Sleep stages classification using cardio-respiratory variables

Asma Gasmi, Vincent Augusto, Paul-Antoine Beaudet, Jenny Faucheu, Claire Morin, Xavier Serpaggi, Franck Vassel

Sleep stages classification using cardio-respiratory variables

Asma GASMI[1,2], Vincent AUGUSTO[1] , Paul-Antoine BEAUDET[2],
Jenny FAUCHEU[3] , Claire MORIN[4] , Xavier SERPAGGI [5] and Franck VASSEL [2]

***Abstract*— Analysis of sleep is important in order to detect health issues and try to prevent them. In particular, sleep dysfunctions may be the first signs of cognitive frailties for elderly persons. The polysomnography (PSG) is considered the golden standard to perform a comprehensive sleep analysis, as it is based on several sensors placements. However, for longitudinal study of sleep that is required to prevent frailty for elderly persons, such medical equipment is not suitable since it is very invasive. Recent technological advances in sensors allow to gather data with a good precision with less intrusive equipment. The main objective of this study consists in developing a new algorithmic approach to analyse sleep using data from low intrusive sensors. In this study we focus on sleep phase detection, i.e. wake, Non-Rapid Eye Movement (NREM) and Rapid Eye Movement (REM). We consider the following sources of data: heart beat rate, as well as user data such as gender, age, etc. The problem is considered as a supervised classification machine learning problem. We propose a benchmark of several machine learning algorithms and compare their performances against the medical gold standard, the PSG. To do so, we use a data-set collected from a published clinical trial. Support Vector Machine (SVM) algorithm globally outperforms all other methods with a 76.5% agreement with the PSG. As a direct perspective of this study, we plan to add other sources of data using custom sensors to improve the performance of the prediction.**

Sleep stages, machine learning, supervised classification, sleep architecture, polysomnography

I. Introduction

Sleep analysis is a promising research field to delay frailty for elderly people and prevent serious health problems. Previous studies [6] underline the good correlation existing between sleep disorders and cognitive frailty, the former being both a source and a consequence of the latter. In this context, polysomnography is acknowledged as the gold standard [12] to assess sleep quality. It allows monitoring brain, heart, muscle, and breathing activity, and therefore deciphering the different sleep stages and the duration of sleep, as well as possible events of sleep disorder (apnea, restless legs syndrome, parasomnia...).

Corresponding author: Asma GASMI

[1]Mines Saint-Etienne, Univ Clermont Auvergne, CNRS, UMR 6158 LIMOS, Center for Biomedical and Healthcare Engineering, F-42023 Saint-Etienne France name.surname@emse.fr

[2]ATELIERS DU HAUT FOREZ, ZAC La Gravoux, 42380 LA TOURETTE France

[3]Mines Saint-Etienne, CNRS, UMR 5307 Laboratoire Georges Friedel, Centre SMS, F - 42023 Saint-Etienne France

[4]Mines Saint-Etienne, Univ Lyon, Univ Jean Monnet, INSERM, UMR 1059 Sainbiose, Centre CIS, F - 42023 Saint-Etienne France

[5]Mines Saint-Étienne, Institut Henri Fayol, Département ISI, F-42023 Saint-Étienne France

However, PSG is a very intrusive and costly tool that cannot allow a longitudinal follow-up of patients. Different substitutes are currently developed and tested for a non-intrusive, accurate and inexpensive assessment of the sleep quality. Among them, ballistocardiograph is a promising tool. It allows a distant monitoring of the patient heart and breathing activity, since the sensor is placed under the mattress of the bed. Several studies report clinical studies and related post-processing to assess the sleep quality provided by the BCG against the gold standard. These studies can be classified depending on the number of raw data and derived parameters included to detect the sleep stages, as well as on the classification chosen for the sleep stages.

