Sono ao Volante - Improvements to detection and prediction of sleep stages

Francisco Friande
Faculdade de Engenharia, Universidade do Porto
Porto, Portugal
up201508213@fe.up.pt

João Bandeira Fidalgo
Faculdade de Engenharia, Universidade do Porto
Porto, Portugal
up201605237@fe.up.pt

***Abstract*—Road collisions lead to over 1.2 million deaths per year and an even higher number of non-fatal, grave injuries. Sleepiness represents 20% of these accidents globally, making this one of the major causes of road deaths. The most susceptible drivers to this problem are professional drivers who spend many hours driving and tend to accumulate more sleep debt. So it is important to have a method that predicts when drowsiness situations will occur. Most of the available approaches focus on detecting drowsiness while driving. So it is important to develop a method to detect the preliminary stages of sleep when they start occurring and also to predict them because it can be too late to prevent an accident when the detection tools detect the driver's immediate drowsiness state. In this paper, we describe a machine learning model that has been previously developed and try to improve the results obtained. In order to do that, we tested Recurrent Neural Networks and two different classifiers, XGBoost and LightGBM. We implemented hyperparameter tuning in all classifiers used. The results of previously used classifiers increased in all cases. LightGBM and XGBoost showed good results and proven that they can also be used. Neural Networks did not show high results, but this needs further investigation because we addressed simple implementations, and the dataset is limited. This work can be valuable for multiple applications, connected with detecting sleep disorders and integrated into sleep detection and prediction while driving, as well as predicting sleep stages for later detection of sleep diseases.**

***Index Terms*—Sleep stages, Machine learning, Drowsiness prediction, Biometric data, Hyperparameter tuning**

I. Introduction

A. Context

Driving is a demanding task, in which attention is crucial to make appropriate decisions at the right time [2]. Road collisions lead to over 1.2 million deaths per year, and an even higher number of non-fatal, grave injuries [3] and sleepiness accounts for up to 20% of all the accidents worldwide and makes it one of the major causes of road deaths [4]. When drivers decide to drive in drowsy conditions, they increase the risk of crashes significantly. Drowsiness effects can appear at any time of the day, but the chances increase when driving at night, up to a threefold [9], so this is even more critical in drivers that drive at night.

All drivers are susceptible to this problem, but especially professional drivers, like bus and truck drivers, this happens because they tend to accumulate a great deal of sleep debt, also spending long hours on the job [5]. Studies conducted with these workers suggest that more than 50% of them have already fallen asleep while driving [6].

B. Motivation

The number of accidents caused by sleepiness represents a high percentage of the total number of accidents, which means that it is important to develop methods to detect drowsiness while drivers are driving but also to predict when those situations can occur even before drivers start to fall asleep. Most of the available approaches focus on detecting drowsiness while driving [21]. So it is important to develop a method to detect the firsts stages of sleep when they start occurring because it can be too late to prevent an accident when the detection tools detect the driver's drowsiness state.

II. State of the Art

A. Background

Sleep is a major part of human life, taking about one third of the time of the whole lifespan of a person [8]. It is the quintessential part of resting, crucial for our normal cognitive functioning and for our survival [9].

Sleep restriction or deprivation is associated with a number of neurobehavioral deficits that are cumulative. Attention and memory problems, mood swings, and compromised cognitive abilities are among the many dangers [22].

Microsleeps and sleep attacks can emerge in individuals who are sleep deprived [23]. These drowsy-driving and unconscious blank stares episodes involve cognitive impairment, both of omission (lack of response to stimulus) and commission (response to inexistent or non-related stimulus) [24]. Persistent neuropsychological deficits and vigilance impairment are at the core of the problem and are caused when the homeostatic mechanisms related to preliminary sleep phases interfere with the vigilance state. This means concentration and attention are compromised because sleep mechanisms are continually triggered. Microsleeps can last from as little as 0.5 seconds up to 10 or more seconds, presenting a significant risk to a driver [22].

