Estimation of Sleep Stages Analyzing Respiratory and Movement Signals

Maksym Gaiduk, Juan José Perea Rodríguez, Ralf Seepold, Natividad Martínez Madrid, Thomas Penzel, Martin Glos, and Juan Antonio Ortega

***Abstract*— The scoring of sleep stages is an essential part of sleep studies. The main objective of this research is to provide an algorithm for the automatic classification of sleep stages using signals that may be obtained in a non-obtrusive way. After reviewing the relevant research, the authors selected a multinomial logistic regression as the basis for their approach. Several parameters were derived from movement and breathing signals, and their combinations were investigated to develop an accurate and stable algorithm. The algorithm was implemented to produce successful results: the accuracy of the recognition of Wake/NREM/REM stages is equal to 73%, with Cohen's kappa of 0.44 for the analyzed 19324 sleep epochs of 30 seconds each. This approach has the advantage of using the only movement and breathing signals, which can be recorded with less effort than heart or brainwave signals, and requiring only four derived parameters for the calculations. Therefore, the new system is a significant improvement for non-obtrusive sleep stage identification compared to existing approaches.**

***Index Terms*—biomedical signal processing, regression analysis, sleep stages, sleep study.**

I. Introduction

SEVERAL studies show the importance of sleep for maintaining good health [1, 2, 3]. They emphasize its duration as an essential key factor to good physiological function [2] and warn about the harmful consequences resulting from abnormalities in the sleep duration [3]. These can be physical and psychological problems, several disease states, and even higher mortality. In this regard, we look to the formulated conclusion given by [2], whereby sleep quantity and sleep quality determine its restorative function and allow us to maintain a good health state. Several aspects should be quantified to assess the quality of sleep, including identifying different stages, their sequence, and the duration of each of them. A standard procedure allows evaluating each sleep stage. This procedure is called polysomnography (PSG), and it is performed following the guidelines of the American Academy of Sleep Medicine (AASM) [4].

During PSG, electroencephalography (EEG), electrocardiography (ECG), electromyography (EMG), and electrooculography (EOG) signals are recorded continuously during sleep [5]. This set of parameters is established because each of the signals changes during each phase of sleep. Thus, the combination of the obtained results allows classifying the subject's stage at each moment.

Although PSG provides an environment for sleep recording and analysis, its implementation is faced with a set of limitations related to:

- logistical and economic cost problems due to the requirement of at least 22 analysis connectors for usual implementation [5, 6];
- the high effort of time and personnel for processing and analyzing of data [7];
- non-natural sleep environment in a sleep lab as well as discomfort due to several electrodes, sensors, and cables attached to the subject's body affecting sleep pattern [8].

Furthermore, there is a limited number of sleep medicine specialists and sleep laboratories [9], which leads to delays in starting the care of patients with sleep disorders. This care is critical because of the large number of undiagnosed sleep disorders resulting in serious problems, even in death [3].

Because non-invasive sleep stage classification is a relevant research topic, many scientific publications can be found on the

This research was partially funded by the EU Interreg V-Program "Alpenrhein-Bodensee-Hochrhein": Project "IBH Living Lab Active and Assisted Living", grants ABH40, ABH41 and ABH66. This research was also partially funded by EDITH Research Project (PGC2018-102145-B-C21,C22 (AEI/FEDER, UE). Thomas Penzel was partially supported by the Russian Federation government mega-grant Nº 075-15-2019-1885.

M. Gaiduk is with the Ubiquitous Computing Lab at HTWG Konstanz, Alfred-Wachtel-Str. 8, 78462 Konstanz, Germany and with the University of Seville, Avda. Reina Mercedes s/n, Seville, Spain (phone: +49 7531 206-703; Email: maksym.gaiduk@htwg-konstanz.de). ORCID: 0000-0001-9607-5709

J.J. Perea is with the University of Seville, Avda. Reina Mercedes s/n, Seville, Spain (Email: jperea@us.es). ORCID: 0000-0003-1630-7579

R. Seepold is with the Ubiquitous Computing Lab at HTWG Konstanz, Alfred-Wachtel-Str. 8, 78462 Konstanz, Germany and Institute of Digital Medicine at I.M. Sechenov First Moscow State Medical University, Moscow, Russia (Email: ralf@ieee.org). ORCID: 0000-0002-5192-2537

N. Martínez Madrid is with the IoT Lab at Reutlingen University, Alteburgstr. 150, 72762 Reutlingen, Germany and the Institute of Digital Medicine at I.M. Sechenov First Moscow State Medical University, Moscow, Russia (Email: nati@ieee.org). ORCID: 0000-0003-1965-9481

T. Penzel is with the Interdisciplinary Sleep Medicine Center, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany and with the Saratov State University, Russia (Email: thomas.penzel@charite.de). ORCID: 0000-0002-4304-0112

M. Glos, is with the Interdisciplinary Sleep Medicine Center, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany. (Email: martin.glos@charite.de). ORCID: 0000-0002-7819-0444.

J.A. Ortega is with the University of Seville, Avda. Reina Mercedes s/n, Seville, Spain (Email: jortega@us.es). ORCID: 0000-0001-9095-7241

subject (e.g. [10-16]). However, they use different vital signals as the input for the classification algorithm. A relatively small part of algorithms can work with only movement and breathing signals, as it is done in our approach. We will present several publications relevant to this research case. The main focus will be on research studies conducted using body movement and respiration as vital signals to recognize sleep stages. Scientific papers utilizing other input signals (e.g., heart rate variability (HRV) in [12] or EEG in [15]) are beyond the scope of this literature review.

Using body movement data for the identification of sleep stages is proposed in [13]. Sixteen healthy people have participated in the experiment. They slept through the night with electrodes attached to their bodies. NapVIEW with an infrared motion sensor was used as the device to detect body movements, placed about nineteen inches away from the subject's head. With this device's help, the body movement density (BMD) was calculated as the number of body movements per time unit to be used as an index of sleep transitions. The results of this study have confirmed the strong relation of BMD with sleep stage transitions. Besides that, the argument has been made that a BMD cycle is less affected by individual variations, and it is, for this reason, a more accurate index for sleep cycle identification than the absolute value of BMD.

Body movement and breathing can be used to classify sleep stages, as presented in [17]. Twenty-five features based on the respiratory signal (depth-based and volume-based) were designed, and the accuracy of 72.3% with Cohen's kappa of 0.34 was obtained for the classification of Wake/NREM/REM stages without subject-specific normalization. If only fourteen best features were used, the accuracy of 71.7% and Cohen's kappa of 0.32 are reported. A considerable correlation between the respiratory signal and sleep stages was demonstrated in [17], a study that constitutes its importance as research on non-invasive sleep phase identification.

