPG using a lowpass filter and to isolate higher frequency noises using a highpass filter. When calculating the SNR of the data, the signal was filtered using a 2nd order, lowpass butterworth filter with a cutoff frequency of 10 Hz. The noise within the signal was isolated using a 2nd order, highpass butterworth filter with a cutoff frequency of 10 Hz. The SNR equation:

$$\text{SNR} = \frac{P_{\text{signal}}}{P_{\text{noise}}} \tag{3}$$
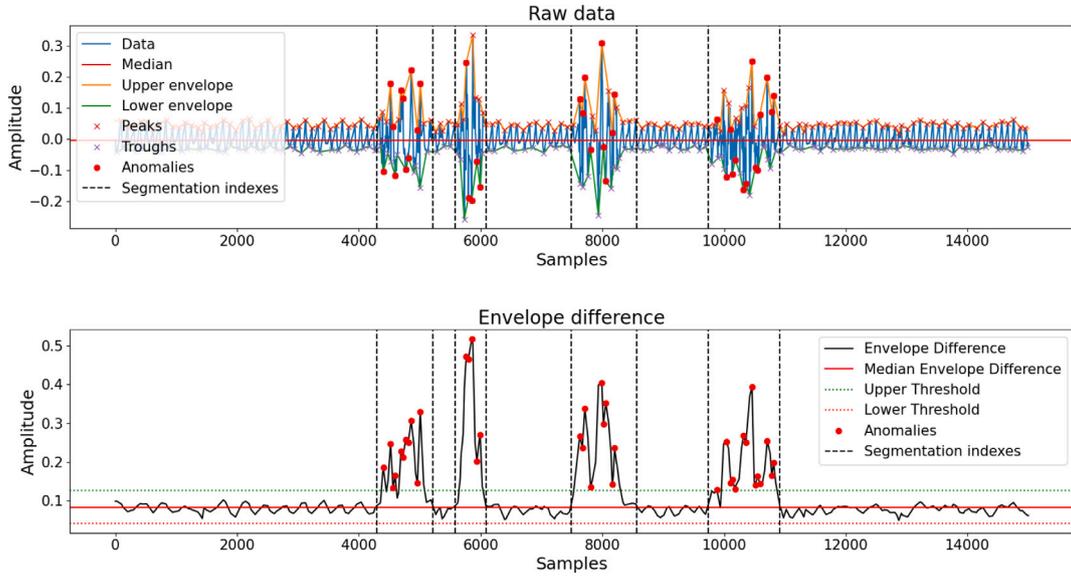
Fig. 5. Figure with two subplots: The top subplot presents the raw data, including peaks, troughs, median values, upper and lower envelopes, anomalies, and calculated segmentation points. The bottom subplot illustrates the envelope difference, median difference, upper and lower thresholds, along with anomalies and calculated segmentation points.
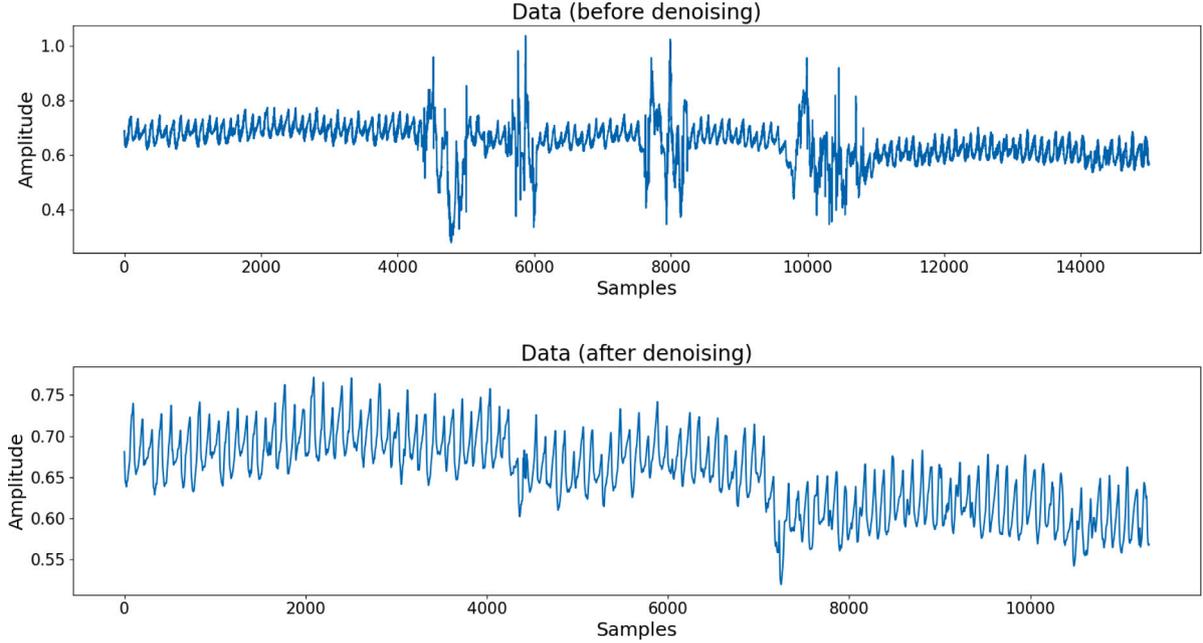
Fig. 6. Figure with two subplots: The top subplot presents the raw data. The bottom subplot illustrates the denoised data.

