Reproducible and accurate subject-wise sleep posture detection by detecting and removing turns

Javier Galvez-Goicuria
Brainguard SL. Madrid, Spain
Universidad Politécnica de Madrid
Madrid, Spain
javier.galvez@upm.es

Josue Pagan
Universidad Politécnica de Madrid
Madrid, Spain
Center for Computational Simulation
Madrid, Spain
j.pagan@upm.es

Lucia Perez
Brainguard SL
Madrid, Spain
lperez@brainguard.es

Julian Catalina-Gomez
Universidad Politécnica de Madrid
Madrid, Spain
julian.catalina.gomez@alumnos.upm.es

Jose M. Moya
Universidad Politécnica de Madrid
Madrid, Spain
Center for Computational Simulation
Madrid, Spain
jm.moya@upm.es

Jose L. Ayala
Universidad Complutense de Madrid
Madrid, Spain
Center for Computational Simulation
Madrid, Spain
jayala@ucm.es

2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS) | 978-1-6654-8356-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/COINS54846.2022.9854954

***Abstract*—Maintaining a good sleep hygiene is an important factor to avoid the symptoms of sleep disorders or worsen the symptoms of other diseases. Polysomnography is the study of sleep performed by professionals during a night at the hospital. On these studies they perform the diagnosis of diseases and patients are not monitored any more. A non-intrusive and low-cost ambulatory monitoring would allow a follow-up of the diagnosed patient. Such studies use numerous and uncomfortable sensors that disturb the patients' rest. One of the sensors on the chest monitors 4 torso postures: prone, supine, left lateral and right lateral. In this work we analyze the reliability of performing posture monitoring during sleep with a wearable device on the wrist. In our methodology we develop classification models to prove that in order to make these models applicable on real data it is necessary to (i) perform a subject-wise training and (ii) detect and eliminate the monitoring periods corresponding to turns of torso or sudden movements. Our methodology improves the state-of-the-art results by more than 0.011 points with F-values on new subjects of 0.966 and 0.989 for Random Forest and k-Nearest Neighbors algorithms respectively.**

***Index Terms*—Sleep, posture, wearable, subject-wise, torso turns**

I. Introduction

Sleep hygiene is defined as a set of behavioral and environmental recommendations intended to promote healthy sleep [1]. The importance of not having a correct sleep hygiene is a public health problem because it leads to sleep disorders, such as insomnia, that affect our daily life and worsen the symptoms of other diseases [2].

In recent years, many studies have been conducted on the factors that promote good sleep quality. Some authors have shown that sleep posture is an important factor in sleep quality [3] [4]. Many studies demonstrate the relationship between sleep posture and sleep disorders or other diseases—physical or neurological—such as: obstructive sleep apnea [5] and how sleeping in lateral decubitus can decrease the symptoms of sleep disorders in people with apnea-hypoapnea syndrome (OSAHS) [6], neck muscle activity [7], non-specific spinal cord symptoms [8], renal diseases [9] [10], and neurological diseases such as migraine [11]. In this work we will focus on demonstrating whether it is possible to use low-cost commercial wearable devices for high-quality and reproducible sleep posture detection.

The study of posture during sleep is performed professionally in hospitals, in sleep units. In these spaces, patients who come for suspected illnesses are monitored overnight: restless leg syndrome, insomnia, sleep apnea, etc. The sleep study is called polysomnography (PSG), and multiple biomedical signals are recorded, such as blood pressure, electrocardiogram (ECG), electroencephalography (EEG), respiration, temperature, or sleep position.

Sleep studies suffer from two problems that we can mention: (i) on the one hand, patients only spend one night in the hospital, and from that night a diagnosis is made without the possibility of continuous ambulatory and recurrent monitoring for follow-up. And (ii) on the other hand, all the sensors used are uncomfortable and disturb the patient's rest, in a sleep session that is not representative. It would be very interesting for professionals, and would open up a new avenue of study, to be able to use comfortable non-invasive ambulatory monitoring devices that do not disturb patient's sleep. Many studies are being carried out along these lines. Above all, and

This project has been partially funded by the Consejería de Educación e Investigación (Grant Number: IND2019/BMD-17276) and Ministerio de Ciencia e Innovaciónn (Grant Number: PID2019-110866RB-I00).

in particular, on the detection of posture during sleep, which is what concerns us.