Moreover, ultra-low-power reflected radiofrequency waves could be used to determine sleep/wake states according to [16]. The proposed algorithm is based on an analysis of movement and, to a lesser extent, respiration signals. The system's evaluation was performed on 113 subjects (94 males, 19 females with an average age of 53±13years). The accuracy of 78% and Cohen's kappa of 0.38 compared to PSG measurement were achieved.

One study [14] presents the comparison of wrist and chest actigraphy combined with HRV to identify sleep stages. Even if this approach uses HRV and movement signal, it is relevant for our work because of a movement role in identifying sleep stages. A support vector machine (SVM) was proposed as the approach for the classification of sleep and wake stages. The methodology was tested on a group of 18 healthy adult subjects. The accuracies of about 77% for wrist and 78% for chest actigraphy combined with HRV were achieved.

As the respiration signal is analyzed to identify sleep stages, it is essential to know if obstructive sleep apnea (OSA) affects identification accuracy. Otherwise, an algorithm providing accurate results for the patients without OSA could have decreased accuracy for the persons with OSA. This research point was investigated in [18], where the accuracy of 70.9% was achieved, and no significant differences in the identification of sleep stages for non-OSA and OSA groups were noticed. This matter allows transferring the classification results obtained with non-OSA test groups to OSA patients. However, the difference cannot be excluded entirely due to different possible sleep phase identification algorithms.

In [11], an approach of identification of Wake/NREM/REM states is presented. The input signals are respiration and movement, and the extracted out of these signals features used for the classification of sleep stages are the following: respiration rate variability, respiration rate, leg movement, body movement, posture, and body orientation features. The signals used for the algorithm's functioning can be obtained in a non-obtrusive way, for example, using the pressure-sensitive e-textile bed sheet, as it was done in [11]. The experiment with seven subjects (3 male, 4 female, age: 21-60 years old) was performed to evaluate the approach using three different classifiers: K-nearest Neighbor, Support Vector Machines, and Naïve Bayes. The best classification results were achieved using the Naïve Bayes classifier with the precision of 70.3%, recall of 71.1%, and total accuracy of 72.2%.

To overcome the limitations and the problems of a classical PSG approach mentioned above and of alternative presented solutions, the main research aim of this study is defined as a development of techniques for pre-processing and analyzing data in order to detect sleep-related pathologies, with simplified procedures apt even for personnel who are not especially qualified and without the need of expensive laboratory installation. In sum, the diagnostic system under examination is low-cost, and its minimally invasive data collection methods help maintain normal sleep conditions, which in turn increase the accuracy of results.

Considering these premises, we have developed a technique to analyze and classify the sleep stages. The main innovations of our proposal are:

1) reduce the directly recorded signals exclusively to breathing and body movement. Thus, we can reduce the invasive action (compared to EEG or ECG measurement) on the subject to a minimum. Furthermore, these signals may be obtained without necessity of involving trained staff;
2) only four parameters, derived from measured signals, to characterize the sleep stages. This aspect represents a quantitative advantage over existing techniques in which up to more than ten parameters are required [5, 10];
3) develop a new derived parameter, based on the premises of logistic regression that relates two of the primary biological signals: the patient's breathing as well as movement.

The Materials and Methods section contains three subsections. In the Statistical Methods subsection, we will explain the quantitative basis to evaluate the parameters characterizing sleep stages. Once we have described our proposal's quantitative scope, we describe our proposed method in the Proposed Model subsection, justifying both the selection

of the parameters and the coefficient proposed to optimize the results obtained. In the Implementation subsection, we describe the details of the implemented model. The obtained results are then presented in the Results section and discussed in the Discussion section. Finally, we mention the possible improvement options and plans for future research in the Conclusion and Outlook section.

II. METHODOLOGY

A. Statistical Methods

This section describes the statistical methods used in our sleep analysis model and justifies this study's preference for statistical methodology. Furthermore, a qualitative explanation of the statistical method is carried out, and the equations used in our model are defined.

The proposed software system is designed to identify sleep stages out of breathing and movement signals. For that, a set of sensors measuring both these signals (e.g., as described in [19, 20]) can be used to quantify a subject's biosignals. These sensors generate numerically quantifiable electrical impulses, *the sample values*, which change at each of the sleep stages. However, although the values change at each stage, it is only possible to perform a sleep analysis after numerical processing using statistical methods. After processing, a set of *the derived parameters* is obtained, with which it is possible to analyze the evolution of sleep.

From a statistical point of view, the derived parameters are *independent variables*, while the sleep stages are identified as *dependent variables*. A research review was conducted to select the appropriate statistical approach. It showed that several studies had proposed different statistical methodologies to determine the correlation between independent variables (derived parameters) and dependent variables (sleep stages), with regression-based analysis providing the best results compared to other statistical approaches [21, 10]. Notably, in regression-based analysis, the dependent variable is predicted or obtained from a modulated weighting of the available independent variables [22].

Regardless of whether a specific correlation can be established between independent variables, the main objective in the field of regression analysis is to establish a mathematical relationship between dependent and independent variables [22]. Depending on the numerical characteristics of the independent variables, there are different statistical methods to obtain this relationship between dependent and independent variables. In our study, we are going to focus on multinomial logistic regression (MLR). MLR is an extension of the logistic regression with which it is possible to manage independent variables (*the derived parameters*) in multiple categories. This choice is due to two factors: the first is that the outcome of a multicategory variable (a variable that can achieve a limited number of categories) needs to be predicted as a function of the independent variables; the second is that these independent variables form a random conditional field or set [23]. In the case of sleep studies, a set is related to vital functions such as breathing or movement of the subject under study. Another advantage of MLR compared to many other algorithms is the fact that the output is not presented as just one value with the detected sleep stage but as a set of probabilities for all sleep stages for each epoch. This allows a further processing and adjustment of the results.

Quantitatively in MLR, the odds ratio is the logistic regression coefficient's antilogarithm, which simplifies the calculations and allows a better obtaining of the dependent variables. It should be noted that the odds ratio presents information on how the change of independent variable influences the probability of being in a category versus being in the reference group. If an odds ratio value is greater than 1, then as an independent variable in the interval increases, so does the probability value. Otherwise, if the odds ratio is less than 1, then there is an opposite scenario - the probability value decreases as an interval-independent variable increase [24].

To quantify the description made of MLR, we will introduce the equations that we will implement in our study. To carry out this task, we will designate the results of the categorical random dependent variables as $Y \in \{0, 1, 2 \dots k\}$.