Where P_{signal} is the power of the signal, calculated as the mean of the squared signal and P_{noise} is the power of the noise within the signal, calculated as the mean of the squared isolated noise. A higher SNR value indicates a stronger, more distinguishable signal compared to noise, whereas a lower SNR value indicates that the signal is relatively weak compared to the noise. A desirable outcome after denoising would be the increase in SNR, indicating effective removal or reduction of noise while retaining signal content.

Variance captures the spread or dispersion of data points around the mean, reflecting the average squared difference of each data point from the mean value, formally represented as:

$$\text{Var}(X) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2 \tag{4}$$

Where N represents the total number of data points in the dataset and x represents each individual data point in the dataset. TV measures the overall variations in signal intensity across the entire duration by summing the absolute differences between adjacent data points:

$$\text{TV} = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \tag{5}$$

Where N represents the total number of data points in the dataset and x represents each individual data point in the dataset. We hypothesise that a reduction in variance and TV values before and after denoising signify smoother and less variable signals, indicating efficacious noise suppression and enhanced signal regularity.

Shannon entropy characterises the complexity or randomness of a signal by quantifying the average information required to describe the signal's values:

$$H(X) = -\int_{-\infty}^{\infty} f(x) \log_2(f(x)) \, dx \tag{6}$$
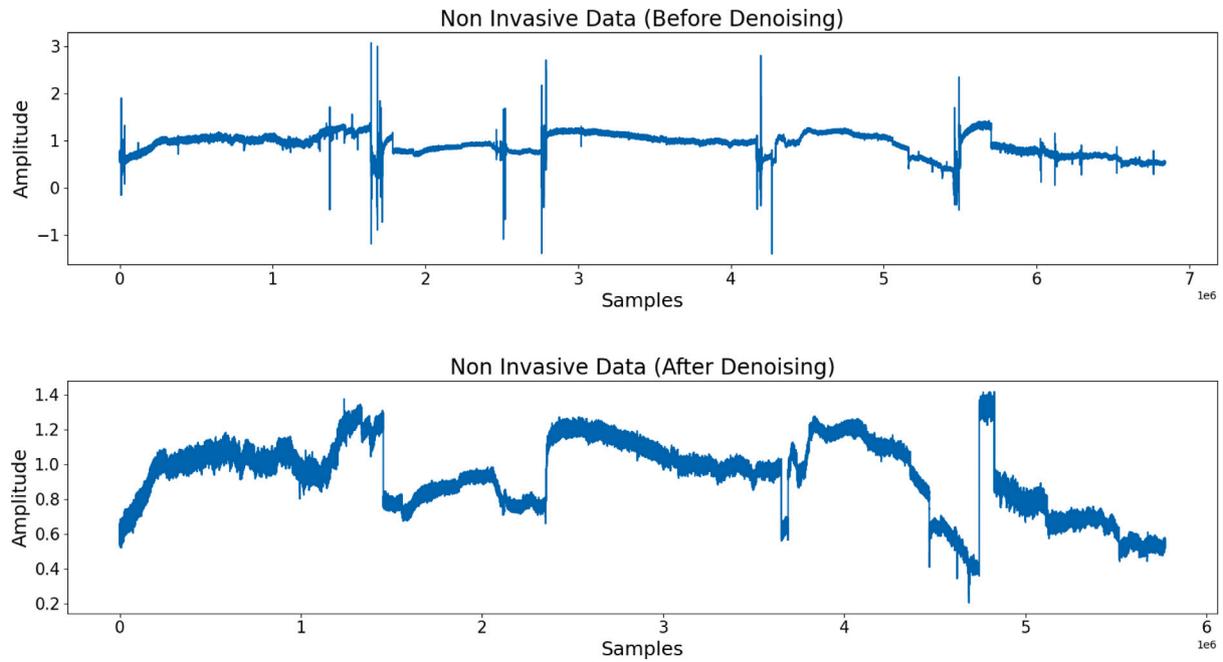
Fig. 7. Figure with two subplots: The top subplot presents the raw data of an entire patient. The bottom subplot illustrates the denoised data.

Where $f(x)$ denotes the probability density function (PDF) of the continuous random variable x, $\log_2(f(x))$ represents the logarithm base 2 of the probability density function. A higher Shannon entropy value indicates greater complexity or randomness, while a lower value suggests more predictability or regularity. We posit that a decrease in Shannon entropy post-denoising indicates successful noise reduction.

