2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 979-8-3503-2447-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/EMBC40787.2023.10340587

Automatic Detection of Chronic Insomnia from Polysomnographic and Clinical Variables Using Machine Learning

Umaer Hanif[1], Ulysse Gimenez[1], Alyssa Cairns[2], Daniel Lewin[2], Nisar Ashraf[2], and Emmanuel Mignot[3]

***Abstract*— Insomnia is defined subjectively by the presence and frequency of specific clinical symptoms and an association with distress. Although sleep study data has shown some weak associations, no objective test can currently be used to predict insomnia. The purpose of this study was to use previously reported and relatively crafted insomnia-related polysomnographic variables in machine learning models to classify groups with and without insomnia. Demographics, diagnosed depression, Epworth Sleepiness Scale (ESS), and features derived from electroencephalography (EEG), arousals, and sleep stages from 3,407 sleep clinic patients (2,617 without insomnia and 790 insomnia patients based on responses to a set of questions) were included in this analysis. The number of features were reduced using pair-wise correlation and recursive feature elimination. Predictive value of three machine learning models (logistic regression, neural network, and support vector machine) was investigated, and the best performance was achieved with logistic regression, yielding a balanced accuracy of 71%. The most important features in predicting insomnia were depression, age, sex, duration of longest arousal, ESS score, and EEG power in theta and sigma bands across all sleep stages. Results indicate potential of machine learning-based screening for insomnia using clinical variables and EEG.**

I. INTRODUCTION

Insomnia is defined as having difficulties initiating or maintaining sleep, despite adequate opportunity for sleep, and is associated with daytime consequences [1]. According to the International Classification of Sleep Disorders (ICSD), insomnia is classified as chronic when it occurs at least three times a week for at least three months [2]. 5-10% of the population suffer from chronic insomnia, which imposes considerable economic burdens on society [3]. Chronic insomnia is associated with increased age, being female, and having comorbid psychiatric disorders, notably anxiety and depression [4]. Insomnia patients have impaired functioning and increased risk of depression, cardiovascular diseases, and hypertension [5], [6]. There is no well-established objective test to diagnose insomnia. The diagnostic approach typically involves clinical and subjective measures including: 1) a physical exam to check for medical conditions that cause insomnia, 2) sleep diaries and psychometrically validated questionnaires to analyze sleep schedules and disruption, sleep habits, subjective feeling of sleep quality, and daytime impairment [4]. Although polysomnography (PSG) is used to evaluate efficacy of hypnotics (typically sleep efficiency, sleep latency, wake after sleep onset, and total sleep time) in clinical trial, PSG is not considered standard of care for insomnia diagnosis [4]. Of note, PSG abnormalities used for clinical trial assessment are only weakly associated with insomnia symptoms in the general population [7].

Currently, limited research efforts have focused on identification of insomnia-related features and utilization of these for machine learning-based prediction of insomnia. Shahin et al. [8] extracted statistical and spectral features from EEG and utilized a deep learning model to distinguish between healthy subjects and insomnia patients. Zhao et al. [9] demonstrated through a comprehensive review that insomnia patients display different behavior in most EEG bands during wakefulness and sleep compared to healthy subjects. Kim [10] evaluated performance of several machine learning models in classifying subjects with insomnia, showing classification accuracies between 56-98%.

Despite impressive performance in classifying insomnia, most studies include only a small sample of insomnia patients ($<$100), limiting the statistical power and representativeness of results. Additionally, the accuracy reported in most studies does not consider the imbalance between the number of insomnia patients compared to healthy subjects, which skews performance. Taking these limitations into account, our study evaluates a large, real world sample of treatment seeking patients. We evaluate discriminatory features associated with insomnia patients and subsequently utilize these to classify insomnia patients based on a combination of self-reported symptoms/complaints and PSG features. The purpose is to identify insomnia based on objective features that may increase understanding of the pathophysiology.

