Sensor Agnostic Photoplethysmogram Signal Quality Assessment using Morphological Analysis

Shahnawaz Alam, Shreyasi Datta, Anirban Dutta Choudhury, Arpan Pal
TCS Research
Kolkata, India
{shahnawaz.alam,shreyasi.datta,anirban.duttachoudhury,arpan.pal}@tcs.com

ABSTRACT

In this article, we propose a method to assess the clinical usability of fingertip Photoplethysmogram (PPG) waveform, collected from medical grade oximeter (train data) and smartphone (test data). We introduce a set of novel Signal Quality Indices (SQIs) to represent the noise characteristics of the PPG waveform. The SQIs are presented to a random forest classifier to discriminate between clean and noisy signals. The proposed method was evaluated on datasets annotated by four experts, resulting into a sensitivity and specificity of (92 $\pm$ 4.7 % , 95 $\pm$ 3 %) and (82.6 $\pm$ 4.6 % , 95.4 $\pm$ 3.1 %) on train and test data respectively. Further we applied the proposed method on PPG waveform of clinically proven control and disease population of Coronary Artery Disease (CAD), which resulted into (77 %,77 %) of sensitivity and specificity respectively.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Ubiquitous and mobile computing**;

KEYWORDS

Photoplethysmogram, Artefact, Morphology, Wearable Sensors, Noise Detection, Mobile Health

1 INTRODUCTION

Photoplethysmogram (PPG) signal obtained at the fingertip indicates the volumetric fluctuation of blood in the finger arterioles [1]. The PPG waveform is periodic in nature and its fundamental frequency indicates the heart rate. Over the last two decades, there has been a growing interest in the real-time monitoring of vital signs using PPG signal captured using wearables, as PPG has important information which can be correlated to certain symptoms and diseases [1]. However, PPG signals are prone to noise, due to motion artefacts, poor blood perfusion and changes in ambient light [2]. These artefacts impede the usability of PPG signal in real-time assessment/monitoring applications like false alarms in an Intensive Care Unit (ICU), inaccuracy in vital measurements like blood pressure, and screening of other diseases [3]. Hence, it is extremely necessary to check the quality of continuous PPG signal before measuring vitals. Also, in home care or rural setup, where people are not trained to collect PPG data, a real-time feedback of signal quality will help to capture clean signal.

Pulse oximeter, a mobile monitoring device is the ubiquitous non-invasive device to extract PPG signal used for measuring heart rate (HR) and oxygen saturation levels [4]. However, recent increase in the use of smartphone and wearable devices with inbuilt camera and IR sensors has broadened the scope of sensor exploitation. Researchers are employing these sensors for extraction of PPG for mobile healthcare solutions. Nevertheless, the main challenge lies in the heterogeneity of these sensors. Particularly, smartphone cameras, when compared to pulse oximeter, are non-medical grade and have varied set of hardware specifications and efficiency, which impacts the PPG waveform vastly. Their low and vacillating sampling frequency (frames per second) and presence of inherently unstable reflective sensing (wherein transmitter and receiver are on the same side) degrades the accuracy of the captured PPG signal [5], resulting in the PPG samples obtained from smartphones being more prone to noise and artefacts than pulse oximeter counterpart as shown in Fig. 1. This, therefore, poses a need for generalized PPG signal assessing technique, portable across devices.

In recent times, PPG signal quality check has been an active area of research, with diverse studies centering on determination of optimal Signal Quality Indices (SQI) for assessing PPG signal. Li et al [3] has proposed a system wherein four SQIs, having thresholds set empirically through trial and error, are fused into one (termed as "qSQI") and used to classify each beat's quality in the dataset. Nevertheless, initial 30 seconds of every session are considered for its online template creation which are not being employed for analysis, thus making the system infeasible to be implemented in real-time scenarios, where there is limited time for analysis and estimation. In [6], Sukor et al has described two levels of hierarchical rule based classification to eliminate inadequate and fallacious samples. However, thresholds of the SQIs are not provided, and in addition, SQIs like pulse width and pulse amplitude threshold used in [6] vary a lot depending upon the pressure exerted on the sensor, sample resolution of sensor and heart rate of subject; hence cannot be generalized. In [7], Orphanidou et al has also made use of hierarchical rule engine to obviate bad PPG samples using rigorous constraints where in the sample is predicted "bad" on failure of even one criterion. Moreover, thresholds on template matching correlation SQI used in [7] varies across devices, for example smartphone based PPG signal tends to have lower correlation coefficient than pulse oximeter counterpart due to low sampling rate and sensitivity of the camera sensor. In [8], Elgendi et al has taken a statistical approach to find out the optimal SQI for assessing PPG samples

MobiQuitous 2017, November 7–10, 2017, Melbourne, VIC, Australia

ACM ISBN 978-1-4503-5368-7/17/11…$15.00
https://doi.org/10.1145/3144457.3144487

in which out of 8 SQIs stated, "skewness" index is said to outperform. But such indices are validated on a small number of PPG samples collected using the same sensor and hence are liable to fail across different types of sensors. Most of these techniques use heuristically determined thresholds to assess PPG signals which might limit their use to a specific device or system. Hence, what is needed, is a robust sensor agnostic PPG quality checker, focusing on stable PPG features that are sensor independent.

A signal with good signal quality refers to a signal whose morphological properties remains stable over time without having artefacts. Nevertheless, fluctuations of the signal due to variations in physiological vitals like Heart Rate (HR) or Blood Pressure (BP), might alter the signal morphology, but this should not perturb the decision of the noise detection algorithm. Thus, the algorithm should be sensitive only to artefacts, and not to the other variations in the PPG waveform, occurring due to pathological conditions. In this article, we describe a method to detect PPG signals that are free from artefacts and can be used for clinical purposes, using an ensemble based classifier over some novel PPG waveform based SQIs. For analyzing the performance of our algorithm, predictions are compared with manual annotations provided by expert annotators.