These 4 metrics, which can be calculated without a "clean" reference dataset, capture the dispersion, randomness and strength of a signal and thereby provide a credible means to evaluate the efficacy of a denoising algorithm. To assess the statistical significance of the differences in metrics calculated before and after denoising, a statistical testing approach was employed. Firstly, the normality of the calculated metrics was evaluated using the Shapiro–Wilk test. If the data in both groups exhibited a normal distribution ($p > 0.05$), a parametric paired t-test ($a = 0.05$) was chosen for further analysis. In cases where one or both groups displayed a non-normal distribution ($p > 0.05$), the non-parametric Wilcoxon signed-rank test ($a = 0.05$) was opted for. This dual testing strategy ensured a robust assessment of the statistical significance of the observed differences in metrics. The choice of parametric or non-parametric test was based on the underlying distribution of the data, enhancing the reliability and validity of the statistical analysis.

The computational speed of the EPDA was assessed by recording the run time of the algorithm on each data sample. From this a 5th metric, instances-per-second was derived which gives a representation of the speed of the EPDA. The analysis of the algorithm's speed enables an assessment of the EPDA's suitability for big datasets and real-time applications. This evaluation approach allows for a holistic understanding of the denoising algorithm's effectiveness.

3. Results

Fig. 7 provides a representative example of the complete dataset for one patient, both before and after denoising. Upon visual inspection, the data after denoising exhibits reduced extreme amplitude variations, which are believed to be associated with motion artifacts. Additionally, the range and dispersion of data following denoising appear narrower, suggesting a more uniform distribution. These observations suggest that the denoising process effectively mitigates motion-related artifacts and leads to a more consistent and refined dataset

Table 1 provides an overview of the EPDA's impact on various metrics, namely the SNR, variance, total variation, and entropy, for each patient. The data includes measurements before and after denoising, along with the computation time of the denoising algorithm and the derived metric IPS.