In this study, we chose the most common sleep classification, split in three stages [3], [4], defined as:

1) wake (representing less than 2% of the total sleep duration), characterised by slow muscle activity and slow eye movement.
2) NREM, characterised by a decreased heart rate, an absence of eye movement, muscle relaxation, and a slow breathing, as well as specific brain activities. It usually represents 75-80% of the total sleep duration;
3) REM, characterised by an intense brain activity, an absence of muscle activity, rapid and irregular breathing, increased heart and breathing rates, and quick eye movements. It represents 20-25% of the total sleep duration.

Table 1 summarises studies which aim at characterising the sleep using cardio-respiratory variables from the BCG or ECG and these three sleep stages. In past clinical trials, most of the methods were tested on a small number of subjects (less than twenty). Most of studies achieve around 75% precision comparing to PSG results. Only [18] has been able to reach a fairly good precision (85%) but it was only tested on 5 subjects while using as many as 78 features in their method. [14] also performs well (76.10%) but the method could not be implemented within a commercial product. Similar conclusions can be drawn for [19].

The main objective of this work is to design a non-invasive way of sleep stages classification that has a precision exceeding 70% while using the minimum number of features possible. We propose in this paper a supervised classification method in association with data that could be driven from ballistocardiograph (BCG).

This paper is organised as follows. We present in Section II the methods based on the use of cardio-respiratory variables that could be driven from BCG. Results are given and

TABLE I

PREVIOUS RELATED WORK

Article	*Num. of subjects*	*Classes*	*Sensor Type*	*Classifier*	*Row Input*	*Accuracy*
Migliorini et al. 2010 [14]	11	REM/NREM/WAKE	Hydraulic signals+ 4 load cells signal	Quadratic-wavelet discrete transform (frequency mode)	LF/HF/ Ratio (HF/LF)	76.10%
Hrihara & Watanabe 2012 [13]	21	REM/NREM/WAKE	ECG	Threshold	respiration/ and heart rate time domain features/ abdominal force	51,60%
Kasper et al. 2012 [9]	8	REM/NREM/WAKE	ECG	Threshold	HR spectral power	57,80%
Park et al. 2014 [15]	20	REM/NREM/WAKE	BCG (with 3 different types of sensors)	Analytic compression	HF/LF/DFA/HRV	66,24%
Kagawa et al. 2016 [16]	10	REM/NREM/WAKE	BCG (radar)	DWk-NN	image proceeding	57,10%
Surantha et al. 2017 [17]	16	REM/NREM/WAKE	ECG	SVM and PSO	SDNN / RMSSD / SDSD / NN50/ pNN50 / AVNN Geometrical / HRV Triangular Index Poincare / SD1 / SD2 / ratio SD1/SD2	73,26%
Chen et al. 2018 [19]	15	REM/NREM/WAKE	ECG	Hidden Markov Models	row ECG channels (60 features)	79,90%
Yi et al. 2019 [18]	5	REM/NREM/WAKE	BCG (hydraulic bed sensor)	SVM and KNN	body position, abdominal position, heart rate variation, respiration rate	85%

discussed in Section III. Section IV gives some conclusions and perspectives.

II. SLEEP STAGES CLASSIFICATION

A. Empirical method approach

Each stage of sleep is characterised by a certain variation on the cardio-respiratory variables. From a medical point of view, the variation of the heart rate is observable from the distance between two R-peaks (also noted the RR interval). Different cardio-respiratory features can be derived from the RR interval, the heart beat and the respiration rate, as summarised in Table II. These features serve as basis for defining a series of conditions allowing the classification of sleep stages. The latter follows from an empirical algorithm that we developed based on the work of Kurihara and Watanabe [13].

This algorithm is detailed in Figure 1. It combines the conditions of Table II to define the occurrence of each sleep stage, hence the importance of a correct determination of the a_i coefficients of Table III. A sensitivity analysis was conducted, and revealed a limited influence of coefficients a_3 and a_5, which were therefore set to the values reported in the literature. For the three other coefficients, a design of experiment was conducted where the sensitivity of each remaining coefficient was tested individually, since the coefficients are independent from each other. Finally, three sets of parameters were kept, and are reported. Note that the coefficients a_i were determined from tests and results in the literature.

