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/EMBC40787.2023.10340630

Toward a Holistic Computational Representation for Sleep Quality and its Support for Explainability

Clauirton A. Siebra, Lais S. Amorim, Jonysberg P. Quintino, Andre L. M. Santos, Fabio Q. B. da Silva

Abstract— **Sleep quality is recognized as one of the main factors that affect human health. Thus, several studies have been encouraged to analyze features, such as stress level and female menopause, which are directly related to sleep quality. While these works rely mostly on reductionism as the philosophical framework, we approach this problem from a holist perspective, using a model with 10 features that could provide more reliable explanations for inductive conclusions. We demonstrate the principles of this hypothesis by analyzing the data regarding the day before a sleep episode of 1736 volunteers. This analysis shows, for example, the performance of each feature when they are jointly used along prediction tasks. Moreover, we evaluate the readability and accuracy of explanations, given as description logic sentences and based on a knowledge representation that considers the 10 features as elements that compose a sleep quality ontological definition.**

Clinical Relevance— **Patient symptoms, such as sleep issues, and the consequence of their constructors must be explained as evidence in diagnosis and clinical trials. This mainly helps the resolution of potential disagreements between systems and human experts.**

I. Introduction

Sleep quality is recognized as one of the main factors that affect human health [1], and several works analyze the relationship between poor sleep quality and diseases. For example, Magee et al. [2] review epidemiological evidence for an association between chronic sleep restriction and obesity in adults. Similarly, the study of Lao et al. [3] suggests that poor sleep quality increases the risk of coronary heart disease, while the meta-analysis of Hertenstein et al. [4] provided evidence that poor sleep quality (e.g., insomnia) increases the risk for psychopathologies. These and other works agree that sleep plays a critical role in brain function and systemic physiology, including metabolism, appetite regulation, and the functioning of immune, hormonal, and cardiovascular systems. Thus, as also emphasized in other papers [1], [5], the promotion of public strategies to improve sleep quality can improve the population's health, working as a method for preventing diseases. One of the limitations of current applications relies on their use of simple inductive models to analyze sleep quality and, mainly, the reasons causing such a quality. For example, most of these models are univariate and based on the relation between sleep and physical exercise [6]. In other words, they rely on the philosophical concept of reductionism that tries to isolate variables to understand the process of cause and effect. This approach brings advantages for scientific research, breaking down complex learning into simple stimulus-response links within the lab. On the other hand, reductionism has been accused of oversimplifying complex phenomena leading to loss of validity. Thus, it is not proper to explain real-life situations, such as the rationale for poor sleep quality of individuals.

This paper shifts toward a holistic over a reductionist approach as a way to better understand and model the complex phenomena that affect sleep quality [7]. Therefore, our contributions are composed of three parts: (1) identification of features that affect sleep quality and use of these features to compose a background knowledge of the sleep quality domain; (2) demonstration of the importance of holist representations using these features; and (3) use of this background knowledge for generating explanations. While the first part of our study creates an initial model that represents various features that affect sleep quality, the experiments described in the second and third parts are restricted by the available information in our sleep quality dataset.

II. Method

A. Selection of Features

We initially considered that the physiological sleep-wake system is regulated by the interplay of two major processes, one that promotes sleep and another that maintains wakefulness [10]. Then, we identified features that could explain modifications in such processes using secondary research found in two main databases: PubMed and ScienceDirect. A premise of our method was to only consider features that present enough quantitative and/or qualitative information to justify their influences on sleep quality. Then, we incrementally consolidated the results using an ontological description, which is a knowledge representation strategy based on description logics. In this ontology, features are concepts that establish relations between them. This model was then simplified according to the dataset introduced in the next section so we could employ instances of this dataset in our experiments.