Examining the results, we observe the affects of the denoising algorithm on the data. The SNR indicates an average improvement from 53.3 (SD: 8.62) to 74.69 (SD: 7.63) after denoising. This represents a 40.16% increase in SNR suggesting that denoising has significantly enhanced the signal power, leading to a more accurate representation of underlying information.

Moreover, the variance, total variation, and entropy, which respectively capture the amount of variability, structural complexity, and information content in the data, exhibit a reduction of approximately 45.59%, 85.84%, and 2.16% on average after denoising. This signifies a successful reduction in unwanted noise and random fluctuations in the data.

The observed reduction of approximately 85.84% in total variation after denoising indicates a success in simplifying the data distribution and eliminating unnecessary variations.

Motion artifacts often introduce irregularities into the data usually manifesting within the signal as large amplitude variance. The decrease in total variation observed in our results suggests that the EPDA has effectively reduced the impact of motion artifacts, leading to a more uniform and smoother representation of the data. Motion artifacts which cause sudden changes in the amplitude of the data lead to higher values of total variation. When these artifacts are successfully removed, the abrupt changes caused by the motion artifacts are replaced with smoother transitions, resulting in a reduction in total variation.

To assess the impact of denoising on the calculated metrics, a statistical analysis was conducted. The normality of the data distribution for each metric was first evaluated using the Shapiro–Wilk test. The results revealed that the data of the SNR was normally distributed with a p-value of 0.057 and 0.055 for the "before denoising" and "after denoising" groups respectively. The data of the remaining 3 metrics was found to be non-normally distributed. Therefore a non-parametric Wilcoxon signed-rank test was chosen to determine if there were significant differences in all 4 metrics before and after denoising, using a significance level of 0.05. By using a single non-parametric test

Table 1

Table presenting the Signal-to-noise Ratio (SNR), Variance, Total variation and entropy of the data for each patient before and after denoising. The table also includes the computation time of the algorithm for each patient and the derived metric Instances-per-second.

Patient	SNR before	SNR after	Variance before	Variance after	Total variation before	Total variation after	Entropy before	Entropy after	Computation time (s)	Instances per second
1	31.189	65.903	0.048	0.013	140 773.859	10 909.716	16.527	16.352	36.899	411 135.192
2	47.007	65.737	0.386	0.235	21 739.31	4100.025	16.506	16.2	37.388	411 730.976
3	46.194	72.751	0.051	0.046	37 075.196	5405.411	15.734	15.601	16.071	425 686.583
4	61.008	78.478	0.343	0.084	9166.518	3364.145	16.16	15.897	31.84	339 511.046
5	48.193	66.293	2.982	0.601	71 059.79	6086.925	16.032	15.444	31.152	422 664.52
6	60.696	73.918	1.036	0.795	23 131.757	5004.932	16.575	16.194	41.182	409 219.627
7	41.05	67.459	0.423	0.256	192 933.432	19 756.286	16.689	16.501	43.945	417 796.334
8	54.202	73.708	0.324	0.277	10 639.948	1929.227	15.449	15.312	14.546	417 460.696
9	62.55	83.328	0.16	0.163	17 511.129	7775.109	16.695	16.588	45.844	400 679.008
10	44.915	72.867	0.445	0.394	83 002.505	18 462.602	16.642	16.458	38.469	449 186.729
11	50.319	68.012	1.405	0.553	72 820.156	10 622.558	16.592	15.763	36.006	479 913.608
12	44.613	69.338	1.009	0.551	171 845.39	14 079.652	16.568	16.259	38.774	441 844.878
13	59.473	77.376	0.656	0.57	35 550.527	6366.344	16.553	16.337	39.746	403 186.178
14	56.559	72.382	1.719	0.956	31 979.635	4606.092	16.558	16.004	47.846	359 541.675
15	51.393	67.601	1.497	0.885	36 806.149	10 484.826	16.54	16.289	47.248	364 712.38
16	45.931	62.772	2.355	0.702	144 723.747	14 598.065	16.459	15.975	36.288	476 012.512
17	55.117	74.773	0.57	0.153	17 258.795	3767.316	16.646	16.381	47.626	379 509.037
18	62.904	84.446	0.026	0.026	17 806.375	4194.915	16.639	16.512	43.365	390 169.125
19	66.553	89.476	0.015	0.009	6909.931	1307.428	16.614	15.562	40.313	407 831.278
20	61.429	91.887	0.016	0.008	28 334.381	1578.581	16.668	16.268	42.294	410 127.987
21	54.656	72.106	0.299	0.292	34 702.886	11 940.3	16.625	16.376	38.504	448 506.658
22	43.825	71.005	0.215	0.185	192 317.583	16 779.758	16.652	16.501	40.587	425 845.848
23	62.883	88.807	0.015	0.012	18 877.106	2112.399	16.659	16.128	37.575	457 268.441
24	51.973	70.694	0.356	0.159	29 089.716	4875.059	16.633	15.723	45.736	372 933.237
25	40.743	67.083	0.022	0.022	280 400.591	13 161.378	16.598	16.038	34.196	473 043.35
26	62.51	74.459	0.223	0.208	35 619.511	23 093.016	16.628	16.257	30.057	559 857.53
27	60.102	75.865	0.351	0.327	13 403.344	7517.603	16.169	15.973	25.66	417 419.782
28	59.835	86.582	0.027	0.021	18 288.607	3848.823	16.663	16.535	42.04	410 817.225
29	44.86	73.132	0.505	0.498	63 663.937	5187.896	15.949	15.612	19.78	438 663.051
30	59.41	76.07	0.122	0.085	31 772.232	20 243.942	16.656	16.476	46.778	371 073.191
31	60.352	80.996	0.717	0.879	28 295.027	8401.34	16.826	16.344	43.524	483 475.432