According to MLR, the conditional mean of the dependent categorical variables is defined as a logistic function, of a related combination, of independent variables, usually called x. The relationship is defined as

$$E[Y|x] = \sigma(c^T x), \tag{1}$$

where c is the vector of unknown coefficients and where the logistic function, σ, satisfies:

$$\sigma(x) = \frac{1}{1+e^{-x}}. \tag{2}$$

Considering (1) and (2), it is possible by means of MLR to determine the coefficients vector c that maximizes the probability of observations. This operation is defined as

$$\prod_{i=1}^{n} Pr(Y = Y_i|X_i), \tag{3}$$

which is equal to the maximization of the probability of registration, which is estimated as

$$\sum_{i=1}^{n} log\ P\,r(Y = Y_i|X_i). \tag{4}$$

This expression is used to formulate the simplified function to maximize the probability:

$$l(c) = -\sum_{i=1}^{n} log\left(1 + e^{-1^{y_i} * c^T x_i}\right). \tag{5}$$

Additionally, in MLR, the standard error coefficient is of particular importance, which is calculated as

$$se(c_i) = ((X^T AX) - 1)_{ii}, \tag{6}$$

where the Hessian of $H = -X^T AX$ and $A = diag(a_1, \dots, a_n)$ is the diagonal matrix with $a_i = \sigma(c^T x) * \sigma(c^T x)$ [25].

Finally, to know the maximum probability, it is necessary to test the predictors' meaning in the logistic regression [23, 10]. For this purpose, the *Wald z-test* [25] defined by the equation is usually used:

$$z_i = \frac{c_i}{se(c_i)}. \tag{7}$$

Knowing this last coefficient, it is necessary to demonstrate the predictors' meaning in the logistic regression. The value of the so-called *Wald p-value* for the coefficient i is used. The probability of obtaining a value at least as extreme as the observed one is determined, for which the null hypothesis has

to be imposed, that is, $(c_i = 0)$.

The formula for the value of the *Wald p-value* for the coefficient i is

$$p_i = Pr(|Z| \geq |z_i|) = 2 * \left(1 - F(|z_i|)\right), \quad (8)$$

where F is the cumulative density function of a standard normal distribution [25].

B. Proposed model

The characteristics of the proposed model are described here. We will justify the choice of the parameters derived for identifying each of the three sleep stages and how we can drastically reduce the number of these parameters from 10 to 4 compared to the previous research [10]. We will also introduce a new coefficient that allows us to compare two sources of information, breathing and body movement; we will justify the formulation of this coefficient and describe how this new coefficient can improve the accuracy of the results obtained.

Several investigations have revealed a relevant fact from the analytical point of view: *there is a direct correlation between biosignal patterns, quantifiable using electronic sensors, and the different sleep stages*. Thus, from this correlation of sleep stages-biosignal patterns, it is possible to detect and quantify the wakefulness-sleep situation in which a subject of study can be found. To carry out this task, it is necessary to implement recognition algorithms analyzing parameters obtained from biosignals, quantify and detect the different sleep stages. Some of the essential characteristics of biosignals are summarized in Table 1 extracted from [26-29].

TABLE I
RELEVANT CHARACTERISTICS OF SLEEP STAGES [26-29]

Stage	Characteristics
Wake	Respiration is typically stable and more frequent than in the NREM (especially in Deep Sleep) stage. Typically, more movement than in NREM (especially in Deep Sleep) and REM stages.
NREM	When sleep deepens from the WAKE stage, body movements become smaller and less frequent. The deeper the sleep, the less frequent the heart rate and respiration. After some instability in respiration during the Light Sleep stage, respiration becomes more stable when sleep deepens.
REM	Heartbeat and respiration become more frequent and less rhythmical. Body movement is typical for the epochs just before and directly after the REM stage. Anti-gravity muscles lose the tension, and therefore, the movement is typically absent.

As a result, we have developed a minimally invasive identification approach in which we only need two sources of biosignal data: (a) body movement, (b) respiration.

This selection is based on specific changes in physiological features of the human body caused by sleep. When moving from the wake stage to deeper sleep, both a body's movement (amplitude and frequency) and, at the same time, a heart rate are being decreased [26], so only one of these two data sources needs to be used. To achieve our goal of using signals that can be obtained with less effort and more comfort, we have decided to use a movement signal and refrain from using a heart signal. Furthermore, after a transition from NREM to REM stage, body movement is typically decreased even more or absent [27]. The absence of movement in the REM stage also helps distinguish it from the WAKE stage with a high level of accuracy.

Additionally, to obtain more accurate results, we use a respiratory signal and have developed a new derived parameter, integrating body movement and breathing. This will be described later and presented in (16). It is known from [26] that respiration becomes more frequent and less rhythmical in the REM stage, leading to changes in respiratory volume and supports the differentiation between NREM and REM stages.

To quantify body movement in our model, which builds on the research developed by [10], the average value of the position changes for all three axes will be calculated according to the equations:

$$X = \frac{1}{n}\sum_{i=1}^{n}|x_i - x_{i-1}| \quad (9)$$

$$Y = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_{i-1}| \quad (10)$$

$$Z = \frac{1}{n}\sum_{i=1}^{n}|z_i - z_{i-1}|, \quad (11)$$

where x_i, y_i, z_i and x_{i-1}, y_{i-1}, z_{i-1} are the sensor coordinates in the current and correspondingly previous moment.

From the (9), (10), (11), we calculate the instantaneous position of the body at the moment i as the modulus of the changes in coordinates, i.e.:

$$Body_i = \sqrt{X^2 + Y^2 + Z^2}, \quad (12)$$

where X, Y and Z are the signal values for each axis of a 3D activity sensor used in the study. If the signal provided by the sensors measuring movement is 1-dimensional, it can be directly used as the input for (13).

The derived body movement parameter is defined as the mean value of the body movement (BM) signal. It is defined as

$$BM(k) = \frac{1}{n}\sum_{i=0}^{n-1} Body_i, \quad (13)$$

where n is the number of body movements and $Body_i$ is described in (12).

Additionally, to quantify respiration, we will consider the recommendations given by [17, 10]. In these studies, up to four parameters are formulated from the data provided by the breathing sensors. The innovation of our proposal is that we only need two of these parameters to obtain satisfactory results. Performed tests with different sets of breathing-related parameters have indicated no significant improvement of the classification rate in increasing the number of used derived parameters. Therefore, in our proposal, we will only need

- Mean respiratory depth of exhalation (T_{sdm})
- Median respiratory volume during inhalation (V_{in})

The choice of these parameters is conditioned by several clinical studies in [17, 26] in which the importance of considering mean respiratory depth, mainly in terms of exhalation, is highlighted, so we have chosen T_{sdm}, as well as the Median respiratory volume during inhalation, should already reflect the changes of respiratory effort signals and these, in turn, are related to the stages of sleep [26]. Selecting only two parameters instead of the four proposed by [26] does not imply loss of information since the biosignal data are related to each other [30], so a proper selection allows knowing all the information, as we will demonstrate in the Results section.