It is well known that PPG signal annotation is not a straightforward or simple process [8]. Recently, Orphanidou et al. [7] has proposed an annotation methodology based on Heart Rate (HRs). Their annotation is a binary decision - "good" (i.e., having reliable HR) and - "bad". In other words, if a heartbeat derived from a sample is found to be in a particular range, the sample is classified as clinically fit. This, however, will only work in cases where the goal is to detect HR, but for evaluating complex physiological process it provides an incomplete view of the PPG signal. PPG waveform can be morphed easily due to it's sensitivity towards body motion and sensor error, wherein signal might alter with the introduction of non-stationary variation. Therefore, in such cases HR component might be visible in frequency domain, however, the component of artefacts will also be prominent as seen in Fig. 1, including higher frequency noises. Therefore, to tackle these difficulties we will also address the waveform morphology to annotate the PPG signals.

The rest of the paper is organized as follows: Section 2 describes the datasets used and the annotation technique, methodology of proposed method is presented in Section 3. Section 4 shows the results and observation, while section 5 describes case study and finally section 6 concludes the article.

2 DATASET DESCRIPTION & ANNOTATION

Though there are few publicly available PPG database (e.g. Physionet [1]), to the best of our knowledge, none of them includes signal quality annotation. Thus, in this work, we take a manual annotation approach. Moreover, we would be requiring PPG waveforms from multiple sensor sources to find sensor agnostic SQIs. We create our own datasets comprising 1) the ICU unit of an urban hospital in India and 2) a basic healthcare unit of an Indian village. In the first case, PPG signals are collected using Pulse oximeter (Contec CMS 50D+ [2]) at 60 Hz of sampling rate while the 2^{nd} PPG dataset are collected using an in-house smartphone video camera app to capture red component of the rgb frames at fixed frame per second of 24 Hz. All PPG signals are collected from the right hand index finger of every subjects, as shown in Fig. 2. A total of 85 subject's (Pulse oximeter : 35 & Smartphone : 50) data are captured having diverse age, weight, height (shown in Fig. 3) and clinical conditions in a stable sitting posture. For collecting data, the permission from the Institutional Review Board (IRB) and the written consent of the subjects are taken.

Analyzing the datasets, annotators found that smartphone camera based PPG signal is more prone to jitter caused by random noise and motion artefacts. While pulse oximeter based PPG signals have significantly clean morphology, however with plenty of flat and clipped portions due to intermittent signal loss. Considering prevalent noise pattern differences in the datasets, finding out a common set of SQIs for quality check is a challenge. Smartphone data being significantly prone to noise, as formerly explained and shown in Fig. 1, are considered as hidden test dataset for evaluating our algorithm while pulse oximeter data is used as training dataset.

For a real-time PPG signal quality assessment, we take into account fixed small window based quality check instead of longer samples to reduce data drop and computation loss. However, for a PPG sample to retain sufficient information for analysis, a minimum window length must be adhered to. Considering the lowest human heart rate to be 40 beats (cardiac cycles) per minute i.e. 1.5 second per cycle and a minimum 4 such cycles (chosen heuristically) to observe morphological variation with maximum loss of 2 incomplete cycles (from start and end of the window), we keep the window length to be 8 sec. Therefore the collected PPG datasets are sub-divided into multiple chunks of 8 seconds window length which resulted into 1553 samples from pulse oximeter and 243 samples in Smartphone dataset. These 8 sec PPG signal chunk will be cited as "sample" throughout this paper.

A graphical user interface is built on Matlab GUIDE software as shown in Fig. 4 for manual annotation. Four expert annotators are presented with filtered PPG samples to be assessed one at a time. Bandpass filtering is done on the PPG samples with the cutoff frequencies of 0.5 to 6 Hz using Butterworth filter. Thereafter, PPG waveform patterns are analyzed and classified into two discrete quality levels: 'Clean' and 'Noisy'. For scoring purposes, a 'radio button' facility is provided in the GUI as a tool to assist the scorers in assigning the pulses to an appropriate quality class. Following are the set of criteria annotators considered for a PPG sample to be clean

- every cardiac cycle in the sample have similar morphological pattern,
- every cardiac cycle in the sample have distinct primary peak and troughs,
- sample does not contain clipping, and
- absence of any broken and unwanted non PPG signal.

A total of 1796 samples are randomly shuffled (to reduce bias over consecutive waveform while annotation) and annotated 'Clean' or 'Noisy'. We however, know that annotation is a subjective decision making process, thus final annotation is developed based on majority voting of annotations done by experts. A sample annotated clean and noisy twice is not taken into consideration. However, we do not encounter a such tie situation in the dataset. Out of 1553