B. Dataset

The *SleepHealth Public Researcher Portal* dataset [9] was used to create the inductive models. This dataset was chosen since it presents data associated with many of, but not all, ontological concepts identified in our research. Therefore, we only used the subset of ontological concepts present in this dataset. We are using data from 1736 volunteers. These data

* This paper follows the principles outlined in the Helsinki Declaration of 1975, as revised in 2000. Research supported by SABIA - Centro de Excelencia em Saude e Bem-Estar e Inteligencia Artificial", financed by Samsung Eletronica da Amazonia Ltda., under the auspices of the Brazilian Federal Law of Informatics no. 8.248/91

C. Siebra is with Federal University of Paraiba and University of Geneva, Switzerland (e-mail: clauirton.dealbuquerque@unige.ch).

L. Amorim and J. Quintino are with CIn/Samsung Project, Brazil.

A. Santos and F. da Silva are with Centro de Informatica, Universidade Federal de Pernambuco, Recife, Brazil.

are divided into two groups. The first group comprises data assessed only once over the assessment to define the user profile (age, gender, body mass index - BMI, menopause, cancer diagnostic, and diabetes diagnosis). The second group contains data daily assessed (step count, stress level, number of naps, and heart rate average). Categorical data were codified using integer values, while numeric values were normalized. The prediction feature is *sleep quality*, which presents four classes with the following number of instances: poor (295), fair (433), good (637), and very good (371). Therefore, the classes are unbalanced, and we can use the most frequent class approach to define our prediction accuracy baseline as 0.3669.

C. Setup and Experiments

The experiments aim to verify if a holistic model brings advantages for a prediction model in terms of accuracy and explainability compared to univariate models. Therefore, we first created univariate (using each of the ten features separately) and multivariate models comparing their accuracies. Three methods of different complexities (logistic regression, random forest, and neural networks) were used to demonstrate the method-independent aspect of this experiment. Secondly, we used the *Radom Forest model feature importanc*e to identify the most important features of this model, together with the *partial dependence plot* to verify the relationship of the features that could be controlled (e.g., steps) to influence sleep quality. As neural networks do not offer any direct interpretation of feature importance, unlike the coefficients available from a logistic regression model or the built-in feature importance for tree-based models like random forests, we employed the SHAP (SHapley additive exPlanations) method [10] to explain how each feature affects the model. Finally, as these methods are mostly focused on the importance of features, we also employed the DL-Learner framework [11] to generate explanations, in the form of description logics sentences, which can be evaluated regarding their readability and accuracy.

III. RESULTS

A. Sleep Quality Model

Table I summarizes the features identified in the literature as mainly associated with sleep promotion, a short description of their physiological actions, and related references.

TABLE I. FEATURES RELATED TO SLEEP PROMOTION

Feature	Physiological description	Refs.
Stress	Stress results in problems to falling asleep since it disrupts the hypothalamic pituitary adrenal (HPA) axis, which produces the stress hormone cortisol and helps coordinate the sleep cycle.	[12]
Cancer diagnosis	Learning the cancer diagnosis often causes sleeping problems due to stress (cortisol)	[13]
Heart rate (HR)	HR must be evaluated with others features since the reasons for high average HR can be good or bad for sleep.	[14]
Naps	They are associated with sleep inertia when longer than 30 minutes	[15]
Screen time*	Screen light stimulates the brain and suppresses melatonin production resulting in increased sleep latency.	[16]
Ambient air pollution*	Pollutants are a chronic source that can change levels of neurotransmitters and subsequent inflammatory responses.	[17]

* Features that are not used during the experiments

Table II summarizes the features mainly associated with wakefulness maintenance. For example, hydration does not affect sleep promotion. However, dehydration reduces the levels of essential aminoacids that are needed to produce melatonin. On the other hand, too much fluid intake can cause excess urination that may lead to sleep interruptions.