Currently, posture during PSG is measured with a sensor placed on the chest. Four postures of the torso are recorded: prone, supine, left lateral and right lateral. Numerous studies are exploring alternative, less obtrusive methods for posture detection with different technologies.

- Works such as the one of Enayati *et al.* [12], or the study of Hsiao *et al.* [13] use pressure sensors on the bed. The solution proposed by Hsiao *et al.*reaches an accuracy of 0.881 for posture detection using resistive force sensors placed on the top of the bed to measure and an array of infrared sensor.
- Other authors have implemented video image analysis. The authors in [14] achieve an overall accuracy of 0.916 using 13 body postures (not only torso postures). In another study [15], the authors achieve an average accuracy of 0.909 distinguishing between 10 body postures. The use of cameras has the drawback of compromising user privacy and consuming more computational resources and energy than other technologies. On the positive side, by looking at the movements of the whole body, they are able to detect more postures than the aforementioned torso postures used in clinical practice. In this study we will focus only on the four torso postures.
- Other studies analyze inertial signals such as accelerometry with single or multiple wearable devices [16] [17] [18]. In the work shown by Parastoo Alinia *et al.* [19] it has been studied four body positions and nine locations for accelerometer placement have been assessed. The results have shown F-values between 0.952 to 0.978, with the best positions for accelerometer placement on the thighs and the chest.

A. Problem statement

The market size of wearable devices is growing year by year. The use in any population segment is generalized, which makes them an enabling tool to bring health to home through ambulatory monitoring. Generally all commercial wearable devices include an accelerometer; an inexpensive and energy efficient technology. Therefore, in this paper we will focus on the use of wrist wearable devices. This work is based on the need to implement a torso posture detection algorithm for real life. After searching the state of the art (SoA), and seeing that it is feasible to do it using wearable devices, we realized that we were not able to replicate the results. The main problem encountered was that testing with new subjects on machine learning models created by training-test split data on a randomization over all the data did not work.

Consequently, we decided to create our own experimental dataset and investigate the reasons why we were not able to replicate the results of the state-of-the-art. In this work we: (i) evaluate classical machine learning algorithms to detect 4 torso postures during sleep, (ii) analyze the best features for these models based on clinical literature; (iii) show that the best solution for a real implementation is to create trained models with a subject-wise data partitioning; and (iv) demonstrate that it is essential to detect and isolate turns and sudden wrist movements to achieve a satisfactory classification.

Our experimental methodology is based on two hypotheses about why the models of the SoA are not replicable:

1) The system fails when new subjects are presented to models trained with data splitting techniques based on stratified random data splitting, or cross-validation (CV) over the entire dataset. This was demonstrated by Saeb *et al.*in [20]. The authors demonstrated that in order to bring machine learning models closer to real—clinical—practice, a subject-wise data splitting is needed. A subject-wise CV method does not overestimate the ability of machine learning models to predict clinical outcomes as a record-wise CV method does.
2) Motion signals are very stable during sleep. The evaluation of posture during movement of the subjects generates much noise that does not allow a good generalization of the models. Therefore, we suggest labeling these movements which, in general, are due to body turns during sleep. Once identified, we are going to evaluate if it is feasible and worth classifying the turns, or on the contrary it would be better to detect them and override the posture classification during the time when the subject moves.

We consider that our proposal to identify and isolate movement periods is a key point of differentiation in our work that, combined with a subject-wise training strategy, has led to improve the metrics of the state of the art and make them reproducible. We are comparing our results with the ones referred in Table I which we have considered the most relevant.