[1]https://physionet.org/
[2]http://www.contecmed.com/

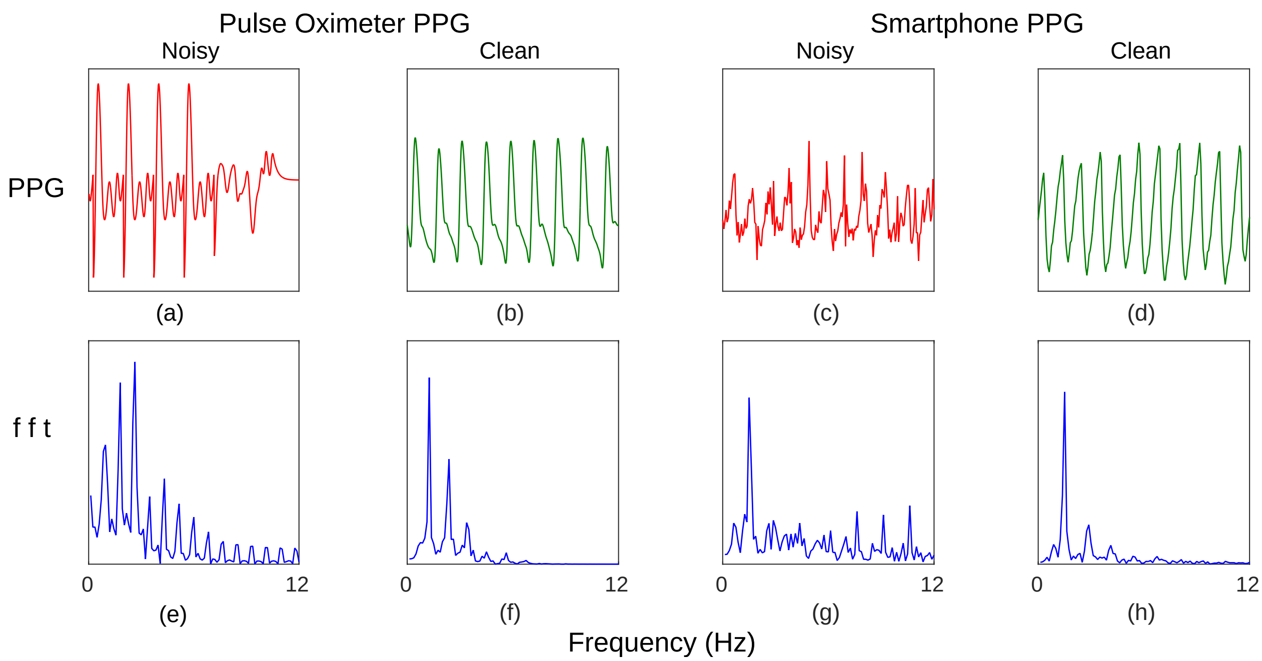
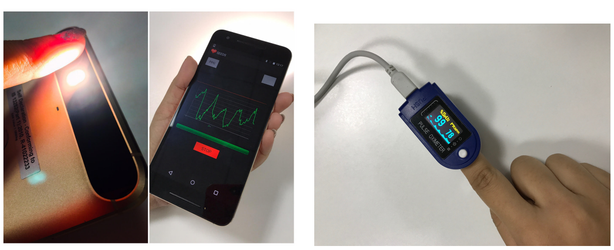


Figure 1: Clean & Noisy PPG samples captured from Pulse Oximeter & Smartphone with corresponding FFTs



(a) (b)

Figure 2: Data collection using (a) Smartphone camera with flash light on, and (b) Pulse Oximeter

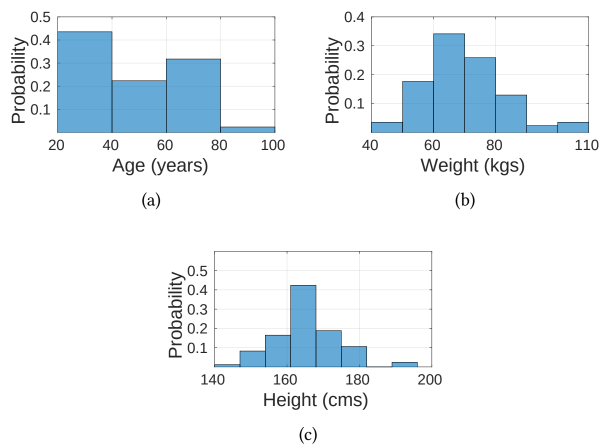


Figure 3: Histogram of (a) Age, (b) Weight and (c) Height of the subjects

samples 273 are labeled as corrupt or noisy in oximeter dataset while in smartphone dataset 66 out of 243 samples are labeled as noisy, also depicted by a pie chart in Fig. 5. Inter Rater Agreement - shows the correlation of 0.9359 ± 0.05 from all the 6 combinations of 4 sets of annotations.

3 METHODOLOGY

Post development of the annotation process an algorithm is developed, summarized in Fig. 6, to classify clean and noisy PPG samples. We introduce hierarchical decision making approach by combining heuristics and machine learning technique, ordered below:

- Sufficiency check of the sample, and
- Sample prediction using two class classification.

Later we compare our method to the ground truth for performance evaluation. Following sections will describe the algorithm in details.

3.1 Signal Sufficiency Check

The Signal Sufficiency does the sanity check on the sample to ensure it has sufficient information for further processing. Following are the criteria for a sample to be considered sufficient:

3.1.1 Clipping & flat signal. Clipped and flat signals are a form of waveform distortion that occurs when signal amplitude goes beyond the range restricted by a chosen representation and have constant amplitude level, respectively. A typical flat and clipped signals are represented in Fig. 7. Presence of clipped or flat signal is undesirable while applying morphological operation, thus needs

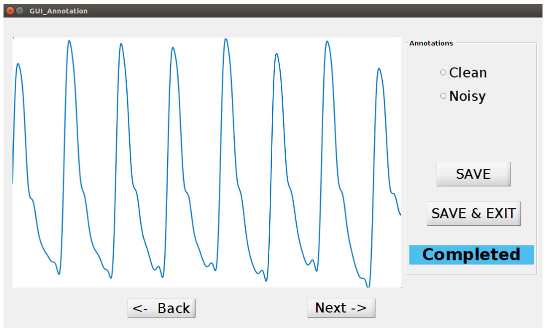


Figure 4: Annotation GUI

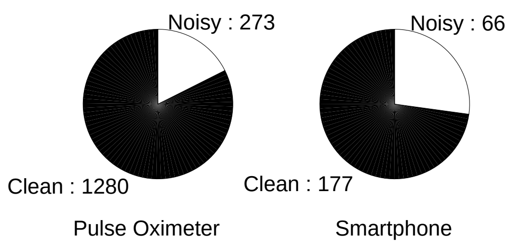


Figure 5: Distribution of Clean and Noisy Samples

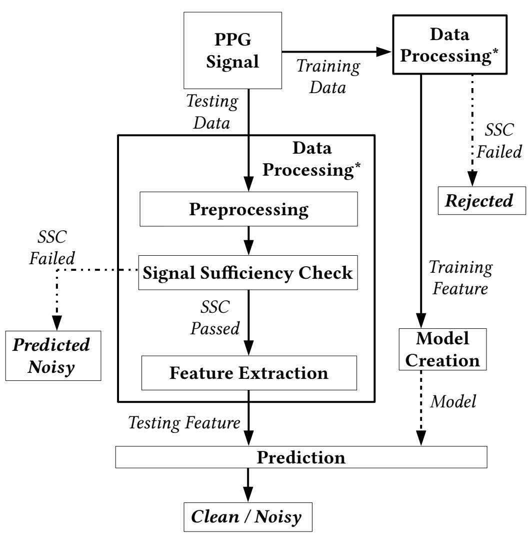


Figure 6: End-to-end block diagram of the proposed methodology

to be detected. Portions of signal having maximum, minimum or constant amplitude value are found. If any such portion contributes more than 10% of the sample length then the sample is considered as noisy.

