Prediction of Microsleeps from EEG: Preliminary Results

Reza Shoorangiz, *Student Member, IEEE,* Stephen J. Weddell, *Senior Member, IEEE,*
Richard D. Jones, *Fellow, IEEE*

Abstract— Brief episodes of momentarily falling asleep – microsleeps – can have fatal consequences, especially in the transportation sector. In this study, the EEG data of eight subjects, while performing a 1-D tracking task, were used to predict imminent microsleeps. A novel algorithm was developed to improve the accuracy of microsleep identification from two independent measures: tracking performance and face-video. The uncertain labels of gold-standard were then pruned out. Additionally, the state of microsleep at 0.25 s ahead was continuously predicted. Log-power spectral features were then extracted from EEG data. The most relevant features were selected by mutual information. Leave-one-subject-out was performed to test the classifier on an independent subject and this procedure was done for all the subjects. Two oversampling methods, synthetic minority oversampling technique (SMOTE) and adaptive sampling (ADASYN), were utilized to improve the training in the presence of imbalanced data. The best average area under the curve of receiver operating characteristic (AUC$_{ROC}$) of 0.90 was achieved using SMOTE oversampling over a 5.25 s window length, with a corresponding geometric mean (GM) of 0.74. ADASYN oversampling achieved the best sensitivity of 0.76 (cf. 0.70 for SMOTE), but with a lower specificity of 0.77 (cf. 0.86 for SMOTE).

I. INTRODUCTION

Microsleeps are brief episodes ($\sim$0.5–15 s) of loss of consciousness in which a person unintentionally stops responding and seems to momentarily fall asleep [1], [2]. Microsleeps are usually associated with behavioural sleep cues such as slow eye closure, droopy eyes, head nods, and increased duration of eye blinks [1], [3]. This phenomenon is a safety hazard to active and monotonous task operators [4], such as truck drivers, pilots, and air traffic controllers. Vanlaar et al. used a public opinion poll to collect data from 750 Ontario drivers in which 58% of participants admitted to having driven while fatigued or drowsy, and 14% acknowledged experiencing nodding off or falling asleep while driving [4]. Sleep related motor vehicle accidents have been estimated to account for 2% (Norway) to 25% (Australia) of car crashes [5].

Reza Shoorangiz is with the Department of Electrical and Computer Engineering at University of Canterbury, Christchurch Neurotechnology Research Programme, and New Zealand Brain Research Institute, Christchurch, New Zealand (email: reza.shoorangiz@nzbri.org).

Stephen J. Weddell is with the Department of Electrical and Computer Engineering at University of Canterbury and Christchurch Neurotechnology Research Programme, Christchurch, New Zealand (email: steve.weddell@canterbury.ac.nz).

Richard D. Jones is with the Department of Electrical and Computer Engineering at University of Canterbury, Christchurch Neurotechnology Research Programme, and New Zealand Brain Research Institute, Christchurch, New Zealand (email: richard.jones@nzbri.org).

Peiris et al. [2] showed that non-sleep-deprived subjects experienced frequent microsleeps while conducting a 1-hour 1-D continuous task with an average rate of 15.2 (0.0–72.0) h^{-1} and average duration of 3.2 (1.8–4.6) s. Poudel et al. [6] found that propensity of falling asleep increases with sleep restriction. However, they found no correlation between the number of microsleeps when normally rested and after sleep restriction [7]. They also found that normally-rested subjects had 11.4 (0–85) microsleeps on a 20-min 2-D tracking task with a mean duration of 2.8 s. These results highlight that even healthy normally-rested people are susceptible to microsleeps. They also highlight the benefits of being able to predict microsleeps and being able to prevent fatal accidents, especially in the transportation sector.

