IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received August 18, 2019, accepted September 20, 2019, date of publication September 30, 2019, date of current version October 11, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2944801

Computational Sleep Behavior Analysis: A Survey

SARAH FALLMANN[1], (Student Member, IEEE), AND LIMING CHEN[2], (Senior Member, IEEE)
[1]School of Computer Science and Informatics, De Montfort University, Leicester LE1 9BH, U.K.
[2]School of Computing, Ulster University, Belfast BT15 1ED, U.K.

Corresponding authors: Sarah Fallmann (sarah.fallmann@gmail.com) and Liming Chen (l.chen@ulster.ac.uk)

This work was supported by the European Union's Horizon 2020 Research and Innovation Program through the Marie Sklodowska-Curie Grant under Agreement 676157.

ABSTRACT Sleep is a key marker of health, as it can either be a cause or a consequence. It is traditionally studied in clinical environments using dedicated medical devices. Recent technological developments, e.g., in sensing and data analysis, have led to new approaches for sleep monitoring and assessment, which are attracting increasing attention in the emerging domain of personalized smart healthcare. Nevertheless, a high-level overview of technology-enabled research on sleep that can inform related communities of the latest developments is lacking. In this paper, we present a comprehensive review to examine the current status of various aspects of technology-based sleep research. We first characterize sleep behavior and key areas of sleep assessment, and we introduce a general review of the methodologies used in this domain. We review the major technological methods and trends associated with sleep monitoring, data collection and sleep behavior analysis, from which strengths and weaknesses are highlighted. Finally, we also discuss challenges and promising directions for future research.

INDEX TERMS Sleep behavior analysis, home environment, wearables, polysomnography, actigraphy, sleep stage classification, sleep positions, sleep disorders, disease recognition, data mining, machine learning, deep learning, sleep monitoring, sleep parameters.

I. INTRODUCTION

Sleep influences people's lives but still remains mysterious in many ways. It is a recovery mechanism in which heart rate and breathing are slowed, approaching a state of paralysis of the body, while the brain processes experiences from the day and relaxes. Sleep is necessary for life, although the evolutionary reasons for this process have not yet been fully explored. Sleep status can be assessed using physical or physiological parameters, such as respiration rate, heart rate, temperature and body movement [1]. Based on the features extracted from these parameters, sleep behavior can be determined in terms of sleep time, duration, latency, arousal, wake after sleep onset (WASO) and sleep efficiency (SE) [2].

Sleep behavior is traditionally studied in clinical environments and is still the commonly accepted method for sleep assessments. Polysomnography (PSG) is the main medical gold standard used for sleep disorder classification, such as sleep-related breathing disorders [2]. This system calculates SE, sleep latency, arousal index, sleep stages and other sleep-disorder-related factors from the measurement data. In addition to PSG, another important tool is the noninvasive actigraphy, which measures acceleration to extract information from movement-related changes [1]. These technologies are mainly used by clinicians and have certain restrictions. The restrictions for PSG include short-term sleep monitoring and expensive equipment. Furthermore, it is well known that a first-night effect is present during laboratory recordings, and at least two consecutive nights of data collection are necessary, especially for patients with insomnia [3]. For actigraphy, the available sleep information is restricted based on the collected movement data. Medical research is interested in automating processes to provide faster and earlier diagnoses of sleep disorders and exploring night behavior.

Computational methods are trending and able to address more complex problems. These include the diagnosis of sleep disorders, investigation of the areas that are influenced by sleep and recurrent sleep patterns. Moreover, IoT devices are advancing, creating new opportunities and attracting increasing attention in home-based sleep assessments. Home-based monitoring allows self-assessment and self-management of sleep status on a day-to-day basis within a person's natural home environment. The devices used in this field vary, e.g., smartwatches, radio signals or Doppler radar devices.

The associate editor coordinating the review of this manuscript and approving it for publication was Junchi Yan.

The collected data are analyzed and assessed by applying data mining and machine learning techniques to extract key sleep parameters and indicators. Home-based sleep assessments with sensor technology can not only help individuals assess and manage their sleep but also help researchers find connections and correlations between, e.g., day and night behaviors [4].

Many studies have investigated the use of various technologies for sleep monitoring and assessment, as well as the influence factors of sleep. Although substantial progress has been achieved, challenges and gaps remain in terms of (1) the accuracy and validity of the proposed methods towards gold standards and (2) correlations between sleep and daily behavior. Review papers regarding sleep mainly concentrate on the medical viewpoint using medical devices [2], [5], [6], excluding the interesting computational component, especially in home applications. The signals used in sleep analysis are discussed in [1], but over the past years, technologies have advanced, and new devices and areas have been developed. This includes automatic sleep disorder detection and new developments in techniques such as deep learning. Nevertheless, a comprehensive literature review that can inform researchers and practitioners of the state-of-the-art in this emerging research field and highlight research opportunities and directions is lacking.

This paper is structured as follows: Section II characterizes common sleep behavior considering movement, stable states, abnormal behavior, and how they relate to sleep disorders. In section III, sleep monitoring and data collection are discussed, introducing important sleep parameters and devices from medical and research perspectives. The focus is on section IV, which reviews computational methods for sleep behavior analysis in the fields of sleep stage classification, sleep position recognition, and disease investigation.

II. SLEEP BEHAVIOR CHARACTERIZATION

Sleep behavior can basically be divided into movement and stable states. Movement contains information about sleep and wake episodes from which sleep stages can be extracted. Conversely, stable states mainly describe sleep positions during periods without movement. Accordingly, normal and abnormal behaviors can be characterized, possibly leading to the diagnosis of sleep disorders and chronic diseases.

A. MOVEMENT BEHAVIOR

Sleep is a relaxed state that still contains self-induced movements, mainly to prevent pressure ulcers [7]. These self-induced movements are defined as movement states that create behavior over time. Movement is the main information source for most sensors. An exception is PSG, which is based on a combination of motion and non-motion information.

