A Hierarchical Framework for Apnea Detection and Respiration Pace Assessment Using Seismocardiogram Signals

Berke Kizir and Beren Semiz[a]
Department of Electrical and Electronics Engineering, Koc University, Istanbul, Turkey

Keywords: Seismocardiogram, Respiration Rate, Apnea, Health Monitoring.

Abstract: Sleep constitutes one-third of human life and plays a critical role in physical repair, mental functioning, and memory consolidation. Although polysomnography (PSG) has been used to assess sleep performance; this test requires participants to visit a sleep clinic and have multiple sensors attached to their bodies. Hence, there is a need for alternative methods which can provide sleep monitoring outside clinical settings, but with clinical standards. In this work, a novel hierarchical framework was built to leverage the seismocardiogram (SCG) signals in apnea detection and respiration pace assessment using a simulated data collection protocol. In the first step of the framework, a binary Light Gradient-Boosting Machine (LGBM) model was trained to detect the breath-holding (apnea) episodes. If the prediction was not a breath-holding state, the data was fed into a multi-class LGBM model to distinguish between normal, slow and fast breathing episodes. Overall, the binary LGBM resulted in an accuracy, recall, precision and f1-score of 0.99, 0.95, 0.87 and 0.91, respectively; whereas for the multi-class case all metrics were 0.96. Additionally, the optimum window length to achieve real-time detection was determined as 5 seconds. The results show that the SCG signals hold substantial information regarding the changes in breathing patterns, thus could potentially be leveraged in the design of wearable systems as an alternative to the PSG test.

1 INTRODUCTION

Sleep constitutes one-third of human life and plays a critical role in physical repair, mental functioning, and memory consolidation (Kwon et al., 2021). Recent surveys have shown that the 44% of adults have experienced a decline in the quality of their sleep over the past five years, and eight out of every ten adults expressed a desire to improve their sleep quality (Philips, 2019). The most commonly observed sleep problems include insomnia, sleep apnea, and narcolepsy. Moreover, deteriorations in sleep efficiency due to these problems are associated with secondary conditions such as depression, obesity, diabetes, heart diseases, and neurocognitive disorders (Altevogt et al., 2006).

Traditionally, polysomnography (PSG) has been used to assess sleep performance; however, this test requires participants to visit a sleep clinic and have multiple sensors attached to their bodies. Although this system provides a thorough assessment of sleep quality and related problems, assessing sleep in natural settings (at home and without multiple sensor attachment) could potentially provide a more realistic evaluation. On the other hand, there have emerged watch-like systems (Apple Watch, etc.) providing sleep monitoring and staging, however these systems only use customized algorithms based on movement and heart rate, and they cannot provide detailed information regarding other vital parameters, such as respiration rate. Hence, there is a need for alternative methods which can provide sleep monitoring outside clinical settings, but with clinical standards.

Recent studies have shown that the seismocardiogram (SCG) signal collected from the thoracic region can provide information about various hemodynamic parameters (Inan et al., 2014). SCG signal corresponds to the chest micro-vibrations occurring due to the ejection of blood and contraction of the heart in each cardiac cycle. The part of the SCG below 1 Hz corresponds to the chest movements associated with respiration, the frequencies between 1 - 20 Hz includes the cardiac vibrations, and the component above 20 Hz represents the heart sounds (Pandia et al., 2012). As SCG signals can assess the thoracic region from different perspectives, they are widely

[a] https://orcid.org/0000-0002-7544-5974

Kizir, B. and Semiz, B.
A Hierarchical Framework for Apnea Detection and Respiration Pace Assessment Using Seismocardiogram Signals.
DOI: 10.5220/0012446400003657

In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2024) - Volume 1*, pages 793-798
ISBN: 978-989-758-688-0; ISSN: 2184-4305

used in wearable systems aiming to monitor cardiovascular and cardiopulmonary diseases (Hayirlioglu and Semiz, 2023). Indeed, previous studies have shown that the SCG signals could potentially be used in the detection of heart failure (Inan et al., 2018), aortic stenosis (Yang et al., 2019), and atrial fibrillation (Hurnanen et al., 2016), as well as predicting stroke volume values (Semiz et al., 2020), estimating systolic time intervals (Shandhi et al., 2019), classifying valvular heart disease locations (Erin and Semiz, 2023) and assessing respiration phases (Imirzalioglu and Semiz, 2022; Pandia et al., 2012).