for all the metrics, despite the normality of one we ensure consistency across the analysis. Allowing for a reliable comparison of the denoising effect across all the metrics while accommodating potential data non-normality.

The Wilcoxon signed-rank test which compares the distribution of the differences between paired observations suggests that there was a significant difference in the SNR (p-value = $9.313225746154785e-10$), variance (p-value = $9.955838322639465e-07$), total variation (p-value = $9.313225746154785e-10$) and entropy (p-value = $9.313225746154785e-10$) before and after denoising. Showing a significant difference in all 4 metrics. Fig. 8 contains the boxplots representing the values for each of the metrics before and after denoising.

In addition to the analysis and interpretation of the 4 metrics, the IPS was also calculated for each patient's data. The calculated average IPS value across patients amounted to 421,833.003, signifying a substantive computational speed. This IPS can be presented as processing 70.3 min of data per second which suggests the EPDA may have potential for real-time applications and scenarios necessitating efficient data processing such as large retrospective studies.

This metric, when considered in conjunction with the analysis of the four other evaluation metrics, provides a comprehensive assessment of the algorithm's overall effectiveness. These findings demonstrate significant improvements in SNR, variance, total variation, and entropy metrics following denoising, as supported by both statistical summary of the results, statistical testing through the use of Wilcoxon signed-rank test and visual analysis of the boxplots.

4. Discussion

The evaluation of denoising algorithms in the absence of clean reference data is challenging. In our evaluation through the use of SNR, variance, total variance, entropy and IPS we have endeavoured to capture the dispersion, randomness and strength of the signal before and after denoising in addition to its computation speed to help evaluate the EPDA and its utility.