EEG-based detection of microsleeps has been investigated [3], [8], [9]. Peiris et al. extracted power spectral features, fractal dimension, approximate entropy, and Lempel-Ziv complexity of EEG [8]. They used PCA to create meta-features and stacking of 6 linear discriminant analysis (LDA) classifiers to detect microsleeps. All of the principal components (PCs) were fed to the classifiers, as overfitting was not observed. Their best results of AUC$_{ROC}$ = 0.86, Pearson's correlation coefficient (ϕ) = 0.39, and AUC$_{PR}$ = 0.43 were achieved by using spectral features only. Davidson et al. [3] used the same procedure, but with long-short-term-memory (LSTM) recurrent neural networks. PCA was used to reduce the number of features to 30 to avoid overfitting and reduce computational complexity. They were able to achieve ϕ = 0.38, AUC$_{ROC}$ = 0.81, and AUC$_{PR}$ = 0.43. Ayyagari et al. [9] used the same features and applied them to echo state networks with leaky neuron. They achieved ϕ = 0.51, sensitivity (S_n) = 0.85%, specificity (S_p) = 0.94, and selectivity (S_l) = 0.53 with stacked generalization. All of the aforementioned studies used EEG segments of 2 s with 50% overlap.

The previous studies all showed promising results in terms of detection of microsleeps. Microsleeps were defined as a flat response in tracking together with eye-closure identification independently from face-video. Logical operators were used to combine tracking performance, and the expert rated face-video. This however introduced mistakes into the gold-standard. For instance, a person might exhibit eye-closure leading to a video-lapse rating from an expert, while tracking performance was satisfactory. This situation requires a logical AND to correctly identify microsleeps. However, there are other situations where logical OR is needed to correctly identify microsleeps. Therefore, the term microsleep is more centred on the subject's performance. We have decided to

redefine the gold-standard with more strict measures so as to increase the accuracy of microsleeps. In addition, previous research focused on the detection of microsleeps. To the best of our knowledge, prediction of microsleeps has not been investigated in the literature. In the current study, our aim was to *predict* the occurrence of microsleeps.

The aim of this work was to (1) improve accuracy of the behavioural gold-standard for microsleeps, and (2) carry out preliminary work toward prediction.

II. METHODOLOGY

A. Data

Fifteen healthy non-sleep-deprived volunteers were recruited. No neurological or sleep disorder was reported by subjects. The average sleep duration for the previous night was 7.8±1.2 h [2].

Each subject took part in two sessions of one hour to perform a 1-D random preview tracking task. This task was to keep the tracking cursor as close as possible to the pseudo-random target. EEG, tracking performance, and facial video of subjects were recorded while performing the task. EEG was recorded with sampling frequency of 256 Hz from 16 channels placed according to international 10-20 system, namely Fp1, Fp2, F3, F4, F7, F8, C3, C4, O1, O2, P3, P4, T3, T4, T5, and T6. Video of facial behaviour features was recorded at 25 fps. Tracking performance was sampled at 64 Hz.

B. EEG Preprocessing

Following band-pass filtering of the EEG between 1 Hz and 45 Hz, artefact subspace reconstruction (ASR) [10] was employed to remove artefacts with z-score over 5. ASR finds a clean template of data and uses that to find and remove artefacts in the rest of EEG. However, EEG is non-stationary and therefore it might not be appropriate to find a clean portion of 1-hour data and use it to find artefacts in the rest. Therefore, we segmented the EEG into 2-min epochs with 50% overlap. ASR was then applied to each epoch independently. Clean data was found within each epoch which was then used to clean the same epoch. The epochs were then concatenated together to form a cleaned set of original data. In this process, the overlapping parts of consecutive epochs were averaged to avoid discontinuity. Muscle artefacts were removed from EEG using canonical correlation analysis [11].

C. Improving the Gold-standard and Microsleeps

Identification of microsleeps was done by fusion of independent measures of video analysis and tracking performance. The rating of behavioural clues was done by three experts using video recordings without knowledge of the tracking performance. This analysis was done based on 6-scale ratings: alert, distracted, forced eye closure, light drowsy, deep drowsy, and microsleep [8].

Peiris et al. [2], [8] analysed tracking performance to find flat responses longer than a specific time span. The gold-standard was then estimated by applying a logical operator, i.e., AND or OR, to the video ratings and tracking analysis. This approach, however, does not take the transition from responsive to microsleep, or vice versa, into account. Furthermore, flat tracking responses might have been due to the slow target velocity. In such situation, a logical OR might introduce false microsleeps into gold-standard while the participant was in fact responsive. In order to reduce the uncertainty of the gold-standard, we developed a rigorous algorithm using more strict measures to account for slow tracking velocity, transition from one state to another, and uncertain parts of gold-standard.

