Parkinson’s Disease Recognition based on Sleep Metrics from Actigraphy and Sleep Diaries

M. Mikulec¹, J. Mekyska¹, and Z. Galáž¹

¹Department of Telecommunications
Faculty of Electrical Engineering and Communication
Brno University of Technology
Brno, Czech Republic

E-mail: xmikul56@vutbr.cz, mekyska@vut.cz, galaz@vut.cz

Abstract—Parkinson’s disease is accompanied by sleep disorders in most cases. Therefore, patients with Parkinson’s disease could be identified according to the proper sleep metrics. The study aims to train a classifier and identify proper sleep metrics, that could distinguish patients with Parkinson’s disease from subjects in the control group based on data from actigraphy and sleep diaries. The study sample consisted of 23 patients with probable Parkinson’s disease and 71 control subjects resulting in 654 nights of actigraphy and sleep diary data, with 26 unique features per night. XGBoost classifier was trained to distinguish the groups, scoring 80% accuracy and 52% F₁ on test data. Actigraphy-based parameters targeted at wake analysis during sleep were marked as most important. The study provided a classifier and obtained the most important parameters to identify patients with Parkinson’s disease based on actigraphy and sleep diary data.

Keywords—actigraphy, machine learning, Parkinson’s disease, SHAP values, sleep diary, sleep disorders, XGBoost

1. INTRODUCTION

Changes in sleep are thought to be a common process associated with aging. Secondary, it could be further compromised by medical or psychiatric diseases (e.g., dementia, depression) [1]. Idiopathic Parkinson’s disease (PD) is associated with sleep disorders in 74-98% of patients [2]. Patients with mild cognitive impairment (MCI) are at risk of sleep disorders in 65% of cases [3]. The spectrum of sleep disorders is wide and includes, for example, insomnia, sleep-disordered breathing, REM behavior disorder, or restless legs syndrome [3]. Generally, there is over 60% probability that patient with dementia or mild cognitive impairment would have some sleep disorder, and therefore could be identified according to sleep parameters [3].

Although this study used data from actigraphy and sleep diary, polysomnography (PSG) is considered a gold standard measure of sleep. PSG is a really complex examination, it utilizes various sources of data, for example, electroencephalogram (EEG), electrooculogram (EOG), electrocardiogram (EKG), pulse oximetry, and pulse and respiratory effort [1, 4]. PSG is possible to distinguish phases of sleep and create a hypnogram, its results are precise and reliable, but the costs of the physical examination are significant and the examination is usually limited to one night per patient. In contrast to the fact that the metrics are measured precisely, they can be changed by an unusual laboratory environment and other unusual external influences.

Actigraphy (ACG) is a non-invasive monitoring method. An actigraph unit is a bracelet that is worn mainly on a non-dominant wrist [5]. It records the occurrence and the degree of movement of the limb, as well as the temperature, blood pressure, light intensity, etc. Compared to PSG, ACG is used in decentralized clinical studies with lower budget and it is generally targeted on home usage. ACG is eligible to collect a larger amount of data, since the unit can continuously record up to one month of data ¹. On the contrary, ACG needs specialized software to identify sleep/wake intervals and due to various manufacturers and proprietary software, the results can vary significantly. It leads to initiatives to create and standardize an open access method [6, 7].

¹GENEActiv original, actigraph that was used in this study, is capable to record up to one month of continuous data depends on measurement frequency setting.
Sleep diary (SD) is a widely used method for estimating sleep parameters without specialized equipment. The percentage agreement between subjective sleep diary and polysomnographic data is acceptable according to Rogers et al. (kappa = .87, Sensitivity = 92.3% and Specificity = 95.6%) [8]. Low cost and straightforward usage are the main advantages of sleep diaries; however, the results depend heavily on the subject’s ability to fill the diary precisely. This study was aimed at patients with a probability of PD; therefore, diary entries could produce misleading results.

Sleep parameters obtained through PS, ACG or SD might be able to distinguish subjects with sleep disorders and identify diseases such as PD. However, this field of science has many challenges and knowledge gaps. For example, Maglione et al. stated that PSG and ACG can vary dramatically in crucial parameters, and there is a significant degree of variability in precision between individual patients. It recommends ACG for the measurement of mean total sleep time (TST), sleep efficiency (SE) and wake after sleep onset (WASO) [9]. Stavitsky et al. suggest ACG as an appropriate method to measure sleep quality in PD. Furthermore, it states that SD correlates with actigraphy-derived estimates of sleep quality for patients with PD, in contrary there were no variables that were significantly related to any of the SD measures in the control group [10].

Our goal is to train a classifier that would be able to distinguish nights of patients with Parkinson’s disease from control participants of the study, especially:

1. select proper algorithm for actigraphy sleep/wake recognition,
2. select and calculate sleep features,
3. train binary classifier to distinguish between patients with Parkinson’s disease and control participants,
4. analyze and validate trained classifier and its selection of characteristics.