1) Sleep pattern: Human sleep divides into two high-level states: non-rapid eye movement (NREM) and rapid eye movement (REM). Sleep cycles, which usually take about 90 minutes, involve both these states. A typical good night of sleep usually has 4 to 5 cycles, but the states vary in length in

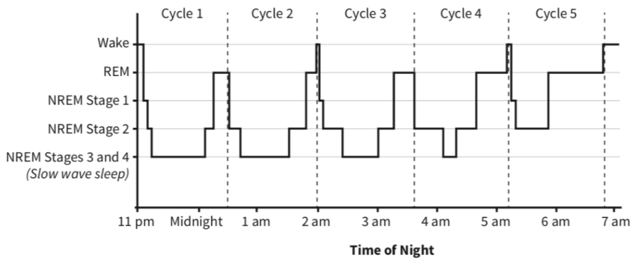

Fig. 1. Sleep Cycles Ref: [26]

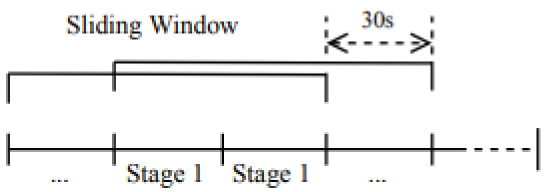

Fig. 2. Representation of the sliding window from [1]

every cycle [25]. The sleep cycle begins with the REM state and then NREM state, which divides into three stages:

- First stage - where the sleep state begins and most of the cognitive ativities decrease.
- Second stage - where it is lighter
- Third stage - corresponds to the deeper stage of sleep

The figure 1 represents the sleep cicle and the stages that a person goes through during the night.

Overnight polysomnography is the most indicated method for the diagnosis of various sleep disorders [27]. This method includes electroencephalogram (EEG), electro-oculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), oximetry and breathing.

B. Previous work

This work is an extension of the Sono ao Volante – Machine Learning para Previsão e Deteção de Sonolência Masters Thesis [1], involved in the project "Sono ao Volante 2.0" with the cooperation between FEUP, Instituto Politécnico do Cávado e do Ave, and the company Optimizer e o Instituto de Sono. This Thesis performs the pre-processing and first measurements that we used as an object of comparison with our results. It uses a public dataset that provides data from 11 subjects. The pre-processing of the dataset synchronizes the ECG signal with the polysomnography results, considering 30 seconds segments. It divides the signal into 1.5, 2.5, 3.5, and 4.5 minutes centered in each 30 seconds segment, and this means that it discards the first and last 30 seconds segments. Figure 2 represents this approach for the window segments.

After that, there is the preparation phase where the data is balanced using SVMSMOTE and Standard Scaler to reduce the features to the same scale. Following this, it divides the dataset into ten folds using a Stratified K-Folds cross-validator.

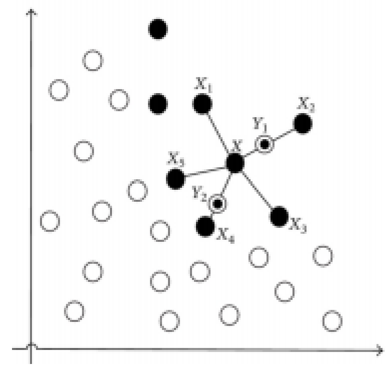

Fig. 3. Representation of SMOTE algorithm with k=5 [7]

It uses the mean of the results obtained in each fold to get the most approximate results of predicting an unseen dataset. There is two feature selection techniques, where the first one is using the features that had a correlation value greater than 0.9 to the objective, and the second technique uses seven features that presented the highest importance in Random Forest Classifier.

In the prediction phase, it uses Random Forest (RF), Support Vector Classifier (SVC), K-nearest neighbors (KNN), and Linear Discriminant Analysis (LDA) as classifiers, but with the default parameters for each one of these. The work enables the prediction of three or four phases of sleep. In the case of three phases, it predicts NREM, REM, and Wake, and in the case of four phases, it predicts Light, Deep, REM, and Wake.

III. Proposal

This work is an extension of [1], where the main objective is to improve the results obtained by implementing different classification methods and performing hyperparameter tuning [10]. To do this, we implemented methods that had not been tested in the previous work. We selected the state-of-the-art tree-based methods eXtreme Gradient Boosting (XGBoost) [11], and Light Gradient Boosting Machine (LightGBM) [12], as well as Recurrent Neural Networks (RNN) [17]. After defining which new methods to test, we created combinations of parameters for each classifier (except RNNs) to run in GridSearchCV to find the best parameters that influenced the results positively. We also defined sets of parameters for each of the methods tested in the previous work to try and improve the results.