Movement behavior can be used to distinguish sleep from wake episodes [8]–[11]. Based on this information, objective sleep features can be extracted, e.g., sleep continuity [12], efficiency [11], and time [13]. Combinations of motion and non-motion information lead to sleep motion behavior, which can be further correlated to sleep stages. Sleep stages describe different levels of sleep, provide hints about patient health and are one of the major aspects considered during a PSG visit.

Many disorders exhibit correlations with specific anomalies in sleep cycles or amount of time in specific sleep stages. Normally, sleep stages are measured and defined over brain-wave data but are considered difficult to classify and therefore need trained technicians to be distinguished [1]. Abnormal movement during sleep helps to diagnose certain diseases. This includes movement from the eyes, chin, limbs, chest wall, and upper abdomen [5]. Based on irregular movements, sleep-related movement disorders can be diagnosed, such as periodic limb movement disorder, restless legs syndrome, and sleep-related bruxism [5]. Furthermore, irregular movements can help to diagnose disorders, such as rapid eye movement (REM) disorders or sleep apnea. Sleep apnea can be measured by respiration effort over abnormal abdomen movement during apneas [5]. Abnormal wake-sleep behavior during the day is used to diagnose circadian rhythm sleep-wake disorders [5]. This knowledge can potentially help investigate the severeness of insomnia based on the wake and sleep periods during the night. From movement behavior, various information can be obtained, but it can still be extended by investigating the periods with no motion.

B. STABLE STATE BEHAVIOR

Stable states investigate the periods in rest with no movement. These stable states are mainly related to sleep positions. Sleep postures are independent from sleep stages [14]; therefore, they provide additional insights into sleep behavior [15].

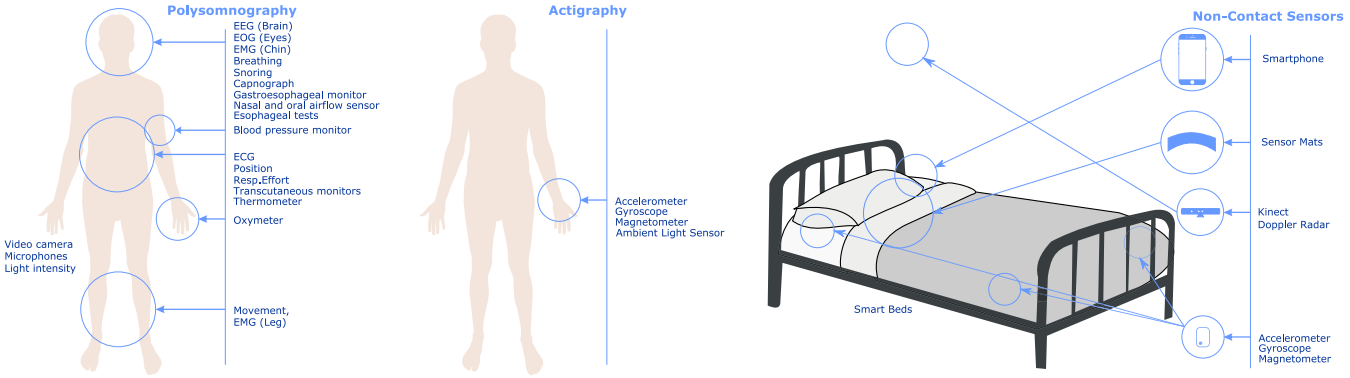
During periods without movement, four basic sleep postures can be distinguished, i.e., supine, prone, right, and left lateral. Sleep position tracking is predominately motivated by the prevention of pressure ulcers [15]–[17] or based on the influence on sleep apnea [15], [18]. For sleep apnea, sleeping on the back, i.e., supine position, relates to a higher apnea/hypoapnea index (AHI) compared to laying on the side [18]. Moreover, sleep parameters such as sleep quality are influenced by different sleep positions [19].

Research mainly concentrates on monitoring the four basic sleep postures, but postures with a higher granularity, including leg positions, are also of interest [15]. Sensors in this field can be (1) applied in or on the bed [16], [17], [20]–[22], (2) wearables [19], [23] or (3) imaging devices [24]. These approaches will be discussed in more detail later.

C. SLEEP AND HEALTH CORRELATIONS

Sleep behavior is manifested in established sleep parameters. These parameters have proven useful in investigating abnormal sleep behavior. Consequently, abnormal behavior can classify sleep disorders and is related to certain chronic diseases [25].

Certain sleep patterns are used to define sleep disorders and have already been investigated for sleep apnea [26] and



(a) Polysomnography (b) Actigraph Unit (c) Nonwearable Sensors

FIGURE 1. Application area of sensors in the medical field for (a) polysomnography and (b) actigraph device. Figures adapted from [44]. (c) Placement of nonwearable sensors in home environments. Figure of bed adapted from [45].

insomnia [27]. The diagnosis of sleep apnea relies on sensors in a sleep clinic, whereas the diagnosis of insomnia is often based on subjective sleep questionnaires [2]. Sleep apnea is diagnosed using AHI, which represents the apnea and hypopnea events per hour [28]; see table 3. An AHI of less than 5 is interpreted as healthy, whereas an AHI between 5 and 15 is classified as mild obstructive sleep apnea, an AHI between 15 to 30 is classified as moderate sleep apnea, and an AHI higher than 30 is classified as severe sleep apnea [25]. For insomnia, the Insomnia Severity Index and Bergen Insomnia Scale can be used for the assessment [2]. Medical history and physical exams can be used in combination to quickly and accurately make diagnoses, such as explored for sleep apnea diagnosis in [25]. Sleep disorders are generally related to poor sleep quality. They are common in the population and add costs to health care and the economic system [1]. A chronic lack of sleep can lead to impulsive behavior, depression, and chronic illnesses. However, when people rest well, their well-being benefits, and they can better handle pressure and stress [29]. Additionally, sleep quality has a high impact on physical and mental well-being [30]. Therefore, monitoring social and sleep behaviors can help the early diagnosis of, e.g., major depressive disorder [31].