The first step of tracking analysis was to find any consistent lead or lag between the tracking and target using cross correlation and compensate it. This is due to the fact that the subjects could have seen the preview of the following 8 s of target [8] and therefore a small lead or lag between tracking and target might have been introduced. The next step was to find the responsive parts of tracking where the subject was tracking the target accurately. At first, fractions of target with slow velocity were temporarily removed from the analysis, since it was inaccurate to estimate whether or not the subject was actually tracking while the target was moving slowly. The absolute error between the target and corresponding tracking output was then computed using a moving window of 2 s. Windows with a mean absolute error less than 9.6 mm/s were added to a responsive template. To account for abrupt attention lapses, consecutive responsive windows in the template that were closer than 0.5 s were merged. Next, the slow velocity parts of data was analysed. A slow velocity window was considered to be responsive if the length of such window was less than 2 s and the subject was tracking accurately 4 s prior and after that.

The parts of the tracking task with a mean absolute error > 3 cm/s for longer than 1 s were labelled 'deviated-regions'. Flat-spots were also defined as the regions of tracking task that velocity of tracking drops below 1.1 mm/s while the target's velocity is higher than 2.6 mm/s (10th percentile of target velocity). In addition, a flat-spot must last for at least 1 s and the mean absolute deviation from the target be larger than 1.5 cm. Slow velocity regions of the target were not considered in the analysis of either flat-spots or deviated-regions.

The gold-standard was then formed by fusion of video ratings, responsive template, deviated-regions, and flat-spots. The gold-standard corresponds to 'responsive' if the subject is tracking, regardless of video ratings. As a result, responsive templates were directly added to the gold-standard. Microsleeps were defined as non-tracking (i.e., union of deviated-regions and flat-spots), in conjunction with a video-rating of deep drowsy or lapse. The remainder of the gold-standard was defined as 'uncertain' and pruned out. Figure 1 illustrates the gold-standard, tracking, and target.

D. Feature Selection and Reduction

Power density estimation was done using Welch's method [12] based on 2-s windows and 75% overlap. The log-power of various frequency bands of EEG were ex-
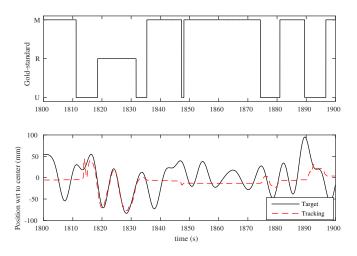
Fig. 1. Illustration of gold-standard with respect to target and tracking. The states of gold-standard are uncertain (U), responsive (R), and microsleep (M).

Fig. 2. Correspondence of gold-standard and EEG segments.

tracted, namely: delta (1–4.5 Hz), theta (4.5–8 Hz), alpha1 (8–10.5 Hz), alpha2 (10.5–12.5 Hz), alpha (8–12.5 Hz), beta1 (12.5–15 Hz), beta2 (15–25 Hz), beta (12.5–25 Hz), gamma1 (25–35 Hz), gamma2 (35–45 Hz), gamma (25–45 Hz), overall (1–45 Hz). This led to 12 features per EEG channel, and a total of 192 features per epoch. Due to the large number of features per epoch, the curse of dimensionality might exist. To overcome this, we used a greedy forward feature selection algorithm based on mutual information of features and the class labels [13]. The aim of this procedure was to add the most informative feature with respect to the gold-standard at each step. Mutual information can be written as follows:

$$I(\mathbf{f}, G) = H(G) - H(G|\mathbf{f}) \tag{1}$$

where $\mathbf{f}$ is a feature set, G is the gold-labels. $H(G)$ is the entropy of gold-standard and is defined as

$$H(G) = - \sum_{G \in \{R,M\}} P(G) \log(P(G)) \tag{2}$$

where $P(G)$ is the probability function of gold-standards. $H(G|\mathbf{f})$ is the conditional entropy and is defined as

$$H(G|\mathbf{f}) = - \sum_{G \in \{R,M\}} P(G|\mathbf{f}) \log(G|\mathbf{f}). \tag{3}$$