XGBoost is a decision-tree-based ensemble Machine Learning algorithm. It is an efficient and scalable implementation of the Gradient Boosted Trees algorithm [13]. XGBoost has some desirable features such as data sparsity awareness and weighted quantile sketch for approximate learning, that, alongside regularization insurance, built-in cross-validation, and efficiency optimization, make it a big player in ML nowadays [11].

Like XGBoost, LightGBM is also a gradient boosting framework that uses tree-based learning algorithm. LightGBM distinguishes itself from others mainly due to the training speed improvements, good approximation ratio and accuracy of split point determination [14]. LightGBM is ill-advised in small-datasets circumstances because overfitting is a danger to be aware of [12], so we were hesitant about using it, but good results were acheived in the end.

Recurrent Neural Networks are powerful in modeling sequential data, such as time-series problems [18], like the one we have in hands. Particularly, Long Short-Term Memory (LSTM-RNN) networks are widely used in an extensive set of problems, being time-series the most common type [19]. LSTM networks generalize well and, due to their forget gate in each neuron, have an inherent ability to bridge long time lags [20].

IV. Results

This section gives an overview of what was done in terms of combinations of parameters tested, the best results for each classifier, and the results obtained by the previous work.

We used the same measurement metrics as [1] to compare and evaluate the results, but we only used 4.5 minutes signal divisions.

A. Running for XGBoost

For XGBoost, we used five sets of combinations and tested the results for each of those combinations. Table I represents the results obtained by each set of parameters' combinations. We can see that the best results are with default parameters, first combination, and fifth combination. The variation of results between the best ones and the others is caused by the boosting method chosen.

TABLE I
Sets of combinations used and results for XGBoost

Parameter Grid	Accuracy	Kappa	Recall
Default Parameters	0.90 (0.01)	0.72 (0.04)	NREM = 0.92 (0.01) REM = 0.74 (0.06) WAKE = 0.84 (0.03)
First Combination	0.91 (0.01)	0.70 (0.03)	NREM = 0.90 (0.01) REM = 0.77 (0.06) WAKE = 0.87 (0.02)
Second Combination	0.87 (0.01)	0.66 (0.03)	NREM = 0.88 (0.01) REM = 0.79 (0.07) WAKE = 0.86 (0.03)
Third Combination	0.59 (0.08)	0.20 (0.07)	NREM = 0.61 (0.13) REM = 0.52 (0.19) WAKE = 0.51 (0.18)
Fourth Combination	0.59 (0.08)	0.20 (0.07)	NREM = 0.61 (0.13) REM = 0.52 (0.19) WAKE = 0.50 (0.18)
Fifth Combination	0.90 (0.01)	0.71 (0.03)	NREM = 0.91 (0.01) REM = 0.75 (0.07) WAKE = 0.85 (0.03)

B. Running for LightGBM

For LightGBM, we used four sets of combinations to tune using GridSearch, and tested in previous work implementation. Table II represents the results obtained by each set of parameters' combinations. Unlike XGBoost, LightGBM kept constant results, changing from 0.85 to 0.89 in terms of accuracy, except for the Fourth Combination that had lower results, 0.79, and it is caused by the different boosting type used.

TABLE II
Sets of combinations used and results for LightGBM

Parameter Grid	Accuracy	Kappa	Recall
Default Parameters	0.89 (0.01)	0.70 (0.03)	NREM = 0.90 (0.01) REM = 0.78 (0.07) WAKE = 0.87 (0.02)
First Combination	0.85 (0.01)	0.63 (0.02)	NREM = 0.85 (0.01) REM = 0.76 (0.06) WAKE = 0.86 (0.04)
Second Combination	0.88 (0.01)	0.68 (0.03)	NREM = 0.88 (0.01) REM = 0.78 (0.06) WAKE = 0.87 (0.02)
Third Combination	0.89 (0.01)	0.70 (0.03)	NREM = 0.90 (0.01) REM = 0.77 (0.06) WAKE = 0.87 (0.02)
Fourth Combination	0.79 (0.02)	0.53 (0.04)	NREM = 0.78 (0.02) REM = 0.78 (0.07) WAKE = 0.84 (0.04)

C. LSTM

For LSTM, the approach was slightly different since no oversampling operations were executed, and data was not scaled. The architecture chosen for the network was an Emsemble, fully connected RNN, containing an LSTM layer as the top layer, followed by two Dense layers (containing a Dropout layer in-between). K-fold cross-validation was not performed to maintain temporal coherence in the dataset.