II. DATA DESCRIPTION

A total of 15,662 subjects were initially included in the study; they were extracted from a larger cohort, collected by Bioserenity between 2012-2015, in which subjects were referred for PSG to rule out obstructive sleep apnea or other sleep disorders. The Institution's Ethical Review Board approved all experimental procedures involving humans.

The dataset contained subject demographics, sleep-related questions, and PSG data. From the sleep questionnaire, six questions were chosen to identify insomnia as outlined in Table I. The questions were answered by subjects considering the past six months, fulfilling the time duration criterion for

[1]Umaer Hanif and Ulysse Gimenez are with Data Science, Bioserenity, 75013 Paris, France umaer.hanif@bioserenity.com, ulysse.gimenez@bioserenity.com

[2]Alyssa Cairns, Daniel Lewin, and Nisar Ashraf are with Sleep Research, Bioserenity, Atlanta, GA 30339, USA alyssa.cains@bioserenity.com, daniel.lewin@bioserenity.com, nisar.ashraf@bioserenity.com

[3]Emmanuel Mignot is with the Stanford University Center for Sleep and Circadian Sciences, Stanford University, Palo Alto, CA 94304, USA mignot@stanford.edu

TABLE I

TWO SETS OF QUESTIONS FOR DEFINING INSOMNIA. THE FIRST SET IS SLEEP-RELATED, AND THE SECOND SET IS DAYTIME-RELATED. 0: NONE, 1: SLIGHT, 2: MODERATE, 3: OFTEN, 4: SEVERE.

How much of a problem do you have:	0	1	2	3	4
1. with going to sleep at night? 2. because of waking up during the night? 3. with having restless, disturbed sleep?					
1. not feeling rested no matter sleep amount? 2. with tiredness during the day? 3. with sleepiness during the day?					

chronic insomnia (≥three months) [2]. Furthermore, a score ≥ 3 for a question indicated that the person experienced the symptom at least three times a week, fulfilling the frequency criterion for chronic insomnia [2]. The first three questions were related to sleep complaints, and a total score ≥ 10 was required for insomnia. Additionally, the last three questions were related to daytime complaints, and a total score ≥ 9 was required, given that the first criterion was fulfilled, for insomnia. The first threshold was determined to identify subjects with the most severe form of insomnia, who experience a combination of all sleep-related symptoms at a moderate-to-severe level. The second set of questions were used to exclude insomnia patients not experiencing daytime consequences, in order to satisfy ICSD definition [2].

Fig. 1 shows the distributions for both set of questions, which were used to determine both thresholds combined with the fact that chronic insomnia patients should constitute 5-10% of the dataset. To identify subjects without insomnia, a score ≤ 2 was used as threshold based on the first three questions only. The second set of questions was not used because other sleep disorders, such as sleep apnea, may cause daytime complaints with no relation to insomnia. Patients with other sleep disorders were not excluded because the severity of such disorders could potentially have importance in predicting insomnia. Using these criteria, the dataset was reduced from 15,662 to 3,407 subjects, for whom demographics are shown in Table II.

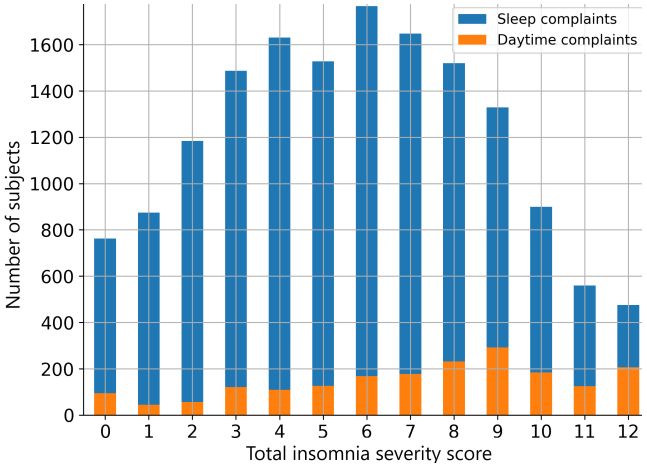

Fig. 1. Distribution of total scores based on three sleep-related questions and three daytime-related questions. Each question is given a score between 0 (no complaint) and 4 (severe complaint). The total scores for daytime questions is shown for subjects who have a sleep-related score ≥ 10.