TABLE II

VARIABLES DEFINITION

Main Features	Definition
$R - RInterval$	RR interval for every second
$R - RInterval_{mean}$	mean value of RR interval for the whole night
HBM	Heartbeat mean value for every 30s
HB_{mean}	Heartbeat mean value for the whole night
$HBstakaHRV$	Heart rate variability for every 30s
$HBst_{mean}$	The mean value of the heart rate variability for the whole night
R	R-algorithm given in [13]
RR	Respiration rate
RR_{mean}	Respiration rate mean value for the whole night
$RRst$	Respiration rate variation
$RRst_{mean}$	Respiration rate variation mean value for the whole night

TABLE III

CONDITIONS FOR CLASSIFIERS

C_1	$HBM < a_1 \times HB_{mean}$
C_2	$R(k) \in [I_{REM}]$
C_3	$RR < RR_{mean}$
C_4	$HRst \leq a_2 \times HBst_{mean}$
C_5	$RRst \leq a_3 \times RRst_{mean}$
C_6	$HBM < a_4 \times HB_{mean}$
C_7	$RRst \leq a_5 \times RRst_{mean}$
C_8	$RRI \geq RRI_{mean}$

We consider that the NREM phase could be detected if the conditions C_1, C_3, C_4 and C_5 described in Tableé III

are satisfied. Otherwise, if conditions C_2, C_6, C_7 and C_8 are satisfied, the REM phase is confirmed. We assume that every other state is wake.

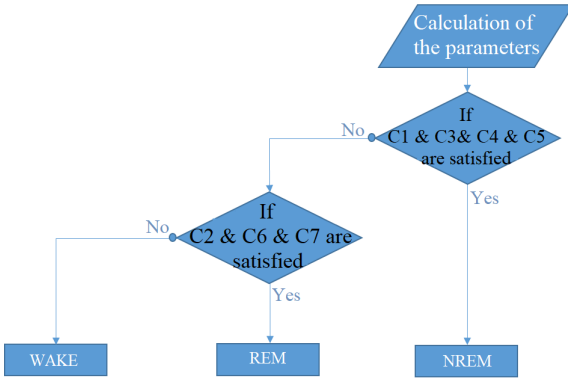

Fig. 1. The 3 phases classification algorithm

Finally three sets of parameters were kept and are reported:

- Set1: $a_1 = 1.1$, $a_2 = 0.3$, $a_3 = 1.3$, $a_4 = 1.2$, $a_5 = 1.4$
- Set2: $a_1 = 0.8$, $a_2 = 0.3$, $a_3 = 1.3$, $a_4 = 1.4$, $a_5 = 1.4$
- Set3: $a_1 = 1.1$, $a_2 = 0.3$, $a_3 = 1.3$, $a_4 = 1.4$, $a_5 = 1.4$

We have combined the results of the three algorithms by creating a vote system. This voting system respects the majority rule. In the case of a different result, the results from the one with the highest precision for detecting the REM phase (which is the less represented stage), is chosen. Since REM is the least detected phase, and since algo2 is the best one in detecting it, we added a condition in case of conflict, if algo2 recognises the epoch as a REM epoch then algo2 prevails.

B. Classification Method Using Machine Learning Approaches

Five classical machine learning algorithms were selected to improve the sleep stages determination, namely: support vector machine (SVM), Multinomial Logistic regression (Rlogic), k-nearest neighbours (KNN), Classification And Regression Trees (CART) and Naive Bayes (NB).

We considered that the features were the variables defined earlier in Table II. The target is the phase detected by the PSG for each epoch. Hence, the considered problem is a supervised classification problem with 3 outcomes (wake, NREM, REM). We use 80% of the collected data was used to train the machine learning algorithm. Figure 2 shows the steps of the classification algorithm we use in this paper.

The sleep algorithms resulting from these five machine learning processes were validated on the same set of data that they were trained in.