Furthermore, sleep behavior is related to chronic diseases. The diagnoses of these diseases are generally based on (1) invasive methods, e.g., blood sugar screening; (2) clinical history, including symptoms, and risk factors; or (3) sensor-based data, e.g., with blood pressure measurements [25]. Current research investigates early disease detection by marker-based clinical analysis or sensor-based behavioral analysis [25]. Sensor-based analysis uses data mining on sensory data, such as actigraphy, to investigate Alzheimer's disease [32], Parkinson's disease [33], diabetes, hypertension, and chronic kidney disease (CKD) [25], [27].

III. SLEEP MONITORING AND MEASUREMENTS

In this section, home-based technologies, medical devices for sleep assessment, and sleep parameters are discussed.

A. SLEEP MONITORING IN THE MEDICAL DOMAIN

The gold standards for assessing sleep disorders and issues from a medical expert's perspective are PSG and actigraphy.

1) POLYSOMNOGRAPHY

PSG is a method that collects sensory data from devices applied to the human body and within the environment. The wearable devices that can be used in a PSG setting are electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), pulse oximetry, respiratory monitors, capnography, transcutaneous monitors, thermometers, esophageal tests, nasal and oral airflow sensors, gastroesophageal monitors, and blood pressure monitors [2], [34]. The application area of these specific sensors is shown in fig. 1a. The nonwearable sensors that are applied in the environment are microphones, video cameras, and light intensity sensors [2].

PSG monitors brain and heart signals and movement. This method is predominately used for assessing sleep disorders such as sleep apnea and restless legs syndrome. Factors such as sleep stages, SE, sleep latency, and arousal index can be extracted from the data [35]. To extract knowledge from the data, scoring methods are applied. The clinical gold standards for sleep scoring with PSG are the Rechtschaffen and Kales (R&K) method [4], [36] and an alternative method presented by the American Academy of Sleep Medicine (AASM) [37]. The scoring is generally based on 30-second epochs [35]. Originally, six sleep stages [36] were assessed, wake, REM, S1, S2, S3, and S4, whereas AASM [37] provides five stages: wake, REM, N1, N2, and N3. In general, R&K stages can be interpreted as AASM stages by combining S3 and S4 as N3. In table 1, a description of the sleep stages and difference between the guidelines are presented.

The higher number of body-attached sensors has the drawback of falsified sleep behavior, which does not represent the natural habit. This leads to the advice that at least two consecutive nights of data collection should be performed [3]. Another issue is that the method is expensive due to the laboratory setting and the fact that an observer is needed to check

TABLE 1. Sleep stages for AASM [37] and R&K [36].

R&K	AASM	Description
Wake		Alert Wakefulness to Drowsiness
REM		Rapid Eye Movement, Few Movements, Dreams
NREM		
S1	N1	Indicates Sleep Onset, Shallow, Quick Transition
S2	N2	Spindle Sleep, Light Sleep, Lower Heart Rate and Body Temperature [38]
		High Amplitude
S3	N3	Moderate Amount Slow Wave-, Deep-Sleep, Body Relax,
S4		Large Amount Rebuilds and Repairs [38], [39]

the functioning of the applied devices during the night [7]. Research on PSG is currently focusing on how to reduce the number of sensors while automatically recognizing the main sleep behavior measurements. This leads to studies with single-channel EEG in sleep stage classification [40]–[42], which can more easily be applied at home.

2) ACTIGRAPH UNIT

Actigraph devices are able to measure activities during a 24-hour time period and are therefore also used in sleep assessments. The device is widely accepted for objective sleep quality measurements. An actigraph is a wearable device attached to the nondominant wrist, providing information about sleep-wake patterns during the night [7]. The device can be used at home, which has the advantage of interfering less with natural sleeping behavior.

Actigraph monitors are mainly based on accelerometer data but can also include gyroscopes and magnetometers. Ambient light is generally collected to help in wake-sleep recognition. Furthermore, personal inputs can be given that tell the device when the sleeping period starts. The usual assessment consists of seven consecutive days to obtain a representative picture of the patient's sleep. This can also be considered the recommended amount of days [43]. This method is an unsupervised wearable approach for sleep analysis. The application area of the sensor is shown in fig. 1b. Current research is concentrating on extracting knowledge from the already processed data coming from medically approved devices such as ActiGraph or Actiwatch. These systems provide activity levels with wake and sleep labels. Activity levels provide the intensity of movement within a usual 30-second interval.

B. HOME-BASED SLEEP MONITORING

Sleep monitoring at home is generally accomplished through wearable or nonwearable devices.

1) WEARABLE TECHNOLOGY

Wearable devices are attached to the human body. Small sensors are typically attached to one of the following areas: wrist, chest, ankles or hip. The advantage of these sensors is their low cost and easy application at home. Data are gathered from sensors such as 3-axis accelerometers, thermostats or photoplethysmography (PPG). Knowledge is generally obtained by applying data mining techniques for sleep position detection [15], [23], [46], sleep stage classification [4], [9], [11], [47], heart rate [29] and respiration rate [46] analysis, and body temperature monitoring [29], [48].