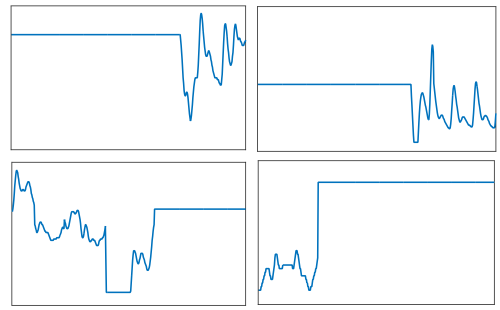


Figure 7: PPG samples having flat signal

3.1.2 Heart Rate. The HR extrapolated from the sample must lie between 40 and 150 cardiac cycles per minute (beats per minute). Though ideally it is possible to have HRs beyond this range, however this is the ideal physiological range of HR for the population likely to use screening devices.

If the signal sample passes both the criteria, it is then proceeded for template creation and further processing. However, if the sample fails to pass SSC in training phase, it is not considered for training the model, and in testing phase is predicted as noisy.

3.2 Template Creation

Individual PPG samples are segmented into cardiac cycles using subsequent trough positions for template creation. Peaks and troughs of filtered samples are detected using a threshold based approach as described in [9]. Superimposing cardiac cycles centering around the peaks reveal that despite the differences in base amplitude and pulse width, clean pulses taken from a short duration are almost identical morphologically, except when corrupted with artefacts [6]. The proposed approach computes the median of all the cycles to create online template *T*. This ensures that *T* retains the most frequently occurring morphological features in the sample as shown in Fig. 8

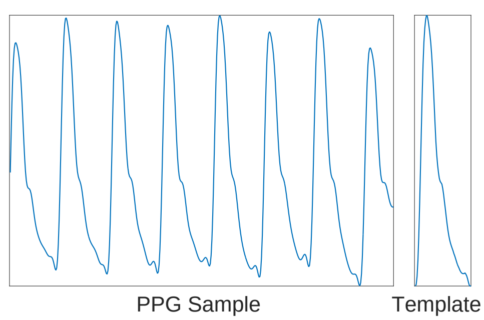


Figure 8: PPG sample with corresponding template

3.3 Proposed Novel Signal Quality Indices (SQIs)

A detailed description of the set of SQIs considered for our algorithm are defined below. The novel SQIs introduced in this article are numbered 1 to 4. SQIs numbered 5 and 7-10 are derived from [8] [6] [3], and SQI_6 is taken from [7].

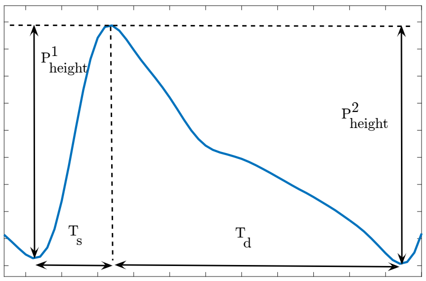


Figure 9: PPG sample with defined time domain features (t_s, t_d, P^1_{height}, and P^2_{height})

3.3.1 SQI_1. Ratio of Cardiac diastole duration (t_d) to Ventricular systole duration (t_s) are computed for every PPG cycle in the sample, shown in Fig. 9. PPG cycle must have $t_d > t_s$ [3]. The $Cycle_{t_d > t_s} \subset Cycle$ are those cycles, thus $SQI_1 = |Cycle_{t_d > t_s}| \,/\, |Cycle|$. The intuition behind using this SQI is to find that ratio of actual PPG cycles present in the sample.

3.3.2 SQI_2. Presence of signature which are not present in clean signal such as aperiodic signal and amplitude fluctuations (which may happen due to coronary perfusion pressure, motion artefacts, or variation in ambient lighting) are undesirable and thus need to be detected. The ratio of summation of every complete PPG cycle length extracted from sample to the sample length would be the clear indicator for the presence of such unwanted signal. $SQI_2 = \sum duration(\ Cycle\ )/duration(Sample\ )$

3.3.3 SQI_3. PPG being a quasi periodic and inherently repetitive signal, Dynamic Time Warping (DTW) can be used to find an optimal match between two time series as described in [3]. The sequences are aligned or "warped" non-linearly in the time dimension to determine an optimal path. $DTW_{distance}$, the length of the optimal path, is calculated for every cycle in the sample from the template T. The mean of $DTW_{distance}$ of the sample is SQI_3.

3.3.4 SQI_4. Baseline variation is very common in PPG signal, where beat-to-beat low frequency fluctuation reflects mechanical consequence of respiration on venous return [10]. The variation of the amplitudes due to baseline affects the morphology of the signal, thus for the stability check we take into account the peak heights from both preceeding and succeeding trough of every cardiac cycles. Considering a PPG cycle having two troughs (t_1 and t_2) and a peak pk in between, as shown in Fig. 9, two pulse heights of a cycle P^1_{height} (absolute height difference between t_1 and pk) and P^2_{height} (absolute height difference between t_2 and pk) are calculated. Variance of the ratio $P^1_{height}/P^2_{height}$ of every cycle in a sample is SQI_4.

Probability distribution of SQIs numbered 1 to 4 over clean and noisy samples on training dataset are depicted in Fig. 10.