We built the concordance matrix to compare the classification between the PSG and the tested algorithm by counting the number of times each phase appeared for an epoch of 30s, compared to the ones of PSG. The percentage of agreement by $phase_i$ equal to numbers of epochs that the algorithm predicted that it is the $phase_i$ over the numbers of epochs that the $phase_i$ was given by the PSG. The percentage of agreement is obtained by dividing the number of epochs that the algorithm was right, by the total number of epochs.

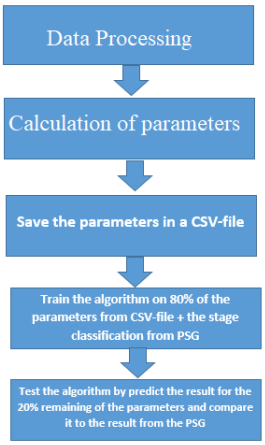

Fig. 2. Classification method using Machine learning

We consider the recall. We compute the **F1-Score** which is a measure that combines precision and recall with their harmonic mean, called F-measure or F-score
such that the precision here is the percentage of agreement with PSG.

III. Results and Discussion

A. Data

We used data collected from the clinical trial "St. Vincent's University Hospital / University College Dublin[1]". This database contains 25 full overnight polysomnograms from adult subjects with suspected sleep-disordered breathing. The subjects were all aged over 18 with no heart disease or dysfunction, and they did not take any kind of drug that could trouble the heart rate.

The first step consists in reassembling all the data collected to same sampling frequency (in this case we use 1Hz as frequency). Then a check on the data must be done to see if one of the main three variables is not available but could be calculated. In our case, the respiration was not available, so a calculation based on the thorax movements was performed. If all the 3 inputs are available, we proceed to the estimation of the derived parameters. The parameters are stored chronologically for every 30s epoch. We have chosen an epoch of 30s because the PSG that gave the classification divided the night into 30s epochs.

B. Results of tests using the empirical method

The results of the empirical algorithm detailed in Figure 1 together with each set of parameters (see Table II) are gathered in Table IV. One notes the poor global capacity of this algorithm to properly detect to NREM stage (only half of the NREM epochs are correctly classified), and the picture is even worse for the REM phase, with an agreement being far less than 25%.

[1]doi: https://doi.org/10.13026/C26C7D

A graphical representation is proposed in Figure 3. The occurrence of each sleep stage as detected by the PSG is plotted with black dots. Then, the results provided by each set of parameters is represented by one colour line of markers. The marker is a dot in case the sleep stage is correctly detected, a cross otherwise. One notices that this empirical algorithm tends to detect the NREM phase instead of the REM, while the wake stage is properly identified.

Due to the clinical importance of the REM phase, one chooses the second set of parameters as the option to be chosen in conflict cases for the classification method based on a voting system.

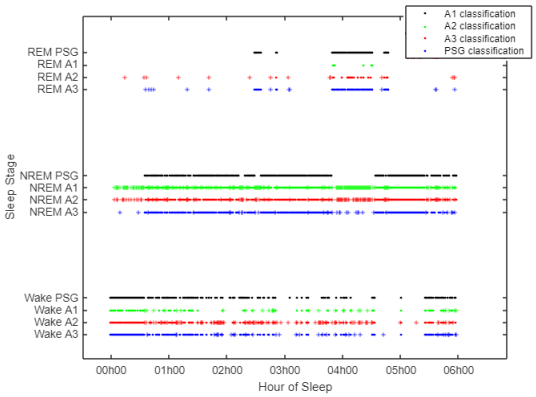

Fig. 3. Hypnograms of different algorithm in comparison with the one got by the PSG based on the brain activities

Figure 3 shows that none of the three algorithms has succeeded in detecting the REM phase correctly. Most of them either detect it wrongly or detect nothing at all. By considering all the nights, the precision is 57.89% , 50.72% and 60.11% respectively for A1, A2 and A3. Although A1 and A3 have higher total precision in respect of A2, it was showed that the latter performs better in detecting REM phase.