TABLE II. FEATURES RELATED TO WAKEFULNESS MAINTANANCE

Feature	Physiological description	Refs.
Menopause	Low levels of hormones (e.g., estrogen) lead to hot flashes, which are unpleasant sensations of extreme heat. Nighttime hot flashes are often paired with unexpected awakenings.	[18]
Diabetes	Nocturnal hypoglycemia can lead to sleep disruption due to low glucose levels at night.	[19]
Physical activity (Steps)	Increases the secretions of brain-derived neurotrophic factor and growth hormone, among others, which are important to sleep.	[20]
Hydration*	Reduces kidney function and affect the production of essential aminoacids for melatonin	[21]

* Features that are not used over the experiments

The following schema (Fig. 1) employs an ontology to organize these features. Additional features (age, gender, body mass index - BMI) complement such an ontology. The white boxes indicate features that are not present in our current dataset. Thus, this part of the ontology was not used during the experiments.

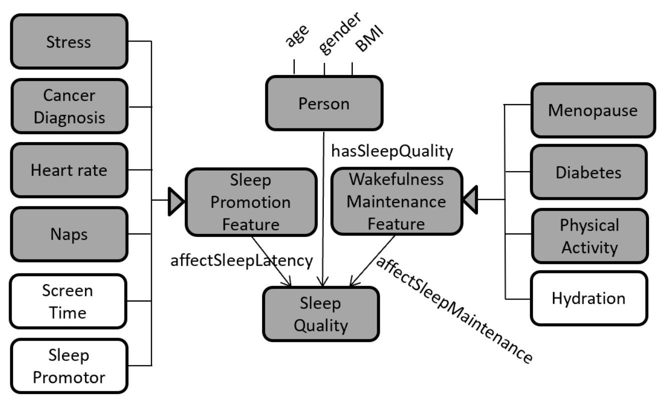

Figure 1. Ontology for the sleep quality domain.

B. Accuracy: univariate vs multivariate models

The following results (Table III) represent the means and standard deviation (*M*/*SD*) of 10 rounds. The top part shows the accuracy of each of the ten features using the logistic regression method, while the bottom part shows the results for the multivariate approach using three different learning strategies (logistic regression, random forest, and neural network). These results show that the accuracy of the multivariate model is better than any univariate model, independently of the learning strategy.

TABLE III. ACCURACY OF UNIVATE AND MULTIVARIATE MODELS

Feature	Logistic Reg.	Feature	Logistic Reg.
Age	0.438 / 0.012	Heart rate	0.367 / 0.016
Gender	0.356 / 0.014	Naps	0.368 / 0.017
BMI	0.357 / 0.015	Menopause	0.369 / 0.014
Stress	0.359 / 0.017	Diabetes	0.425 / 0.013
Cancer	0.418 / 0.014	Phys. activity	0.367 / 0.013
	Logistic Reg.	**Random For.**	**Neural Net.**
Multivariate	0.481 / 0.012	0.493 / 0.019	0.5 / 0.023

C. *Explaining the Features Importance*

The simplest way to explain an inductive model is to indicate the importance of the features for the prediction task. The results presented in Fig. 2 rely on the random forest model. They show that age and the presence of cancer and diabetes diseases are the main features for predicting sleep quality.

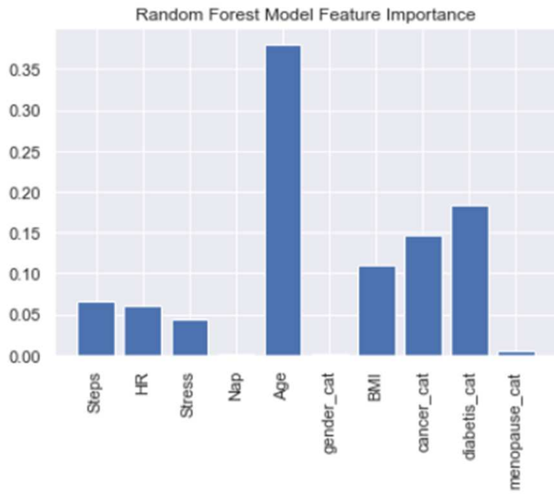

Figure 2. Importance of features for prediction.

This traditional form of minimally explaining the results can be augmented using partial dependence plots. Fig. 3 illustrates the results of such plots for Age, Steps, and HR features. We see that, as age increases, the probability of an individual presenting poor sleep quality also increases. Consider now Steps and HR features, which are easier to modify through simple interventions. While HR presents an expected directly proportional relation regarding sleep quality, the dependence plot of Steps presents an inverse behavior.