Sleep position detection is usually investigated using either accelerometers [15], [46] or wearable wireless devices [23] on the chest or ankles. In general, sleep stage classification uses either only 3-axis accelerometers [9], [11], [47] or in combination with, e.g., chest strap on wrist and ankle [4]; alternatively, such classification can use PPG [8], [49]. Other sensors include thermometers that measure body temperature to extract sleep and wakefulness [48]. Research also applies commercial wearables that are available on the market, typically smartwatches, e.g., for sleep behavior analysis, determining that automated self-management tools reduce burdens and increase efficiency [50]. Various devices are available and have already been analyzed in terms of performance and accuracy [13], [35], [51]–[55] and user perception [35], [56]. The devices considered in the performance investigations are Actiwatch, ActiGraph GT3X+, FitBit Flex/One, FitBit Charge 2, FitBit Alta HR, Misfit Shine, Basis Health Tracker, Withings Pulse O2, GENEactiv, Jawbone Up3 [13], [35], [51], [55]–[60], ResMed S+ [54], EarlySense [53], Smart Eye Masks [61] and Microsoft Band I [34]. A summary of the Bland-Altmann mean difference and error percentage is shown in table 2, excluding values that are not significant ($\alpha = 0.05$). The output shows that most of the sleep devices overestimate the total sleep time (TST), where Actigraph GT3X+ (4 min) provided the most relevant output compared to the Z-machine [13] and FitBit Charge 2 (-9 min) compared to PSG [55] considering healthy participants. For SE measurements, Actiwatch (4.8%) performed with the best accuracy [60] for healthy individuals, and FitBit Alta HR (2%) performed the best for individuals with specific diseases. FitBit Charge 2 (24.5 min) represents WASO most accurate [59] for healthy individuals, whereas Actiwatch (−21.6 min) [60] is the most accurate for participants with medical conditions. FitBit Charge 2 (4 min) can represent sleep onset latency the best [55] for healthy individuals, whereas FitBit Flex (2.4 min) [60] does so for individuals with medical conditions. The differences between these investigations are also dependent on the participants included, which means that for healthy individuals and those with medical conditions, agreement changes [62]. For Misfit Shine and Basis, light and deep sleep can be distinguished, and a comparison reveals that Basis performs more accurately than Misfit Shine [52]. Additionally, accelerometer data extraction is over- and underestimating certain sleep parameters compared to PSG [12]. Ravichandran *et al.* [35] reviewed different sleep-sensing devices and their performance in relation to the opinions of experts and end users. They investigated Misfit Shine, Jawbone UP3, FitBit One, and FitBit HR. Overall, experts are concerned that sleep quality information is inaccurately transported to end users. Design recommendations based on

TABLE 2. Wearables validated against PSG if not otherwise indicated with Z-machine (Z). The Bland-Altmann-Mean-Difference is given in percentage (%B) or minutes (m).

Wearable	TST*	SE*	SOL*	WASO	Research
Healthy					
FB* Charge 2	-9.0 m		4.0 m		[55]
	-12.3 m		-11.1 m	24.5 m	[59], [62]
Actiwatch	17.8 m	4.8 %B			[60], [62]
	5.8 %*	10.4 %*			[52]
FB* Flex$_{S*}$	3.0 %*				[52]
Basis Health	7.8 %*				[52]
Pulse O2	6.0 %*				[52]
Misfit Shine	15.3 %*				[35], [52], [56]
GT3X+	4.0 m/Z				[13]
Disease-Affected					
Actiwatch	40.6 m	7.0 %B	-13.5 m	-27.1 m	[51]
	43.9 m	7.5 %B	-12.9 m	-33.9 m	[58], [62]
				-21.6 m	[60], [62]
FB* Flex$_{N*}$	32.9 m	7.9 %B	-2.4 m	30.5 m	[60], [62]
	46.0 m	8.1 %B		-44.0 m	[51], [56]
FB* Flex$_{S*}$	-86.3 m	-16.0 %B	11.5 m	74.8 m	[51], [56]
FB* Alta HR	11.6 m	2.0 %B			[57]
Jawbone Up3	39.6 m	6.8 %B	-5.1 m	-34.3 m	[58], [62]

*N-Normal; S-Sensitive; %-Percentage of Error; TST-Total Sleep Time; SE-Sleep Efficiency; SOL-Sleep Onset Latency; FB-FitBit

the level of automation, understandable visualization, and emotional influence factors were investigated in [56] considering Polar Loop, Jawbone Up3, Misfit Shine, and FitBit Flex.

2) NONWEARABLE TECHNOLOGY

Nonwearable devices are not attached to the human body and therefore are the least interfering sleep assessment method, not disturbing the person's regular sleeping habits. In general, techniques are based on either single devices, e.g., Kinect sensors, or multiple devices, e.g., integrated in smart beds. Movement investigations use data collected from sleep trackers. There are already applications on the market for sleep self-management with smart devices, providing insights into users' sleep.

Nonwearable sensors are widely applied, such as load cells, force sensors, air cushions, pressure pads, water-filled vinyl tubes [7], smartphones [31], [63], Shimmer sensors [10], Doppler radar signals [64], [65] also with sound signals [65], [66], pressure sensors [16], [17], [17], [20]–[22], [67], and radio signals [68]. Air cushions under the bed collect data such as respiration rate, heart rate, and body movement to estimate sleep stages. Equally, a water-filled tube under a pillow can be used to record these three features. Additionally, pressure pads can be used to evaluate the heart and breathing rates and even snoring, body movements, and sleep apnea events [7]. However, load cells are employed to detect movement and sleep-wake patterns and estimate deep sleep stages. Pressure sensors integrated in mats, beds or bedsheets typically report good performance in detecting body locations and positions in the bed [15]–[17], [17], [20]–[22], [67], and even sleep stages can be extracted [23]. However, imaging devices are also used to detect sleep postures, such as Kinect sensors [24] and depth cameras [30], [69]. In sleep-wake and sleep stage recognition, Shimmer sensors on the bed [10], Doppler radar [64], and sound signals [65], [66] were investigated.

Sleep trackers that are applied within the environment include smartphones, smart mats or whole beds. Smartphone applications are the most easily accessible for users and therefore an inexpensive method for sleep tracking. Smartphones monitor behavior during sleep, including noise, audio, ambient light, and movement [63], where less mobility and phone usage relate to better sleep [31]. A large number of applications attempt to provide insights into daily sleep cycles, SE, and duration [70], [71]. Additionally, some applications provide the opportunity to self-report moods and daily habits [70].

Smart mats are a good source of information, are easily applicable, and provide higher accuracy. There are various devices on the market that use smart mat technology to collect movement, heart rate, and respiration rate [7]. Different sensors can be integrated into smart beds, such as force, piezoelectric, and pressure sensors [7]. Others also combine this technology with environmental sensors [7]. Based on these data, specific sleep-related information, e.g., bed exits, and sleep statistics, such as sleep quality and movement [7], can be provided to the users, which is useful in, e.g., pain management and fall prevention [7]. The locations of nonwearable devices inside a sleeping environment are depicted in fig. 1c.