The results showed consistency, having achieved 75% off accuracy and Cohen's Kappa coefficient of 0.2, with losses surrounding 1 to 1.5.

The Kappa measure is the most unsatisfactory metric since it shows a possible low level of generalization. We attribute the problem to the minimal dataset (keep in mind no oversampling was performed in this approach). The dataset is neither big enough nor varied enough for this kind of approach, but still, it showed great potential.

D. Overall results

In terms of results for the classifiers used in [1], we performed a hyperparameter tuning only with one set of combinations and obtained good results. Table III presents the best results obtained in each classifier. It is important to note that were tested both feature selection techniques used in [1], and all the results, except for LDA, were significantly better using seven features with the higher feature importance in Random Forest. For LDA, using the importance in Random Forest decreased the results, so we used features that had a correlation value higher than 0.90 with the objective. Comparing these results with the obtained in previous work:

- Common classifiers with the previous work had their results increased.

- Random Forest still has the best results.
- XGBoost show great results, in the case of XGBoost, it has the same accuracy, but worst kappa and recall's values has Random Forest
- Also, LightGBM had high accuracy, 0.90
- KNN presents the best results in REM recall, as all the other classifiers showed 0.77 or lower values, and it also has high accuracy, 0,90.

TABLE III
BEST RESULTS OBTAINED AFTER GRIDSEARCHCV

Method	Accuracy	Kappa	Recall
LighGBM	0.89 (0.01)	0.70 (0.03)	NREM = 0.90 (0.01) REM = 0.77 (0.06) WAKE = 0.87 (0.02)
XGBoost	0.91 (0.01)	0.73 (0.03)	NREM = 0.93 (0.01) REM = 0.76 (0.09) Wake = 0.81 (0.03)
RF	0.91 (0.01)	0.74 (0.02)	NREM = 0.93 (0.01) REM = 0.76 (0.07) Wake = 0.84 (0.03)
KNN	0.90 (0.01)	0.71 (0.02)	NREM = 0.92 (0.01) REM = 0.82 (0.06) Wake = 0.82 (0.03)
SVC	0.85 (0.01)	0.62 (0.02)	NREM = 0.86 (0.01) REM = 0.80 (0.06) Wake = 0.83 (0.03)
LDA	0.74 (0.02)	0.41 (0.03)	NREM = 0.76 (0.03) REM = 0.55 (0.11) Wake = 0.69 (0.03)

Tests performed involving four sleep stages showed worse results, but they improved comparing with the previous work. In this case LDA had 0.51 accuracy, RF resulted in 0.80 accuracy and good recall values, KNN obtained 0.78, and SVC had 0.71, which is a low value. The new classifiers introduced obtained 0.77 and 0.80 accuracies for LightGBM and XGBoost, respectively.

V. CONCLUSIONS AND FUTURE WORK

We think that we were able to reach the goals with this work, where we wanted to improve the accuracy, and other metrics, obtained in [1], but also test new methods. The overall results were higher for all methods previously tested, and the new Classifiers showed great values. We had some limitations in terms of time and dataset available, where the dataset we used is considerably small and cannot relate to everyone in the world. This work opens opportunities for further investigations regarding detection of sleep disorders, integration with smartwatches for warning drivers when they are driving, and the watch detects a drowsiness state.

REFERENCES

[1] C. C. C. Rodrigues, "Sono ao Volante - Machine Learning para Previsão e Deteção de Sonolência," Faculdade de Engenharia da Universidade do Porto, Porto, 2020.

[2] M. Q. Khan and S. Lee, "A comprehensive survey of driving monitoring and assistance systems," Sensors (Basel), vol. 19, no. 11, p. 2574, 2019.

[3] J. Rolison, S. Dorchin-Regev, S. Moutari, and A. Feeney, "What are the factors that contribute to road accidents? an assessment of law enforcement views, ordinary drivers' opinions, and road accident records. Accident Analysis Prevention," 115, 06 2018.

[4] J. Harvy, N. Thakor, A. Bezerianos, and J. Li, "Between-frequency topographical and dynamic high-order functional connectivity for driving drowsiness assessment," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 27, no. 3, pp. 358–367, 2019.