TABLE II

MEAN ± STANDARD DEVIATION OF DEMOGRAPHICS FOR DATASET CONSISTING OF SUBJECTS WITH AND WITHOUT INSOMNIA. N: NUMBER OF SUBJECTS, F: FEMALES.

Group	**N**	**Sex**	**Age (years)**	**BMI (kg/m^2)**
No insomnia	2,617	F: 45%	53.7 ± 18.5	31.6 ± 7.58
Insomnia	790	F: 62%	48.0 ± 15.0	32.6 ± 7.68
All	3,407	F: 49%	52.3 ± 19.9	31.8 ± 7.6

The Epworth Sleepiness Scale (ESS) [11] was also obtained from each subject. For PSG variables, aggregated metrics such as total sleep time (TST), sleep onset latency (SOL), wake after sleep onset (WASO), sleep efficiency (SE), sleep maintenance efficiency (SME), time in bed (TIB), latency to persistent sleep (LPS), sleep onset REM (SOREM), apnea-hypopnea index (AHI), respiratory disturbance index (RDI), arousal index (AI), periodic leg movement index (PLMI), and desaturation index (DI) were available for each subject. However, the original sleep event annotations performed by technicians were lost at the time of processing the PSGs for clinical reports. Thus, it was necessary to run automatic scoring algorithms on the PSG recordings from this dataset.

III. METHODS

A. Automatic Scoring of Events

Two automatic deep learning-based scoring algorithms were applied: sleep staging and arousal detection. For sleep staging, the U-Sleep algorithm developed by Perslev et al. [12] was used to score sleep stages every 30 seconds. U-Sleep inputs electroencephalography (EEG) and electrooculography (EOG) signals to a convolutional neural network (CNN) [13] for sleep stage prediction, and it was trained and validated on 19,924 PSGs from 21 different datasets, obtaining an overall F1-score of 0.79. U-Sleep was chosen over other algorithms because it performed as accurately as the best human experts and generalized well across all cohorts included in the study [12].

For arousal detection, the Multimodal Arousal Detector (MAD) developed by Brink-Kjaer et al. [14] was utilized to obtain location and duration of each arousal in a recording. MAD inputs EEG, EOG, chin electromyography, and electrocardiography signals to a CNN followed by a long short-term memory network [13] for arousal detection, and it was trained and validated on 3,915 PSGs from 4 different datasets, obtaining an overall F1-score of 0.76. MAD was chosen over other algorithms for similar reasons as for U-Sleep, and because it predicts arousals and subsequent awakenings for any duration, as opposed to other detectors which detect arousals according to the American Academy of Sleep Medicine definition (microarousal $<$15 s, wake epoch $>$15 s, with dependence on how the signal is presented epoch by epoch for manual scoring) [15].

B. Feature Extraction

Three categories of features (F) were included for subsequent machine learning purposes: clinical information, arousal duration features, and sleep stage features.

1) Clinical Information (20 F)*:*

- Sex, age, BMI, diagnosed depression (4 F)
- TIB, TST, SOL, LPS, WASO, SE, SME, SOREM (8 F)
- AHI, RDI, AI, PLMI, DI, ESS score, number of snoring episodes, lowest oxygen saturation (8 F)

2) Arousal Durations (22 F)*:*

- Mean, standard deviation, minimum, maximum, percentiles (5%, 25%, 50%, 75%, 95%) (9 F)
- Percentage distribution (3-5 s, 5-15 s, 15-30 s, 30-60 s, 1-1.5 min, 1.5-2 min, 2-2.5 min, 2.5-3 min, 3-5 min, 5-10 min, 10-30 min, 30-60 min, >60 min) (13 F)

3) Sleep Stages (110 F)*:*

- Percentage distribution (N1, N2, N3, REM) (4 F)
- Sleep Fragmentation (1 F)
- Relative power in EEG bands (delta, theta, alpha, sigma, beta, gamma) + total absolute power in each brain area (frontal, central, occipital) in each sleep stage (wake, N1, N2, N3, REM) ((6+1)x3x5 = 105 F)

Sleep fragmentation was calculated by dividing the combined number of awakenings and transitions from N3 and REM to N1 by the total sleep duration in hours [17].