C. DISCUSSION AND SUGGESTIONS ON SLEEP MONITORING

There are advantages and disadvantages for all sensors that must be carefully considered based on the area of exploration.

Actigraph units and PSG are used for different areas of investigation. PSG provides the highest accuracy but is expensive, can only be performed in a supervised laboratory setting, and interrupts the person's sleep. PSG requires two days of data collection and applies more than three sensors. In comparison, home-based actigraphy needs at least 7 days of collection and only one sensor, resulting in low costs, not intervening with the natural sleeping habit, and providing intermediate accuracy. Nonwearable devices applied in the environment at home are the least interfering method but have less accurate outcomes, e.g., for smartphones [72], and are often immobile compared to wearables.

Although imaging can generate good position recognition outcomes, it leads to privacy concerns and cannot accurately recognize movements through blankets [15], [24]. In sleep position recognition, nonwearable devices generally cannot distinguish supine and prone positions. In contrast, a wearable approach provides a mobile solution with limited privacy issues and is able to distinguish between multiple people in one bed. However, it still relies on body-worn sensors, which can create discomfort [15]. Sleep stage recognition

depends on more advanced technology measuring EEG data. Currently, research is advancing wearable technologies such as actigraphy to be easier to apply at home. Details can be found in IV. Self-management becomes much more accessible for users at home. Thus, current research suggests different devices depending on the sleep factors of interest; for details, refer to table 2. The easiest devices to start with are smartwatches, as they have a reasonable accuracy compared to smartphone applications and are more accessible than smart beds. Although numerous ready-to-use applications are available with different collection processes and outputs, open issues such as accuracy and validity remain. When choosing the ideal wearable device for investigation, it is suggested to consider the target group, as this can influence the accuracy of wearable devices [62]. Another usual generalization problem comes with validating small datasets, which does not allow overall conclusions such as in [31].

D. MEASURED SLEEP PARAMETERS

The main features describing sleep are respiration rate, heart rate, temperature, body movement [1] and brain waves [2]. Considering various hypotheses, different features are important and are weighted more strongly than others. Depending on the aim, objective or subjective measurements are explored, which present different insights into the sleep of individuals; for details, refer to table 3.

1) OBJECTIVE SLEEP MEASUREMENTS

Objective sleep parameters are concluded from sensor data. These parameters include sleep stages, disturbances, sleep regularity, SE, duration, latency, arousals, spindles, and many more. For example, sleep stages are investigated by heart or respiration rate using the knowledge of existing relations between them [73], whereas skin temperature can be used for estimating disturbances [48]. Sleep continuity is based on the percentage of TST in each sleep stage, SE, and the arousal index [12]. Sleep quality, regularity, sleepiness level, and chronotype are considered new insight indicators investigated with wearable devices compared to traditional parameters, such as time falling asleep, number of awakes, and sleep duration [74]. Explicitly, sleep regularity measures the affinity between sleeping periods from consecutive days [74], [75].

2) SUBJECTIVE SLEEP MEASUREMENTS

Subjective sleep parameters are typically assessed by sleep questionnaires that extract information from users by asking questions to assess sleep issues [2]. These sleep parameters include, e.g., nightmares, bedtime, and rise time; see table 3 for details [74]. The extracted factors can be assessed with different techniques, such as the SATED assessment [35], Consensus Sleep Diary, Pittsburgh sleep quality index, Mini Sleep Questionnaire, Epworth Sleepiness Scale, Insomnia Severity Test, and Sleep Disorders Questionnaire [5]. We refer to the work of Ibáñez *et al*. [5] for a thorough review of subjective methods.

TABLE 3. Subjective and objective sleep parameters adapted from [2], [74]–[80].

Factors of Investigation		
Subjective		
Pains in the Night	Fall Asleep	Rise Time
Real Sleep Duration	Feel Cold	Awakes
Mood and Energy Developing Activity	Bad Breath	Snoring
User-Perceived Sleep Quality	Feeling Hot	Nightmares
Somnolence Developing an Activity	Drug Ingestion	Bedtime
Awakes to the Toilet	# Co-Sleepers	
Objective		
Time in Each Stage (TiES)	Spindles	Slow Waves
Light Off Time	K-Complex	Arousals
Total Sleep Time (TST)	Light On Time	
Factors	**Description**	
Sleep Stages	Wake, REM, S1-S4/N1-N3	
Total Recording Time (TRT)	Lights out to Lights on.	
Sleep Efficiency (SE)	$100\frac{\text{TST}}{\text{TRT}} = 100\frac{\text{len(Sleep Period)} - \text{WASO}}{\text{len(Sleep Period)} + \text{SL}}$	
Sleep Latency (SL)	Lights out to first sleep stage.	
REM Latency	Sleep onset to first epoch of REM.	
Wake-After-Sleep-Onset (WASO)	WASO = TRT - SL - TST	
Arousal (AR)	Wake period <10s	
Awakening (AW)	Wake period >10s	
Arousal Index	#AR $\times$ 60/TST	
Fragmentation Index (FI)	$\frac{\text{1 min. scored sleep bouts}}{\text{\# Sleep bouts of any length}} \times 100$	
Movement Index (MI)	$\frac{\text{Scored awake minutes}}{\text{Time in bed in hours}} \times 100$	
Sleep Fragmentation Index (SFX)	MI + FI	
Significant Limb Movement (LM)	Duration 0.5-10 sec	
Periodic LM of Sleep (PLMS)	$\geq$4 consecutive LM events	
PLMS Index	PLMS $\times$ 60/TST	
Apnea (A)	Cessation of breathing	
Hypopnea (H)	Shallow/low-frequency breathing	
Apnea-Hypopnea Index (AHI)	(#A + #H) $\times$ 60 / TST	
Respiratory-Effort Related Arousals (RERAs)	Arousals from sleep do not meet the definitions of apneas or hypopneas but do disrupt sleep.	
Respiratory Disturbance Index (RDI)	RERAs + As +Hs $\times$ 60/TST	
Sleep Onset Time (Actigraphy)	The first of 15 uninterrupted sleep min. after reported bedtime.	
Sleep Awakening Time (Actigraphy)	The last of 15 sleep min. followed by 30 min. of movement.	
Sleep Regularity Index (Actigraphy)	$-100 + \frac{200}{M(N-1)} \sum_{j=1}^{M} \sum_{i=0}^{N-1} \underbrace{\delta(s_{i,j}, s_{i+1,j})}_{=1 \text{ if } s_{i,j}=s_{i+1,j}}$	$s = 1$ if sleeping M daily epochs N days of recording

