Development of a Machine Learning Model for Screening Sleep Apnea in Heart Failure Patients Using Sleep Sensor Data

Gunasegarama, Mathushan; Dinesen, Birthe; Müller Larsen, Nikolaj; Ghamari Gilavai, Ghazal; Røge, Kristine; Kirk Østergaard, Mathias; Rovsing Jochumsen, Mads

Published in:
Good Evaluation - Better Digital Health

DOI (link to publication from Publisher):
10.3233/SHTI251496

Publication date:
2025

Document Version
Publisher's PDF, also known as Version of record

Good Evaluation - Better Digital Health
U.H. Hübner et al. (Eds.)

doi:10.3233/SHTI251496

Development of a Machine Learning Model for Screening Sleep Apnea in Heart Failure Patients Using Sleep Sensor Data

Mathushan GUNASEGARAMA[a,1], Birthe DINESEN[a], Nikolaj MÜLLER LARSEN[a], Ghazal GHAMARI GILAVAI[a], Kristine RØGE[a], Mathias KIRK ØSTERGAARD[a] and Mads ROVSING JOCHUMSEN[b]

[a] *Laboratory Welfare Technologies – Digital Health & Rehabilitation, ExerciseTech, Department of Health Science and Technology, Aalborg University (AAU), Denmark*
[b] *Neural Engineering and Neurophysiology, Department of Health Science and Technology, AAU, Denmark*
ORCiD ID: Mathushan Gunasegaram https://orcid.org/0009-0002-8041-9508

Abstract. Sleep apnea (SA) is a prevalent disorder among individuals with heart failure (HF), often leading to complications. Early identification is essential for timely interventions and better outcomes. This study explores the feasibility of developing a screening tool for SA in patients with HF using data from the Future Patient Telerehabilitation program. A random forest classifier was used to develop a predictive model, achieving a promising receiver operating characteristic area under the curve (ROC-AUC) of 0.85, suggesting that the random forest classifier has the potential as a SA screening tool for HF patients. However, the study lacked key variables, such as oxygen saturation, that are strong predictors for SA assessment according to current literature; this limits the model's generalizability. Despite this, the findings indicate that the ML model shows promise for screening SA in HF patients, highlighting the need for high-quality, standardized data from future clinical trials to enhance its accuracy and clinical utility.

Keywords. Sleep Apnea, Heart Failure, Machine Learning, Telemonitoring, Screening tool, Sleep Sensor

1. Introduction

Heart failure (HF) is a clinical syndrome that results from any cardiac dysfunction leading to reduced cardiac output and symptoms such as dyspnea, swelling, and fatigue. It affects 1–2% of the global population, with around 2 million new cases annually. [1, 2]. Studies have established an association between sleep apnea (SA) and HF. SA is characterized by repeated interruptions in breathing during sleep and has been shown to worsen HF symptoms, increase hospitalization, and reduce quality of life [3].

Sleep apnea (SA) is traditionally diagnosed using polysomnography (PSG), the gold standard for sleep studies [4]. Despite its high accuracy, PSG often suffers from low adherence (7–81.7%) due to limited access, discomfort, and insufficient support [5]. As an alternative, telemonitoring technologies such as wearable devices and smart beds

[1] Corresponding Author: Mathushan Gunasegaram, mgu@hst.aau.dk

enable remote patient monitoring [6, 7]. One such technology is the EMFIT QS sleep sensor (EMFIT), a non-contact device placed under the mattress that uses ballistocardiography to capture heart rate (HR), respiratory rate (RR), movement, and bed presence. It operates passively, transmits data wirelessly, and is suitable for long-term use at home. At a cost of 300–400 EUR, it offers a more accessible alternative to PSG. [11]

EMFIT data may facilitate the development of a machine learning (ML) model for screening SA, enabling early identification of high-risk patients. This could address PSG limitations, enhance accessibility, and support earlier and more effective SA management in patients with HF and atrial fibrillation (AF). In the Future Patient Telerehabilitation Program, the EMFIT sensor has been tested by HF and AF patients. This study aims to investigate how data from the EMFIT can be utilized to develop a pre-screening algorithm of SA for patients with HF and AF.

2. Methods

2.1. Participants

This study includes data from two clinical trials in the Future Patient Telerehabilitation Program, involving adults with HF (n=28) or AF (n=20). Participants were excluded if they had a pacemaker or dropped out during the trial period, resulting in a final sample of 48 participants used to develop the ML model. The participants differ in age and disease severity, as presented in Table 1. The study was conducted in accordance with the Helsinki Declaration, and all participants provided informed consent. The study was approved by the North Denmark Region Committee on Health Research Ethics N-20200037 and N-20220056. [9, 10, 11]

Table 1. Baseline characteristics for included patients [9, 11]

Variable	Value
Age (Mean ± STD)	
Men (n = 35)	62.6 ± 11.5 years
Women (n = 12)	60.7 ± 11.9 years
Men and women (n = 47)	62.1 ± 11.6 years
New York Heart Association Class	
Class I	3
Class II	28
Class III	13
Class IV	0
Ejection Fraction (Mean ± STD)	34.79 ± 11.47

2.2. Experimental setup

Participants included in the study used home telemonitoring equipment, including the EMFIT, blood pressure monitor, scale, and activity tracker. Data was submitted to the HeartPortal, where the patient could get an overview of all monitored values. Each patient collected data for 4 months. [9, 11]

2.3. Machine learning

This project applies ML techniques to develop a model for detecting SA in HF patients. The process is illustrated in Figure 1.