Figure 4 shows the results of the Voting system algorithm. We outline some improvement regarding the detection of the REM phase. By using such Voting system, we have increased the total precision by 3%. Table IV summarises the precision of each algorithm per phase and the total one.

TABLE IV
AGREEMENT PERCENTAGE OF THE DIFFERENT ALGORITHMS

	Algo1	Algo2	Algo3	Voting Algorithm
% of agreement on NREM	74,66	75,17	74,66	75,72
% of agreement on WAKE	45,28	2,91	53,57	46,74
% of agreement on REM	1,52	23,02	0,05	10,32
% of total agreement	57,89	50,72	60,11	63,41

C. Results of tests using Machine Learning methods

As explained earlier, the data set was collected during a clinical trial done over 25 subjects having a high probability to suffer from breathing sleep issues. This data set counts 20,708 epochs of 30 seconds. Only 4,142 Epochs were kept

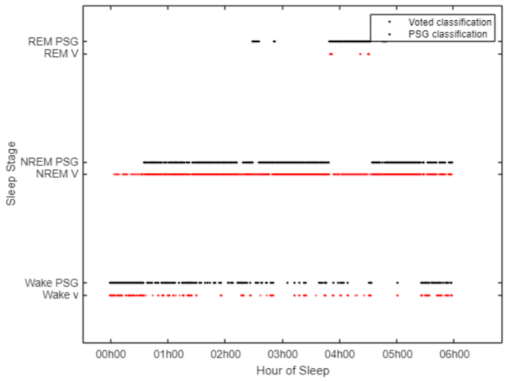

Fig. 4. Hypnograms of the voting algorithm in comparison with the one got by the PSG based on the brain activities

to test the prediction algorithm. The aim was to find a better precision than existing works from the literature using less features.

Figure 5 shows that the best algorithm was SVM with 67.55%. To better understand this result, an agreement by phase matrix is given in Figure 10. This matrix presents the agreements between the PSG and the tested algorithm. Although this method performs well, especially for the NREM phase, we see that it failed to detect perfectly the minor classes. For instance, the total of REM phase did not exceed 12% of the total agreement.

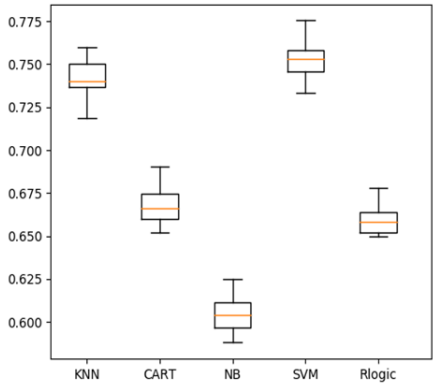

Fig. 5. Precision of each method

The precision, F1-score and recall resulting from the predicted classification are presented in Table V.

Against the previous works, our method is very satisfying since it has more than the average number of studied subjects and it has one of the best precision compared to the results given by PSG. We find out that [18] performs better for sleep stage classification. However, the latter was only tested on five subjects in contrast to our work that was tested on 25 subjects. This method used 78 features while we used only 17.

The total precision of the predicted classification is 76.50%. It is considered as an acceptable percentage comparing to the literature by practitioners. However, we aimed

