

Prediction of Obstructive Sleep Apnea using Machine Learning Technique

National Taiwan University Hospital, Sleep Center

National Taiwan University, Graduate Institute of Biomedical Electronics and
Bioinformatics, and MediaTek Incorporation

Pei-Lin Lee, Wen-Chi Huang*, Yu-Tin Liu, and Feipei Lai

Introduction: Many questionnaires or prediction models tried to identify patients with obstructive sleep apnea (OSA) to prioritize sleep study. However, the performance was not unified which was related to the prevalence and definition of OSA, characteristics of participants, and feature selected. Therefore, the present study proposed a novel model to predict OSA with the minimal features.

Methods: We collected the clinical features and polysomnographic parameters from 5,301 patients (mean age 47.5 ± 14.4 y/o, men 76.5%) referred for suspect OSA [apnea-hypopnea index (AHI) ≥ 10 /hr]. Among 5,301 patients, the mean AHI was 29.7 ± 26.1 /hr and 70.1% had OSA. The patient numbers of training set and testing set for building model were 3,048 and 2,253, respectively. Thirty-five features including comorbidities, anthropometric information, and symptom were utilized to investigate the relevance to OSA. AHI ≥ 10 /hr were considered to suffer from OSA. Support Vector Machine (SVM) was employed for the judgement of feature effectiveness during selection procedure. To eliminate class-imbalance effect, training samples were re-sampling by Synthetic Minority Over-sampling Technique. Eventually, a model to predict OSA was designed based on SVM and the selective set of features. The performance of OSA prediction models were assessed with area under receiver operating characteristic curve (AUROC), sensitivity, specificity, accuracy, Likelihood ratio-positive [LR (+)], and Likelihood ratio-negative [LR (-)].

Results: With 35 features, the AUROC, sensitivity, specificity, accuracy, LR (+), and LR (-) was 0.818, 85.6%, 58.2%, 77.2%, 2.05 and 0.25, respectively. Nine features had been selected based on model performance non-inferior to that in 35-features AUROC. The 9-feature model gave AUROC, sensitivity, specificity, accuracy LR (+), and LR (-) as 0.818, 86.5%, 57.4%, 77.5%, 2.03 and 0.24, respectively.

Conclusion: The 9-feature model had a good performance for identifying patients with OSA. It could be applied to prioritize patients for polysomnography which may bring earlier treatment.

Support: The study was supported by grants from National Taiwan University (number NTU-ERP-104R8951-1, 105R8951-1, 106R880301) and NTU-NTUH-MediaTek Innovative Medical Electronics Research Center.

中文題目：機器學習應用於預測阻塞性睡眠呼吸中止症

作者：李佩玲¹ 黃文圻^{*2} 劉宇庭³ 賴飛熊²

服務單位：¹國立臺灣大學附屬醫院睡眠中心 ²國立臺灣大學生醫電子與資訊
研究所 ³聯發科技股份有限公司