Features were assumed to have normal distribution and joint distributions have full covariance matrix. A 'greedy method' was used to add the most informative feature at every iteration. At every iteration, the mutual information of union of selected features and each of the remaining ones was calculated. The feature resulted in the highest mutual information was added to the selected features. This procedure continued until a stopping criterion was met, such as maximum number of features or the improvement in the mutual information. In this study, we chose 40 as the maximum number of features to be selected.
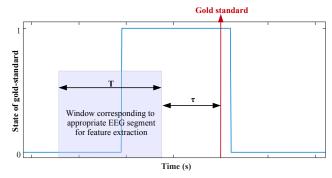
E. Classification

We evaluated the state of the gold-standard every 0.25 s and compared it to the EEG segment τ prior to the gold-standard sample (Figure 2). The minimum duration of microsleep is defined as ∼0.5 s, so the gold-standard at 4 Hz does not miss any microsleeps. A τ of 0.25 s corresponds to one-step-ahead (0.25 s) prediction of the gold-standard.

Classification was done using a single LDA. Such a simple classifier was chosen to investigate the efficacy of prediction with a simple method before exploring more complex classifiers. LDA also ties covariance matrices for both classes which avoids singularity, especially when the number of microsleeps might be low. The training phase was done based on nested 5-fold cross-validation on the number of features and window length of EEG segments. The optimal classifier was then applied to the independent test data to measure the performance of the system.

To reduce the effect of imbalanced data on the training of the classifier, two oversampling methods, synthetic minority oversampling (SMOTE) [14] and adaptive synthetic sampling (ADASYN) [15], were applied to the training data, while the validation set was unchanged. The results of such methods were compared with training without oversampling. AUC_{PR} was used to select the best model for each training set, as the microsleep dataset is substantially imbalanced [16]. For each window length, the simplest model with only one feature was initiated. At every iteration, one feature was added to the model and mean AUC_{PR} of cross-validation was examined until there was no further improvement on three consecutive iterations. This process was carried out for all window lengths and the model corresponding to the highest AUC_{PR} was selected.

III. RESULTS

Of the 15 participants in this study, our analysis was limited to 8 who experienced at least one microsleep in both sessions. Evaluation of microsleep prediction was done by leaving the whole data of one subject (both sessions) for testing, while training was done on the other seven. This process was repeated eight times and the average of performance measures were computed.

The new gold-standard had an average participant responsive time per 60-min of 35.8 (10.8–47.7) mins and definite

	No resampling	SMOTE	ADASYN
AUC_{ROC}	0.90	0.90	0.90
AUC_{PR}	0.38	0.36	0.34
GM	0.58	0.74	0.70
Phi (ϕ)	0.33	0.33	0.29
Sensitivity	0.46	0.70	0.76
Specificity	0.94	0.86	0.77
Precision	0.37	0.27	0.22

microsleep time of 2.2 (0.1–8.4) mins. The remainder of the gold-standard was marked as uncertain. Definite microsleeps occured at a rate of 15.3 (2–30) h^{-1}.

The performance of microsleep prediction is presented in Table I. Prediction using ADASYN gave the highest sensitivity but at the cost of low precision. On the other hand, maximum specificity (0.94) was achieved without resampling but with a low sensitivity (0.46). It must be noted that precision, ϕ, and AUC_{PR} depend on imbalance ratio [16] of the gold-standard which was not consistent across the subjects. As a result, the average value of such measures may not be accurate. The geometric mean (GM) is based on both sensitivity and specificity at the operating threshold of the classifier which might be a better measure of performance. Nevertheless, selection of the best classifier depends on the application and the relative importance of high sensitivity versus high specificity.

The number of features needed for classification and the best EEG window length were chosen based on each training set and oversampling method. The best model without over-sampling had 37 features on average and an EEG window length of 5.5 s. The SMOTE-based classifier had less features on average (40) and an optimal EEG window length of 5.25 s. A minimum number of features was needed for the classifier with ADASYN (40), with a corresponding EEG window length of 6.0 s.