[5] A. T. McCartt, J. W. Rohrbaugh, M. C. Hammer, and S. Z. Fuller, "Factors associated with falling asleep at the wheel among long-distance truck drivers," Accid. Anal. Prev., vol. 32, no. 4, pp. 493–504, 2000.

[6] W. Wang and G. Wets. Computational intelligence for traffic and mobility. In Atlantis Computational Intelligence Systems, 2013.

[7] Y. Xie, Y. Liu, and Q. Fu, "Imbalanced data sets classification based on SVM for sand-dust storm warning," Discrete Dyn. Nat. Soc., vol. 2015, pp. 1–8, 2015.

[8] M. J. Aminoff, F. Boller, and D. F. Swaab, "We spend about one-third of our life either sleeping or attempting to do so," Handb. Clin. Neurol., vol. 98, p. vii, 2011.

[9] J. He, W. Choi, Y. Yang, J. Lu, X. Wu, and K. Peng, "Detection of driver drowsiness using wearable devices: A feasibility study of the proximity sensor," Appl. Ergon., vol. 65, pp. 473–480, 2017. [Accessed: 18-Jan-2021].

[10] J. Badvelu, "Hyperparameter tuning for Machine learning models - Towards Data Science," Towards Data Science, 31-Oct-2020. [Online]. Available: https://towardsdatascience.com/hyperparameter-tuning-for-machine-learning-models-1b80d783b946. [Accessed: 11-Jan-2021].

[11] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 2016.

[12] "Welcome to LightGBM's documentation! — LightGBM 3.1.1.99 documentation," Readthedocs.io. [Online]. Available: https://lightgbm.readthedocs.io/.

[13] J. Friedman. Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5):1189–1232, 2001.

[14] Ke, Guolin et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." NIPS (2017). Rezaei, Mohammad; Mohammadi, Hiwa ; Khazaie, Habibolah (2017), "EEG/EOG/EMG data from a cross sectional study on psychophysiological insomnia and normal sleep subjects", Mendeley Data, V4, doi: 10.17632/3hx58k232n.4

[15] "sklearn.preprocessing.StandardScaler scikit-learn 0.24.1 documentation," Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. [Accessed:10-Jan-2021].

[16] "sklearn.model_selection.StratifiedKFold scikit-learn 0.24.1 documentation," Scikit-learn.org. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html. [Accessed: 10-Jan-2021].

[17] "Recurrent neural networks (RNN) with keras," Tensorflow.org. [Online]. Available: https://www.tensorflow.org/guide/keras/rnn. [Accessed: 12-Jan-2021].

[18] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," arXiv [cs.LG], 2018.

[19] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM – a tutorial into long Short-Term Memory Recurrent Neural Networks," arXiv [cs.NE], 2019.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[21] C. Collet and O. Musicant, "Associating vehicles automation with drivers functional state assessment systems: A challenge for road safety in the future," Front. Hum. Neurosci., vol. 13, 2019.

[22] S. Banks and D. F. Dinges, "Behavioral and physiological consequences of sleep restriction," J. Clin. Sleep Med., vol. 3, no. 5, pp. 519–528, 2007.

[23] N. Goel, H. Rao, J. S. Durmer, and D. F. Dinges, "Neurocognitive consequences of sleep deprivation," Semin. Neurol., vol. 29, no. 4, pp. 320–339, 2009.

[24] L. N. Boyle, J. Tippin, A. Paul, and M. Rizzo, "Driver performance in the moments surrounding a microsleep," Transp. Res. Part F Traffic Psychol. Behav., vol. 11, no. 2, pp. 126–136, 2008.

[25] D. Aliakseyeu, J. Du, E. Zwartkruis-Pelgrim, and S. Subramanian, "Exploring interaction strategies in the context of sleep," in Human-Computer Interaction – INTERACT 2011, Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 19–36.

[26] Matthew Walker. Why we sleep: Unlocking the power of sleep and dreams. Simon and Schuster, 2017.

[27] P. Smallwood, D. K. Quinn, and T. A. Stern. 22 - patients with disordered sleep. In Theodore A. Stern, G. L. Fricchione, N. H. Cassem, M. S.

Jellinek, and J. F. Rosenbaum, editors, Massachusetts General Hospital Handbook of General Hospital Psychiatry (Sixth Edition), pp. 289 – 302. W.B. Saunders, Saint Louis, sixth ed., 2010.