The power spectrum in a given sleep stage s, $s \in \{\text{W, N1, N2, N3, R}\}$, derived from a given EEG channel c, $c \in \{\text{F3, F4, C3, C4, O1, O2}\}$, was calculated as follows: first, all segments with the same s for a given c were concatenated, yielding the signal $x_{c,s}(n)$, where n denotes the sample index. Subsequently, the power spectrum $S_{c,s}(k)$, where k denotes the frequency index, was computed using Welch's method with a sliding Hamming window of length 4 seconds and 50% overlap [18]. Welch's method was chosen over theoretically more robust approaches, such as the Multitaper method [19], due to significantly lower computation time per recording. To obtain power in one of the bands b, $b \in \{\delta, \theta, \alpha, \sigma, \beta, \gamma\}$, $S_{c,s}(k)$ was integrated in the defined frequency interval using Simpson's composite rule, yielding $P_{b,c,s}$. Finally, relative power in a band was calculated as:

$$\hat{P}_{b,c,s} = \frac{P_{b,c,s}}{\sum_b P_{b,c,s}}, \tag{1}$$

where the denominator represents the total power in a given sleep stage and channel, which was also used as a feature. The relative and total power were averaged for channels F3 and F4, C3 and C4, and O1 and O2, yielding aggregated channels $\bar{c}$, $\bar{c} \in \{\text{F, C, O}\}$ with respect to spectral features. Spectral features were computed using YASA toolbox [16].

Each extracted feature vector was normalized to the range [0, 1] using min-max normalization, which is given by:

$$\mathbf{f} = \frac{\mathbf{f} - \min(\mathbf{f})}{\max(\mathbf{f}) - \min(\mathbf{f})}, \tag{2}$$

where $\mathbf{f} = (f_1, f_2, ..., f_N)$ is a given feature vector containing N examples, one for each subject in the dataset. Feature normalization was performed to ensure faster convergence during subsequent training of machine learning models.

C. Feature Selection

The feature extraction process generated a total of 152 features. Reducing the number of features was deemed necessary for two reasons: eliminating highly correlated, redundant features, and enhancing the interpretability of the selected machine learning model. Initially, a correlation matrix was generated by computing the pairwise correlation between all features. The correlation between two feature vectors $\mathbf{g}$ and $\mathbf{h}$ was calculated using the Pearson correlation defined as:

$$r_{gh} = \frac{\sum_{i=1}^{N}(g_i - \bar{g})(h_i - \bar{h})}{\sqrt{\sum_{i=1}^{N}(g_i - \bar{g})^2}\sqrt{\sum_{i=1}^{N}(h_i - \bar{h})^2}}, \tag{3}$$

where $\bar{\cdot}$ denotes the mean of a feature vector. A feature $\mathbf{g}$ was removed if, for any other feature $\mathbf{h}$, $|r_{gh}| > 0.9$. This approach yielded 25 features to remove. Subsequently, recursive feature elimination (RFE) was utilized to further reduce the number of features. RFE is a feature selection technique, which fits a model to the data and recursively removes features according to their importance, as determined by the fitted model, until a desired number of features is reached. RFE was carried out using logistic regression due to its simplicity and intuitive nature of calculating feature importance based on its coefficients. Furthermore, RFE was performed with 5-fold cross-validation to ensure robust feature selection. RFE yielded 55 additional features to remove. Thus, applying pairwise correlation and RFE reduced 152 features to 72 features.