3) DISCUSSION ON SLEEP PARAMETERS

Objective methods monitor and measure individuals' sleep behavior in a specific setting, e.g., PSG in a hospital. In comparison, subjective sleep analysis has the advantage that experts are not necessary and are not location dependent, but the disadvantage of inaccuracy remains [74]. These methods are occasionally difficult to compare, as definitions are not consistent in objective and subjective feature calculations. In table 3, features and their formulas are presented. Some of the formulas are for PSG and can be translated to actigraphy; others are developed specifically for actigraph data.

IV. COMPUTATIONAL ANALYSIS METHODS

The main methods for sleep behavior analysis are data mining techniques, such as artificial intelligence, and

statistical analysis. Statistical analysis is a well-developed method, whereas artificial intelligence has recently become more popular in the fields of health and medicine. It has proven to be a good performing method for analyzing more difficult scientific problems, such as sleep behavior and disease detection. Various techniques are available and perform best for specific sleep problems. Prominent methods are random forests (RFs), decision trees (DTs), support vector machines (SVMs), k-nearest neighbors (kNNs), hidden Markov models (HMMs), Bayesian classifiers (BCs), neural networks (NNs), and deep learning methods. Specifically, deep learning can be described as an NN with more than three layers. Methods such as recurrent NN (RNN), e.g., long short-term memory network (LSTM) and convolutional NN (CNN), are adapted in sleep research.

We will discuss computational sleep behavior analysis with a focus on machine learning approaches for the main problems, such as sleep stage, sleep position, and sleep disorder investigations.

A. CHARACTERISTICS OF COMPUTATIONAL METHODS

The applied computational methods in sleep research must be chosen carefully as they influence the expected results, limitations and discussions.

DTs extract rules to split data into subsets represented in a tree structure [81]. RFs are a collection of multiple DTs that can counteract the issue of DTs easily overfitting the training dataset due to the law of large numbers [82]. SVMs represent data in a higher-dimensional feature space to separate classes by hyperplanes, which makes SVMs slow to train on large datasets as they solve a quadratic problem where the number of variables is equivalent to the quantity of training data [83], [84]. The performance is sensitive to the choice of the kernel and parameters [83]. KNNs use the k-nearest neighbors within a metric space to decide to which class new data points belong. The model calculates the distance (commonly the Euclidean distance) to every neighbor for each prediction step and performs slowly when many predictions are made [85]. HMMs are based on Markov chains, where the current state depends on the previous states, to represent transitions and observations [86]. Naive BCs, which assume that features are independent of each other following the Bayes' theorem, are a very simple approach that require little training data [87], [88]. NNs are based on layers of artificial neurons. Functions are applied to the layer inputs, and the outputs are sent to the next layer. The training process is based on weighting, from which predictions can be made even for incomplete information. NNs need long training times because of the large number of parameters that are best determined empirically and are difficult to interpret for humans [88]. NNs with more than three layers are considered to be deep learning methods and are extensively used currently due to the good performance and the advantage that even without preprocessing, good features can be generated from raw data. However, it is not possible to see which features are important or how the outcome is produced, and a large quantity of training data is necessary, resulting in high computational cost. In sleep research, time-series data are often collected; therefore, specific methods, such as LSTM, are designed to incorporate the time aspect. However, in certain cases (e.g., sleep position detection) good outcomes can still be achieved if the time component is ignored.

Descriptive statistics, e.g., principal component analysis (PCA) [89], draw conclusions based on the data itself, whereas inferential statistics, e.g., logistic regression (LR), draw conclusions based on samples from the population. Classic LR is often used to analyze accelerometer data as LR is easy to interpret but is subject to limitations. LR is usually unsuitable for learning complex patterns from noisy accelerometer data; therefore, higher-level features are often extracted, which requires expertise and is potentially time consuming. Furthermore, these methods do not use task labels for feature construction and therefore cannot learn task-specific features [78].

B. VALIDATION OF SLEEP ANALYSIS METHODS

Most technologies and approaches, such as home-based sensors and single-channel EEG, are tested against the gold standard PSG. Researchers should be aware that the interscorer agreement of human-scored parameters does not have a perfect agreement but rather 82.6% [90]. This can result in a bias towards a rater's style if only one person scores sleep data. It also means that discussions need to take this into account. Seldom, data from actigraph units are used for validation [10]. The most often applied performance measures in computer science are accuracy, recall (=sensitivity), specificity, precision, and Cohen's kappa (κ). These measures are also used for validation purposes in sleep behavior analysis. Occasionally, the receiver operating characteristic (ROC) curve, area under the curve (AUC) or F1 score is given.

Accuracy is the percentage of predictions that a specific classifier correctly makes. Sensitivity describes the capability of the classifier to recognize true positives, and specificity indicates that it does not generate a false negative [26]. Specificity is defined over the number of false positives (FPs) and true negatives (TNs), whereas precision is the positive predictive value [42]. The F1 score based on precision and recall can be calculated with macro- and micro-averaging methods. The multiclass F1 score is based on the weighted individual class scores. The macro F1 score is uniformly weighted, whereas the micro F1 score is measured by calculating the overall number of false negatives (FNs), true positives (TPs), and FPs [42]. Accuracy and precision (recall) are not able to completely describe the situation in multiclass classification. Additionally, for imbalanced classes, accuracy can be misleading; therefore, precision and recall are of importance, for example, represented by the F1 score. Regardless, the drawback is that the F1 score has no good intuitive explanation [91]. Therefore, Cohen's kappa statistic was introduced for imbalanced and multiclass classifications. Cohen's kappa statistic compares the classifier performance to random guessing [91], measuring the agreement between

annotators for categorical items statistically [92]. An ROC curve represents a classifier's performance at different classification thresholds in a graph, providing a global estimation of the classification ability [26]. It is based on the precision and FP rate (1-specificity). The larger the AUC is, the better the classification performs [26]. For multiclass problems, multiple numbers of graphs are needed.