TABLE I
RELEVANT WORKS FOR COMPARISON

Authors	Devices used	# postures	# subjects	Hours monitored
Hsiao *et al.* [13]	FSR sensors and one infrared sensor array	6	?	?
Ostadabbasn *et al.* [14]	pressure sensor array	3 (13)	9	-[a]
Li *et al.* [15]	depth sensor and RGB camera	10	22	-[a]
Van Laerhoven *et al.* [16]	Porcupine wearable wristband	3-5	4	63.0
Jeng. *et al.* [18]	1 wristband	4	2	0.3
Alinia *et al.* [19]	9 (2) accelerometers	4	22+8	0.2+1.3
This work	1 wristband	4 (+2)	13	7.4

[a]No time, one image per posture per subject.

II. METHODOLOGY

Figure 1 shows the experimental scheme followed along this work. In this section we will go through each of the following

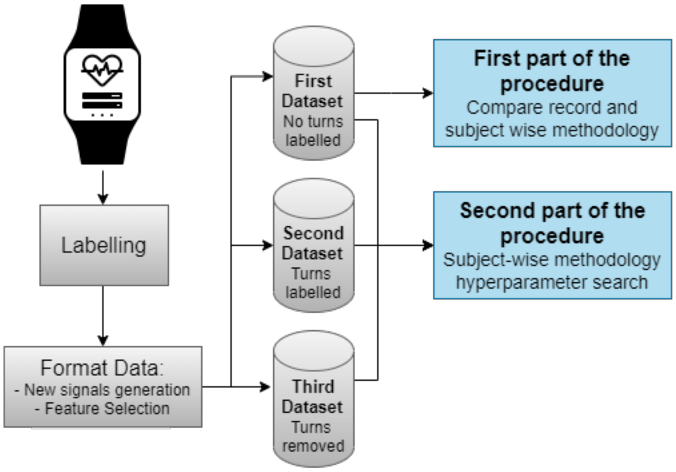

Fig. 1. Methodology diagram.

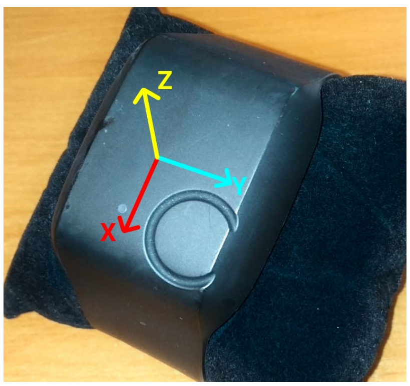

Fig. 2. Empatica E4 wristband indicating the axes of the accelerometer.

blocks: (i) data acquisition, (ii) data labeling (with and without torso turns), (iii) feature generation and feature selection, and (iv) modeling, where we compare record-wise and subject-wise methodology with and without labeling of the turns.

A. Data acquisition

We conducted a controlled monitoring experiment with 13 individuals: 7 women (53.8%) and 6 men (46.2%), on average 32.1 ± 14.6 years old. The monitoring was performed using the Empatica E4 wristband shown in Figure 2. This wristband has a 3-axis accelerometer sensor in the range $[-2, 2] \times g_n$ of 8 bit resolution and 32 Hz sampling frequency. The wristband has an Event Mark Button to indicate timestamps.

As mentioned above, the four postures monitored have been: prone, supine, left lateral and right lateral. The posture will be estimated with respect to the torso, *i.e.*, with the torso at rest, no matter the movement of the extremities and/or head.

During the experiment, the participants wear the bracelet on the non-dominant wrist, always leaving the button on the thumb side. The experiment lasted about 34 minutes and was divided into three rounds of 360º turns. Starting from the supine position, the subject rotated four times to the right, passing through the other three postures and concluding in the starting position. Each posture was held for 5 minutes. The next round was analogous to the former, changing the direction of the turn to the left and holding each posture for 3 minutes. Finally, the last round, repeats the first one holding each posture for 30 seconds. The Event Mark Button was pressed at the end and the beginning of each posture.

