2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Objective Pain Assessment Using Wrist-based PPG Signals: A Respiratory Rate Based Method

Rui Cao[1,*], Seyed Amir Hossein Aqajari[1], Emad Kasaeyan Naeini[1], and Amir M. Rahmani[2,3,4]

***Abstract*— Pain, as a multivalent, dynamic and ambiguous phenomenon is difficult to objectively quantify, in particular, in real clinical settings due to several uncontrollable factors. Respiratory rate is one of the bio-signals whose fluctuations strongly correlates with pain, however, it has been often neglected due to its monitoring difficulties. In this paper, to the best of our knowledge for the first time, we propose an objective pain assessment method using respiratory rate derived from wristband-recorded Photoplethysmography (PPG) signals collected from real post-operative patients (in contrast to the existing studies analyzing stimulated pain). We first derive respiratory rate from post-operative patients' PPG signals using an Empirical Mode Decomposition (EMD) based method and extract several statistical features from it. We then implement a feature selection method to identify the top most significant features, and exploit a weak supervision method to address the unbalanced nature of the collected labels in real settings. Several machine learning algorithms are applied to perform binary classification of no pain (NP) vs. three distinct pain levels (PL1 through PL3). We obtain prediction accuracy of up to 81.41% (NP vs. PL1), 80.36% (NP vs. PL2) and 79.48% (NP vs. PL3) which outperform the results reported by the state-of-the-art, despite obtained from data collected from real post-operative patients.**

I. INTRODUCTION

Assessment of the presence and intensity of pain is the key to adequate pain management [1]. Under- or over-treatment of pain can lead to multiple critical health problems [2], [3]. However, pain as a multivalent phenomenon is difficult to quantify [4]. Recently, automatic and continuous objective pain intensity assessment methods using physiological signs have gained interest [5]. Through monitoring Electrocardiography (ECG), Photoplethysmography (PPG), Electrodermal Activity (EDA), and Electromyography (EMG) (often using wearable devices), researchers have studied the association between pain and the autonomic nervous system activity [6], [7]. However, the existing solutions have only focused on stimulated pain collected from healthy volunteers. This motivated us to develop the UCI iHurt Database [8] as the first multimodal dataset collected from post-operative patients suffering from real pain in an acute pain unit.

Recently, a number of studies have also suggested that pain influences respiration by increasing its frequency, flow, and volume [9], [10], however, to date, there are limited efforts to objectively assess pain intensity using respiratory signals. For example, Thiam *et al.* [11] and Kessler *et al.* [12] assessed (stimulated) pain intensity using video-extracted raw respiratory signals collected from healthy volunteers. However, their method shows a rather low estimation accuracy (%50-%60) when classifying pain into three different intensity levels.

In this paper, to the best of our knowledge, for the first time, we propose an objective pain intensity assessment method for post-operative patients by only using RR. It should be noted that a pain assessment study on real patients is associated with several challenges (e.g., unbalance labels distribution, missing data, motion artifacts, etc.) since several parameters such as the intensity, distribution, frequency, and time of the pain as well as the environment can not be controlled by researchers. Our method first derives respiratory signals from wrist-based PPG signals and performs feature extraction from them. Then, we apply a feature selection method to find the most contributing features to reduce our algorithm's complexity. A machine learning based weak supervision method [13] is implemented to increase the number of pain intensity labels and balance their distribution to enhance the model training performance. We use five different machine learning algorithms and evaluate their performance compared with the state-of-the-art [11]. Our method obtains a higher pain assessment accuracy using respiratory features extracted from PPG-recording wristbands despite being tested in real clinical settings. The contribution of this work is three fold:

- We propose a pain assessment method using a dataset collected from post-operative patients (UCI_iHurtDB) while obtaining a higher accuracy compared with the existing works.
- We assess patients' pain levels using RR derived from PPG signals from a wristband showing the promises of our method to be used in everyday settings using wearable sensors as well.
- We provide a novel method to enhance the sparsely of the labeled dataset based on a weak supervision method.