IV. CONCLUSION

We investigated continuous prediction of microsleep in a one-step-ahead (0.25 s) configuration. Using log-power spectral features, mutual information feature selection, and a single LDA as classifier, we were able to achieve an AUC_{ROC} of 0.90, but a relatively low sensitivity. However, applying oversampling methods to the training substantially improved sensitivity, with a small drop in specificity. By applying SMOTE at the training stage, we were able to achieve the same AUC_{ROC} with moderate sensitivity and specificity.

Future work will be centred on incorporating more complex classifiers and features. LSTM and hidden Markov network methods can be used to integrate the dynamics of features to increase the prediction performance. Other features such as complexity measures, time-frequency domain features, and cortical connectivity [17] might also increase accuracy of prediction of microsleeps.

REFERENCES

[1] R. Jones, G. Poudel, C. Innes, P. R. Davidson, M. Peiris, A. Malla, T. Signal, G. Carroll, R. Watts, and P. Bones, "Lapses of responsiveness: Characteristics, detection, and underlying mechanisms," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Aug. 2010, pp. 1788–1791.

[2] M. T. R. Peiris, R. D. Jones, P. R. Davidson, G. J. Carroll, and P. J. Bones, "Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects," *J. Sleep Res.*, vol. 15, no. 3, pp. 291–300, 2006.

[3] P. Davidson, R. Jones, and M. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, May 2007.

[4] W. Vanlaar, H. Simpson, D. Mayhew, and R. Robertson, "Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors," *J. Safety Res.*, vol. 39, no. 3, pp. 303–309, 2008.

[5] T. Zhang and A. H. Chan, "Sleepiness and the risk of road accidents for professional drivers: A systematic review and meta-analysis of retrospective studies," *Saf. Sci.*, vol. 70, pp. 180–188, 2014.

[6] G. R. Poudel, C. R. H. Innes, and R. D. Jones, "Cerebral perfusion differences between drowsy and nondrowsy individuals after acute sleep restriction," *Sleep*, vol. 35, no. 8, pp. 1085–1096, aug 2012.

[7] C. R. H. Innes, G. R. Poudel, and R. D. Jones, "Efficient and regular patterns of nighttime sleep are related to increased vulnerability to microsleeps following a single night of sleep restriction," *Chronobiol. Int.*, vol. 30, no. 9, pp. 1187–1196, 2013.

[8] M. T. R. Peiris, P. R. Davidson, P. J. Bones, and R. D. Jones, "Detection of lapses in responsiveness from the EEG," *J. Neural Eng.*, vol. 8, 016003, 15 pp, 2011.

[9] S. Ayyagari, R. Jones, and S. Weddell, "Optimized echo state networks with leaky integrator neurons for EEG-based microsleep detection," in *Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Aug 2015, pp. 3775–3778.

[10] T. Mullen, C. Kothe, Y. Chi, A. Ojeda, T. Kerth, S. Makeig, T.-P. Jung, and G. Cauwenberghs, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2553–2567, Nov 2015.

[11] W. D. Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2583–2587, Nov 2006.

[12] M. Naderi and H. Mahdavi-Nasab, "Analysis and classification of EEG signals using spectral analysis and recurrent neural networks," in *Iran. Conf. of Biomed. Eng.*, Nov 2010, pp. 1–4.

[13] M. Arvaneh, C. Guan, K. Ang, and C. Quek, "Mutual information-based optimization of sparse spatio-spectral filters in brain-computer interface," *Neural Comput. Appl.*, vol. 25, no. 3–4, pp. 625–634, 2014.

[14] N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.

[15] H. He, Y. Bai, E. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. Int. Jt. Conf. Neural Netw.*, June 2008, pp. 1322–1328.

[16] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, p. e0118432, 2015.

[17] J. Toppi, L. Astolfi, G. R. Poudel, C. R. Innes, F. Babiloni, and R. D. Jones, "Time-varying effective connectivity of the cortical neuroelectric activity associated with behavioural microsleeps," *NeuroImage*, vol. 124, pp. 421–432, 2016.