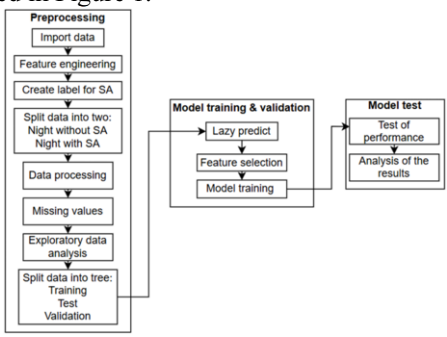

Figure 1. A graphical representation of the ML process

2.3.1. Data preprocessing

The raw dataset included EMFIT files (4-second resolution), Fitbit step counts, blood pressure, weight, and patient demographics. These data were pre-processed and synchronized for algorithm development. EMFIT provided detailed sleep-related data, while additional variables, such as age, weight, New York Heart Association (NYHA) class, and baseline HR were added to capture both physiological signals and individual patient characteristics. Each EMFIT file represented a sleep cycle, and all measurements were aligned to create 4,399 measurement data points, each representing one night of sleep. Sleep apnea events were labeled using the following formula, a calculation model detecting apneas (complete airflow cessation) and hypopneas (partial airflow reduction):

$$T_{RR} = RR_{stable} - 10 * \frac{RR_{stable}}{60}$$

Where T_{RR} is the detected respiratory event and RR_{stable} is the baseline RR. Following Apnea-Hypopnea Index (AHI) is applied to categorize SA severity:

$$AHI = \frac{Hours\ of\ Sleep}{Number\ of\ Apneas\ and\ Hypopneas}$$

This method assigns SA labels even without a documented diagnosis in patient records.

2.3.1.1. Handling missing values

Missing values in the datasets were handled with imputation methods to ensure accurate analysis. For individual missing values, linear interpolation was used to estimate data based on nearby values. For columns with more missing data, the median or mean value was imputed to preserve dataset integrity.

2.3.2. Model selection and hyperparameter Tuning

After preprocessing, the dataset was split into 85% for training/validation and 15% for testing. The Lazy Predict method was used to compare multiple classifiers. Based on ROC-AUC and F1-score, a random forest model was chosen. Hyperparameters were optimized via grid search and 5-fold cross-validation, resulting in a model with 200 trees, a maximum depth of 6, and a minimum of 10 samples to split nodes. This configuration provided the best performance while ensuring simplicity and generalizability.

2.3.3. Interpretability and clinical relevance

In the development of predictive models for clinical applications, the interpretability of results by healthcare professionals is important. Alongside the optimization of performance metrics, particular attention was given to the selection of clinically significant features. Features such as HR and RR possess well-documented relevance in the context of SA and HF, thereby enhancing the interpretability of the model's output.

3. Results

The random forest model was trained on 85% of the data, with 15% used for testing. Forward feature selection was applied to optimize performance, and ROC-AUC was used to assess the model's ability to distinguish between SA and no SA. Based on this process, eight features were selected for the final model: age, height, HR and RR during sleep, NYHA class, duration in bed before sleep onset, sleep onset duration, and baseline weight. The model was evaluated with a classification threshold of 0.41, achieving an ROC-AUC of 0.85. Based on 206 true positives, 294 true negatives, 41 false positives, and 49 false negatives, it performed with a sensitivity of 0.81, specificity of 0.88, positive predictive value of 0.83, negative predictive value of 0.86, and an F1-score of 0.82.

4. Discussion

This study aimed to develop an ML-based screening tool for detecting SA for patients with HF and AF using data from non-invasive home monitoring devices. The final random forest model achieved an ROC-AUC of 0.85, with a sensitivity of 0.81 and a specificity of 0.88. These results suggest the model is a promising screening solution. Features such as age, HR, and RR are known to be associated with SA [12]. NYHA was also included as HF severity correlates with higher SA prevalence [12]. Features like height may have been selected due to bias from the small dataset. Additionally, key physiological variables such as oxygen saturation (SpO_2), which are strong predictors for SA, were not available. Previous studies have shown that adding SpO_2 could improve the model's performance [13]. This study has had limitations such as: The SA labels utilized in this study were not validated through PSG; The dataset was relatively small and imbalanced, with only six participants confirmed to have SA, potentially introducing bias and limiting the model's generalizability. Since data were collected at home, adherence played a critical role in data quality. Moreover, it is crucial to address the potential risk of overfitting, especially considering that the model was developed using a relatively small dataset with a limited number of confirmed SA cases. However cross-validation and hyperparameter tuning have been employed to mitigate these risks. The interpretability of the model's predictions is limited due to the inherent complexity of the random forest algorithm. While the random forest model achieved high performance, its complexity can limit clinical interpretability. Future studies may consider more transparent models (decision trees and logistic regression), which offer insight into variable relationships and can enhance clinician trust [14].