Here is the mathematical representation of the derived breathing parameters,

$$T_{sdm}(k) = \frac{median(t_1,t_2,\dots t_n)}{IQR(t_1,t_2,\dots t_n)}, \quad (14)$$

$$V_{in} = (k)median\left(\sum_{s_x\in\Omega_1^{in}} s_x, \sum_{s_x\in\Omega_2^{in}} s_x, \dots, \sum_{s_x\in\Omega_k^{in}} s_x\right), \quad (15)$$

where $t = t_1, t_2, \dots, t_n$ are the sequences of peaks and troughs of a selected time window, IQR is the interquartile range for the given sequences of peaks, the k^{th} inhalation and exhalation cycle is defined through Ω^{in}_k with k consecutive breathing cycles ($k = 1, 2, \dots, K$) and S_x is the respiratory effort value.

It is important to note that not absolute values of the derived from respiratory signal parameters are significant for the correct functioning of the proposed algorithm, but their changes over time. Furthermore, all signals are being subject-normalized and do not contain raw values. Therefore, even not calibrated sensors (e.g., inductive plethysmography) can be used for signal recording.

However, with these three parameters, defined in (13), (14), and (15), it is impossible to obtain precise results that allow the determination of the sleep phase. In order to avoid this problem and achieve a higher precision from the three derived parameters, we formulated a new derived parameter BV_{in} which represents the combination of the inhalation parameters V_{in} and the $BM(k)$ according to the equation,

$$BV_{in} = ln\frac{BM(k)}{BM(k)+V_{in}}. \quad (16)$$

This choice is not arbitrary but is based on two fundamental conclusions: the first is that when the subject under study evolves from the waking stage to that of a deep sleep, the movement of the body decreases, and the depth of breathing increases [17, 26] so that the combination in the form of a coefficient will improve the accuracy of detecting the sleep stage. The amplitudes of the breath and body movement signals can vary significantly between subjects. To standardize these deviations, $BM(k)$ is divided by $BM(k) + V_{in}$ The second is that using the coefficient as an argument for a logarithmic function reduces the large fluctuations associated with body movement, as highlighted by [26].

Tests were also conducted in which all selected parameters were excluded from the algorithm one by one. This resulted in a reduction in accuracy of about 10-15%, with no clearly detectable differences depending on which parameter was excluded.

Thus, by calculating three derived parameters and a fourth obtained from them, we will demonstrate that accurate results are obtained, as presented in the section Results.

C. Implementation

In our study, we have implemented a classification model based on the statistical technique of MLR, in which we identify as dependent variable (Y) the various stages of sleep, while the independent variables $(X_1, X_2, \dots, X_n)$ are the biosignal parameters which in our case are derived out of breathing and body movements, as described in the section Proposed Model. To calculate the derived parameters, the data obtained with a sampling rate of 32Hz was split into epochs of 30 seconds each.

The general structure of the system modules is presented in Figure 1. First, the input signals are pre-processed to facilitate further elaboration, then the derived parameters are calculated, and the calculation of probabilities of sleep phases for every epoch is performed. After each step, the results of the

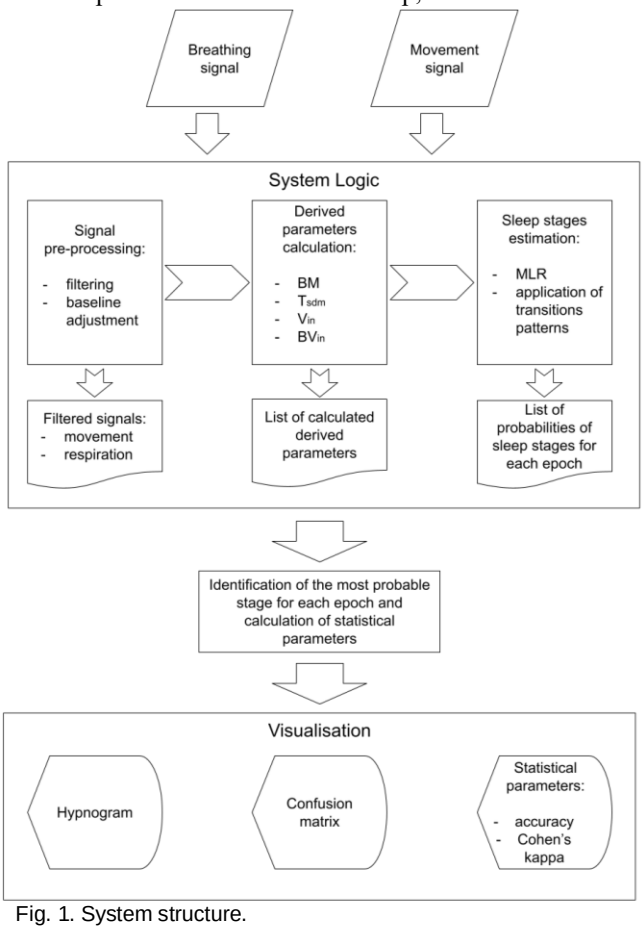

Fig. 1. System structure.

calculations are saved in a separate file to enable a subsequent analysis of the operations carried out and simplify the exchange of system modules if required. The next step is to determine the most probable sleep phase for each epoch and finally to visualize the results.

The input signals can be recorded independently or as a part of a polysomnography study. For breathing signals, a chest inductive plethysmography (RIP) record may be used. Other sensors that can record respiratory signals, including changes in their amplitude, can also be applied [e.g. 19, 20]. Body movement signals can be replaced by a signal monitored by a 3D acceleration sensor in a recording device placed on the subject's chest. Another option for the measurement of movement signal could be pressure sensors placed under the mattress [19]. The baseline wandering should be excluded from the lower frequency range and not be significant for the final results, as only the changes in the signal will be analyzed and not its absolute value. Also, the baseline shift has a much smaller amplitude than body movements. A median filter was applied to the respiratory signal to remove short, high-amplitude body movements. Afterward, a rolling mean filter was applied to get a cleaner signal, which was beneficial for calculating the selected derived parameters.