Note that validation is performed with (1) k-fold cross-validation (k-CV); (2) leave-one-out cross-validation (LOOCV), which is favorable; or (3) one specific data split, from which no general conclusions can be drawn. All measurements are not always provided; therefore, it is not always possible to directly compare the results of different methods. Generally, user-independent classification is ideal, i.e., users who are trained on should not be tested on or else the generalizability of the method cannot be guaranteed. This can be realized in methods (1) and (2) but needs to be addressed to ensure that the results are trustworthy.

C. SLEEP STAGE CLASSIFICATION

The main goals in sleep stage classification are to (1) automate the process that is normally performed by trained technicians and to (2) make home-based assessment possible. Validation is usually performed against trained human classification, which is not always the best because human classification includes known variability. This is based on the fact that technicians classify specific epochs differently [93], [94]. It is especially important to consider the performance on healthy subjects and subjects with medical conditions. In the following paragraphs, approaches using sensory monitoring at home and in the medical domain are discussed. The different technologies and their performances and details can be found in table 4 for home-based sensory data and in table 5 for medical devices, focusing on single-channel approaches such as EEG and ECG as these can potentially be applied at home. The values are recalculated from the confusion matrix or by averaging the given outcome parameters.

1) HOME-BASED SLEEP STAGE ANALYSIS

Wearable and nonwearable devices are investigated for sleep stage classification.

a: FIVE-STAGE CLASSIFICATION

An ideal approach would be able to distinguish 6 or 5 sleep stages depending on the chosen guideline. Currently, research is mainly focusing on the 5 AASM stages. In [4], MSR accelerometer data from the wrist and ankle and Zephyr BioHarness 3^2 data from the chest were collected in a sleep lab from 26 individuals to investigate sleep stages in comparison to PSG measurements. The data were analyzed using RF and deep learning. Deep learning was used for unsupervised feature learning, followed by a deep belief network (DPN) built from stacked restricted Boltzmann machines. The DBN approach achieved a 10-fold CV accuracy of 77.6% for accelerometer data only. The RF method, which fused the data sources of the chest strap and accelerometers, classified 80.7% correctly. A commercial Microsoft Band I sensor was used in [34] to collect heart rate and actigraph recordings from 39 healthy subjects. They proposed a method using multilevel feature learning and an RNN. LOOCV resulted in a precision of 64.5%, recall of 65%, and F1 score of 60.5% in the comprehensive group, where resting and nonresting sleep were included. The performance of the RNN approach is affected by the dataset size, which with 37k epochs is likely too small. We will see later that EEG home systems are the most promising as sleep stages were originally defined over brain waves. Using sensors in a home environment that interfere less with sleeping habits are showing good results for using accelerometer data and a chest strap with an RF method. The limitations in 5-stage classification are that (1) only healthy participants are considered and (2) the validation datasets are relatively small, especially for deep learning. Overall good outcomes over all classes can be seen when comparing Cohen's kappa.

b: FOUR-STAGE CLASSIFICATION

Because it is difficult to distinguish N1 and N2, the stages are occasionally fused to light sleep and compared to deep sleep (N3) [96]. This results in 4 stages: wake, REM, light, and deep sleep. In [96], continuous positive air pressure (CPAP) flow signals from 400 subjects were analyzed to detect sleep stages. High-level features were extracted with CNN and RNN, which were further used in a conditional random field (CRF). An accuracy of 74.1% was reached, but only with a weak Cohen's kappa of 0.57. We can conclude that the dataset is highly imbalanced and cannot detect minority classes sufficiently based on only one split for training and testing. Likewise, radio waves can be analyzed by combining CNN and RNN [68]. This approach reached a moderate Cohen's kappa of 0.70 with up to 79.8% accuracy for 25 healthy subjects that participated, but investigated only one data split. Ultimately, the advantages of both main deep learning methods were used. Specifically, the CNN was able to separate wake and REM stages, whereas RNN could separate deep and light sleep [68]. Instead, accelerometer data were collected in [47] from the nondominant hand for 36 individuals. They analyzed different classifiers and performed feature selection, concluding that RC performs the best. The results showed an accuracy of 80% for light sleep and 90% for wake, REM, and deep sleep for 10-CV. The subjects' health status was not provided, but the study is most likely based on healthy participants. In contrast to machine learning approaches, equation threshold-based approaches are also investigated, such as in [95], by recording data for 100 participants from a wrist-worn device that includes a 3-axis accelerometer and a reflective photoelectric volume pulse sensor. The system reached an accuracy of approximately 68.5%, which is generally lower compared to other accelerometer approaches but is validated on a larger database. The commercial ResMed S+ device based on an ultra-low-power radio-frequency sensor was analyzed in [54]. Respiration amplitude, frequency, and body movement were extracted from the signal.

TABLE 4. Home-based sleep stage classification in 30-second intervals following AASM [37]; validated against PSG except [10] against Actiwatch and [61] against a wrist-worn device. Certain performances were recalculated from confusion matrices.