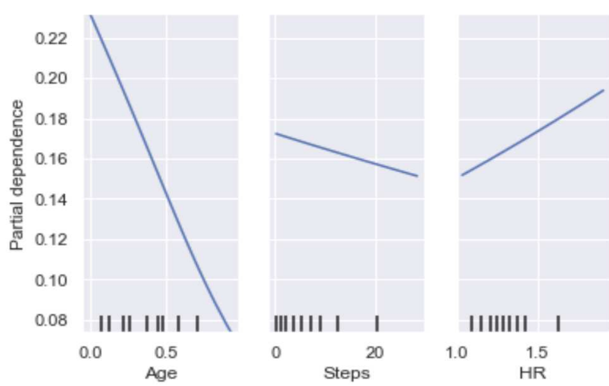

Figure 3. Age, Steps and HR partial dependence map.

Our third analysis regarding the influence of features on prediction used the SHAP method, which generates the Bee swarm plot in Fig. 4. This visualization allows observing the influence of each input point on the result. For example, most of the *age* feature input points with high values positively affect the results. Meanwhile, low values have negative influences. An interesting result is obtained with *Naps*. According to Fig. 1., this feature does not affect the prediction. However, Fig. 4 shows that points with a high value negatively affect the result. As there are very few samples of individuals with such high values of Naps in the dataset, its importance is not correctly indicated in the previous graph (Fig. 2).

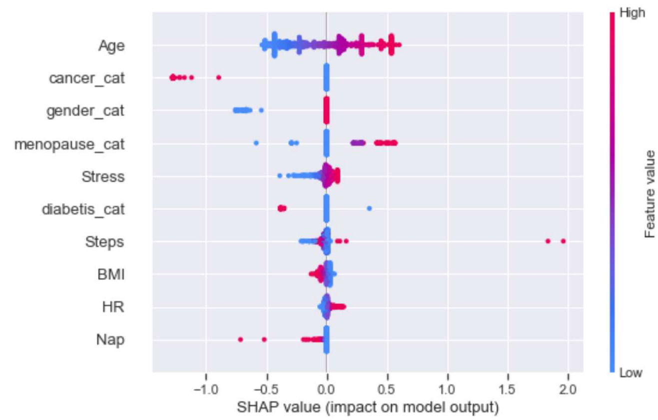

Figure 4. Bee swarm plot for the SHAP values of the features

D. *Explanations with Description Logics*

The previous strategies are important to have a general view of the importance of features for the prediction task. However, they do not support the idea of holistic analysis. Therefore, we modeled the previous ontology (Fig. 1) using the Protégé editor and imported the dataset content as instances of this ontology. After that, we configured the DL-Learner framework to employ this ontology as background knowledge, together with the *Class Expression Learning for Ontology Engineering* (CELOE) algorithm [22], to generate explanations. CELOE builds a search tree based on refinement operators, which define a mapping from a given input concept description to a set of derived or refined concept descriptions. Moreover, heuristics are also employed to find good candidates to look at.

CELOE is very sensitive regarding the level of noise. Moreover, it is well-recognized that medical datasets are often noisy and incomplete due to the difficulties in data collection and integration [23]. Thus, we conducted the experiments using values of 0%, 10%, and 20% as the percentage of noise allowed within the dataset. The execution time of the algorithm was 600 seconds since no significant improvements were returned after this point. Table IV shows the results (accuracy and F-measure) for these three noise values.