In this study, we developed a novel hierarchical framework to leverage the SCG signals in apnea detection and respiration pace assessment using a simulated data collection protocol including breath-holding, slow-breathing, normal-breathing and fast-breathing episodes. Instead of focusing on the component below 1 Hz, we leveraged the vibration and acoustic components of the signal (1-40 Hz) as all apnea types are not necessarily associated with halted chest movements. In the first step of the framework, as the most crucial aspect of the study was to determine the breath-holding states, a binary classification model was trained to detect the breath-holding (apnea) episodes. If the prediction was not breath-holding state, the data was fed into a second model, which was designed as a multi-class classification model to distinguish between normal, slow and fast breathing episodes. In addition, the optimum window length for achieving real-time apnea detection and breathing rate assessment was studied. Overall, the results show that the SCG signals hold substantial information regarding the changes in breathing patterns, thus could potentially be leveraged in the design of wearable systems as an alternative to the PSG test.

2 METHODS

2.1 Data Collection Protocol

This study was conducted under a protocol approved by the Koc University Institutional Review Board and all subjects provided written consent. A total of 8 healthy subjects (6 females and 2 males) participated in the study (Age: 21.9 $\pm$ 3.0, height: 171.6 $\pm$ 7.7, and weight: 65.4 $\pm$ 12.2). The signals were simultaneously collected using the BIOPAC system (BIOPAC Systems, Inc. Goleta, CA, USA) at 2 kHz. The electrocardiogram (ECG) and respiration signals were acquired using three gel electrodes and a respiration transducer, respectively. The signals were transferred to the BIOPAC system using wireless Bionomadix RSPEC-R module. To record the SCG signal, a tri-axial low noise analog accelerometer (ADXL354, Analog Devices, Inc., Norwood, MA) was used. It was placed on the mid-sternum of the subject using hypoallergenic transparent medical tape. The sensor locations are detailed in Fig. 1. The X, Y, Z axes of the accelerometer were corresponding to the vibrations in lateral, head-to-foot and dorso-ventral directions, respectively.

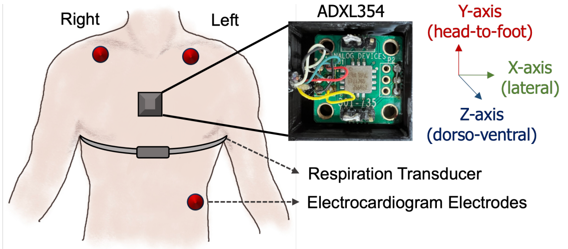

Figure 1: The locations of the sensors.

Data collection protocol is detailed in Fig. 2. The subject was first asked to breathe normally for 1 minute, followed by Valsalva maneuver where the subject held breath for 15 seconds. It was followed by another 1 minute-long normal breathing. The subject was then asked to perform slow breathing at 40 beats-per-minute (bpm) controlled by an online metronome. After 1 minute-long normal breathing, the subject performed fast breathing at 120 bpm, again controlled by a metronome. The last steps were including an additional 1 minute-long normal breathing, 15-seconds-long Valsalva maneuver and another normal breathing phase for 1 minute. The signals were recorded continuously during the protocol and the timestamps for each transition were recorded thoroughly.

2.2 Preprocessing

SCG signals were filtered with a Kaiser window finite impulse response filter to remove out-of-band noise. The cut-off frequencies were selected as 1-40 Hz for all three axes. No other preprocessing steps were applied not to lose any information (spikes, oscillations, etc.), which might be useful in determining different breathing states.