D. Machine Learning

Three machine learning models were implemented to evaluate the predictive value of each model in classifying subjects with and without insomnia: logistic regression, neural network, and support vector machine (SVM). Hyperparameter values were chosen based on validation set performance. All models were trained and evaluated using 5-fold cross-validation to obtain predictions on the entire dataset.

1) Logistic Regression: Logistic regression outputs a probability between 0 and 1 for insomnia and is given by:

$$f(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{w}\cdot\mathbf{x}}}, \tag{4}$$

where $\mathbf{x} = (x_1, x_2, .., x_M)$ is a vector containing M features and $\mathbf{w} = (w_1, w_2, .., w_M)$ is a vector containing their associated coefficients or weights. The weights are optimized by minimizing the binary cross-entropy cost function:

$$J(\mathbf{w}) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot \log(f(\mathbf{x}_i)) + (1 - y_i) \cdot \log(1 - f(\mathbf{x}_i)), \tag{5}$$

where y_i is the label for the i^{th} training example (0: no insomnia, 1: insomnia). A L2 regularization term was added:

$$J(\mathbf{w}) + \frac{\lambda}{2N}\sum_{j=1}^{M} w_j^2, \tag{6}$$

where λ is called the regularization parameter and has the purpose of shrinking the weights to prevent overfitting. λ was set to 0.1 to introduce a small amount of regularization, and

stochastic average gradient was used for optimization [20]. Weights were introduced for both classes in the loss function due to imbalance between the number of insomnia patients compared to subjects without insomnia and were given by:

$$w_k = \frac{N}{2N_k}, \tag{7}$$

where N_k is the number of examples for class k, $k \in \{0,1\}$.

2) Neural Network: A neural network with two layers was implemented, given by the equation:

$$f(\mathbf{x}) = \sigma^{(2)}(\mathbf{w}^{(2)} \cdot \sigma^{(1)}(\mathbf{w}^{(1)} \cdot \mathbf{x} + b^{(1)}) + b^{(2)}), \tag{8}$$

where $\mathbf{w}^{(l)}$ and $b^{(l)}$ are the weights and bias of the l^{th} layer, and $\sigma^{(l)}$ are non-linear activation functions. Layer 1 consisted of 128 neurons, followed by a ReLU activation function, batch normalization [21], and dropout (probability of 0.1) for regularization [13]. 128 neurons proved to be the optimal number for addressing the complexity of this level, since incorporating additional neurons led to overfitting. Layer 2 consisted of one neuron followed by a sigmoid activation function to output a probability for insomnia. The cost function was given by Eq. 5, and penalty weights were introduced for each class in a similar manner as for logistic regression. The network was optimized using the Adam optimizer [22]. The learning rate was set to $1 \cdot 10^{-4}$, batch size was 128, and the network was trained for 100 epochs. Early stopping was applied when the validation error did not decrease for 10 consecutive epochs.

3) SVM: SVM creates a decision boundary using hyperplanes in a high-dimensional space by maximizing the margin (distance between the decision boundary and the nearest data point of each class) [23]. The output is given by:

$$f(x) = \sum_{i \in \text{SV}} y_i \alpha_i K(x_i, x) + b, \tag{9}$$

where SV are support vectors, which are data points that lie within the margin, α_i, $0 \leq \alpha_i \leq C$, are called dual coefficients, and $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$ is a kernel that maps the data into a higher-dimensional space by the function ϕ. C was set to 1.15 and acts as a penalty term for samples that are at a distance from their correct margin boundary. A polynomial of degree 3 was used as kernel function.

E. Performance Evaluation and Feature Importance

Due to subjects without insomnia being overrepresented compared to insomnia patients, balanced accuracy was applied as performance metric, which is given by:

$$\text{Acc}_{\text{bal}} = \frac{\text{Sensitivity} + \text{Specificity}}{2} = \frac{\frac{\text{TP}}{\text{TP+FN}} + \frac{\text{TN}}{\text{TN+FP}}}{2}, \tag{10}$$

where TP are the true positives, FN are the false negatives, TN are the true negatives, and FP are the false positives. Feature importance was calculated using the Shapley Additive Explanations (SHAP) algorithm [24]. The SHAP value for the j^{th} feature using N data points and given a logistic regression model (Eq. 4), is calculated as:

$$\phi_j = w_j x_j - w_j \frac{1}{N} \sum_{i=1}^{N} x_{ji}, \tag{11}$$

where x_j and w_j are the features and weights, respectively.