The subsequent sections are organized as follows. Section 2 describes the study design and data collection. Section 3 presents our proposed method and a full pipeline of deriving RR signal from PPG signal to the weak supervision algorithm. Section 4 presents the experimental results. Section 5 describes the analysis and discussion. Section 6 concludes the paper.

II. STUDY DESIGN & DATA COLLECTION

This study is a prospective observational data collection from 25 post-operative patients likely having mild to mod-

[1]Dept. of Electrical Engineering and Computer Science, [2]Dept. of Computer Science, [3]School of Nursing, [4]Institute for Future Health (IFH), University of California, Irvine, (*correspondence e-mail: caor6@uci.edu)

erate pain. The full detail of the study and data collection protocol is explained in [8].

In this study, PPG signals were continuously recorded for approximately 30 minutes using a commercial wristband (Empatica E4). At the beginning of the study, we used a Transcutaneous Electrical Nerve Stimulation (TENS) device to induce pain and collect baseline signals from the subjects. TENS unit can stimulate multiple levels of acute pain by delivering small electrical impulses through electrodes attached to the subject's skin with adhesive pads. Participants were asked to gradually increase the TENS unit's intensity to their tolerable level and hold for at least 10 seconds. After that, we decreased the stimulus intensity back to level 0. Then, participants were instructed to perform some low-intensity activities such as walking, coughing, sitting up, or lifting legs that caused a degree of pain. This process was repeated several times to enhance the data reliability. The person's self-reported pain levels were measured using the Numeric Rating Scale (NRS), which is a segmented numeric version of the Visual Analog Scale (VAS). NRS quantifies the pain intensity to 10 levels, where 0 indicates *no pain* while numbers from 1 to 10 represent different pain levels, with 10 being the highest pain imaginable. Pain scores were recorded by making a mark on a 10-cm line representing a sequence between "no pain" and "worst pain".

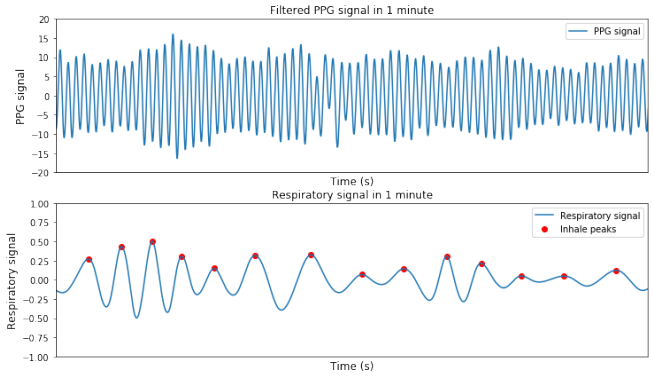

Fig. 1: Filtered PPG signal and the corresponding respiratory signal in one minute

TABLE I: Extracted features and their descriptions

Feature	Description
Peaks	The number of inhale peaks
Mean	The mean value of the signal
Max	The maximum value of the signal
Min	The minimum value of the signal
Range	The difference between the maximum and the minimum value of the signal
STD	Standard deviation of the signal
AVPI	The average value of the inhale peak intervals
SDPI	The standard deviation of the inhale peak intervals
RMS	The root mean square of successive differences between adjacent inhale peak intervals
COE	Standard deviation of inhale duration/average inhale duration

III. Method

A. Derived Respiratory Signals From PPG and Feature Extraction

We used an Empirical Mode Decomposition (EMD) based method proposed by Madhav et al. [14] to derive respiration signals from PPG. This method was proved to derive RRs from a PPG signal with high accuracy (99.87%). Figure 1 shows a filtered PPG signal and its corresponding respiratory signal. As can be seen from this Figure, the RR signal derived from a PPG signal only include inhale peaks, which is different from a regular respiratory signal. Therefore, extracting other types of respiratory features from this signal is not feasible. The respiratory features extracted in this study are briefly described in Table I.