In this work, one of the fundamental aims was to assess the effect of window length in detection performance. To that end, the analysis pipeline was repeated using different window lengths (1, 2, 3, 4 and 5 seconds). Between the consecutive windows, 500 milliseconds-long (0.5 seconds) overlap was employed. By this way, the number of instances to be used during training was increased, additionally the subject's state could be updated every 0.5 seconds.

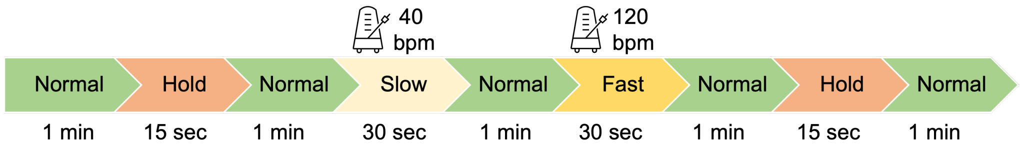

Figure 2: Experimental protocol.

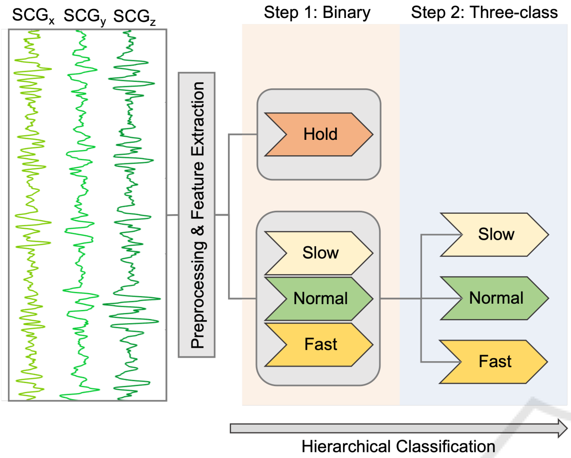

Figure 3: Study pipeline.

It should be noted that the ECG was not used as the reference since the real-time apnea detection scenario will necessitate the use of continuous data streaming, which can be achieved through a sliding window.

2.3 Feature Extraction

Statistical, temporal and spectral features were extracted from each SCG window in all three axes (Table 1). As the statistical features, mean, variance, skewness (asymmetry) and kurtosis (tailedness) were calculated. As the temporal feature, signal energy, i.e. the squared sum of the samples, were computed. On the other hand, spectral domain features were including centroid, spread, rolloff and bandpowers. Rolloff indicates the frequency at which a specific percentage of the signal energy is accumulated. On the other hand, centroid and spread relate to the center of mass and the distribution of frequencies in the spectrum, respectively (Giannakopoulos and Pikrakis, 2014). Finally, bandpower frequency intervals were selected as logarithmically spaced frequency bands between 1 and 40 Hz. After extracting all features, a dataframe was generated where columns including the feature values and rows indicating the corresponding SCG frames. The whole feature extraction and data frame generation steps were repeated for different window lengths.

Table 1: Feature groups.

Statistical Features	Temporal Features	Spectral Features
Mean	Signal Energy	Centroid
Variance		Spread
Skewness		Rolloff
Kurtosis		Bandpowers (logarithmically spaced between 1 - 40 Hz)

2.4 Model Training and Feature Importance Analysis

2.4.1 Hierarchical Model Framework

In this study a hierarchical model framework was built (Fig. 3). Since the most crucial aspect of the study was to determine the breath-holding states, a binary classification model was trained first. In this model, breath-holding states were labeled as 1, while the others (normal, slow, fast) were all labeled as 0. Test data was first fed into the binary classifier. If the prediction was not breath-holding state, the data was fed into a second model, which was designed as a multi-class classification model. In this secondary step, the aim was to distinguish between the normal, slow and fast breathing episodes. This hierarchical pipeline was built so that the breath-holding states (i.e. apnea periods) could be determined as fast as possible regardless of the changes in breathing pace in-between.

2.4.2 Model Selection and Validation

As the superior performance of the tree-based methods are well known in SCG-related applications, four different tree-based models were trained and compared: decision tree (DT), random forest (RF), extreme gradient boosting trees (XGB) and light gradient-boosting machine (LGBM). For all classification models, 5-fold cross validation was applied.