IV. RESULTS

Table III summarizes performance for each model described in Section III D in classifying subjects with and without insomnia. Table IV shows the 15 most important features in classifying insomnia patients based on SHAP importance values calculated by Eq. 11. Table V summarizes performance using different features for classification.

V. DISCUSSION

For all three models, balanced accuracy was ∼20% higher than if the model predicted all subjects to not have insomnia (balanced accuracy of 50%). Although the best three models were presented, many more were investigated (including random forest and XGBoost), but results for these were intentionally left out. For further analysis of the results, logistic regression was chosen because it displayed highest sensitivity in detecting insomnia, and for its simple nature.

Table IV shows that in this dataset, insomnia is associated with depression, decreased age, female sex, increased ESS score, decreased absolute EEG power in N3, longer duration of arousals, increased relative theta and sigma power during N1 and N3, decreased relative theta and sigma power during REM, decreased relative theta power during N2, and increased relative sigma power during N2. The association between insomnia and depression, sex, and ESS score is well-known [4], however not with decreasing age. Table II shows that insomnia patients have a lower mean age versus subjects without insomnia, while the average age of the dataset indicates that included subjects are young. Perhaps if older subjects were included, the inverse relationship between insomnia and age would not have been present. The increased theta and sigma activity during NREM was similarly reported by Zhao et al. [9], however the decreased activity in the same bands during REM was opposite. Increased alpha activity during REM was also reported in [9].

Table V shows that using the top 15 features obtained by the SHAP algorithm gives a balanced accuracy of ∼69%. Thus, the remaining 58 features contribute only a 2% increase in overall performance. Table V further highlights importance of the relative power in theta and sigma bands across all sleep stages, and the overall power in N3, for distinguishing between subjects with and without insomnia. It is also evident when using only arousal features that although the longest arousal is most important, the remaining arousal features still contribute a 10% increase in sensitivity. When using only a single feature for classification, depression is most important with a balanced accuracy of

TABLE III

PERFORMANCE FOR MACHINE LEARNING MODELS IN CLASSIFYING INSOMNIA PATIENTS VS. SUBJECTS WITHOUT INSOMNIA.

Model	Acc (%)	Sen (%)	Spec (%)
Support Vector Machine	71.3	65.1	77.5
Logistic Regression	70.9	67.5	74.3
Neural Network	70.0	64.7	75.4

TABLE IV

TOP 15 FEATURES FOR CLASSIFYING SUBJECTS WITH AND WITHOUT INSOMNIA BASED ON SHAPLEY ADDITIVE EXPLANATIONS VALUES.

Features	No insomnia	Insomnia	Coef	p-val
Depression (0, 1)	0.15 ± 0.36	0.49 ± 0.50	1.44	10^{-30}
Age (years)	53.7 ± 18.5	48.0 ± 15.0	-2.77	10^{-16}
Sex (0: F, 1: M)	0.55 ± 0.50	0.38 ± 0.49	-0.58	10^{-9}
ESS score (0-24)	9.02 ± 5.49	10.7 ± 6.05	0.93	10^{-6}
N3 power (μV^2/Hz)	48.5 ± 19.7	44.8 ± 20.7	-3.85	10^{-4}
Longest arousal (min)	40.5 ± 30.9	48.5 ± 41.9	1.96	0.01
Theta in N3 (%)	6.74 ± 2.76	6.36 ± 3.02	2.49	0.02
Sigma in N3 (%)	1.19 ± 0.69	1.42 ± 0.93	3.66	0.06
Sigma in N1 (%)	6.34 ± 2.33	6.89 ± 2.84	2.56	0.07
Theta in N1 (%)	20.1 ± 5.96	19.2 ± 6.03	2.76	0.11
Alpha in REM (%)	7.91 ± 3.57	7.09 ± 4.02	1.74	0.12
Sigma in REM (%)	4.29 ± 2.12	4.11 ± 2.56	-1.86	0.16
Theta in REM (%)	15.9 ± 5.81	13.8 ± 6.76	-1.68	0.19
Theta in N2 (%)	13.0 ± 3.94	12.0 ± 4.24	-1.51	0.27
Sigma in N2 (%)	4.56 ± 2.03	5.41 ± 2.75	2.32	0.99