While searching for applicable evaluation metrics we noted that there is a dearth of research which focuses on PPG denoising evaluation, with some work on the quality assessment of pulses through the use of a number of signal quality indexes [17–19]. These indices include, skewness, kurtosis and perfusion index (PI). We suggest that these indices while they appear to be credible for the assessment of individual PPG pulses may not directly capture the impact of denoising. We propose that while skewness, which measures data asymmetry, and kurtosis, which measures peakedness or flatness, are relevant for characterising the morphology of individual PPG pulses, they may not offer meaningful insights into the effectiveness of denoising methods. Denoising techniques aim to reduce noise in the PPG signal, but their success may not always be easily interpretable using skewness and kurtosis alone. For instance, if a dataset contains substantial high-amplitude noise at the beginning, causing overall skewing, and this noise is successfully removed by denoising, it might appear as if the denoising was effective based on the restoration of symmetry. Conversely, if noise is distributed evenly throughout the dataset, denoising might effectively reduce the noise without significantly impacting the overall skewness. As a result, skewness and, similarly, kurtosis do not appear to be reliable indices for evaluating the efficacy of denoising algorithms.

Motion artifacts in PPG signals can introduce high-amplitude variance, leading to fluctuations in the signal that are not related to changes in perfusion. We posit that these motion-related fluctuations may artificially increase the amplitude of the PPG waveform, positively biasing the calculations of perfusion index (PI). PI is often calculated as the ratio of the pulsatile blood flow to the non-pulsatile blood flow, and motion artifacts can falsely elevate the pulsatile component, resulting in an inflated PI value. Therefore, when evaluating denoising algorithms on PPG datasets containing motion artifacts, denoising can effectively remove the noise and reduce the artificially inflated amplitude variance
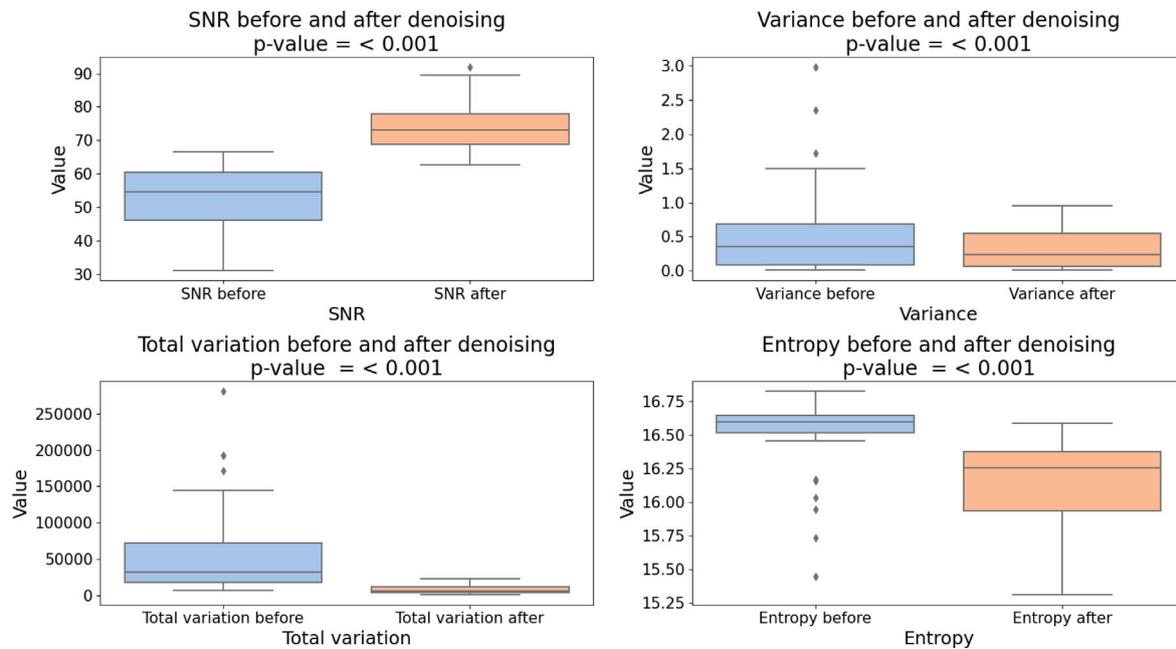
Fig. 8. Figure with four subplots, each subplot contains 2 boxplots representing the values for the signal-to-noise ratio, variance, total variation and entropy of the data before and denoising.

caused by motion, leading to a counterintuitive reduction in PI. This reduction may be misinterpreted as a deterioration in the signal quality, while in fact, it reflects the removal of motion artifacts and the restoration of a more accurate representation of the underlying physiological information.

