engineering proceedings

MDPI

Proceeding Paper

Developing and Validating Ensemble Classifiers for At-Home Sleep Apnea Screening †

Zilu Liang [1,2]

[1] Ubiquitous and Personal Computing Lab, Faculty of Engineering, Kyoto University of Advanced Science (KUAS), Kyoto 615-8577, Japan; liang.zilu@kuas.ac.jp
[2] Institute of Industrial Science, The University of Tokyo, Kyoto 153-8505, Japan
† Presented at the 10th International Electronic Conference on Sensors and Applications (ECSA-10), 15–30 November 2023; Available online: https://ecsa-10.sciforum.net/.

Abstract: In this paper, we developed ensemble classifiers with SpO2 signals for sleep apnea screening. The ensemble classifiers (eclf) were built on top of five base classifiers, including logistic regression (LR), random forest (RF), support vector machine (SVM), linear discrimination analysis (LDA), and light gradient boosting machine (LGMB). Performance evaluation showed that when heavier weights were assigned to the LR and SVM classifiers, the ECLF achieved a better balance between sensitivity (0.81 ± 0.02) and specificity (0.80 ± 0.02) while maintaining the overall performance as measured by AUC (0.81 ± 0.01).

Keywords: sleep apnea; ubiquitous computing; SpO2; ensemble learning

check for updates

Citation: Liang, Z. Developing and Validating Ensemble Classifiers for At-Home Sleep Apnea Screening. *Eng. Proc.* **2023**, *58*, 49. https://doi.org/10.3390/ecsa-10-16184

Academic Editor: Stefano Mariani

Published: 15 November 2023

Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Due to lifestyle changes, more and more people are suffering sleep disorders nowadays. Among all the different types of sleep disorders, sleep apnea is a most common one around the globe. Sleep apnea is characterized by brief stops of breathing during sleep. If that happens, on average, more than five times per hour throughout the night, then the person is clinically considered sleep apnea-positive. Sleep apnea is often accompanied by snoring during sleep and excessive daytime sleepiness. The long-term consequence of untreated sleep apnea includes increased risks of cardiovascular diseases, metabolic disorders, mental problems, and cognitive impairment [1]. Developing sleep apnea screening methods that are easy to use at home has been the objective of many studies. The main challenge of home sleep apnea screening is that most of the signal modalities in medical settings, such as EEG, ECG, and breath effort, are not readily measurable with home-use sensors. To address this challenge, several studies have attempted to develop apnea screening methods based solely on the blood oxygen levels, or SpO2, during sleep. This signal modality is ideal for home sleep apnea screening because it can be collected using consumer sleep tracking gadgets with reasonable accuracy [2]. This study aims to develop ensemble classifiers that takes the SpO2 signals as input to predict whether a user is sleep apnea-positive or -negative.

2. Related Work

Polysomnography (PSG) is the gold standard for sleep apnea diagnosis. A PSG test measures many physiological signals which allow doctors to calculate the apnea-hypopnea index (AHI), defined as the average number of obstructive respiratory events occurred per hour during sleep, for diagnosing sleep apnea. An AHI of larger than 5 events/h is considered out of the normal range and thus sleep apnea-positive. Other AHI cutoff thresholds, such as 15 and 30, are also widely used to determine the severity of sleep apnea [3]. The PSG test is expensive, invasive, and not always available. Many studies have attempted to develop alternative methods for sleep apnea screening using information stored in medical records. Instead of using physiological signals, those methods rely on

information such as demographic characteristics, blood tests, and questionnaire scores to assess the likelihood of sleep apnea [4]. In a related vein, machine learning and deep learning models that use SpO2 signals for sleep apnea screening were also developed. While early works have used only small datasets for model development [5], a very recent publication trained and tested their models on several large datasets [6]. In this study, we developed and validated our models using one of the largest sleep datasets.

3. Methodology

In this section, we present the pipeline for developing the ensemble classifier for sleep apneas screening and the dataset used. Signal processing and machine learning were performed using Python 3.10.5. Several Python modules were used, including Numpy, Pandas, Matplotlib, Scikit-learn, and Scipy.

3.1. Dataset

The Sleep Heart Health Study (SHHS) dataset used in this study consists of multi-modal physiological signals collected from 5786 subjects during sleep [7]. Permission was granted from the National Sleep Research Resource (NSRR) to export the SpO2 signals as well as some demographic information (i.e., age, gender, BMI) from the dataset. Data usage in this study was compliant with the Data Assess and Use Agreement (DAUA). The SpO2 signals were preprocessed to remove readings of zeros, below 50% or above 100%. Sudden changes between consecutive readings (>4%) were also removed. The '*nsrr_ahi_ph3r_aasm15*' variable was used as the ground truth values for AHI.