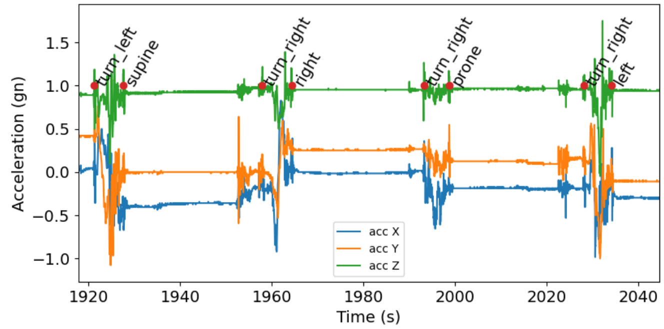

Fig. 3. Example with the accelerometry values of 4 postures and turns.

TABLE II
DATA LABELING DISTRIBUTION

Label (position)	Number of 5-second windows	Label distribution (%)
up (supine)	6663	25.32%
left	6472	24.60%
down (prone)	6443	24.49%
right	5947	22.60%
turn right	525	**01.99%**
turn left	260	**01.00%**

B. Data labeling

As discussed at the beginning of this manuscript, we were not able to reproduce the results of the models of the state of the art. One of the hypothesis for which we have not been able to reproduce the results is: the signal values during the turns in the bed add noise preventing a good performance of the model. Thus, using the timestamps, we were able to delimit the time intervals in which the turns occurred. An example is shown in Figure 3, where the 3 axes of accelerometry are shown for the four postures and turns.

Three different data sets have been used for experimentation, as shown in Figure 1: (i) the first dataset does not consider the labeling of the turns but only the four postures of the torso. The turns are labeled with the value of the previous posture; (ii) the second dataset contemplates the labeling of the right and left turns, resulting in a total of 6 classification labels; and, (iii) the third dataset marks and eliminates the time periods of the turns. Table II shows the percentage distribution of the data for each of the postures and turns. Note that, in our experiment, turns constitute approximately 3% of the total data.

C. Feature generation and feature selection

For our first experiments we relied on the features used by Jeng *et al.*in [18], which were the mean value of every axis of accelerometry. As the results were not satisfactory, we also decided to add other features that we considered to be of interest, as well as other features commonly used in the state of the art of clinical sleep studies [21].

After some preliminary tests, it was decided to use a 5-second window with 4-second time overlap, thus yielding results every 1 second. These values allow us (i) to cover the time duration of torso turns as measured in 6.3 ± 2.0 seconds, and (ii) to have a richer dataset for training. A final market implementation would not have such a high resolution according to the time standards of clinical studies which are 30 seconds [22].

Thus, 42 features have been calculated in each window. For each one of the 3 accelerometry axes, the Euclidean Norm Minus One (ENMO), the Z-angle, and the Energy signals there were computed as defined in [21]: the maximum, minimum, mean (μ), standard deviation (σ), mean value of the first derivative (μ_δ), the area of the absolute value of the signal ($A_{|si|}$), and the counts ($c_{|si|}$). The counts are defined as the number of crossings per zero crossover, or number of crossings for a threshold delta for positive signals only.

Previously to the modeling step, a feature section has been performed based on Pearson's correlation. The features with a correlation higher than 0.8 have been removed.

D. Modeling

In our modeling methodology we are going to present several experiments named as *procedures* in Figure 1.

With the first procedure we aim to show that the state-of-the-art models are not reproducible when used on data gathered from new individuals. The models are trained using a record-wise cross-validation and then tested on new subjects. With this procedure we corroborate the results of Saeb *et al.* in [20] and proceed to continue with a subject-wise split methodology on the training data.

In the second part of the procedure, we intend to demonstrate that delimiting and removing the time intervals in which torso turns—or abrupt wrist movements occur—improves the performance of the models. To do so, we perform subject-wise cross-validations on the three different labeled datasets: (i) without marking the turns, (ii) labeling the turns, and (iii) removing the data acquired during turns.