TABLE V

PERFORMANCE USING FEATURE SUBSETS TO CLASSIFY SUBJECTS WITH AND WITHOUT INSOMNIA USING LOGISTIC REGRESSION.

Features	Acc (%)	Sen (%)	Spec (%)
Top 15 features	68.8	64.4	73.3
Depression, age, sex, ESS	66.9	59.5	74.4
Depression	66.9	49.1	84.7
All spectral features	61.3	56.0	66.6
Theta, sigma, abs power N3	59.8	52.8	66.8
All arousal duration features	58.0	51.3	64.8
Longest arousal	54.5	41.8	67.3

∼67%. However, the sensitivity is only ∼50%, which means that half of the insomnia patients are misclassified because they lack a depression diagnosis. Insomnia is a common and important symptom of depression, and it is possible that the insomnia patients reflect primary depression instead of primary insomnia, given the strong association. Further adding age, sex, and ESS score to the model increases the sensitivity by 10%, correctly classifying some insomnia patients that are wrongly classified based on depression alone. The benefit of these features is that no overnight PSG is required, and the information is gathered in few minutes.

A limitation of this study is that only severe insomnia patients and controls were chosen based on the criteria presented in Section II, so this study is not representative of the whole population and patients with partial phenotypes of insomnia with e.g., only sleep initiation or sleep maintenance problems. Furthermore, important variables with strong association to insomnia, such as anxiety or taking medications, were missing. In the sleep-related questionnaire, a question was related to anxiety about falling asleep, but was left out of the analysis, because the answer is subjective and highly correlated with insomnia. Future work should investigate application of deep learning on PSG signals for automatic feature extraction, which could lead to new biomarkers.

This study demonstrates that machine learning of PSG can identify features of insomnia and may be used as an objective screening method. Accuracy may be increased with multiple nights of recording with home devices, more consistent with the definition of symptoms occurring ≥3 times per week.

REFERENCES

[1] M. Sateia, D. Buysse, A. Krystal, D. Neubauer, and J. Heald, "Clinical practice guideline for the pharmacologic treatment of chronic insomnia in adults: an American Academy of Sleep Medicine clinical practice guideline", *J Clin Sleep Med*, vol. 13, no. 2, pp. 307-349, 2017.

[2] M. Sateia, "International classification of sleep disorders", *Chest*, vol. 146, no. 5, pp. 1387-1394, 2014.

[3] M. Daley, C. Morin, M. LeBlanc, J. Grégoire, and J. Savard, "The economic burden of insomnia: direct and indirect costs for individuals with insomnia syndrome, insomnia symptoms, and good sleepers", *Sleep*, vol. 32, no. 1, pp. 55-64, 2009.

[4] S. Schutte-Rodin, L. Broch, D. Buysse, C. Dorsey, and M. Sateia, "Clinical guideline for the evaluation and management of chronic insomnia in adults", *J Clin Sleep Med*, vol. 4, no. 5, pp. 487-504, 2008.

[5] C. Baglioni, G. Battagliese, B. Feige, K. Spiegelhalder, C. Nissen, U. Voderholzer, et al., "Insomnia as a predictor of depression: a meta-analytic evaluation of longitudinal epidemiological studies", *J Affect Disord*, vol. 135, n. 1-3, pp. 10-19, 2011.