3.2. Ensemble Classifier

Ensemble is a method in machine learning that combines multiple base classifiers to improve the performance of the classifications. In this study, we used five base classifiers, including linear regression (LR), random forest (RF), support vector machine (SVM), linear discrimination analysis (LDA), and light gradient boosting machine (LGMB). The final output of an ensemble classifier is the most frequently occurring label (i.e., the mode) of the weighted outputs of the based classifiers. Let f_{LR}, f_{RF}, f_{SVM}, f_{LDA}, f_{LGMB} be the 5 base classifiers, and the final output of the ensemble classifiers f_{eclf} is calculated using Equation (1). We built two ensemble classifiers: f_{eclf1} uses a weight vector of [1, 1, 1, 1, 1], while f_{eclf2} uses a weight vector of [2, 1, 2, 1, 1]. We assigned heavier weights to the LR and SVM classifiers in the latter case because those two classifiers yielded better performance in our previous study [8].

$$f_{eclf} = \text{mode}\left(w_{LR} \cdot f_{LR},\ w_{RF} \cdot f_{RF},\ w_{SVM} \cdot f_{SVM},\ w_{LDA} \cdot f_{LDA},\ w_{LGMB} \cdot f_{LGMB}\right), \quad (1)$$

Following the common practice in machine learning, 80% of the dataset was used for model training and tuning, and the rest was used for model testing. Hyper-parameter tuning was achieved through 5-fold cross validation. Classifier performances were measured using AUC, sensitivity, specificity. We repeated the dataset split 50 times with different random seeds. Boxplots were generated to show the distribution of the performance measures over the 50 repetitions.

4. Result

The AUC, sensitivity, and specificity of the base classifiers and the ensemble classifiers are shown in Figures 1–3. It is shown that among the five base classifiers, LR achieved the highest AUC and specificity. RF had the highest sensitivity but the lowest specificity, which translated to the lowest AUC among all base classifiers. Compared to ECLF, ECLF2 had similar AUC to LR but a better trade-off between sensitivity and specificity.

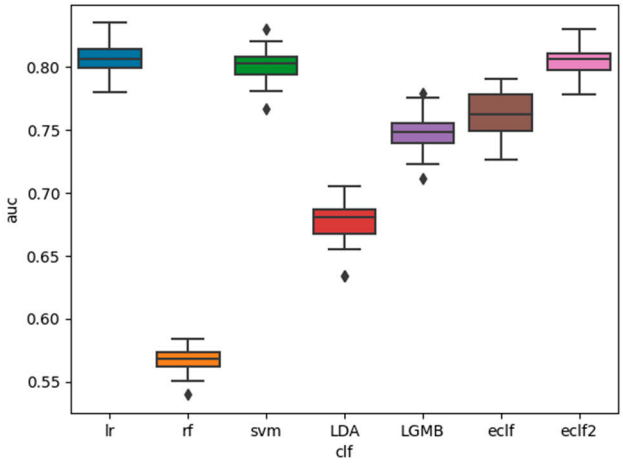

Figure 1. Performance of the classifiers for sleep apnea screening (AUC).

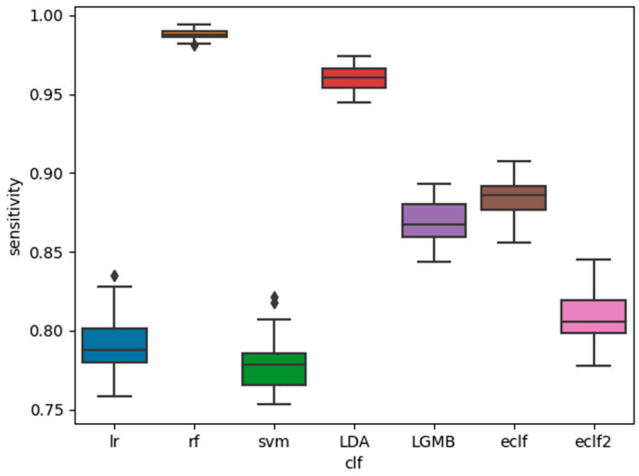

Figure 2. Performance of the classifiers for sleep apnea screening (sensitivity).