To increase the accuracy of the classification results, the

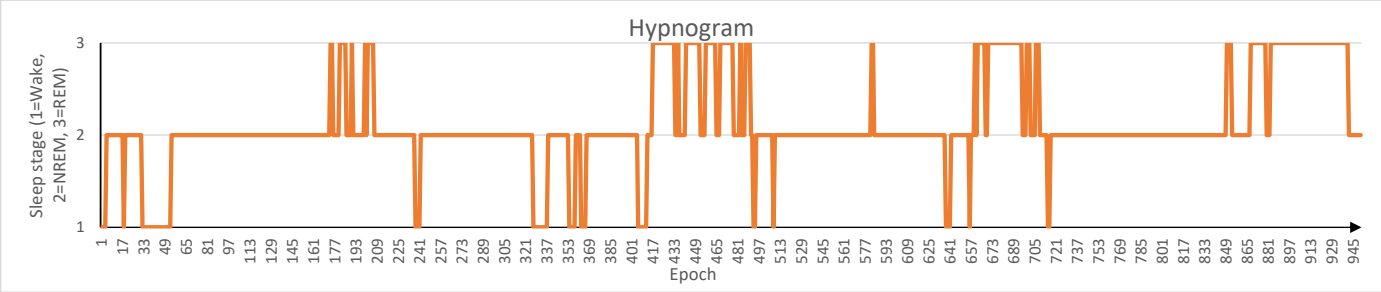

Fig. 2. Results of classification of sleep stages for one subject visualized as a Hypnogram.

transition patterns between the different sleep stages were taken into account in implementing the algorithm described in this work. As shown in [31], some transitions between sleep stages are much more likely than others, and at the same time, some transitions have a very low probability. In the first implementation of the algorithm, the emphasis was placed on non-probability transitions. As the algorithm works with probabilities, it allows additional adjustments, increasing accuracy without having to comply with strict rules. In this case, the algorithm reduced the probability of the examined sleep stage by 10-15% (a higher percentage would negatively affect the total classification accuracy if the previous stage were incorrectly classified) if its probability of occurrence according to the transition pattern is almost zero. This approach has led to an increase in the proposed algorithm's accuracy by up to 3% compared to implementation without considering transition probabilities.

III. Results

The evaluation of the algorithm's work was performed with the use of the dataset provided by the Charité[1] clinic in Berlin. The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board of the Charité-Universitätsmedizin Berlin (application number: EA1/320/114). The overnight recordings of 35 persons (with a total length of about 260 hours) were available for evaluation. The participants' average age was 38.6 +/- 14.5 years old, and the BMI averaged 24.4 +/- 4.9 kg/m². The number of male and female subjects was similar, and no significant health disorders were known for the test persons. The used PSG recordings were previously manually analyzed by sleep medicine physicians, and every 30-second epoch was tagged with the corresponding sleep stage (Wake, N1, N2, N3, REM). As the classification algorithm operates with three stages (Wake, NREM, REM), the stages N1, N2, and N3 from the PSG recordings were merged to the evaluation's NREM stage.

The available dataset was strictly separated into training and test subsets. The training dataset consisted of sleep data of randomly selected subjects (after separation male/female) with a total amount of about 100 hours of recordings. Typically, the NREM sleep stage is the prevalent one during sleep, resulting in classification errors due to unbalanced classes in the training dataset. Therefore, it was necessary to balance the subset used for the system training, which was performed during the evaluation preparation phase. The test dataset included about 160 hours of overnight recordings of 20 subjects, which corresponds to 19324 epochs, 30 seconds each. Both subsets had a similar male/female ratio, and the average BMI of the included subjects did not have a significant discrepancy.

Visual representation of sleep stages estimation is presented in Figure 2. It can be seen that there are multiple rapid transitions from one state to another at some points in time. It happens because there is no algorithm for averaging the results of sleep stage classification implemented, and every single epoch is directly visualized according to calculated values.

Table 2 represents the classification results of the developed algorithm compared to the expert classification. The rows of the table contain the number of epochs corresponding to each sleep stage according to the expert's evaluation, whereas the columns represent the results of classification by the developed algorithm. The main diagonal indicates the number of sleep epochs, where the results of both classifications (by experts and algorithm) are in accordance.

The achieved general accuracy is equal to 73% following the expert's classification. More statistical measurements are presented in Table 3.

In addition to the classification accuracy, we estimated Cohen's kappa parameter, which is commonly used to measure agreement between several observers (in this case – methods). Its calculated value for the developed classification algorithm compared to the results of experts' classification is 0.44.

TABLE II
CLASSIFICATION RESULTS

Stage Expert	Stage developed system			
	Wake	NREM	REM	Total
Wake	1256	862	604	2722
NREM	1104	11081	1691	13876
REM	211	730	1785	2726
Total	2571	12673	4080	19324

TABLE III
STATISTICAL MEASUREMENTS

Overall accuracy	73.0 %	
	Recall	Precision
Wake	46.1 %	48.9 %
NREM	79.9 %	87.4 %
REM	65.5 %	43.8 %

[1] Initial study was carried out in Charité - Universitätsmedizin Berlin Center of Sleep Medicine Charitéplatz 1, D-10117 Berlin (Germany).

IV. Discussion

This manuscript research aimed to provide a scientific base for the development of a sleep study system that could be used in medical or home environments. Monitoring movement and breathing of recumbent subjects in a non-obtrusive way is less challenging than monitoring heart signals (especially HRV) [32]. It was imperative to develop an algorithm that could classify sleep stages relying on only these two signals for input, the target of the performed research.

The achieved general accuracy of algorithm function of 73% indicates that the goal of this challenging task was met. The developed algorithm recognized the NREM stage with high recall of about 80%, which is higher than for both other stages – Wake and REM. The overestimation of the NREM phase can partially explain its prevalence in a typical sleep pattern. Even using a balanced training dataset did not completely solve this problem. Another critical point is that human respiration and movement during the Wake stage just before falling asleep and during the NREM1 stage, which is a part of the NREM stage performed in this work evaluation, are very similar [33]. It may lead to misclassification of the Wake stage as the NREM stage and vice versa. Significant differences in accuracy for different sleep stages are also typical for other systems (e.g. [11, 40]).

Another factor that has a strong influence on the results is the quality of the input signals. When analyzing the recognition rate per person, it was detected that there are some recordings where the results are significantly worse than the average. Accordingly, such exceptional cases have driven down the average accuracy and average Cohen's kappa value. After a thorough analysis of such recordings, it was found that the signals had significant differences from other typical recordings. Figure 3 shows a comparison of two recordings of approximately 130 minutes each of the respiratory signal compressed in time. The upper graph in Figure 3 illustrates a typical recording (accuracy of classification 76%) in which the amplitudes of the respiratory signal have few outliers and can therefore be adequately analyzed by the algorithm. However, when looking at the lower graph in Figure 3 (classification accuracy 49%), one can notice that the signal is volatile and contains many disturbances that make it almost impossible to analyze the amplitude of the signal (and also its volume) in numerous signal sections.