During the whole modeling process we:

- Compare the results of three different machine learning algorithms: a Random Forest (RF), a Support Vector Machine (SVM), and a k-nearest neighbors (KNN). We perform 5-fold cross-validations, and our evaluation metric is the F-value;
- Keep aside data from 2 randomly chosen subjects as if they were real life test data;
- Use a Bayesian hyperparameter optimization in order to search the best combination of hyperparameters for each model [23]. For that purpose, we use the Python libraries Scikit-learn and Tune (Scalable Hyperparameter Tuning and TuneSearchCV).

The list of hyperparameters is as follows:

TABLE III
LIST OF FEATURE SELECTED

feature\signal	**X**	**Y**	**Z**	**ENMO**
minimum				✓
μ	✓	✓	✓	
σ		✓	✓	
μ_δ	✓	✓	✓	✓
A_{s_i}		✓	✓	
c_{s_i}	✓	✓	✓	✓

- RF: *# of trees*: [2000, 1000, 500, 250]; *maximum depth*: [500, 300, 100, 50]; *minimum samples per split*: [50, 20, 10, 5]; *minimum samples per leaf*: [50, 20, 10, 5]
- KNN: *weights*: [$uniform$, $distance$]; *algorithm*: [$auto$, $ball - tree$, $kd - tree$, $brute$]; *p*: [1, 2] being 1 the Manhattan distance, and 2 the Euclidean distance.
- SVM: *regularization parameter (C)*: [1, 2, 3, 4]; *kernel*: [$linear$, $poly$, rbf, $sigmoid$]; *gamma*: [$scale$, $auto$]; *maximum number of iterations*: without limit.

For the sake of repeatability, all training processes have been performed by setting a seed. In the following section we discuss an analyze the results obtained during our experiments.

III. Results

A. Analysis of feature selection

The Pearson's correlation-based feature selection has reduced the number of features for modeling from 42 to 16. The list of features selected is shown in Table III. The reader should notice that the features related to the Z-angle and the Energy have not been chosen. It seems that the axes Y and Z are the most important signals, and the *mean value of the first derivative* (μ_δ) as well as the *counts* ($c_{|si|}$) are the most repeated features. On the other hand, the feature *maximum* has never been selected.

Since there are not a large number of features, we will continue with these features while modeling. A more comprehensive analysis of the features should be done in the future.

B. Record-wise methodology on new subjects data

Table IV shows the results of the experiment on the *first procedure*. As suspected, the system fails when new cases of subjects are fed to a system trained on a record-wise fashion. As can be seen, the F-value decreases as much as 0.46, 0.08 and 0.47 for the RF, SVM and KNN algorithms respectively. Given such a difference in the results, and as mentioned above, the strategy of dividing data by records is not a valid solution and therefore we continue with a strategy of dividing data by subjects in training [20]. As a result, it is not even worth analyzing the architecture of these models and they are discarded.

C. Subject-wise models and detection of torso turns

In the second part of the procedure, based on a modeling methodology with subject-wise data splitting, we aim to demonstrate the need to detect and remove torso turns in order

TABLE IV
MEAN F-VALUE OF RECORD-WISE METHOD ON NEW SUBJECTS DATA

Tested on:	**Model**		
	RF	***SVM***	***KNN***
Record-wise folds	**0.94**	0.63	0.93
New subjects	0.48	0.55	0.46

TABLE V
AVERAGE F-VALUES OF SUBJECT-WISE 5-CV OVER 10 RUNS OF DIFFERENT TEST SUBJECTS

F-value	**Random Forest**		**SVM**		**KNN**	
	mean	***std***	***mean***	***std***	***mean***	***std***
Non-labeled turns	0.946	0.029	0.564	0.175	0.985	0.017
Turns labeled	0.945	0.031	0.544	0.129	0.980	0.020
Turns removed	**0.966**	**0.031**	**0.623**	**0.097**	**0.989**	**0.013**

to improve the performance of the models. The results of such experiments on the three datasets created are shown in Table V. The standard deviation is small, which means that the performance of all the models is similar. One is chosen for each algorithm and their hyperparameters are shown in the Tables VI through VIII. Testing the models on new subjects data leads to average F-values of 0.966, 0.623 and 0.989 for RF, SVM and KNN algorithms respectively. It can be said that the KNN and RF algorithms perform better than SVM for all the experiments.