B. Feature Selection

To ensure generalization and avoid overfitting of the pain assessment model, we implemented a feature selection method. Feature selection reduces the training time and overfitting and improves the accuracy of the classification. In this study, we used a filter-based feature selection method as it is less computationally intensive and has a lower risk of overfitting compared with other methods [13]. This method statistically determines the relationship between input features and target labels. Gini impurity gain is used in our filter-based method to select the most informative features for the classification model. A decision-tree based random forest classifier is used to output the feature importance vector. Inside the decision tree model, every node is a condition on one of the features, and these nodes supposed to separate the data into two different sets. The data with the same labels will be separated into one group in an optimal scenario. The splitting condition depends on the impurity of the features chosen in every node. During the training process, the contribution to the decrease in the impurity of each feature is calculated. Then, the importance of features is ranked according to this measurement.

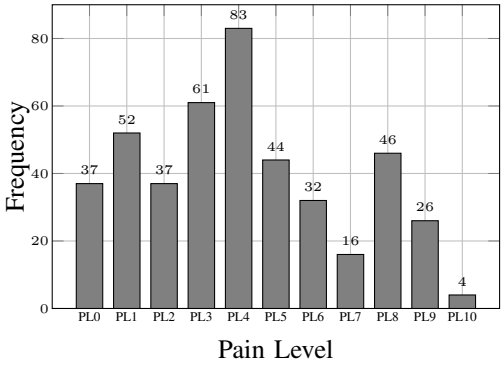

Fig. 2: The distribution of 11 classes NRS labels

C. Features Labeling Method

Figure 2 shows the distribution of 11 NRS pain labels reported by participants during the clinical trials. As can be seen from this figure, unbalanced label distribution is an

inherent challenge for further classification since these NRS labels were recorded from real post-operative patients during clinical trials. For example, there are 97 pain labels "four", but there are only 4 pain labels "ten" among all patients. Such unbalanced distribution of pain levels is unavoidable because of the subjective and realistic nature of this study and the presence of different intensities and/or pain sources among the patients.

Respiratory signal related features are usually calculated per one minute. However, considering that the time interval of pain labels reported by patients might be less than one minute, we chose 20 seconds as the feature extraction window to avoid data overlap during the labeling process. This indicates that all the pain labels are matched to their nearest 20 seconds feature window. After that, we used the Snorkel weak supervision method [13] to label other feature windows for which the NRS pain labels does not exist. Snorkel is an end-to-end method that leverages a weak supervision method to label a training dataset when limited ground truth data is available. This method is in particular beneficial for our case to address the unbalanced nature of our labels collected in a real setting. In this study, we considered all the NRS pain labels collected directly from the patients as "strong" labels to train the labeling function. Each patient's strongly labeled data was only used to mark their unlabeled windows. These remaining data points labeled by Snorkel are called "weak" data. We only use these weakly labeled data in our training process (not in the validation process) to ensure a fair evaluation of the pain assessment accuracy. In other words, our algorithm's final performance is assessed using only real data collected from post-operative patients. After the labeling process, the number of each pain label gained varying degrees of growth resulting in 58 pain levels '0' to use as the baseline.

To compare the performance of our pain assessment algorithm with the state-of-the-art [11], we downsampled the pain labels from 11 classes (0-10) to 4 classes (0 to 3). All the labels of pain level 0 were taken as the baseline while the remaining 10 classes were grouped into 3 classes. We considered pain level 1-3 as a new pain level 1 (PL1), pain level 4-5 as a new pain level 2, and pain level 6-10 as a new pain level 3 (PL3). These downsampling thresholds were chosen carefully to minimize the unbalanced label distribution problem. The new labels distribution is shown in Figure 3.

D. ML-based Predictive Models for Pain Assessment

We used a machine learning based approach to build predictive models for pain assessment. Five different classification methods were implemented, including ADABoost, XGBoost, random forest, support vector machine (SVM), and k-nearest neighbor (KNN) classifiers.

To evaluate the performance of our classification models in terms of generalizability, the Leave-one-subject-out cross-validation method was used. For each iteration of the cross-validation, we considered all the data points with only strong labels as the test set and trained our pain model using the data points including strong labels as well as weakly supervised labels using Snorkel for the rest of the patients.