Nevertheless, the developed algorithm has proven that the accuracy of the classification of sleep stages remains adequate even with the inclusion of such limiting cases. If we would exclude this kind of cases, we could significantly increase the accuracy. However, the presented work aimed to test the functionality of the developed system under natural conditions, which also means the inclusion of low-quality signals.

Cohen's kappa value of 0.44 (which is “fair to good” according to [34]) may be considered a good result because a very high value can be challenging to achieve for epoch-by-epoch sleep stage identification. To clarify, even for the AASM scoring standard, the evaluation by different experts has an overall level of agreement of 82% with Cohen's kappa equal to 0.76 [35]. It is, however, important to mention that in [35], five stages were considered. It is expected that a sleep stage identification with only movement and breathing signals as input will have lower levels of accuracy and Cohen's kappa than the gold standard, which uses PSG signal as input.

The main novel point proposed in this study's approach is that it uses only the signals that can be obtained in a non-obtrusive way: movement and breathing [19, 20], for which a unique set of parameters was designed. For that purpose, comprehensive literature research and statistical analysis were performed. The new combination of derived parameters and MLR-approach, extended by algorithms for considering transition probabilities from one sleep stage to another, has led to the development of a new unique software solution for identifying sleep stages.

A significant amount of research on automatic sleep stage classification is based on using the EEG signal as the algorithm's input [36-38]. As EEG can be recorded only in an obtrusive way with electrodes attached to the subject’s head, these researches are not directly comparable with that presented in this article’s approach.

Another large subset of sleep research considers the identification of sleep stages, having the heart rate or ECG signal as the input. In [39], the authors used the ECG signal and respiration effort. The accuracy of 80% for the classes Wake,

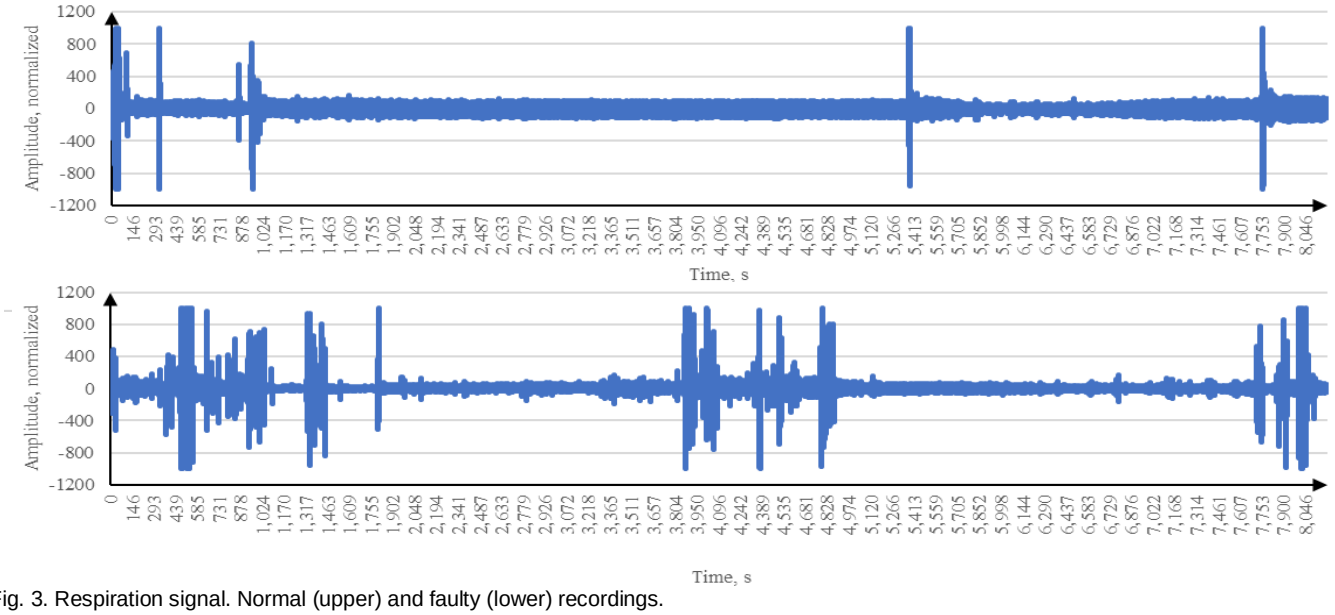

Fig. 3. Respiration signal. Normal (upper) and faulty (lower) recordings.

NREM, and REM is slightly higher than in our approach. However, it would not be possible to perform the measurement described in [39] in a non-obtrusive way because the ECG signal is being measured directly by placing electrodes on the human body. Therefore, our method has a significant advantage as the necessary signals may be measured in a contactless way as described in [19, 20], while reporting only a slight reduction of accuracy. Another crucial advantage of our approach is using only four features, compared to 80 in [39], which means faster processing and less computing power needed, which is essential for the embedded systems.

To continue, the approach described in [40] uses the HRV signal derived from the ECG. The achieved accuracy of 72,6% for subject independent classifier is slightly lower than the accuracy of the method presented in our paper. Though the accuracy is comparable, our method is preferable since it uses only respiration and movement signals (which can be obtained contact-free). Furthermore, the algorithm's number of features used is significantly lower than proposed in [40] (4 against 41).

Breathing and body movement signals are used as the input in [16] approach, making it similar to our method concerning analyzed signals. However, only Wake/Sleep states are recognized in [16], whereas our system identifies Wake, NREM, REM sleep stages. The accuracy of 78%, with a Cohen's kappa of 0.38 stated in [16], compared to 73% of accuracy with a Cohen's kappa of 0.44 in our experiment, confirms the quality of our results. The number of recognized stages in our method is higher, albeit with a marginal decrease in accuracy, and it achieved a better Cohen's kappa value.

The reported results in [17] regarding the accuracy of 72.3% with Cohen's kappa of 0.34 (25 features), and accuracy of 71.7% with Cohen's kappa of 0.32 (14 features) for the classification of Wake/ NREM/REM stages without subject-specific normalization is lower than in our work. Our work also presents data using fewer used features (4 against 25/14), another advantage over other studies. The respiratory signal is used in [17] as the input for the algorithm.