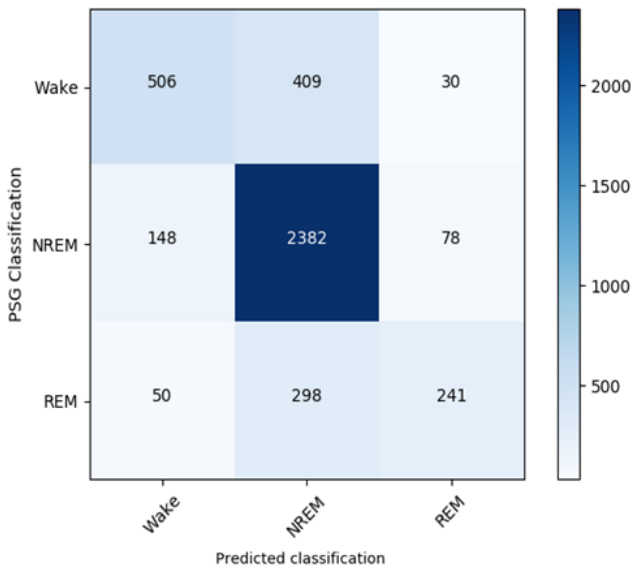

Fig. 6. Confusion matrix without normalisation for the SVM

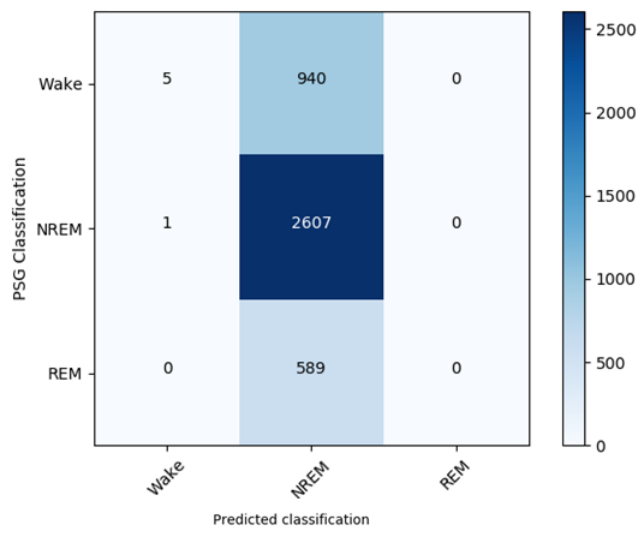

Fig. 7. Confusion matrix without normalisation for the Logistical Regression

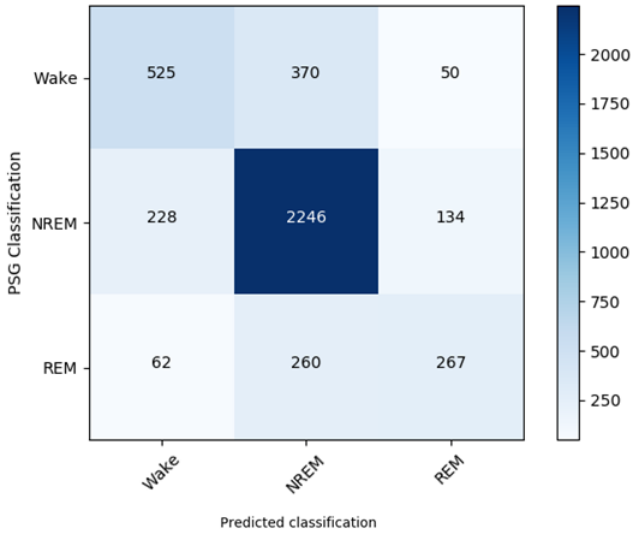

Fig. 8. Confusion matrix without normalisation for the KNN method

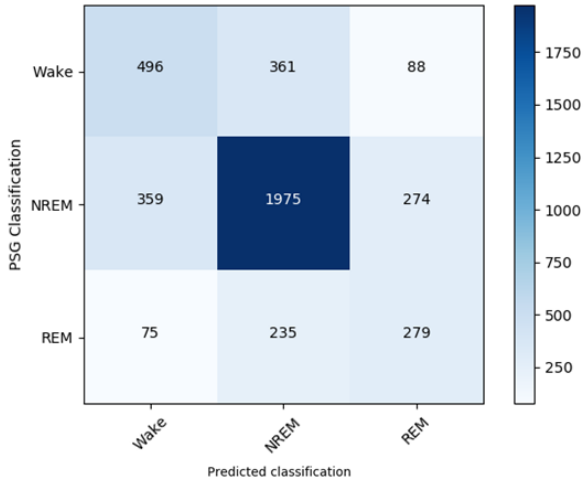

Fig. 9. Confusion matrix without normalisation for the CART method

to improve our results by creating a voting system that considers the best three methods.

