

Construction and Validation of a Machine Learning-Based Risk Prediction Model for Sleep Quality in Patients with OSA [Letter]

Hejia Wan¹, Yifan Li², Fei Xu³

¹School of Nursing (Nursing School of Smart Healthcare Industry), Henan University of Chinese Medicine, Zhengzhou, People's Republic of China; ²Joint Institute of Management and Science University at Henan University of Chinese Medicine, Henan University of Chinese Medicine, Zhengzhou, People's Republic of China; ³School of Information Technology, Henan University of Chinese Medicine, Zhengzhou, People's Republic of China

Correspondence: Hejia Wan, Email 2022140071@hactcm.edu.cn

Dear editor

The recently published paper “Construction and Validation of a Machine Learning-Based Risk Prediction Model for Sleep Quality in Patients with OSA” in *Nature and Science of Sleep* has drawn our close attention.¹ We read this study with great interest. It innovatively combines the LightGBM algorithm with social psychological multimodal data (n=400) to develop a sleep quality prediction model (AUC=0.910). Notably, the authors identified six core predictors through the SHAP explainability framework, including depression syndrome, ODI oxygen reduction index, anxiety symptoms, and caffeine intake, among which depression syndrome was the strongest predictor with a contribution rate of 34.2%.¹ This model not only significantly outperforms traditional methods (SVM/RF by about 12–15%), but also demonstrates clinical applicability through calibration curves and decision curve analysis (DCA) - providing significant net benefits within the 0.1–0.8 risk threshold range. This achievement provides an important decision support tool for precise intervention in obstructive sleep apnea (OSA), and its research paradigm of integrating biomarkers with behavioral factors has a significant promoting effect on individualized diagnosis and treatment in sleep medicine.

Despite the significant innovative value of this study, we still find several technical limitations worth discussing. Firstly, the algorithm's extraction of temporal dynamic features is insufficient: the current model only uses static statistics such as the median ODI,² failing to capture the periodic patterns of nocturnal blood oxygen fluctuations (such as oxygen reduction event clusters during REM sleep), which may lead to misjudgment of the risk of moderate to severe patients. Secondly, the LightGBM shows a completely fitted state in the training set (AUC=1.0), and although L2 regularization is applied, the lack of multi-center external validation (single-institution sample) poses a risk of overfitting, which should be carefully evaluated.³ The study has a male proportion as high as 90%, raising doubts about the model's generalization ability. Finally, although SHAP explanations enhance transparency, the clinical translation of feature interaction effects still faces obstacles - for example, the synergistic mechanism of depression and ODI has not been verified through longitudinal studies, and the complex tree structure prediction still relies on professional data analysts.

To optimize these limitations, it is suggested to explore the integration of time-aware architectures and traditional algorithms for improvement. For instance, introducing LSTM neural networks⁴ to process raw PSG signals: inputting the ODI sequence per minute into a bidirectional LSTM layer to capture the dynamic patterns of oxygen reduction in the first half and second half of the night, and then inputting them together with static features into the LightGBM decision layer. This can enhance the sensitivity of dynamic risk warning. At the same time, adopting a transfer learning strategy⁵ based on large open-source databases to pre-train feature encoders can effectively alleviate the problem of sample bias. To promote clinical application, it is recommended to develop an interactive risk dashboard:⁶ integrating SHAP force-directed graphs and patient historical data comparison modules to visually display actionable insights such as dynamic changes. Such improvements not only strengthen the robustness of predictions but also provide more precise time-sensitive guidance for personalized interventions such as CPAP compliance and cognitive behavioral therapy.

Data Sharing Statement

Data sharing is not applicable to this article as no data were created or analysed in this communication.

Author Contributions

Hejia Wan: Conceptualization, Project administration, Writing - Original Draft, Writing - Review & Editing;

Yifan Li: Resources, Writing - Original Draft, Writing - Review & Editing;

Fei Xu: Methodology, Writing - Original Draft, Writing - Review & Editing;

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

No funding was received.

Disclosure

The authors declare that they have no conflicts of interest in this communication.

References

1. Tong Y, Wen K, Li E, et al. Construction and validation of a machine learning-based risk prediction model for sleep quality in patients with OSA. *Nat Sci Sleep*. 2025;17:1271–1289. doi:10.2147/NSS.S516912
2. Koirala N, Perdue MV, Su X, et al. Neurite density and arborization is associated with reading skill and phonological processing in children. *Neuroimage*. 2021;241:118426. doi:10.1016/j.neuroimage.2021.118426
3. Nakajima T, Katsumata K, Kuwabara H, et al. Urinary polyamine biomarker panels with machine-learning differentiated colorectal cancers, benign disease, and healthy controls. *Int J Mol Sci*. 2018;19(3). doi:10.3390/ijms19030756
4. Xie J, Kai H, Zhu M, et al. Bioacoustic signal classification in continuous recordings: syllable-segmentation vs sliding-window. *Expert Syst Appl*. 2020;152:113390. doi:10.1016/j.eswa.2020.113390
5. Kulathilake KASH, Abdullah NA, Sabri AQM, et al. A review on deep learning approaches for low-dose computed tomography restoration. *Complex Intell Systems*. 2023;9(3):2713–2745. doi:10.1007/s40747-021-00405-x
6. María Conejero J, Preciado JC, Jesús Fernández-García A, et al. Towards the use of data engineering, advanced visualization techniques and association rules to support knowledge discovery for public policies. *Expert Syst Appl*. 2021;170:114509. doi:10.1016/j.eswa.2020.114509

Dove Medical Press encourages responsible, free and frank academic debate. The content of the Nature and Science of Sleep 'letters to the editor' section does not necessarily represent the views of Dove Medical Press, its officers, agents, employees, related entities or the Nature and Science of Sleep editors. While all reasonable steps have been taken to confirm the content of each letter, Dove Medical Press accepts no liability in respect of the content of any letter, nor is it responsible for the content and accuracy of any letter to the editor.

Nature and Science of Sleep

Publish your work in this journal

Nature and Science of Sleep is an international, peer-reviewed, open access journal covering all aspects of sleep science and sleep medicine, including the neurophysiology and functions of sleep, the genetics of sleep, sleep and society, biological rhythms, dreaming, sleep disorders and therapy, and strategies to optimize healthy sleep. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/nature-and-science-of-sleep-journal>

Dovepress
Taylor & Francis Group