Compared to the approach presented in [11], our method uses the same input signals but indicates better results: 73% accuracy against 72%, using only four features compared to more than 30 in [11]. Furthermore, a significantly higher number of overnight recordings was used to evaluate our experiment (20 against 7), which illustrates the results' stability.

It is important to note that a reduction in the number of the used features is especially relevant when it does not have any negative impact on the accuracy of detection because the effort required to calculate the features is moderate. The results of the comparison with state-of-the-art solutions presented above confirm that our system can use a small number of parameters and demonstrate high accuracy at the same time.

Although the results of the work carried out are promising, some limitations should be addressed:

- The number of recordings for the training was limited, and therefore it cannot be excluded that with a significantly larger quantity, the training results could be different.
- Changes in the signal in the process of recording (e.g. due to the displacement of the RIP belts) can lead to a reduction in accuracy.
- The available recordings were divided into 30-second intervals, which is also typical for sleep analysis. However, transitions from one sleep stage to another can also occur within these intervals, which was not considered in the current implementation of the algorithm.
- Although the possibility of substituting the devices for recording the movement and respiration signals is pre-planned, it was not possible to use and compare several alternative devices for the recording within the scope of the study conducted. Therefore, it cannot currently be claimed that the results of the algorithm's work will be identical for different signal sources.

V. Conclusion and Outlook

The development of an algorithm for identifying Wake/NREM/REM sleep stages was performed, and the accuracy of 73% with the Cohen's kappa value of 0.44 was obtained. These results confirm the proposed approach's appropriateness for the defined use case with only breathing and movement signals as the input. Further investigation on this method promises the algorithm's improvement and, consequently, more accurate results of system work.

One possible way to enhance the algorithm's performance is to continue investigating the best set of used derived parameters and the development of new ones. However, this needs an in-depth investigation of the importance of different features and values in the general sleep stages recognition algorithm. This work was done, but further improvements should still be possible and are planned for future projects.

Moreover, other sleep stages recognition approaches could be combined with the one proposed in this study to get an extended algorithm, which would obtain the final identification results as a mix of two or more methods, e.g., MLR + neural network. In this case, the epochs with the same identified stages by both approaches could be used as "trust-anchors," and the epochs before and after them could be identified more precisely, considering the transitions between the different sleep states [28]. Having three or more combined approaches, the decision system could be implemented, determining the final sleep stage identification result as the election by the majority of algorithms.

As shown in the Discussion section, the quality of the input signal plays an essential role in the correct classification of sleep stages. Therefore, in the future version of the system, it is planned to label the recordings with the disturbed signals and, if necessary, to remove some parts of the signal from the analysis and to provide a corresponding notification if no signal correction should be possible.

Signal reconstruction for improving the quality of the signal is a topic in itself and could also be included in future research. This would make it possible to perform a good detection of sleep phases even in parts of the signal where no reliable evaluation was initially possible.

Another possible direction of future work is to develop a standalone system tuned to a home environment [41], which requires developing a hardware part of the system to complement the software. This development's significant progress is already made and presented in several scientific publications [19, 42]. The final aim is to gain a system that

would recognize sleep stages and sleep apnea, available for use in a home environment without high financial and personnel costs. Furthermore, the system should be comfortable to use, which has led to selecting the signals that can be obtained in a non-obtrusive way – movement and breathing. This kind of system could be widely used to provide the necessary information to medical professionals, thus enabling early detection of sleep problems and ultimately improving the population's quality of sleep.

ACKNOWLEDGMENT

We thank the Interdisciplinary Center for Sleep Medicine of Charité-Universitätsmedizin Berlin and in particular Beate Diecker and Claudia Biro for supporting the study.

REFERENCES

[1] K. K. Gulia and V. M. Kumar, "Importance of Sleep for Health and Wellbeing Amidst COVID-19 Pandemic," *Sleep and vigilance*, pp. 1–2, 2020, doi: 10.1007/s41782-020-00087-4.

[2] National Heart, Lung, and Blood Institute, *Your guide to healthy sleep*. Washington D.C.: U.S. Dept. of Health and Human Services National Institutes of Health National Heart Lung and Blood Institute, 2011.

[3] T. L. Lee-Chiong, C. Mattice, and R. Brooks, *Fundamentals of Sleep Technology*. Philadelphia: Wolters Kluwer, 2015.

[4] R. B. Berry *et al*., *The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications: V. 2.6*. Darien, Illinois: American Academy of Sleep Medicine, 2020.

[5] S. Chokroverty, R. J. Thomas, and M. Bhatt, *Atlas of Sleep Medicine,* 2nd ed. London: Elsevier Health Sciences, 2013.

[6] R. B. Berry, "Indications for Polysomnography, Portable Monitoring, and Actigraphy," *Sleep Medicine Pearls (Third Edition)*, 2015.

[7] M. Hirshkowitz, "Polysomnography Challenges," *Sleep medicine clinics*, vol. 11, no. 4, 2016, doi: 10.1016/j.jsmc.2016.07.002.

[8] O. Le Bon *et al.,* "The first-night effect may last more than one night," *Journal of Psychiatric Research*, vol. 35, no. 3, pp. 165–172, 2001, doi: 10.1016/S0022-3956(01)00019-X.

[9] J. Singh *et al.,* "American Academy of Sleep Medicine (AASM) Position Paper for the Use of Telemedicine for the Diagnosis and Treatment of Sleep Disorders," *Journ of clinical sleep med: official publication of the AASM*, vol.11, no.10, 2015, doi: 10.5664/jcsm.5098.

[10] M. Gaiduk *et al*., "Automatic sleep stages classification using respiratory, heart rate and movement signals," *Physiological meas*, vol. 39, no. 12, p. 124008, 2018, doi: 10.1088/1361-6579/aaf5d4.

[11] L. Samy *et al.*, "Unobtrusive Sleep Stage Identification Using a Pressure-Sensitive Bed Sheet," *IEEE Sensors J.*, vol. 14, no. 7, pp. 2092–2101, 2014, doi: 10.1109/JSEN.2013.2293917.

[12] M. Radha *et al.,* "Sleep stage classification from heart-rate variability using long short-term memory neural networks," *Scientific reports*, vol. 9, no. 1, p. 14149, 2019, doi: 10.1038/s41598-019-49703-y.

[13] Y. Kambayashi and H. Hagiwara, "Estimating sleep cycle using body movement density," in *2012 5th International Conf. on BioMedical Engin and Informatics*, Chongqing, China, Oct. 2012, pp. 1081–1085.