- *Decision Tree (DT):* Decision tree is a tree-like flowchart which utilizes a divide and conquer strategy, employing a greedy search to find the best split points. This splitting process is performed in a top-down fashion until the data-of-interest has been assigned class labels (Song and Ying, 2015).

- *Random Forest (RF):* Instead of depending on a single tree, RF involves bootstrapping multiple trees by utilizing randomized subsets drawn from the dataset. These trees are trained independently and in parallel. Majority voting is then applied on the outputs obtained from these trees to yield one single class estimation (Breiman, 2001).
- *Extreme Gradient Boosting Trees (XGB):* XGB is one of the popular boosting algorithms. As RF relies of bagging, XGB operates sequentially, i.e., each subsequent tree relies on the the outcome of the previous one. Overall, in the training process, multiple decision trees are trained iteratively, allowing the prediction and adjustment of residual errors from the previous iteration as the training advances (Chen and Guestrin, 2016).
- *Light Gradient-Boosting Machine (LGBM):* LGBM is another popular type of boosting algorithms, however unlike the horizontal, level-wise growth seen in XGB, LGBM follows a vertical, leaf-wise growth pipeline. This approach leads to increased loss reduction, resulting in higher accuracy and faster processing (Ke et al., 2017).

To assess the performance of the models, accuracy and weighted precision, recall and f1-score were used. These equations are presented in Equations 1, 2, 3 and 4, respectively (*TP: true positives, FP: false positives, TN: true negatives and FN: false negatives*). In addition, the area under the receiver operating characteristics curve (ROC AUC) was computed for the binary classification task.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$f_1 score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

2.4.3 Feature Importance Ranking

Feature importance scores were computed from the best performing model (LGBM) to find out the most important features. The procedure was repeated for both the binary and multi-class tasks. After all folds were completed, average of the normalized LGBM scores was calculated and determined as the final importance score. The corresponding scores were then ranked in descending order to determine the feature importance ranking.

Table 2: Performance comparison (binary).

Model	Accuracy	AUC	Recall	Precision	F1
LGBM	0,99	0,99	0,95	0,87	0,91
RF	0,94	0,95	0,95	0,30	0,46
XGB	0,99	0,99	0,95	0,85	0,90
DT	0,92	0,73	0,56	0,49	0,52

Table 3: Performance comparison (multi-class).

Model	Accuracy	Recall	Precision	F1
LGBM	0,96	0,96	0,96	0,96
RF	0,88	0,88	0,94	0,90
XGB	0,95	0,95	0,96	0,96
DT	0,85	0,85	0,85	0,85

3 RESULTS AND DISCUSSION

3.1 Apnea Detection Results

The first step of the hierarchical classification framework was to distinguish between the breath-holding episodes and breathing (normal, fast, slow) periods. To that end, a binary classification model was trained. The results obtained from different models with a 5-sec window were presented in Table 2. Overall, LGBM and XGB outperformed DT and RF in all metrics. Indeed, the precision values obtained from the RF and DT were significantly lower than the other models, which led to again low f1-scores. On the other hand, XGB and LGBM indeed had comparable performance, however LGBM performed slightly better in terms of precision and f1-score. Having an accuracy value of 0.99 revealed that there is almost no carried-error from the binary model to the multi-class task.

3.2 Breathing Rate Assessment

Using the primary model, the SCG windows were labeled as breath-holding or not. If the prediction was not breath-holding state, the data was fed into a second model, which was designed as a multi-class classification model. In this secondary step, the aim was to distinguish between the normal, slow and fast breathing episodes. The performances of different models are presented in Table 3. Similar to the binary case, LGBM and XGB outperformed RF and DT in terms of all metrics, however the precision and f1-score obtained from RF and DT in multi-class task were significantly higher than the ones obtained in the binary task. Overall, the LGBM model performed slightly better than the XGB, similar to the binary case.

Table 4: LGBM results for different window lengths (binary).