[6] S. Javaheri and S. Redline, "Insomnia and risk of cardiovascular disease", *Chest*, vol. 152, no. 2, pp. 435-444, 2017.

[7] J. Foldager, P. Peppard, E. Hagen, K. Stone, D. Evans, G. Tranah, et al., "Genetic risk for subjective reports of insomnia associates only weakly with polygraphic measures of insomnia in 2,770 adults", *J Clin Sleep Med*, vol. 18, no. 1, pp. 21-29, 2022.

[8] M. Shahin, B. Ahmed, S. Hamida, F. Mulaffer, M. Glos, and T. Penzel, "Deep learning and insomnia: assisting clinicians with their diagnosis", *IEEE J Biomed Health Inform*, vol. 21, no. 6, pp. 1546-1553, 2017.

[9] W. Zhao, E. Van Someren, C. Li, X. Chen, W. Gui, Y. Tian, et al., "EEG spectral analysis in insomnia disorder: A systematic review and meta-analysis", *Sleep Med Rev*, vol. 59, pp. 101457, 2021.

[10] S. Kim, "Recent trends of artificial intelligence and machine learning for insomnia research", *Chronobiol Med*, vol. 3, no. 1, pp. 16-19, 2021.

[11] M. Johns, "A new method for measuring daytime sleepiness: the Epworth sleepiness scale", *Sleep*, vol. 14, no. 6, pp. 540-545, 1991.

[12] M. Perslev, S. Darkner, L. Kempfner, M. Nikolic, P. Jennum, and C. Igel, "U-Sleep: resilient high-frequency sleep staging", *NPJ Digital Medicine*, vol. 4, no. 1, pp. 1-12, 2021.

[13] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning", *Nature*, vol. 521, no. 7553, pp. 436-444, 2015.

[14] A. Brink-Kjaer, A. Olesen, P. Peppard, K. Stone, P. Jennum, E. Mignot, et al., "Automatic detection of cortical arousals in sleep and their contribution to daytime sleepiness", *Clinical Neurophysiology*, vol. 131, no. 6, pp. 1187-1203, 2020.

[15] R. Berry, C. Albertario, S. Harding, R. Lloyd, D. Plante, S. Qyan, et al., "The AASM Manual for the scoring of sleep and associated events: rules, terminology and technical specifications. Version 2", *American Academy of Sleep Medicine*, Darien, IL, 2018.

[16] R. Vallat and M. Walker, "An open-source, high-performance tool for automated sleep staging", *Elife*, vol. 10, pp. e70092, 2021.

[17] M. Morrell, L. Finn, H. Kim, P. Peppard, M. Badr, and T. Young, "Sleep fragmentation, awake blood pressure, and sleep-disordered breathing in a population-based study", *Am J Respir Crit Care Med*, vol. 162, no. 6, pp. 2091-2096, 2000.

[18] P. Welch, "The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms", *IEEE Trans Audio Electroacoust*, vol. 15, no. 2, pp. 70-73, 1967.

[19] M. Prerau, R. Brown, M. Bianchi, J. Ellenbogen, and P. Purdon, "Sleep neurophysiological dynamics through the lens of multitaper spectral analysis", *Physiology*, vol. 32, no. 1, pp. 60-92, 2017.

[20] M. Schmidt, N. Le Roux, and F. Bach, "Minimizing finite sums with the stochastic average gradient", *Mathematical Programming*, vol. 162, no. 1, pp. 83-112, 2017.

[21] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *arXiv preprint*, arXiv:1502.03167 [cs.LG], 2015.

[22] D. Kingma and J. Ba, "Adam: A method for stochastic optimization", *arXiv preprint*, arXiv:1412.6980 [cs.LG], 2014.

[23] L. Wang, "Support vector machines: theory and applications", *Springer Science & Business Media*, vol. 177, 2005.

[24] S. Lundberg and S. Lee, "A unified approach to interpreting model predictions", *Adv Neural Inf Process Syst*, vol. 30, 2017.