[14] M. Aktaruzzaman *et al.,* "Performance comparison between wrist and chest actigraphy in combination with heart rate variability for sleep classification," *Computers in biology and medicine*, vol. 89, pp. 212–221, 2017, doi: 10.1016/j.compbiomed.2017.08.006.

[15] N. Michielli, U. R. Acharya, and F. Molinari, "Cascaded LSTM recurrent neural network for automated sleep stage classification using single-channel EEG signals," *Computers in biology and medicine*, vol. 106, pp. 71–81, 2019, doi: 10.1016/j.compbiomed.2019.01.013.

[16] P. de Chazal *et al.,* "Sleep/wake measurement using a non-contact biomotion sensor," *Journal of sleep research*, vol. 20, no. 2, pp. 356–366, 2011, doi: 10.1111/j.1365-2869.2010.00876.x.

[17] X. Long *et al.*, "Analyzing respiratory effort amplitude for automated sleep stage classification," *Biomedical Signal Processing and Control*, vol. 14, pp. 197–205, 2014, doi: 10.1016/j.bspc.2014.08.001.

[18] S. H. Hwang *et al.*, "Unconstrained Sleep Stage Estimation Based on Respiratory Dynamics and Body Movement," *Methods of information in medicine*, vol. 55, no. 6, 2016, doi: 10.3414/ME15-01-0140.

[19] M. Gaiduk, R. Seepold, N. Martínez Madrid, S. Orcioni, and M. Conti, "Recognizing Breathing Rate and Movement While Sleeping in Home Environment," in *Lecture Notes in Electrical Engineering,* Cham: Springer International Publishing, 2020, pp. 333–339.

[20] M. Gaiduk, D. Wehrle, R. Seepold, and J. A. Ortega, "Non-obtrusive system for overnight respiration and heartbeat tracking," *Procedia Computer Science*, vol. 176, pp. 2746–2755, 2020, doi: 10.1016/j.procs.2020.09.282.

[21] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis,* 5th ed. New York, NY: Wiley, J, 2011.

[22] A. von Eye and C. Schuster, *Regression analysis for social sciences*. San Diego, Calif: Academic Press, 1998.

[23] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, *Applied logistic regression*. Hoboken, NJ: Wiley, 2013.

[24] M. H. Katz, *Multivariable analysis: A practical guide for clinicians and public health researchers,* 3rd ed.: Cambridge Univ. Press, 2011.

[25] Garson G. D., *Logistic Regression: Binary & Multinomial: 2016 Edition*: Statistical Associates Publishers, 2016.

[26] Y. Kurihara and K. Watanabe, "Sleep-Stage Decision Algorithm by Using Heartbeat and Body-Movement Signals," *IEEE Trans. Syst., Man, Cybern. A*, vol. 42, no. 6, pp. 1450–1459, 2012, doi: 10.1109/TSMCA.2012.2192264.

[27] A. Sadeh, "The role and validity of actigraphy in sleep medicine: an update," *Sleep medicine reviews*, vol. 15, no. 4, pp. 259–267, 2011, doi: 10.1016/j.smrv.2010.10.001.

[28] H. C. Dringenberg, *Handbook of sleep research*. London: Elsevier Academic Press, 2019.

[29] N. J. Douglas *et al.*, "Respiration during sleep in normal man," *Thorax*, vol. 37, no. 11, pp. 840–844, 1982, doi: 10.1136/thx.37.11.840.

[30] D. A. Belsley, *Conditioning diagnostics: Collinearity and weak data in regression*. New York: Wiley, 1991.

[31] A. Schlemmer *et al.*, "Changes of sleep-stage transitions due to ageing and sleep disorder," *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, vol. 373, no. 2034, 2015, doi: 10.1098/rsta.2014.0093.

[32] I. Sadek, J. Biswas, and B. Abdulrazak, "Ballistocardiogram signal processing: a review," *Health information science and systems*, vol. 7, no. 1, p. 10, 2019, doi: 10.1007/s13755-019-0071-7.

[33] R. D. Ogilvie, R. T. Wilkinson, and S. Allison, "The detection of sleep onset: behavioral, physiological, and subjective convergence," *Sleep*, vol. 12, no. 5, pp. 458–474, 1989, doi: 10.1093/sleep/12.5.458.

[34] J. L. Fleiss *et al.*, *Statistical Methods for Rates and Proportions*. Hoboken, NJ, USA: John Wiley & Sons, Inc, 2003.

[35] H. Danker-Hopfe *et al.,* "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard," *Journal of sleep research*, vol. 18, no. 1, pp. 74–84, 2009, doi: 10.1111/j.1365-2869.2008.00700.x.

[36] X. Li *et al.*, "HyCLASSS: A Hybrid Classifier for Automatic Sleep Stage Scoring," *IEEE journal of biomedical and health informatics*, vol. 22, no. 2, pp. 375–385, 2018, doi: 10.1109/JBHI.2017.2668993.

[37] H. Dong *et al.*, "Mixed Neural Network Approach for Temporal Sleep Stage Classification," *IEEE transactions on neural systems and rehabilitation engineering: a publication of the IEEE Engineering in Medicine and Biology Society*, vol. 26, no. 2, pp. 324–333, 2018, doi: 10.1109/TNSRE.2017.2733220.

[38] M. Diykh, Y. Li, and S. Abdulla, "EEG sleep stages identification based on weighted undirected complex networks," *Computer methods and programs in biomedicine*, vol. 184, p. 105116, 2020, doi: 10.1016/j.cmpb.2019.105116.

[39] P. Fonseca, X. Long, M. Radha, R. Haakma, R. M. Aarts, and J. Rolink, "Sleep stage classification with ECG and respiratory effort," *Physiological measurement*, vol. 36, no. 10, pp. 2027–2040, 2015, doi: 10.1088/0967-3334/36/10/2027.

[40] M. Xiao, H. Yan, J. Song, Y. Yang, and X. Yang, "Sleep stages classification based on heart rate variability and random forest," *Biomedical Signal Processing and Control*, vol. 8, no. 6, pp. 624–633, 2013, doi: 10.1016/j.bspc.2013.06.001.

[41] M. Conti, S. Orcioni, N. M. Madrid, M. Gaiduk, and R. Seepold, "A Review of Health Monitoring Systems Using Sensors on Bed or Cushion," in *Lecture Notes in Computer Science,* Cham: Springer International Publishing, 2018, pp. 347–358.

[42] M. Gaiduk *et al.,* "Embedded system for non-obtrusive sleep apnea detection," *IEEE Engineering in Medicine and Biology Society. Annual International Conference*, vol. 2020, pp. 2776–2779, 2020, doi: 10.1109/EMBC44109.2020.9176075.