2. MATERIALS AND METHODS

2.1. Study sample
Patients were recruited with the help of St. Anne’s University Hospital Brno. There were 94 subjects in total, 23 with probable Parkinson’s disease (24.5%) and 71 control subjects (75.5%). The classification of the subjects was performed by neurology specialists from St. Anne’s 1st Neurology clinic. Demographically, the participants were 64 ± 14 years old, there were 41 men and 53 women.

2.2. Dataset
Final dataset contains ACG and SD data for 6-7 nights from each participant, resulting in 654 nights in total with 160 entries from patients with PD and 494 entries from control subjects (CG).

Sleep/wake recognition was performed using two algorithms on actigraphy data, with first attempt to use the open access method of Hees et al. [6]. However, the metrics calculated using this method did not produce results in the final classification task and the correlation with the SD data was low. In the second final attempt, the XGBoost classifier was used based on [7]. Additionally, the training data set for sleep / wake detection was balanced using SMOTE: Synthetic Minority Over-sampling Technique [12]. The classifier scored 80% accuracy, 86% F₁ and a Mathews correlation coefficient of 0.5 on test set (40% of the Newcastle polysomnography and accelerometer dataset) [13]. In validation against SD data, the algorithm reached 61% accuracy, 73.5% F₁ and a Mathews correlation coefficient of 0.05. 13 features per night were extracted for the sleep / wake classification of actigraphy data and the sleep diaries, resulting in 26 features per night in total. The features are described in Tab. I.

2.3. Data preparation
Dataset was divided into training and testing data. 60% of the data were randomly selected as training data using shuffle split, where 100 entries (25.5%) were marked as positive (probable Parkinson’s disease) and 292 entries (74.5%) were marked as negative (control group). Training data were balanced using SMOTE [12]. The final training dataset consists of 584 entries with 26 features for each entry. The remaining 40% of the original dataset were kept as test data. Those data consisted of 60 positive entries (22.9%) and 202 negative (77.1%). Testing data remained imbalanced to avoid bias from SMOTE synthesis and to reflect real parameters of the dataset.

2.4. Machine learning method
Supervised machine learning method using the XGBoost tree boost classifier was used for the binary classification task of PD patients’s nights and nights of control participants. Labels based on neurology specialist diagnosis were provided for supervised machine learning. Optimal hyper-parameters for
the XGBoost classifier were obtained using hyper-parameters tuning method. Training validation was performed using the cross-validation method. To evaluate the ability of the models to classify PD vs. CG correctly, we employed accuracy, sensitivity, specificity, $F_1$ score, and the Matthews correlation coefficient (MCC), where the MCC provides a trade-off between sensitivity and specificity.

3. RESULTS

3.1. Optimal hyper-parameters

Those parameters were selected as optimal during hyper-parameters tuning phase: the number of gradient-boosted trees = 1000, boosting learning rate = 0.2, minimum loss reduction = 0.5, maximum tree depth for base learners = 8, subsample ratio of columns for each level = 0.6, subsample ratio of columns when constructing each tree = 0.4. The training process was stopped after 36 rounds to avoid overfitting.

3.2. Classification results

Results of classification during training cross-validation phase and model-trained results on test data are shown in Tab. II. The classifier has good results on both train and test data. On the contrary, MCC of 0.4 and Sensitivity of 49% underscore that the classifier can easily mistake patient with Parkinson’s disease as control group patient. However, the Specificity of 89% states that the subject of the control group should not probably be mistaken for a Parkinson’s disease patient.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time in bed (TIB)</td>
<td>Time spent in bed with attempt to sleep</td>
</tr>
<tr>
<td>Sleep onset latency (SOL)</td>
<td>Time before sleep onset</td>
</tr>
<tr>
<td>Sleep onset latency - norm</td>
<td>SOL classified according to NSF norm [11]</td>
</tr>
<tr>
<td>Wake after sleep onset (WASO)</td>
<td>Sum of wake ups between sleep onset and sleep offset</td>
</tr>
<tr>
<td>Wake after sleep onset - norm</td>
<td>WASO classified according to NSF norm [11]</td>
</tr>
<tr>
<td>Wake after sleep offset (WASF)</td>
<td>Time between sleep offset and get up from bed</td>
</tr>
<tr>
<td>Total sleep time (TST)</td>
<td>Total time of sleep during night, $TST = TIB - (SOL + WASO + WASF)$</td>
</tr>
<tr>
<td>Wake bouts (WB)</td>
<td>Number of awakenings during night</td>
</tr>
<tr>
<td>Awakenings &gt; 5 minutes (WB5+)</td>
<td>Number of awakenings longer than 5 minutes</td>
</tr>
<tr>
<td>Awakenings &gt; 5 minutes - norm</td>
<td>WB5+ according to NSF norm [11]</td>
</tr>
<tr>
<td>Sleep efficiency (SE)</td>
<td>Ratio of TST and TIB, $SE = (TST ÷ TIB) · 100$</td>
</tr>
<tr>
<td>Sleep efficiency - norm</td>
<td>SE according to NSF norm [11]</td>
</tr>
<tr>
<td>Sleep fragmentation (SF)</td>
<td>Ratio of WB and TST in hours, $SF = WB ÷ (TST ÷ 3600)$</td>
</tr>
</tbody>
</table>