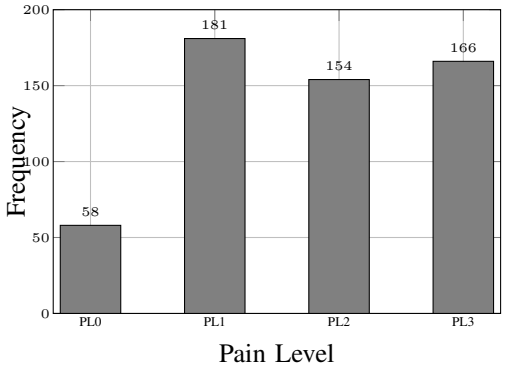

Fig. 3: The distribution of 4 classes labels after using the Snorkel and downsampling

IV. Results

8 out of 10 features, including MAX, MIN, Range, AVPI, Mean, STD and Peaks were selected by our feature selection model. The pain assessment results using 5 classifiers were shown in Figure 4 together with the results from Thiam et al. [11] for comparison. This figure's values are the average accuracy across all subjects resulted from performing three different binary classifications based on pain levels. The final scores are summarized in Table II.

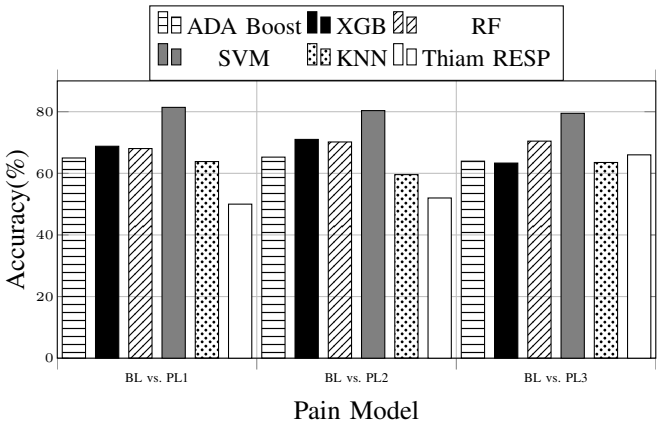

Fig. 4: Validation accuracy of all classifiers on top 8 features

TABLE II: Validation accuracy of our models in comparison with [11]

Pain levels	ADA Boost	XGB	RF	SVM	KNN	[11]
BL vs PL 1	64.98	68.79	68.04	**81.41**	63.82	50
BL vs PL 2	65.26	71.01	70.18	**80.36**	59.58	52
BL vs PL 3	63.95	63.33	70.45	**79.48**	63.52	66

As can be seen from Figure 4, all of our five classifiers achieve a higher accuracy compared with results reported by Thiam et al., for the first two pain levels (BL vs PL 1 and

BL vs PL 2). It should be noted that our models use only 8 features whereas there are 65 features used in the model proposed in [11]. As for pain level 3, only the random forest and SVM classifiers outperform their accuracy. Using the same random forest classifier on respiratory features used by Thiam et al. (65 features), the accuracy of the three classifications is improved by 18.04%, 18.18%, and 4.45%.

Among the models, the SVM classifier achieved the highest performance. The accuracy of the three pain levels are 81.41%, 80.36%, and 79.48% separately. Compared to [11], the differences for three pain levels are 31.41%, 28.36%, and 13.48%, respectively. Our SVM classifier is significantly outperforms their model using only 8 respiratory features compared to the 65 features used in their models.

V. Discussion

To the best of our knowledge, our is the first study to develop an automatic pain assessment tool only using respiratory signal derived from wrist recorded PPG data in post-operative patients. The existing pain assessment methods using video derived RR obtained rather low accuracy from stimulated pain dataset. They cannot discriminate low levels of pain (PL1 and PL2) even with the help of multiple biological features and complicated feature fusion methods. One would expect to observe lower accuracy when moving from stimulated pain in lab-settings to real post-operative pain in hospital settings. However, our models outperform the state-of-the-art despite being trained on a dataset collected in a harsher and more realistic setting (e.g., with environmental noise, motion artifacts due to movements, unbalanced labels, etc.). Our strategy using weak supervision, in particular, addressed the main issue of hospital-settings which is the unbalanced nature of labels.