TABLE IV. DL-LEARNER RESUTS FOR DIFFERENT NOISE VALUES

Noise	**Best Accuracy (Acc)**	**Best F-measure**
0%	0.183	0.293
10%	0.469	0.366
20%	0.559	0.383

As expected, the experiments using 20% as noise value returned the best accuracy. DL-Learner returned the following explanations (E_i) for this case:

- E_i: (Acc = 53.11%): *Person and ((BMI some decimal[>= 26.70889377878]) or (BMI some decimal[<= 21.240991789653]))*
- E_2: (Acc = 53.17%): *Person and (((BMI some decimal[>= 14.808757415228]) and (BMI some decimal[<= 21.240991789653])) or (BMI some decimal[>= 26.70889377878]))*
- E_3: (Acc = 55.9%): *Person and ((BMI some decimal[>= 26.70889377878]) or (BMI some decimal[<= 21.240991789653])) and (age some decimal[<= 65.499999993])*

These results show that the readability of DL sentences is straightforward. For example, the explanation E_1 means that 53.11% of persons with poor sleep problems present BMI higher than 26.7 or lower than 21.24. Considering that the BMI values range from 18.5 to 24.9 indicate normal weight, we can observe a significant relation between weight and good sleep. E_2 only refines BMI ranges to obtain a slightly more accurate value. However, the best accuracy (55.9%) is obtained by E_3, which indicates that E_1 is only valid for persons younger than around 65 years.

The previous explanations show that the holistic analysis of age and BMI present the best accuracy. However, while such explanations assist in interpreting inference results, they are not useful for support, for example, interventions. This means BMI is a feature that is hard to modify in the short term, while age cannot be modified at all. A strategy is to configure the algorithm or change the ontology to avoid ontological elements we are not interested in using as part of the explanations. The next example avoids the age, BMI, and gender data properties:

- E_4: *(Acc = 52.13%): Person and (hasSleepQuality some (isAffectedBySleepLatency some (Cancer or ((averageHR some decimal[>= 74.32928322]) and (averageHR some decimal[<= 102.33092664])))))*

This explanation means that 52.13% of people with poor sleep quality have cancer or had an average heart rate (HR) between 74 and 102 beats per minute. As the normal average resting HR of individuals is around 60 to 100 beats per minute, this result may indicate that such individuals should have an average health rate at least higher than their resting HR. While cancer diagnosis is not a feature easy to modify in the short-term, the average HR could be improved, for example, with a physical exercise program. However, as a side-effect, avoiding features may reduce the explanation accuracy when features that provide gains to the inductive process are no longer used.

IV. Conclusions

This paper mainly shows that (1) multivariate/holistic models can create more accurate results (see Table III) for prediction systems, (2) the use of background knowledge augments the traditional ways to explain the inductive results since it supports the relation of various features in the same explanation, (3) explanations defined as description logic sentences are expressive while still provide good readability, and (4) accuracy and semantic value of explanations tend to be conflicting parameters that must be balanced over the process.

The main limitation of this work was the dataset used to create the models since it is very noisy. Moreover, DL-Learner only works for binary classification problems. Thus, we needed to cluster the sleep quality scores into two classes (poor and good). Our research directions intend to analyze graph-based approaches [24] since they are a more natural way to use domain knowledge defined through ontologies.

References

[1] Y. S. Bin, “Is sleep quality more important than sleep duration for public health?” in *Sleep*, 39(9), 1629-1630, 2016.

[2] C.A. Magee, D.C. Iverson, X.F. Huang, and P. Caputi, “A link between chronic sleep restriction and obesity: Methodological considerations,” *Public Health*, 122(12): 1373-1381, 2008.

[3] X. Q. Lao, *et al*., “Sleep quality, sleep duration, and the risk of coronary heart disease: a prospective cohort study with 60,586 adults”, *Journal of Clinical Sleep Medicine*, 14(1):109-117, 2018.

[4] E. Hertenstein, *et al*., “Insomnia as a predictor of mental disorders: A systematic review and meta-analysis,” *Sleep Med Rev*. 43:96-105, 2019.

[5] L. Hale, W. Troxel, and D. J. Buysse, “Sleep Health: An Opportunity for Public Health to Address Health Equity,” *Annu Rev Public Health*, 41:81-99, 2020.

[6] F. Wang, and S. Boros, “The effect of physical activity on sleep quality: a systematic review,” *European Journal of Physiotherapy*, 23(1):11-18, 2021.