According to our tests, the three most precise algorithms are SVM (76,50%), CART (66,97%) and K-nearest neighbours (72.11%).

IV. Conclusions and Perspectives

This article presents a new approach to estimate sleep stage classification using only cardio-respiratory variables that can be obtained from a contact-less BCG sensor. The method was tested using a collected data set of 25 subjects presenting symptoms of breathing sleep apnea. Inspired from the previous publications on this topic, a choice of parameters was made. These parameters were driven from the cardio-respiratory variables.

We started by creating an empirical decision tree, based on the combination of conditions proven in previous research works. By modifying the coefficients in these conditions, we obtained three algorithms that achieved more than 50% of agreement with the results driven by the PSG using brain activities. A system of vote was designed between these three algorithm resulting in a better score agreement (up to 63.41%).

By using five machine learning methods, we found that SVM is the most precise technique in this application with a total agreement of 76.50%, compared to the classification given by the PSG.

In our future works we aim to improve this method by including other parameters such as body movements and

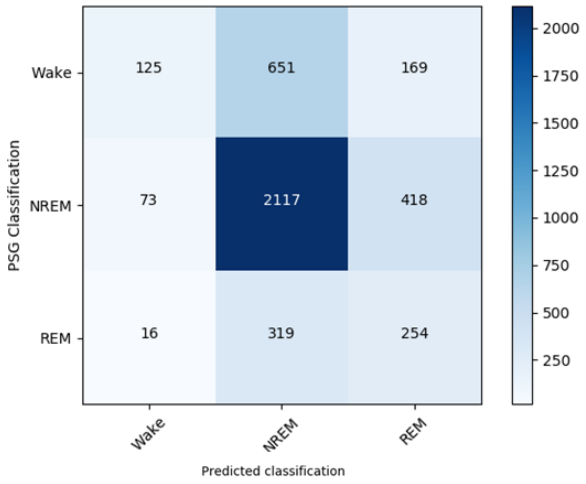

Fig. 10. Confusion matrix without normalisation for the NB method

TABLE V
ACCURACY SCORE FOR ALL THE TESTED MACHINE LEARNING METHOD

		Wake	*NREM*	*REM*	*Global*
KNN	**precision**	0.64	0.78	0.59	0.72
	recall	0.56	0.86	0.45	0,73
	f1-score	0.60	0.82	0.51	0,71
CART	**precision**	0.54	0.77	0.44	0.67
	recall	0.51	0.77	0.47	0,67
	f1-score	0.52	0.77	0.46	0.67
NB	**precision**	0.58	0.69	0.30	0.61
	recall	0.13	0.81	0.43	0.60
	f1-score	0.22	0.74	0.36	0.57
SVM	**precision**	0.72	0.77	0.69	0.76
	recall	0.54	0.91	0.41	0.75
	f1-score	0.61	0.84	0.51	0.74
RL	**precision**	0.83	0.63	0.00	0.59
	recall	0.01	1.00	0.00	0.63
	f1-score	0.01	0.77	0.00	0.49

mass distribution. We plan to conduct tests on healthier subjects: an upcoming clinical trial, whose main aim is to place a BCG sensor in a medical bed alongside with scaling sensors, will to help us reach better results. This trial is part of a project aiming to link the sleep disorders with issues on cognitive impairment and deterioration in memory quality.

REFERENCES

[1] Rechtschaffen Allan and Kales Anthony. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. Brain Research Institute. US Dept. of Health, Education, and Welfare, 1968.

[2] A. M. Bianchi and M O Mendez. Automatic detection of sleep macrostructure based on a sensorized tshirt. 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, pages 3606–3609, 2010.

[3] Emilie Crawford Achour. Physiological and pathological sleep: impact on successful aging and cognitive function. Theses, Universite Jean Monnet - Saint-Etienne, 2013.