In view of the results in Table V, it might appear that there is no substantial difference in the strategy of not labeling the data (first dataset), detecting and classifying them (second dataset), or detecting and eliminating them (third dataset). However, by looking at the confusion matrix of one of the models detecting torso turns (Figure 4 for an RF model), one can see that the detection of turns is not correct. The recall and precision of the *turn_left* and *turn_right* labels are 0.258, 1.000 and 0.777, 0.777, respectively. This shows the need to detect turns, yes, but to remove them and not to classify them. Removing the turns of the torso results in models that, in the worst case, show an F-value improvement of 0.009 with low deviation. The RF and KNN algorithms yield the best results with average F-values in test of 0.966 and 0.989, respectively.

Notice the reader that comparing the results in Figure 4 and those in Figure 5, it seems clear that the inclusion of labeling and detection of torso turns leads the models to decrease the performance on the classification of the postures. This effect might be more accentuated in a real life scenario where more torso turns occur. Therefore it is recommended to implement a subject-wise classification strategy that implements a detection mechanism for turns or sudden movements in order to pause the posture classification until the accelerometry signals become stable. The issue of automatic detection will be addressed in future work.

Finally, Table IX compares the performance of our results with the works that we consider most representative of the state of the art.

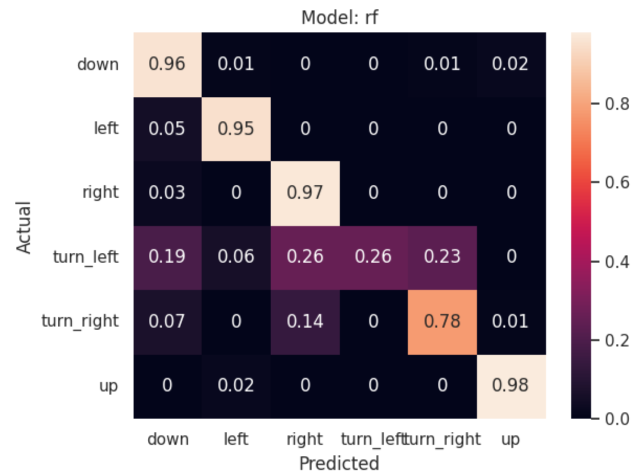

Fig. 4. Confusion matrix of the RF trained on the dataset with turns labeled.

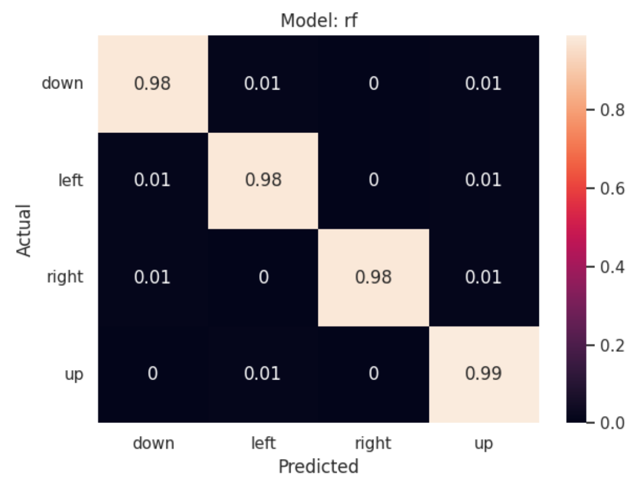

Fig. 5. Confusion matrix of the RF trained on the dataset with turns removed.