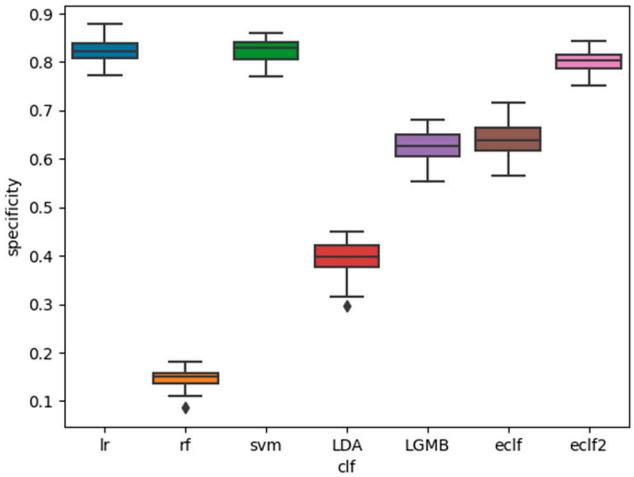

Figure 3. Performance of the classifiers for sleep apnea screening (specificity).

5. Discussion

We have presented the development and validation of ensemble classifiers with hard voting for sleep apnea screening. Performance evaluation analysis showed that when heavier weights were assigned to the LR and SVM classifiers, the ECLF achieved a better balance between sensitivity (0.81 ± 0.02) and specificity (0.80 ± 0.02) while maintaining

the overall performance as measured by AUC (0.81 ± 0.01). RF and LDA achieved high sensitivity (>0.95) at the sacrifice of specificity, while LR and SVM achieved high specificity (>0.80) at the sacrifice of sensitivity. LGMB demonstrated mediocre performance on both sensitivity and specificity. In our future work, we plan to apply other ensemble techniques, such as soft voting, and to tune the weights given to the base classifiers.

Funding: This research was funded by the JSPS KAKENHI, grant number 21K17670.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Review Board of Kyoto University of Advanced Science.

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

Data Availability Statement: Restrictions apply to the availability of the data used in this study. Data were obtained from the National Sleep Research Resource (NSRR) and are available at https://sleepdata.org/datasets/shhs (accessed on 10 October 2023) with the permission of NSRR.

Acknowledgments: The author would like to thank the NSRR for sharing the SHHS dataset.

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

References

1. Franklin, K.A.; Lindberg, E. Obstructive sleep apnea is a common disorder in the population-a review on the epidemiology of sleep apnea. *J. Thorac. Dis.* **2015**, *7*, 1311–1322.
2. Liang, Z.; Chapa-Martell, M.A. Accuracy of Fitbit wristbands in measuring sleep stage transitions and the effect of user-specific factors. *JMIR Mhealth Uhealth* **2019**, *7*, e13384. [CrossRef] [PubMed]
3. Berry, R.B.; Brooks, R.; Gamaldo, C.E.; Harding, S.M.; Marcus, C.; Vaughn, B.V. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications*; Version 2.4 ed.; American Academy of Sleep Medicine: Darien, IL, USA, 2017.
4. Huang, W.C.; Lee, P.L.; Liu, Y.T.; Chiang, A.A.; Lai, F. Support vector machine prediction of obstructive sleep apnea in a large-scale Chinese clinical sample. *Sleep* **2020**, *43*, zsz295. [CrossRef] [PubMed]
5. Mostafa, S.S.; Mendonça, F.; Morgado-Dias, F.; Ravelo-García, A. SpO2 based sleep apnea detection using deep learning. In Proceedings of the 2017 IEEE 21st International Conference on Intelligent Engineering Systems (INES), Larnaca, Cyprus, 20–23 October 2017.
6. Levy, J.; Álvarez, D.; Del Campo, F.; Behar, J.A. Deep learning for obstructive sleep apnea diagnosis based on single channel oximetry. *Nat. Commun.* **2023**, *14*, 4881. [CrossRef] [PubMed]
7. Quan, S.F.; Howard, B.V.; Iber, C.; Kiley, J.P.; Nieto, F.J.; O'Connor, G.T.; Rapoport, D.M.; Redline, S.; Robbins, J.; Samet, J.M.; et al. The Sleep Heart Health Study: Design, rationale, and methods. *Sleep* **1997**, *20*, 1077–1085. [PubMed]
8. Liang, Z. Novel Method Combining Multiscale Attention Entropy of Overnight Blood Oxygen Level and Machine Learning for Easy Sleep Apnea Screening. *Digital Health* **2023**, *9*, 1–19. [CrossRef] [PubMed]

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