[7] J. Sherwin, “ An approach towards holism in science and engineering,” *Journal of Systems Science and Systems Engineering*, 19(3): 285-305, 2010.

[8] Altevogt, B M., and Colten, H. R. eds. Sleep disorders and sleep deprivation: an unmet public health problem. National Academies Press (US) 2006.

[9] S. Deering, A. Pratap, C. Suver, et al, “Real-world longitudinal data collected from the SleepHealth mobile app study,”. *Sci Data* 7, 418, 2020.

[10] S. M. Lundberg, and S. I. Lee, “A unified approach to interpreting model predictions,” In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4765–4774, 2017.

[11] T. Procko, T. Elvira, O. Ochoa, and N. Del Rio. “An Exploration of Explainable Machine Learning Using Semantic Web Technology,” *IEEE 16th International Conference on Semantic Computing*, pp. 143-146, 2022.

[12] S. M. Bassett, S. B. Lupis, D. Gianferante, N. Rohleder, and J. M. Wolf , “Sleep quality but not sleep quantity effects on cortisol responses to acute psychosocial stress,” *Stress*, 18(6):638-644, 2015.

[13] T. Akman, T. Yavuzsen, Z. Sevgen, H. Ellidokuz, A. U. Yilmaz, “Evaluation of sleep disorders in cancer patients based on Pittsburgh Sleep Quality Index,” *European Journal of Cancer Care*, 24(4): 553-559, 2015.

[14] A. Sajjadieh et al, “The association of sleep duration and quality with heart rate variability and blood pressure,” *Tanaffos*, 19(2): 135, 2020.

[15] K. P. Jakubowski, M. H. Hall, L. Lee, K. A. Matthews, “Temporal relationships between napping and nocturnal sleep in healthy adolescents,” *Behavioral Sleep Medicine*, 15(4): 257-269, 2017.

[16] J. Falbe, K.K. Davison, R.L. Franckle, C. Ganter, S.L. Gortmaker, L. Smith, T. Land, E.M. Taveras, “Sleep duration, restfulness, and screens in the sleep environment,” *Pediatrics*, 135(2):e367-75, 2015.

[17] Y. J. Ju, J. E. Lee, D. W. Choi, K. T. Han, S. Y. Lee, "Association between perceived environmental pollution and poor sleep quality: results from nationwide general population sample of 162,797 people," *Sleep Med*, 80:236-243, 2021.

[18] F. C. Baker, L. Lampio, T. Saaresranta, P. Polo-Kantola, "Sleep and Sleep Disorders in the Menopausal Transition," *Sleep Med Clin*, 13(3):443-456, 2018.

[19] S. Barakat, M. Abujbara, R. Banimustafa, A. Batieha, and K. Ajlouni, “Sleep quality in patients with type 2 diabetes mellitus,” *Journal of clinical medicine research*, 11(4): 261, 2019.

[20] A. N. S. Bisson, S. A. Robinson, and M. E. Lachman, “Walk to a better night of sleep: testing the relationship between physical activity and sleep,” *Sleep health*, 5(5): 487-494, 2019.

[21] A. Y. Rosinger, A. M. Chang, O. M. Buxton, J. Li, S. Wu, and X. Gao, “Short sleep duration is associated with inadequate hydration: Cross-cultural evidence from US and Chinese adults,” *Sleep*, 42(2), zsy210, 2019.

[22] T. Procko, T. Elvira, O. Ochoa, and N. Del Rio, “An Exploration of Explainable Machine Learning Using Semantic Web Technology,” *IEEE 16th International Conference on Semantic Computing* (ICSC), pp. 143-146, 2022.

[23] J. Sáez, B. Krawczyk, M. Woźniak, “On the Influence of Class Noise in Medical Data Classification: Treatment Using Noise Filtering Methods,” *Applied Artificial Intelligence*, 30(6): 590-609, 2016.

[24] H. Yuan, H. Yu, S. Gui and S. Ji, "Explainability in Graph Neural Networks: A Taxonomic Survey," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022