[3]http://www.physiologyweb.com/calculators/mean_arterial_pressure_calculator.html

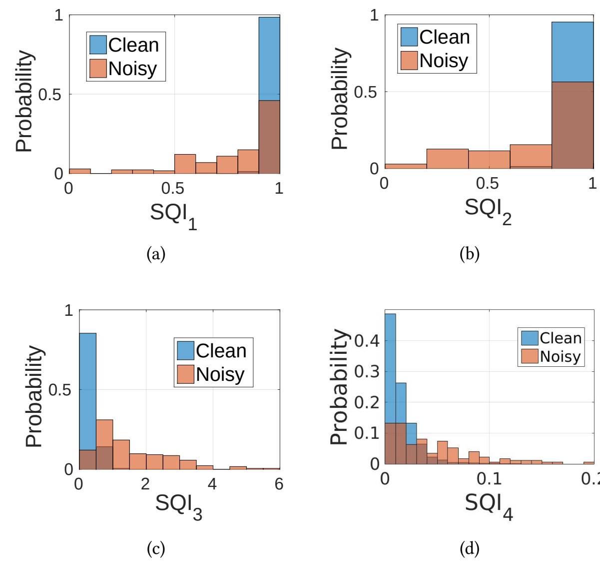


Figure 10: Probability distribution of (a) SQI_1, (b) SQI_2, (c) SQI_3, and (d) SQI_4 over clean and noisy samples

Table 1: SQIs derived from prior arts

SQIs	Description
SQI_5	variance of the kurtosis of every pulse, derived from [8]
SQI_6	ratio of max PP & min PP [7], where PP is peak to peak distance in a sample
SQI_7	variance of amplitude ratio (amplitude of cycle / amplitude of T), derived from [6]
SQI_8	median of direct matching (Correlation Coeficient) between cycle and T, derived from [3]
SQI_9	variance of Pulse Width , derived from[6]
SQI_{10}	variance of root mean square error of cycle and T, derived from [6]

3.4 Classification

Our aim is to develop a robust sensor-agnostic binary classification method for PPG signal quality assessment. Random Forest (RF) [11], an ensemble of decision trees is considered for the assessment of our method. RF classifier creates a bagging of decision trees where

Table 2: Feature Rank (novel SQIs in italics)

SQI Index	*1*	*2*	*3*	*4*	5	6	7	8	9	10
Rank	*3*	*2*	*1*	*6*	10	4	7	5	9	8
STD	*0*	*0*	*0*	*1.5*	0	0.4	1.5	0	0.8	0.8

successive trees do not depend on earlier trees and are independently constructed using a bootstrap sample of the dataset. In the end, a simple ensembling method (majority voting) is executed for prediction. This avoids an overfitted solution.

4 RESULTS AND OBSERVATIONS

4.1 Signal Sufficiency Check Evaluation

The Signal Sufficiency Check (SSC), explained in Section 3.1, is carried out on train and test datasets (described in Section 2), to remove PPG samples which are insufficient to analyze. It is observed that none of the clean PPG samples have failed the SSC block in both datasets, as shown in Fig. 11. Since the annotation of PPG samples is not done on the basis of completeness or sufficiency parameters uniquely, the evaluation is only done on the factor whether any clean sample has been failed by this block or not. Out of 273 noisy samples present in train dataset 99 failed in SSC, and are not considered for training. While on other hand 19 of 66 noisy samples in test dataset have failed SSC block, and are classified as noisy. After removing the SSC failed samples, training dataset is left with 1280 clean and 174 noisy samples, while testing dataset have 177 clean and 47 noisy samples remaining, as shown in Fig. 11.

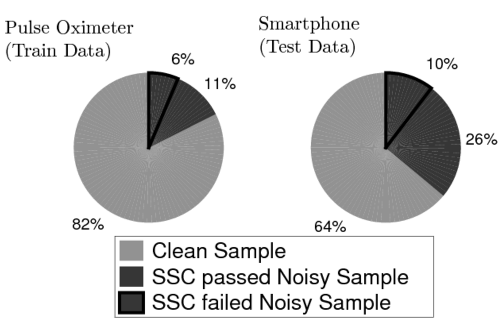


Figure 11: SSC performance on train and test dataset

4.2 Feature Selection

Considering the SQIs described in Section 3.3, we try to find an optimal set of SQIs for assessing the quality of PPG sample, which are not redundant, reduces computation time, produces high and low variance performance. SQIs are therefore ranked based on Minimum Redundancy Maximum Relevance (mRMR) [12] strength computed against the data class on training dataset. However, SSC passed training dataset is highly imbalance (Clean:Noisy class ratio 7:1, Section 4.1), which will not give a holistic view of the SQIs correlation over data class. Thus we create 7 sets of balance class subsets of training dataset each containing all noisy samples and equal number of randomly taken non overlapping clean class samples. For each of 7 subsets we compute rank based on mRMR strengths of SQIs individually.

The final rank shown in Table 2 is calculated by ranking the mean of rank sets over every subset. The final rank is correlated to initial 7 sets of rank series using Spearman's Rank Correlation Coefficients, with mean value 0.96, range [0.94, 0.98] and standard deviation of 0.01.

Selection of optimal set of SQIs is done using internal performance evaluation; here top n SQIs (Table 2), iterated from 1 to 10, are trained and validated using 5-fold in training dataset to find the optimal value of n having high and stable performance metrics. In Fig. 12, 5-fold performances on 7 subsets are depicted (edge of the boxes represent interquartile range), where x-axis represent the top n SQI set taken cumulatively. Proceeding thus, we find that the optimum value of n for best performance, in terms of specificity is at $n = 9$ and sensitivity is at $n = 5$. Since the major focus is to predict noisy class correctly (i.e high specificity) because we cannot afford to predict any noisy sample as clean on field, thus we trade-off sensitivity to choose $n = 9$. Therefore top 9 SQIs from Table 2, which includes all our proposed SQIs, are considered for evaluation hereafter.