TABLE VI
HYPERPARAMETERS OF RF ALGORITHM

experiment	**Random Forest**	
	number of trees	***max depth***
Non-labeled turns	500	100
Turns labeled	2000	50
Turns removed	1000	500
	min samples per split	***min samples per leaf***
Non-labeled turns	10	5
Turns labeled	5	5
Turns removed	10	5

TABLE VII
HYPERPARAMETERS OF SVM ALGORITHM

experiment	**SVM**		
	C	***kernel***	***gamma***
Non-labeled turns	3	*rbf*	*scale*
Turns labeled	4	*rbf*	*scale*
Turns removed	4	*poly*	*scale*

TABLE VIII
HYPERPARAMETERS OF KNN ALGORITHM

experiment	KNN			
	Neighbors	***Weights***	***Algorithm***	***p***
Non-labeled turn	4	$distance$	$brute$	1
Turns labeled	6	$distance$	$ball - tree$	2
Turns removed	4	$distance$	$brute$	1

TABLE IX
RESULTS COMPARISON

Author	Metric	Value	Improvement
P.-Y. Jeng. *et al.* [18]	accuracy	0.833	+0.156
K. Van Laerhoven *et al.* [16]	precision	0.808	+0.180
S. Ostadabbas *et al.* [14]	accuracy	0.916	+0.073
Hsiao *et al.* [13]	accuracy	0.881	+0.108
Y.-Y. Li, *et al.* [15]	accuracy	0.909	+0.080
Parastoo Alinia *et al.* [19]	F-value	0.978	+0.011

It is also worth highlighting that we achieve these results with a low-cost commercial ambulatory monitoring device placed on the wrist. Jeng *et al.* [18] already advanced that using only a wristband device improve the user experience with respect to solutions that require more sensors, cameras or more complex devices such as pressure blankets.

IV. CONCLUSION

It has been verified that it is indeed possible to classify posture during sleep, however, to make it reproducible and accurate it is necessary (i) to detect and remove torso turns, and (ii) to perform a subject-wise training. The performance of our solution improves the results of the state-of-the-art. The test results of our models show F-values of 0.966 and 0.989 for RF and KNN algorithms respectively.

The difference in metrics may seem meaningless between (i) not labeling, (ii) labeling and classifying, and (iii) labeling and removing torso turns. However, it has been demonstrated that the low recall of the *turn_left* and *turn_right* labels when labeling the turns—0.258 and 0.777, respectively—claims to use the methodology proposed. In a real case of monitoring subjects, the number of turns and noisy data will be larger, and the difference between methodologies will show greater.

Our solution outperforms the results of the state of the art from 0.011 to 0.180 points. In future work we will tackle the problem of automating detection of turns to pause the classification in real time. These results are a step ahead towards making ambulatory sleep monitoring with clinical perspective feasible.

REFERENCES

[1] E. J. Stepanski and J. K. Wyatt, "Use of sleep hygiene in the treatment of insomnia," *Sleep medicine reviews*, vol. 7, no. 3, pp. 215–225, 2003.

[2] L. A. Irish, C. E. Kline, H. E. Gunn, D. J. Buysse, and M. H. Hall, "The role of sleep hygiene in promoting public health: A review of empirical evidence," *Sleep medicine reviews*, vol. 22, pp. 23–36, 2015.

[3] E. R. Colson, N. L. Geller, T. Heeren, and M. J. Corwin, "Factors associated with choice of infant sleep position," *Pediatrics*, vol. 140, no. 3, 2017.

[4] M. Muehlhan, M. Marxen, J. Landsiedel, H. Malberg, and S. Zaunseder, "The effect of body posture on cognitive performance: a question of sleep quality," *Frontiers in human neuroscience*, vol. 8, p. 171, 2014.

[5] Y. Itasaka, S. Miyazaki, K. Ishikawa, and K. Togawa, "The influence of sleep position and obesity on sleep apnea," *Psychiatry and clinical neurosciences*, vol. 54, no. 3, pp. 340–341, 2000.

[6] J. B. Lee, Y. H. Park, J. H. Hong, S. H. Lee, K. H. Jung, J. H. Kim, H. Yi, and C. Shin, "Determining optimal sleep position in patients with positional sleep-disordered breathing using response surface analysis," *Journal of sleep research*, vol. 18, no. 1, pp. 26–35, 2009.