Window Length (sec.)	Accuracy	AUC	Recall	Precision	F1-Score
1	0,95	0,91	0,72	0,51	0,59
2	0,97	0,96	0,86	0,71	0,78
3	0,98	0,97	0,91	0,77	0,83
4	0,98	0,99	0,93	0,83	0,87
5	0,99	0,99	0,95	0,87	0,91

Table 5: LGBM results for different window lengths (multi-class).

Window Length (sec.)	Accuracy	Recall	Precision	F1-Score
1	0,85	0,85	0,88	0,86
2	0,90	0,90	0,92	0,91
3	0,93	0,93	0,94	0,93
4	0,95	0,95	0,96	0,95
5	0,96	0,96	0,96	0,96

3.3 Effect of Window Length

As previously discussed, the ECG was not used as the reference since the real-time apnea detection scenario necessitates the use of continuous data streaming, which can be achieved through a sliding window. However, determining the size of the optimum analysis window is an important research question. Hence, different window lengths (1, 2, 3, 4, 5 seconds) were tested to find the optimum length for the current application. Throughout the experiments, the model was set to LGBM.

For the binary and multi-class tasks, the performance values for varying window lengths are presented in Tables 4 and 5. When window length was selected as 1 second, the results got worse significantly for both tasks. For the binary case, the recall value was calculated as 0.72 when 1-sec windows were used, i.e., there was a %24 decrease compared to the case where 5-sec windows were used. Window lengths longer than 5-sec resulted in similar results, and when the length exceeded 10-sec, the performance started to decrease. A similar pattern was also valid for the multi-class task. Based on these observations, the optimum window length was determined as 5-sec.

3.4 Feature Importance Ranking

Feature importance scores were computed from the LGBM classifier. After all folds were completed, average of normalized scores were calculated as the final score. The feature importance results for both task are presented in Figure 4. Overall, bandpowers, kurtosis and skewness were dominating in both binary and multi-class classification tasks. Considering that the kurtosis represents the tailedness and skewness represents asymmetry in the data, they are important indicators of how the data is distributed. Having kurtosis and skewness of multiple axes as the most important features thus reveals that the distributions of different breathing episodes were indeed different than each other.

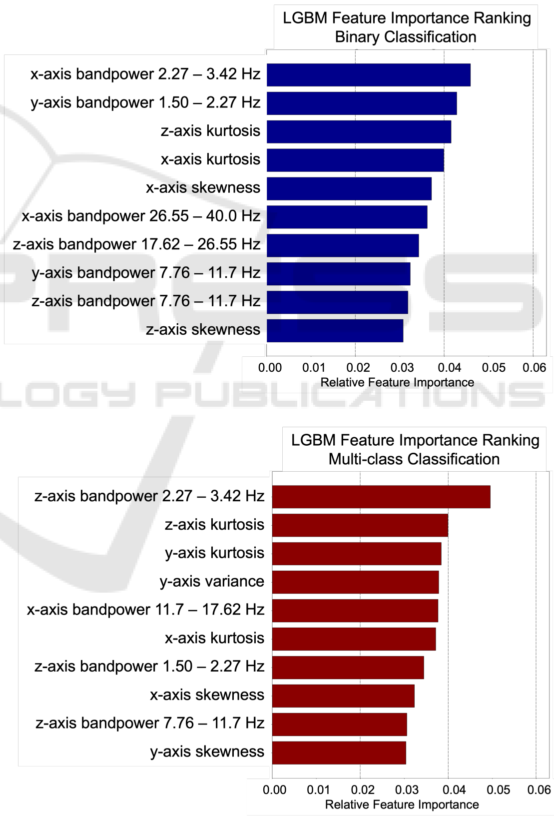

Figure 4: Feature importance ranking for the binary and multi-class tasks.

4 CONCLUSION

In this work, a novel hierarchical framework was built using a simulated data collection protocol for evaluating the potential use of SCG signals in apnea detection and respiration pace assessment. In the first step of the framework, a binary Light Gradient-Boosting Machine (LGBM) model was trained to detect the breath-holding (apnea) episodes. If the prediction was not a breath-holding state, the data was fed into a multi-class LGBM model to distinguish between normal, slow and fast breathing episodes.