4.3 Statistical Analysis of Signal Quality Indices

Considering a null hypothesis that the noisy and clean PPG samples comes from the distributions having equal mean, unpaired two sample t-test (Welch's t-test) is performed on selected SQIs over clean and noisy class population of test dataset, accounting for unequal variance. Differences are considered statistically significant at $P < 0.05$. From Welch's t-test result shown in Table 3, we can observe that all the SQIs except SQI_2 fails the null hypothesis (p-value of $SQI_2 > 0.05$), in other words only SQI_2 is not able to differentiate between clean and noisy class in smartphone dataset within 5 % significance level.

4.4 Performance Evaluation

The proposed method is evaluated in multiple scenarios described below,

4.4.1 Offline. Random Forest models are created for each of the 7 balanced subsets (as explained in Section 4.2) over the selected SQIs. Prediction on the test dataset is done using these 7 models and the result is shown in Fig. 13, where mean and standard deviation of

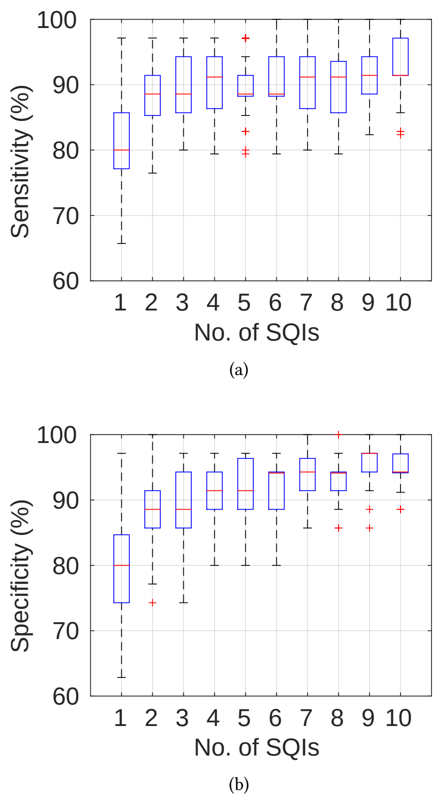


Figure 12: Mean (a) Sensitivity and (b) Specificity across 5-folds over 7 balanced subsets. The number of SQIs are cumulatively added as per ranking in Table 2

Table 3: *p*-value of Welch's *t*-test done on selected SQIs over clean and noisy class

SQIs	*p*-value	SQIs	*p*-value
SQI_1	$1.4 * 10^{-12}$	SQI_7	0.0015
SQI_2	0.32	SQI_8	$3.2 * 10^{-14}$
SQI_3	$5 * 10^{-18}$	SQI_9	0.0121
SQI_4	$8.5 * 10^{-07}$	SQI_{10}	$2.6 * 10^{-04}$
SQI_6	$9 * 10^{-07}$		

sensitivity are 82.6 % and 4.6 % while mean and standard deviation of specificity are 95.4 % and 3.1 %.

The classifier is also trained using the entire unbalanced (7:1) training dataset and then evaluated on the testing dataset to check the influence of majority class. This resulted to sensitivity and specificity of 96% and 87%, respectively. We further perform SMOTE (Synthetic Minority Over-Sampling Technique) [13] analysis to test the stability of classifier by varying the ratio of clean and noisy

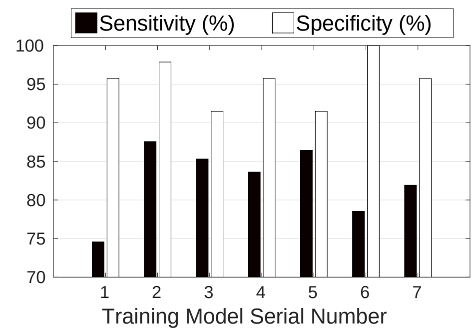


Figure 13: Sensitivity and Specificity of 7 predictor models on the test dataset.

class in training dataset. SMOTE upsamples the minor class from its population distribution using K-Nearest Neighbor (KNN) method. As shown in Fig. 14, when we gradually try to upsample the noisy samples in the training dataset, increase in specificity is observed. However, there is a drop in sensitivity by 10 % while the noisy dataset is increased to 20 times that of clean samples. With mere 5 % of clean instances in new training dataset, our model is still able to correctly classify 140 out of 177 clean samples. This proves that our model is not getting biased to the majority class.

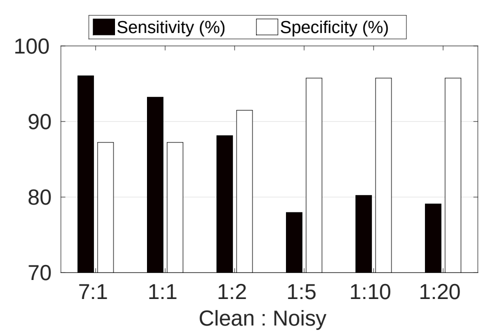


Figure 14: Sensitivity and Specificity of 6 predictor models (trained using multiple class ratios by upsampling noisy class in training dataset) evaluated on test dataset.

4.4.2 Online. Further we do a real-time analysis over the portion of PPG signal recorded separately using smartphone to observe the performance of the classifier. We have taken a 32 seconds long continuous signal as shown in Fig. 15 with experts annotated noisy portion and partitioned into 8 second windows with 50 % overlapping. The 7 predictors are then used for classification of these PPG samples. Posterior probabilities, probability of sample to be classified as clean, are evaluated wherein, the value of more than 0.5 is classified as clean class. The range of predicted posterior probabilities for every PPG sample are observed to be very small in Fig. 15, which shows high confidence of the predictors. Introduction