The ability of the EPDA to return the indexes of identified anomalous data is important when dealing with two datasets that are synchronously collected. In scenarios where two datasets are collected simultaneously for purposes such as implementing supervised predictive modelling, maintaining temporal synchronicity between the datasets is crucial to ensure the production of reliable and accurate results. The EPDA's capability to identify and provide the indexes of anomalous data allows users to remove the corresponding indexes in the second dataset. By doing so, the temporal synchronicity between the two datasets is preserved. This is particularly useful when using PPG data to predict physiological biomarkers and constructing labelled datasets for supervised learning tasks to help insure the accuracy and reliability of the results obtained.

Our proposed algorithm demonstrates potential although we acknowledge there are avenues for improvement. One of these is the implementation of a sliding window approach. The sliding window approach in place of a fixed window size may provide the calculation of more contextually relevant thresholds by considering short-term fluctuations within each window. This is likely to facilitate the algorithm's ability to changes in signal characteristics and morphological quality resulting in a more robust approach.

A limitation of the study is the calculation of SNR without a "clean" reference and in the absence of a standardised approach the SNR calculations of this study were predicated on a cutoff frequency of 10 Hz under the assumption that due to the sedated or low motion state of the patients during data collection the majority of noise present within the data would be of higher frequency outside of the expected frequency band of the PPG. Consequently, the proposed SNR calculation is suitable when comparing what is expected to be within a normal PPG frequency range to higher-frequency noise. Which although may be suitable for our dataset but may not be generalisable. We suggest that future studies should be conducted to validate the EPDA on different datasets with a greater presence of noise within the PPG frequency band. The collection of bilateral PPG signals, with body movement added on one side could be collected to introduce motion artifacts into the signal and offer a reliable reference for SNR calculations.

The denoising of PPG signals plays a pivotal role in advancing non-invasive physiological monitoring. In our research, we have introduced our novel "Envelope PPG denoising algorithm" designed to accomplish the dual objectives of identifying and removing anomalous data while ensuring computational efficiency. Our research contributes to the advancement of PPG signal processing and its application in diverse healthcare domains, ultimately enhancing the potential for accurate patient monitoring.

5. Conclusion

This study aims to contribute to the field of PPG signal denoising, acknowledging its importance in the context of non-invasive physiological monitoring. Our investigation has focused on the development and evaluation of our "Envelope PPG Denoising Algorithm". Our approach has been shown to be efficacious in the denoising of PPG data, reporting significant differences in the SNR, variance, total variation and entropy after denoising. The IPS of the EPDA was calculated as 421,833.003, which can be represented as processing 70.3 min of data per second.

Our research has contributed a denoising algorithm, the EPDA, to the field of PPG signal processing. Our work augments the potential for non-invasive accurate patient monitoring across various healthcare domains, ultimately helping to improve the impact of non-invasive monitoring technologies, such as wearables, in enhancing patient care and wellbeing. Moving forward, we suggest continued research in this area, with a focus on refining denoising algorithms and in particular the detection and mitigation of motion artifact to help unlock the full potential of PPG technology in modern healthcare practices.

CRediT authorship contribution statement

George R.E. Bradley: Conceptualization, Methodology, Software, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Panayiotis A. Kyriacou:** Conceptualization, Investigation, Resources, Data curation, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgement

References

[1] Amy V. Creaser, et al., The acceptability, feasibility, and effectiveness of wearable activity trackers for increasing physical activity in children and adolescents: A systematic review, Int. J. Environ. Res. Public Health 18 (12) (2021).

[2] Denisse Castaneda, et al., A review on wearable photoplethysmography sensors and their potential future applications in health care, Int. J. Biosens Bioelectron 4 (4) (2018) 195–202.