According to our results, our algorithm's accuracy in all three pain levels is higher than the respiratory signal based results presented in [11]. Besides, the accuracy is considerably higher in our SVM classifier for all three pain levels compared to Thiam et al.'s results using 65 automatic extracted respiratory features using a complicated feature fusion method. These comparisons show that our algorithm can assess different intensities of pain with high accuracy from real patients using only respiratory signal derived from PPG data recorded by a wristband despite the existence a variety of noises such as motion artifacts.

The choice and proper use of the machine learning methods significantly contributes to these improved results. The feature selection method helps understand the importance of multiple respiratory signal features and reduce the model's complexity. Our initial assumption was that the number of respiration highly determines the pain assessment. However, our results shows that the number of peaks is less influential compared to other features (eighth among ten) after the feature selection process. The potential reason for this might be the 20 seconds time window used in this study. Due to our realistic monitoring settings for post-operative patient, the frequency and time of the pain reports are uncontrollable, therefore, the time interval for some pain labels are less than 1 minute. For this reason, we were unable to use the standard 1 minute time window to avoid data overlap between different pain labels. To the best of our knowledge, we used the Snorkel weak supervision method for the first time for pain assessment. In addition to help with balancing the dataset, this method also reduces the workload of data collection. We believe the holistic use of these methods enhanced the reliability of our pain assessment results.

PPG signals can be monitored easily using smart wristband or smart ring these days. Our novel method can assess patients' pain levels using RR derived from PPG signal from these devices, which shows high promises of our method to be used in everyday settings too.

VI. Conclusions

In this paper, we proposed a uni-modal pain assessment method using respiratory signals derived from wrist-recorded PPG in post-operative patients. Our method demonstrates considerably higher accuracy improvements in patients assessment compared to other related pain studies despite being conducted in a real setting. Our model can assess pain intensity with less complexity using only respiratory signal derived from wristband recorded noisy PPG data. The assessment accuracy for three pain levels were up to 81.41% (NP vs. PL1), 80.36% (NP vs. PL2) and 79.48% (NP vs. PL3), respectively, showing that the promises of our method in pain assessment application for clinical use. Future studies can focus on adding more modalities into the pain assessment algorithm to further increase the performance.

References

[1] Dolin *et al.*, "Effectiveness of acute postoperative pain management: I. evidence from published data."

[2] Barr *et al.*, "Clinical practice guidelines for the management of pain, agitation, and delirium in adult patients in the intensive care unit."

[3] M. Stites, "Observational pain scales in critically ill adults," *Critical Care Nurse*, vol. 33, no. 3, pp. 68–78, 2013.

[4] W. Raffaeli and E. Arnaudo, "Pain as a disease: an overview," *Journal of Pain Research*, vol. 10, pp. 2003–2008, 2017.

[5] Naeini *et al.*, "An edge-assisted and smart system for real-time pain monitoring."

[6] Werner *et al.*, "Towards pain monitoring: Facial expression, head pose, a new database, an automatic system and remaining challenges."

[7] Walter *et al.*, "The biovid heat pain database data for the advancement and systematic validation of an automated pain recognition system."

[8] Kasaeyan *et al.*, "Prospective study evaluating a pain assessment tool in postoperative environment:protocol for algorithm testing and enhancement."

[9] Kato *et al.*, "Habituation of the early pain-specific respiratory response in sustained pain."

[10] Gélinas *et al.*, "Pain assessment in the critically ill ventilated adult: validation of the critical-care pain observation tool and physiologic indicators."

[11] Thiam *et al.*, "Multi-modal data fusion for pain intensity assessment and classification."

[12] Kessler *et al.*, "Multimodal fusion including camera photoplethysmography for pain recognition."

[13] Ratner *et al.*, "Snorkel: Rapid training data creation with weak supervision."

[14] Madhav *et al.*, "Estimation of respiration rate from ecg, bp and ppg signals using empirical mode decomposition."