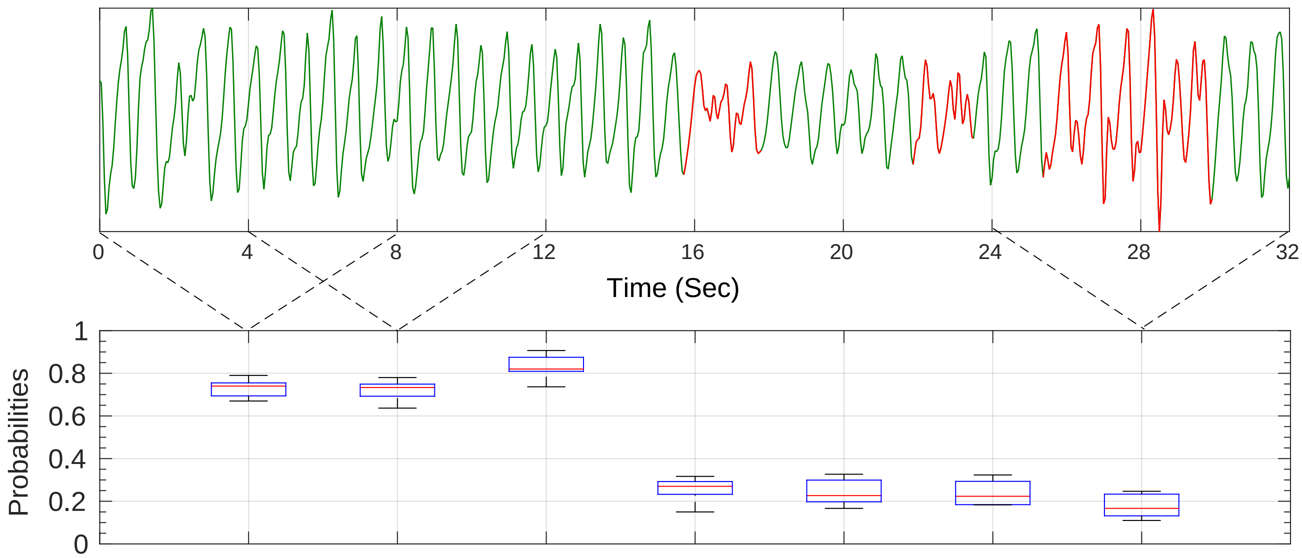


Figure 15: Posterior probabilities (clean : 1 & noisy : 0) of PPG quality computed on smartphone PPG signal, considering 8 sec window length with 50 % overlap

of even small artefact in 4^{th} window lead decision to be in favor of noisy class. 3^{rd} window being visually highly stable and clean in morphology - predicted probabilities are highest too. While the last window, with majority portion annotated as noisy, shows the lowest score. This, thus proves that our system is reliable on field.

4.4.3 Classifier Comparision. We also use some other standard classifiers for analyzing the goodness of the SQIs, including Support Vector Machine (linear SVM), Artificial Neural Network (ANN) with 1 hidden layer having 6 hidden nodes (chosen using internal cross validation), and K-Nearest Neighbor (Parameter : $k = (N)^{1/2}$ [14], where N is number of samples in training dataset). Individually each classifiers are trained on the whole unbalanced (7:1) training data and then tested on the testing data. As seen in Fig. 16, none of the classifiers are giving less than 90 % F_1 score. The high performance metrics shows that our SQIs are true representative of quality indices.

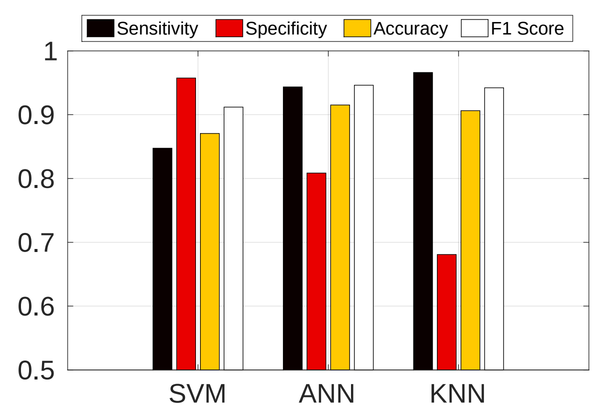


Figure 16: SVM, ANN, and KNN classifiers prediction on test dataset

Table 4: Prior arts result using defined methodology

Methods	Sen (%)	Spe (%)
[3]	100	10.6
[3] ML	100	20
[7]	43	98

4.4.4 Prior Art Comparision. We further compare our proposed method with the prior arts described in Section 1. Evaluation of methods are done in two phases, using the methodology defined in the corresponding prior art; and secondly, using our classification technique (using respective prior art's SQIs) to observe the improvement. In method [8] & [6], heuristically taken threshold for SQIs are not provided, therefore we only evaluate it using our classification method.

The first result, depicted in Table 4, shows the evaluation of the prior arts on smartphone dataset. Here, the methods [3] (both heuristic and machine learning approaches) and [7] have considered

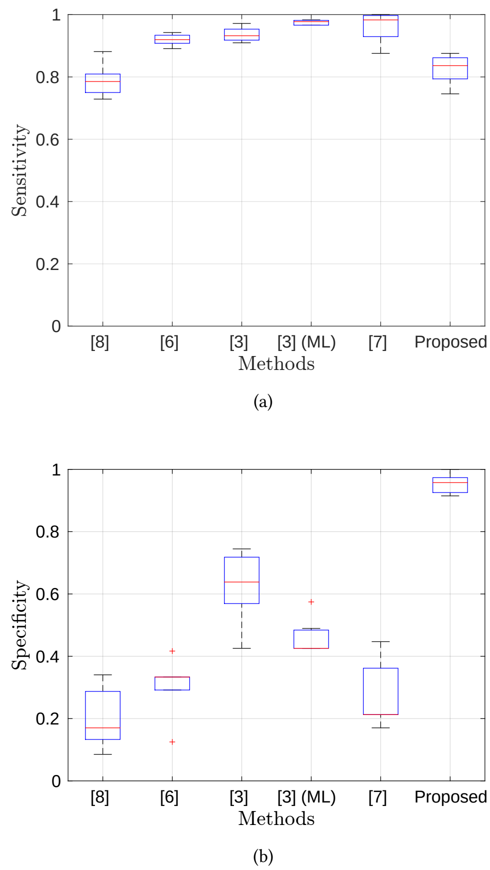


Figure 17: (a) Sensitivity and (b) Specificity of prior arts and proposed method (implemented using RF classifier)

heuristically determined threshold on SQIs to classify the PPG signal as clean and noisy. As expected, without good noise check, all three state of the arts are biased towards a single class, which proves that heuristically taken SQIs threshold values fails to work in sensor agnostic environment.