Table II: Classification results of Parkinson’s disease recognition

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Accuracy [%]</strong></td>
<td>88</td>
<td>80</td>
</tr>
<tr>
<td><strong>Sensitivity [%]</strong></td>
<td>89</td>
<td>49</td>
</tr>
<tr>
<td><strong>Specificity [%]</strong></td>
<td>87</td>
<td>89</td>
</tr>
<tr>
<td><strong>F$_1$ [%]</strong></td>
<td>91</td>
<td>52</td>
</tr>
<tr>
<td><strong>MCC</strong></td>
<td>0.76</td>
<td>0.40</td>
</tr>
</tbody>
</table>

3.3. Feature importance and model interpretation

Feature importance of top 10 features is visualized in Fig. 1. The features marked (A) are obtained from actigraphy, and the features of the sleep diary are marked (D). The features extracted from the
actigraphy data that record the number and duration of awakenings between the sleep onset and offset are the most important according to the model, namely wake bouts and sleep efficiency according to the norm. Similarly, the most important SD feature is the sleep efficiency according to the norm[11].

To explain the output of the model even further, the game-theoretic approach called SHAP was used. SHAP values encode the importance that a model gives to each feature of each data point in contributing to the model’s output. The results are visualized on Fig. 2. The most important features according to SHAP are also based on actigraphy, while WB remained the most important feature at all².

---

4. DISCUSSION

Training dataset for XGBoost classifier used for the detection of sleep/wake was balanced using SMOTE. However, the results were worse than suggested using the model trained on an imbalanced dataset. Therefore, the sleep parameters calculated based on actigraphy sleep/wake recognition could be affected. The experiment should be repeated with the suggested model setting.

The study could be severely affected by the study sample and errors in its diagnosis. The diagnosis determination was made at the early stage of project NU20-04-00294 Diagnostics of Lewy body diseases in the prodromal stage based on multimodal data analysis. Due to some examinations that were incomplete and the methodology to determine the diagnosis was uncertain, the labels provided to the supervised learning algorithm could be wrong or misleading. Therefore, the experiment should be repeated after the final results of the NU20-04-00294 project.

Inaccuracies could be caused by the selection of subjects in the control group. There were patients with mild cognitive impairment diagnosis and patients with Lewy body dementia in the prodromal stage. These diseases are associated with sleep disorders in many cases, and MCI could even be an indicator of Parkinson’s disease. Therefore, to make the classification much more accurate, the subjects in the control group should be healthy controls. On the contrary, the study sample perfectly matches the target group for the real-life scenario of Parkinson’s disease recognition, consisting of elderly adults with sleep problems.

The model has a relatively low sensitivity. But the results are made night-by-night and not subject-by-subject. For example, RBD should not be present on all nights in a patient with probable PD. Therefore, to classify subjects, a threshold for PD-classified nights should be established. Reaching this threshold

---

2To replicate the results of the study and use the classifier, visit https://github.com/BDALab/sleep-analysis-system for source codes, or https://nicelife.utko.fee.vutbr.cz to see a running instance of the sleep analysis system, which contains an implemented classifier. The dataset must be requested via email due to GDPR.

---

Figure 1: Feature importance (PD classification)  Figure 2: SHAP values (features impact, PD classification)
will mark all nights of this patient as PD-classified. In this way, the sensitivity and overall accuracy could be dramatically increased.

5. CONCLUSION

This study proposes a new method for the detection of Parkinson’s disease based on actigraphy and sleep diary data. The method achieved 80% accuracy and a Mathews correlation coefficient equal to 0.4 (49% sensitivity, 89% specificity) when compared to the diagnosis of neurological experts. Although the method has a low sensitivity for clinical use, it can detect patients with a further stage of Parkinson’s disease.

ACKNOWLEDGMENT

This work was supported by grant no. NU20-04-00294 (Diagnostics of Lewy body diseases in the prodromal stage based on multimodal data analysis) of the Czech Ministry of Health and by the European Regional Development Fund no. CE1581 (Interreg niCE-life – Development of an integrated concept for the deployment of innovative technologies and services allowing independent living of frail elderly) of the European Union.

REFERENCES