[7] W.-H. Lee and M.-S. Ko, "Effect of sleep posture on neck muscle activity," *Journal of physical therapy science*, vol. 29, no. 6, pp. 1021–1024, 2017.

[8] D. Cary, K. Briffa, and L. McKenna, "Identifying relationships between sleep posture and non-specific spinal symptoms in adults: A scoping review," *BMJ open*, vol. 9, no. 6, p. e027633, 2019.

[9] B. Shekarriz, H.-F. Lu, and M. L. Stoller, "Correlation of unilateral urolithiasis with sleep posture," *The Journal of Urology*, vol. 165, no. 4, pp. 1085–1087, 2001.

[10] S. A. Ziaee, A. Moradi, M. Fateh, and S. M. Hosseini Moghaddam, "Sleep posture and unilateral renal stone formation," *Scandinavian journal of urology and nephrology*, vol. 42, no. 6, pp. 551–554, 2008.

[11] B. Bag and N. Karabulut, "Pain-relieving factors in migraine and tension-type headache," *International journal of clinical practice*, vol. 59, no. 7, pp. 760–763, 2005.

[12] M. Enayati, M. Skubic, J. M. Keller, M. Popescu, and N. Z. Farahani, "Sleep posture classification using bed sensor data and neural networks," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 461–465.

[13] R.-S. Hsiao, T.-X. Chen, M. A. Bitew, C.-H. Kao, and T.-Y. Li, "Sleeping posture recognition using fuzzy c-means algorithm," *Biomedical engineering online*, vol. 17, no. 2, pp. 1–19, 2018.

[14] S. Ostadabbas, M. B. Pouyan, M. Nourani, and N. Kehtarnavaz, "In-bed posture classification and limb identification," in *2014 IEEE Biomedical Circuits and Systems Conference (BioCAS) Proceedings*. IEEE, 2014, pp. 133–136.

[15] Y.-Y. Li, Y.-J. Lei, L. C.-L. Chen, and Y.-P. Hung, "Sleep posture classification with multi-stream cnn using vertical distance map," in *2018 International Workshop on Advanced Image Technology (IWAIT)*. IEEE, 2018, pp. 1–4.

[16] K. Van Laerhoven, M. Borazio, D. Kilian, and B. Schiele, "Sustained logging and discrimination of sleep postures with low-level, wrist-worn sensors," in *2008 12th IEEE International Symposium on Wearable Computers*, 2008, pp. 69–76.

[17] M. Borazio, U. Blanke, and K. Van Laerhoven, "Characterizing sleeping trends from postures," in *International Symposium on Wearable Computers (ISWC) 2010*, 2010, pp. 1–2.

[18] P.-Y. Jeng, L.-C. Wang, C.-J. Hu, and D. Wu, "A wrist sensor sleep posture monitoring system: An automatic labeling approach," *Sensors*, vol. 21, no. 1, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/1/258

[19] P. Alinia, A. Samadani, M. Milosevic, H. Ghasemzadeh, and S. Parvaneh, "Pervasive lying posture tracking," *Sensors*, vol. 20, no. 20, p. 5953, 2020.

[20] S. Saeb, L. Lonini, A. Jayaraman, D. C. Mohr, and K. P. Kording, "The need to approximate the use-case in clinical machine learning," *Gigascience*, vol. 6, no. 5, p. gix019, 2017.

[21] K. Sundararajan, S. Georgievska, B. H. Te Lindert, P. R. Gehrman, J. Ramautar, D. R. Mazzotti, S. Sabia, M. N. Weedon, E. J. van Someren, L. Ridder *et al.*, "Sleep classification from wrist-worn accelerometer data using random forests," *Scientific reports*, vol. 11, no. 1, pp. 1–10, 2021.

[22] C. Iber, "The aasm manual for the scoring of sleep and associated events: Rules," *Terminology and Technical Specification*, 2007.

[23] M. Feurer, J. Springenberg, and F. Hutter, "Initializing bayesian hyperparameter optimization via meta-learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, 2015.