[3] Panicos A. Kyriacou, John Allen, Photoplethysmography: Technology, Signal Analysis and Applications, Academic Press, San Diego, CA, 2021.

[4] Elisa Mejía-Mejía, et al., 4 - photoplethysmography signal processing and synthesis, in: John Allen, Panicos Kyriacou (Eds.), Photoplethysmography, Academic Press, 2022, pp. 69–146, http://dx.doi.org/10.1016/B978-0-12-823374-0.00015-3, URL https://www.sciencedirect.com/science/article/pii/B9780128233740000153.

[5] Joonnyong Lee, et al., Bidirectional recurrent Auto-Encoder for photoplethysmogram denoising, IEEE J. Biomed. Health Inform. 23 (6) (2018) 2375–2385.

[6] Ju Hyeok Kwon, et al., Preeminently robust neural PPG denoiser, Sensors (Basel) 22 (6) (2022).

[7] Fahimeh Mohagheghian, et al., Noise reduction in photoplethysmography signals using a convolutional denoising autoencoder with unconventional training scheme, IEEE Trans. Biomed. Eng. PP (2023).

[8] Ke Xu, Xinyu Jiang, Wei Chen, Photoplethysmography motion artifacts removal based on signal-noise interaction modeling utilizing envelope filtering and time-delay neural network, IEEE Sens. J. 20 (7) (2020) 3732–3744, http://dx.doi.org/10.1109/JSEN.2019.2960370.

[9] Wan-Hua Lin, et al., A characteristic filtering method for pulse wave signal quality assessment, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC, 2019, pp. 603–606, http://dx.doi.org/10.1109/EMBC.2019.8856811.

[10] Duy Dao, et al., A robust motion artifact detection algorithm for accurate detection of heart rates from photoplethysmographic signals using Time-Frequency spectral features, IEEE J. Biomed. Health Inform 21 (5) (2016) 1242–1253.

[11] M. Raghu Ram, et al., Computation of SpO2 using non-parametric spectral estimation methods from wavelet based motion artifact reduced PPG signals, in: 2011 International Conference on Signal Processing, Communication, Computing and Networking Technologies, 2011, pp. 776–780, http://dx.doi.org/10.1109/ICSCCN.2011.6024656.

[12] M. Raghu Ram, et al., A novel approach for motion artifact reduction in PPG signals based on AS-LMS adaptive filter, IEEE Trans. Instrum. Meas. 61 (5) (2012) 1445–1457, http://dx.doi.org/10.1109/TIM.2011.2175832.

[13] Manuel Merino-Monge, et al., Heartbeat detector from ECG and PPG signals based on wavelet transform and upper envelopes, Phys. Eng. Sci. Med. 46 (2) (2023) 597–608.

[14] Maria Roldan, Panicos A. Kyriacou, A non-invasive optical multimodal photoplethysmography-near infrared spectroscopy sensor for measuring intracranial pressure and cerebral oxygenation in traumatic brain injury, Appl. Sci. 13 (8) (2023) http://dx.doi.org/10.3390/app13085211, URL https://www.mdpi.com/2076-3417/13/8/5211.

[15] J.M. Murkin, M. Arango, Near-infrared spectroscopy as an index of brain and tissue oxygenation, Br. J. Anaesth. 103 (2009) i3–i13.

[16] Pauli Virtanen, et al., SciPy 1.0: Fundamental algorithms for scientific computing in python, Nature Methods 17 (2020) 261–272, http://dx.doi.org/10.1038/s41592-019-0686-2.

[17] Mantas Rinkevičius, et al., Influence of photoplethysmogram signal quality on pulse arrival time during polysomnography, Sensors (Basel) 23 (4) (2023).

[18] Mohamed Elgendi, Optimal signal quality index for photoplethysmogram signals, Bioengineering (Basel) 3 (4) (2016).

[19] G.D. Clifford, et al., Signal quality indices and data fusion for determining clinical acceptability of electrocardiograms, Physiol. Meas. 33 (9) (2012) 1419–1433.