5. Conclusion

This study demonstrates that a machine learning model can effectively screen for sleep apnea in patients with heart failure using non-invasive sleep sensor. The final random forest model achieved a ROC-AUC of 0.85, sensitivity of 0.81, and specificity of 0.88, indicating strong potential for early detection in telerehabilitation. Further work should explore clinical validation, long-term follow-up, and regulatory approval.

6. Acknowledgment

This study is funded by the Aage and Johanne Louis-Hansen Foundation and the Danish Heart Association. Thanks to healthcare professionals at the Regional Hospitals in Silkeborg and Viborg, and the Healthcare Centers in Viborg, Skive, and Silkeborg.

References

[1] Kemp CD, Conte JV. The pathophysiology of heart failure. Cardiovasc. Pathol. 2012 Sep;21(5):365-71- doi: 10.1016/j.carpath.2011.11.007.
[2] Schwinger RHG. Pathophysiology of heart failure. Cardiovasc Diagn Ther. 2021 Feb;11(1):263-276. doi: 10.21037/cdt-20-302.
[3] Mansukhani MP, Kolla BP, Surani S. Management of obstructive sleep apnea. Diagnostic testing for obstructive sleep apnea. Cham: Springer; 2021. p. 75–83. doi:10.1007/978-3-030-54146-0_7.
[4] Patil SP, Schneider H, Schwartz AR, Smith PL. Adult obstructive sleep apnea: pathophysiology and diagnosis. Chest. 2007 Jul;132(1):325-37. doi: 10.1378/chest.07-0040.
[5] Aalaei S, Amini M, Taghipour A, Afsharisaleh L, Rezaeitalab F, Asadpour H, Eslami S. Adherence to prescribed overnight sleep study in patients suspected of sleep apnea: problem size and influential factors. Sleep Breath. 2021 Sep;25(3):1359-1368. doi: 10.1007/s11325-020-02216-9. Epub 2020 Nov
[6] Full KM, Kerr J, Grandner MA, Malhotra A, Moran K, Godoble S, Natarajan L, Soler X. Validation of a physical activity accelerometer device worn on the hip and wrist against polysomnography. Sleep Health. 2018 Apr;4(2):209-216. doi: 10.1016/j.sleh.2017.12.007. Epub 2018 Jan 17.
[7] Labarca G, Schmidt A, Dreyse J, Jorquera J, Barbe F. Telemedicine interventions for CPAP adherence in obstructive sleep apnea patients: Systematic review and meta-analysis. Sleep Med Rev. 2021 Dec;60:101543. doi: 10.1016/j.smrv.2021.101543. Epub 2021 Aug 31.
[8] Saleem K, Rana R, Khan M, Anwar SM, Khan MA. IoT healthcare: design of smart and cost-effective sleep quality monitoring system. J Sens. 2020;2020:8882378. doi.org:10.1155/2020/8882378
[9] Dinesen B, Dittmann L, Gade JD, Jørgensen CK, Hollingdal M, Leth S, Melholt C, Spindler H, Refsgaard J. "Future Patient" Telerehabilitation for Patients With Heart Failure: Protocol for a Randomized Controlled Trial. JMIR Res Protoc. 2019 Sep 19;8(9):e14517. doi: 10.2196/14517.
[10] Dinesen, Birthe B, Huniche L, Nielsen C, Toft E. Future Patient - telerehabilitering af patienter med atrieflimren. Version 1.0. 2019 Oct 28.
[11] Dinesen B, Albertsen AE, Joensen EDR, Spindler H, Jensen KM, Kidholm K, Frost L, Dittman L, Gunasegaram M, Johnsen SP, Jochumsen MR, Svenstrup D. Future Patient-Telerehabilitation of Patients With Atrial Fibrillation: Protocol for a Multicenter, Mixed Methods, Randomized Controlled Trial. JMIR Res Protoc. 2025 Feb 18;14:e64259. doi: 10.2196/64259.
[12] dos Santos Silva M, de Almeida FR, Lima AMJ, Silva BC, de Araújo ML, Rocha RS, et al. Associations of the Severity of Obstructive Sleep Apnea With Age-Related Comorbidities: A Population-Based Study. Front Neurol. 2022;13:802554. doi:10.3389/fneur.2022.802554.
[13] Ganglberger W, Bucklin AA, Tesh RA, Da Silva Cardoso M, Sun H, Leone MJ, Paixao L, Panneerselvam E, Ye EM, Thompson BT, Akeju O, Kuller D, Thomas RJ, Westover MB. Sleep apnea and respiratory anomaly detection from a wearable band and oxygen saturation. Sleep Breath. 2022 Sep;26(3):1033-1044. doi: 10.1007/s11325-021-02465-2. Epub 2021 Aug 18.
[14] Luo Y, Tseng HH, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. BJR Open. 2019 Jul 4;1(1):20190021. doi: 10.1259/bjro.20190021.