Evaluation of a Deep Learning Sleep Staging Algorithm Utilizing a Single Frontal EEG Channel on a Clinical Population with Suspected or Known Obstructive Sleep Apnea

D.J. Wicks, M.A. McMahon
Epworth Hospital, Sleep Disorders Unit, Melbourne, Australia

Introduction

Home Sleep Apnea Testing (HSAT) for the diagnosis of Obstructive Sleep Apnea (OSA) has emerged as a simpler and more cost-effective diagnostic option compared with attended in-lab Polysomnography (PSG). The identification of sleep stages form an essential part of the OSA diagnosis as it allows for proper phenotyping of OSA, specifically the REM phenotype. The manual staging of sleep is arduous and costly, so the development of accurate Deep Learning (DL) algorithms the automatically classify sleep stages forms a crucial role in the diagnosis of OSA with HSAT.

The purpose of this study is to investigate the accuracy of a DL sleep staging algorithm in a new miniaturized sleep monitoring device – Somfit® Compumedics Ltd (see Figure 1). This device is attached to the patient's forehead using an adhesive electrode patch (See Figure 2). This patch records one EEG channel (Fp1 – Fp2), two EOG channels (EOG-R, EOG-L), one EMG channel (frontalis) with additional channels recorded onboard the Somfit ® module (oximetry, PAT, pulse rate, snore sounds, head positioning, actigraphy).

Methods

40 participants were enrolled into the study and involved simultaneous recording of full overnight PSG and Somfit® data. The eligible participants were patients who were referred to the unit for investigation of suspected OSA or reassessment of know OSA.

All PSG recordings were independently scored according to current AASM guidelines by three qualified, experienced sleep scientists with over 10 years experience in sleep science. A consensus hypnogram was generated from the three manual PSG hypnograms and used as the final hypnogram for all subsequent comparisons with the Somfit ® automatic hypnogram.

The Somfit ® sleep staging algorithm is based on DL U-sleep convolutional neural network (CNN) architecture. This architecture was initially trained on 783 Somfit ® studies with the manual Somfit ® hypnograms used as the ground truth. To improve detection of stage N1, another similar architecture was trained on 216 parallel Somfit ® and PSG recordings with the manual PSG hypnograms used as the ground truth.

Results

Table 1 Characteristics of Study Cohort

Participants Characteristics		N=40
OSA status at enrolment	Suspected	37
	Known	3
ESS – Mean (SD)		9.45 (4.71)
OSA Severity	No OSA	11
	Mild	15
	Moderate	6
	Severe	8
Gender	Male	28
	Female	12
Age Group	<65 years old	34
	≥65 years old	6
BMI Category	<25	5
	25-30	17
	≥30	18

Table 2 Percent Agreement Between Manual PSG Hypnograms and Consensus PSG vs Somfit

		Percentage Agreement – Mean (SE)			
		N1-N2-N3-REM-WAKE	Light NREM-N3-REM-Wake	NREM-REM-Wake	Sleep - Wake
Manual PSG Scorers	A vs B	84.45 (0.66)	89.20 (0.56)	95.23 (0.34)	96.60 (0.30)
	A vs C	85.50 (0.64)	89.69 (0.51)	95.14 (0.59)	96.69 (0.24)
	B vs C	84.46 (0.77)	89.71 (0.67)	95.28 (0.29)	96.85 (0.27)
Consensus PSG vs Somfit		75.91 (1.15)	79.87 (1.14)	87.38 (1.15)	89.89 (1.03)

Table 3 Kappa Statistics and Percent Agreement for Somfit Algorithm vs Consensus PSG Hypnogram – Five Sleep Stages (N1-N2-N3-REM-WAKE) – Different OSA Categories

	No OSA (11)	Mild OSA (15)	Moderate OSA (6)	Severe OSA (8)	All (40)
Percent Agreement – Mean (SE)	78.19 (2.36)	76.44 (1.29)	74.82 (3.75)	72.59 (3.10)	75.91 (1.15)
Kappa Coefficient – Mean (SE)	0.684 (0.034)	0.669 (0.019)	0.636 (0.047)	0.580 (0.045)	0.650 (0.017)

Conclusions and Recommendations

Agreement between Somfit and PSG hypnograms is close to that between manual PSG hypnograms, thus confirming the acceptability of the single frontal EEG electrode placement for accurate automatic sleep staging.

Agreement between Somfit and PSG hypnograms was reduced with increased OSA severity, likely due to subsequent increases in micro and macro-sleep fragmentation. No differences in agreements were observed with gender, BMI or age groups.

Steps to further improve DL algorithm accuracy include:

- Integration of a DL arousal detection algorithm and retraining of a DL sub-model for stage N1 on a larger Somfit/PSG data set
- Addition of DL models based on non-EEG Somfit channels (Actigraphy, SaO_2, PAT)

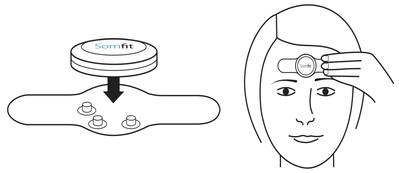

Figure 1

Figure 2