[4] Mary Carskadon, William Dement, Meir Kryger, Thomas Roth, and Timothy Roehrs. Normal Human Sleep: An Overview, volume 2. 2005.

[5] Yuriko Doi, Masumi Minowa, and Toshiro Tango. Impact and Correlates of Poor Sleep Quality in Japanese White-Collar Employees. Sleep, 26(4):467–471, 2003.

[6] Kristine E. Ensrud, Terri L. Blackwell, Susan Redline, Sonia Ancoli-Israel, Misti L. Paudel, Peggy M. Cawthon, Thuy-Tien L. Dam, Elizabeth BarrettConnor, Ping C. Leung, Katie L. Stone, and for the Osteoporotic Fractures in Men Study Group. Sleep disturbances and frailty status in older community-dwelling men. Journal of the American Geriatrics Society, 57(11):2085–2093, 2009.

[7] Maksym Gaiduk, Thomas Penzel, Juan Antonio Ortega, and Ralf Seepold. Automatic sleep stages classification using respiratory, heart rate and movement signals. Physiological Measurement, page 124008, 2018.

[8] Yo-El S. Ju, Jennifer S. McLeland, Cristina D. Toedebusch, Chengjie Xiong, Anne M. Fagan, Stephen P. Duntley, John C. Morris, and David M. Holtzman. Sleep Quality and Preclinical Alzheimer DiseaseSleep Quality and Preclinical Alzheimer Disease. JAMA Neurology, 70(5):587–593, 2013.

[9] Karl Kesper, Sebastian Canisius, Thomas Penzel, Thomas Ploch, and W. Cassel. Ecg signal analysis for the assessment of sleepdisordered breathing and sleep pattern. Med. Biol. Engineering and Computing, 50:135–144, 2012.

[10] Yosuke Kurihara and Kajiro Watanabe. Sleep-stage decision algorithm by using heartbeat and body-movement signals. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, pages 1450– 1459, 2012.

[11] Ibrahim Sadek. Nonintrusive Nocturnal Remote Monitoring of Vital Signs in Ambient Assisted Living Environments. Systems and control, Universite Montpellier, 2018

[12] Rechtschaffen Allan and Kales Anthony. A manual for standardized terminology, techniques and scoring system for sleep stages in human subjects. Brain Research Institute. US Dept. of Health, Education, and Welfare, 1968.

[13] Yosuke Kurihara and Kajiro Watanabe. Sleep-stage decision algorithm by using heartbeat and bodymovement signals. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, pages 1450–1459, 2012.

[14] Migliorini M, Bianchi AM, Nisticò D, et al. Automatic sleep staging based on ballistocardiographic signals recorded through bed sensors. Conf Proc IEEE Eng Med Biol Soc. 2010;2010:3273-3276.

[15] K. S. Park, S. H. Hwang, D. W. Jung, H. N. Yoon and W. K. Lee, "Ballistocardiography for nonintrusive sleep structure estimation," 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, 2014, pp. 5184-5187

[16] Kagawa M, Suzumura K, Matsui T. Sleep stage classification by non-contact vital signs indices using Doppler radar sensors. Conf Proc IEEE Eng Med Biol Soc. 2016;2016:4913-4916.

[17] N. Surantha, S. M. Isa, T. F. Lesmana and I. M. A. Setiawan, "Sleep stage classification using the combination of SVM and PSO," 2017 1st International Conference on Informatics and Computational Sciences (ICICoS), Semarang, 2017, pp. 177-182.

[18] R. Yi, M. Enayati, J. M. Keller, M. Popescu and M. Skubic, "Non-Invasive In-Home Sleep Stage Classification Using a Ballistocardiography Bed Sensor," 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI), Chicago, IL, USA, 2019, pp. 1-4.

[19] Ying Chen, Xin Zhu, Wenxi Chen. Automatic sleep staging based on ECG signals using hidden Markov models. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:530-533.