On the other hand, we use RF classifier for all methods to train on oximeter and test on smartphone data to observe robustness of SQIs across sensors, shown in Fig. 17. All prior arts tends to have biasness over majority class, resulting into very high sensitivity and low specificity. As evident in Fig. 17, the proposed method shows stable and high metrics. Though the sensitivity is marginally low compared to some methods, specificity is however significantly high. On top of that one may notice the small standard deviation in specificity (Fig. 17(b)) which makes the proposed system far more reliable compared to the state of the art methods.

5 CASE STUDY : CORONARY ARTERY DISEASE

The major reason of PPG signal quality assessment is to increase the reliability and accuracy of the vitals prediction, as explained in Section 1. We have done a case study on Coronary Artery Disease (CAD) identification non-invasively using fingertip PPG as described in [15]. Wherein time and frequency domain CAD markers are computed from PPG signal, and discriminated using SVM classifier. However, in [15] manual selection of clean PPG portion is done for CAD classification. Therefore, we replace this manual selection with our signal quality check algorithm to find the clean continuous PPG chunk for re-evaluation. We use same dataset as described in [15] (i.e. 15 CAD and 15 Non-CAD PPG data), which is collected from right hand index finger of each subject for 2 minutes using a commercial pulse oximeter (Contec CMS 50D+) at 60 Hz. Each PPG signals are segmented into 8 sec samples and the longest continuous predicted clean PPG is taken into account for CAD classification. While signal quality evaluation we find that out of 30 data 4 (2 CAD & 2 Non-CAD) are predicted noisy (none of the samples passed the quality check). This thus shows that our method do not detect CAD markers in PPG signal as noise, and is not biased towards either of classes. We discard noisy detected data and evaluate the CAD algorithm on remaining dataset (13 CAD and 13 Non-CAD). For reference, classification results with and without signal quality check are compared, as shown in Fig. 18. Increase in specificity (portion of Non-CAD predicted correctly) is observed using our quality check algorithm, while sensitivity remain the same. Stable result is not there in [15], Fig. 18, as the noisy signals were not rejected and instead got evaluated (Fig. 6).

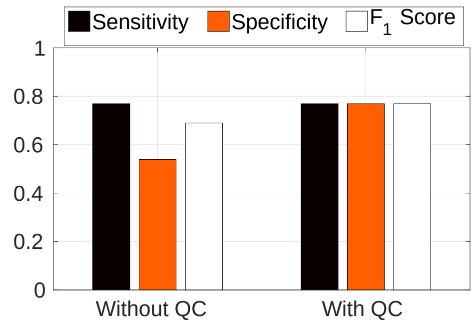


Figure 18: CAD classification comparision

6 CONCLUSION

In this study we propose a sensor agnostic algorithm to automatically classify clean and noisy PPG samples. The proposed method is tested on a large and diverse dataset comprising smartphone PPG signals with varying degrees of noise and motion artefacts. The training and testing datasets are collected using separate sensors thereby validating the robustness of the proposed approach across sensors. The results demonstrate that automated quality assessment of PPG signal using morphological SQIs leads to high performance. Additionally, the use of quality checked PPG for CAD analysis leads to a better classification performance thereby validating the proposed approach. Also, SQIs derived from the proposed ones may prove to be fruitful for the analysis of other physiological signals such as Phonocardiagram (PCG) or Electrocardiogram (ECG).

ACKNOWLEDGMENTS

The authors would like to thank fellow researchers working on PPG signal since multiple years for annotating PPG signals.

REFERENCES

[1] Mohamed Elgendi. On the analysis of fingertip photoplethysmogram signals. *Current cardiology reviews*, 8(1):14–25, 2012.

[2] JM Graybeal and MT Petterson. Adaptive filtering and alternative calculations revolutionizes pulse oximetry sensitivity and specificity during motion and low perfusion. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 2, pages 5363–5366. IEEE, 2004.

[3] Q Li and GD Clifford. Dynamic time warping and machine learning for signal quality assessment of pulsatile signals. *Physiological measurement*, 33(9):1491, 2012.

[4] Gavin D Perkins, Daniel F McAuley, Simon Giles, Helen Routledge, and Fang Gao. Do changes in pulse oximeter oxygen saturation predict equivalent changes in arterial oxygen saturation? *Critical care*, 7(4):R67, 2003.

[5] Mehmet Engin, Alparslan Demirel, Erkan Zeki Engin, and Musa Fedakar. Recent developments and trends in biomedical sensors. *Measurement*, 37(2):173–188, 2005.

[6] J Abdul Sukor, SJ Redmond, and NH Lovell. Signal quality measures for pulse oximetry through waveform morphology analysis. *Physiological measurement*, 32(3):369, 2011.

[7] Christina Orphanidou, Timothy Bonnici, Peter Charlton, David Clifton, David Vallance, and Lionel Tarassenko. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. *IEEE journal of biomedical and health informatics*, 19(3):832–838, 2015.

[8] Mohamed Elgendi. Optimal signal quality index for photoplethysmogram signals. *Bioengineering*, 3(4):21, 2016.

[9] Shreyasi Datta, Rohan Banerjee, Anirban Dutta Choudhury, Aniruddha Sinha, and Arpan Pal. Blood pressure estimation from photoplethysmogram using latent parameters. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–7. IEEE, 2016.

[10] Aymen A Alian and Kirk H Shelley. Respiratory physiology and the impact of different modes of ventilation on the photoplethysmographic waveform. *Sensors*, 12(2):2236–2254, 2012.

[11] Tin Kam Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282. IEEE, 1995.

[12] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

[13] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[14] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[15] Rohan Banerjee, Ramu Vempada, KM Mandana, Anirban Dutta Choudhury, and Arpan Pal. Identifying coronary artery disease from photoplethysmogram. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, pages 1084–1088. ACM, 2016.