Overall, the binary LGBM model resulted in an accuracy, recall, precision and f1-score of 0.99, 0.95, 0.87 and 0.91, respectively; whereas for the multi-class case all metrics were 0.96. Additionally, different window lengths (1, 2, 3, 4, 5 seconds) were tested and the optimum window length was determined as 5 seconds.

The results show that the SCG signals hold substantial information regarding the changes in breathing patterns, thus could potentially be leveraged in the design of wearable systems as an alternative to the PSG test. Future work will focus on validating these results in larger datasets including real data from patients having sleep apnea.

REFERENCES

Altevogt, B. M., Colten, H. R., et al. (2006). Sleep disorders and sleep deprivation: an unmet public health problem.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Erin, E. and Semiz, B. (2023). Spectral analysis of cardiogenic vibrations to distinguish between valvular heart diseases.

Giannakopoulos, T. and Pikrakis, A. (2014). *Introduction to audio analysis: a MATLAB® approach*. Academic Press.

Hayirlioglu, Y. Z. and Semiz, B. (2023). A novel multimodal sensing system prototype for cardiovascular and cardiopulmonary monitoring.

Hurnanen, T., Lehtonen, E., Tadi, M. J., Kuusela, T., Kiviniemi, T., Saraste, A., Vasankari, T., Airaksinen, J., Koivisto, T., and Pänkäälä, M. (2016). Automated detection of atrial fibrillation based on time–frequency analysis of seismocardiograms. *IEEE journal of biomedical and health informatics*, 21(5):1233–1241.

Imirzalioglu, M. and Semiz, B. (2022). Quantifying respiration effects on cardiac vibrations using teager energy operator and gradient boosted trees. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1935–1938. IEEE.

Inan, O. T., Baran Pouyan, M., Javaid, A. Q., Dowling, S., Etemadi, M., Dorier, A., Heller, J. A., Bicen, A. O., Roy, S., De Marco, T., et al. (2018). Novel wearable seismocardiography and machine learning algorithms can assess clinical status of heart failure patients. *Circulation: Heart Failure*, 11(1):e004313.

Inan, O. T., Migeotte, P.-F., Park, K.-S., Etemadi, M., Tavakolian, K., Casanella, R., Zanetti, J., Tank, J., Funtova, I., Prisk, G. K., et al. (2014). Ballistocardiography and seismocardiography: A review of recent advances. *IEEE journal of biomedical and health informatics*, 19(4):1414–1427.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

Kwon, S., Kim, H., and Yeo, W.-H. (2021). Recent advances in wearable sensors and portable electronics for sleep monitoring. *Iscience*, 24(5).

Pandia, K., Inan, O. T., Kovacs, G. T., and Giovangrandi, L. (2012). Extracting respiratory information from seismocardiogram signals acquired on the chest using a miniature accelerometer. *Physiological measurement*, 33(10):1643.

Philips (2019). The global pursuit of better sleep health.

Semiz, B., Carek, A. M., Johnson, J. C., Ahmad, S., Heller, J. A., Vicente, F. G., Caron, S., Hogue, C. W., Etemadi, M., and Inan, O. T. (2020). Non-invasive wearable patch utilizing seismocardiography for peri-operative use in surgical patients. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1572–1582.

Shandhi, M. M. H., Semiz, B., Hersek, S., Goller, N., Ayazi, F., and Inan, O. T. (2019). Performance analysis of gyroscope and accelerometer sensors for seismocardiography-based wearable pre-ejection period estimation. *IEEE journal of biomedical and health informatics*, 23(6):2365–2374.

Song, Y.-Y. and Ying, L. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2):130.

Yang, C., Aranoff, N. D., Green, P., and Tavassolian, N. (2019). Classification of aortic stenosis using time–frequency features from chest cardio-mechanical signals. *IEEE Transactions on Biomedical Engineering*, 67(6):1672–1683.