	Data	P*	Characteristics	Epochs (hpP*)	Analysis	Val.*(%•)	Prec.	Rec.	Acc.	κ
5 Stages - Wake, REM, N1, N2, N3										
[34]	Microsoft Band I	39	H*/30M*/19-64Y*	37.000 (~ 8)	RNN	LOOCV	64.5	65	60.5(F*)	
[4]	2 Accelerometers, CS*	26	H*	~28.080 (~ 9)	RF, DBN	10-CV	78.3	77.9	80.7	0.72
4 Stages - Wake, REM, Light Sleep (N1, N2), Deep Sleep (N3)										
[53]	EarlySense	63	SD*/45M*/17-72Y*	70.965 (~ 9)	Commercial		63.8	60.6	64.5	0.46
[95]	Accelerometer, PWS*	100	H*	108.000 (~ 9)	Thresholds		64.4	59.7	68.5	
[54]	ResMed S+	38	H*/21M*	~41.040 (~ 9)	Commercial	53/47/			70	0.53
[96]	CPAP flow	400	SD*	360.000 (7.5)	C/RNN-CRF	60/20/20	65	71.8	74.1	0.57
[68]	Radio Signals	25	H*/15M*	~90.000 (~30)	CNN-RNN	75/25/	79.6	75.8	79.8	0.70
[47]	Accelerometer	36		~34.560 (~ 8)	Kstar, Bagging, RC, RSS, RF	10-CV	85(S*)	70	87.5	
3 Stages - Wake, REM, NREM (N1, N2, N3)										
[65]	DR*, Microphone	24	SD*/21M*/43.4±13Y*	17.280 (6)	RF	46/54/			64.4	
[66]	Microphone	250	SD*/162M*/19-84Y*	~180.000 (~ 6)	NN	60/40/	82.6	80.9	87.3	0.72
2 Stages - REM, NREM										
[61]	Smart Eye Mask	7	H*/4M*/20-24Y*	~7.920 (~ 6)	RF	11-CV			80	
[49]	PPG*, Accelerometer	15	H*	~16.200 (~ 9)	DT		81.4(S*)	86.3	81.4	
2 Stages - Sleep, Wake										
[8]	PPG*, HRV*	10	SD*/5M*/56±8.79Y*	8.452 (~ 7)	kNN, SVM	50/50/	79(S*)	79	79.4	0.59
[9]	Accelerometer	22	7M*/85.7±3.7Y*	~23.760 (~ 9)	CNN		68(S*)	80		
[10]	5 Shimmer sen.	3	H*/1M*/21-30Y*	24.062 (~67)	RF		86(S*)	93	90	
[97]	Video	10	avg. 22.6 Y*	~6.000 (~ 5)	Thresholds				92.13	
[11]	Accelerometer	81	47M*/20-60Y* 56GSE*, 25PSE*	76.776 (~ 8)	Rules	27/73/	71.3(S*)	95	92.2	0.64

*P-Participants; hpP- Hours per Person; Val.- Validation; CS-BioHarness 3^2 chest strap; PWS-Pulse Wave Sensor; DR-Doppler Radar; PPG-Photoplethysmography; HRV-Heart Rate Variability; H-Healthy; M-Males;Y-Years; SD-Sleep Disorders; RC-Random Committee; RSS-Random Subspace; GSE-Good Sleep Efficiency; PSE-Poor Sleep Efficiency;S-Specificity; F-F1 score; •(training, validation, test) split

The validation was performed with 3 technicians using a majority voting for an overall score. 38 adults were assessed, with an accuracy of 70% compared to an accuracy of 82% for general scorers. In contrast, the Early Sense sensor based on piezoelectric sensors reached only 64.5% accuracy with a weak Cohen's kappa of 0.54 [53], but included 63 subjects with medical conditions.

An RC approach for accelerometers appears to be promising for four stages with an average 87.5% accuracy [47], but a nonwearable device and deep learning method such as in [68] comes with certain advantages but only 80% accuracy for healthy participants.

c: THREE-STAGE CLASSIFICATION

Correspondingly, researchers simplify the problem to wake, REM and N-REM stages. In [65], a sound and Doppler radar sensor were combined to detect sleep-wake episodes followed by NREM-REM classification. Different features for each problem were used, reaching an accuracy of 64.4% with RF for 24 patients with sleep disorders. Each step included a personal-adjustment structure, based on a threshold coming from ordering the likelihood ratios from the RF classification. In [66], audio signals from microphones of 250 participants were analyzed with a one-layer NN, obtaining 87.3% accuracy. Moreover, Smart Eye Masks were applied to determine REM and NREM sleep from photoreflectors and accelerometers for 7 healthy subjects [61]. RF was used, reaching 80% accuracy for 10-CV. In contrast, optical wrist-worn devices were utilized in [49] by applying a trained DT, reaching an overall accuracy of 81.35%. These take into account PPG and a 3-axis accelerometer data from 15 participants.

Although it is difficult to compare different data sources, it appears that for 3 stages, audio signals in combination with NN perform well, including subjects with a medical condition, but providing outcomes of only one training-testing case.

d: TWO-STAGE CLASSIFICATION

Wake-sleep classification is often the first step towards finer granularity sleep stage classification. This classification can be performed with PPG [8], accelerometers [9], actigraph units [11], and Shimmer sensors on the bed [10]. In [8], sleep-wake stages for 10 patients with sleep apnea were classified by kNN and SVM using a PPG from which heart rate variability and PPG features were extracted. The kNN approach achieved an accuracy of 77.35% for 10-CV for HRV, PPG, and feature selection with a small dataset of 8k including participants with sleep disorders. Twenty-two elderly individuals participated in a study to collect accelerometer data. These data were analyzed using a CNN. The CNN approach was compared to a standard sleep-wake

classification approach, increasing the specificity from 54% to 68% while decreasing the sensitivity from 82% to 80% [9]. In [10], five Shimmer sensors were applied on the bed of 3 individuals and validated against a Philips Actiwatch. Undersampling and oversampling were used to prepare the data for RF methods. The overfitting issue was therefore addressed, achieving results with a sensitivity of 93% and specificity of 86%. Alternatively, a rule-based approach was proposed by Kuo *et al*. [11] using an actigraph, reaching an accuracy of 92.16%, specificity of 71.3% and sensitivity of 95.02%. They tested the system with 81 subjects, divided in terms of poor and good SE. Four different rules were introduced from movement density and density thresholds. A descriptive analysis was performed in [97]. For this purpose, camera recordings were analyzed and validated against actigraphy and PSG. Frame difference and motion history were used to classify motion, which is an indicator of wake episodes. Data from 10 subjects reached 92.13% accuracy for the video-based system in comparison to Actiwatch with 91.24% accuracy. Threshold-based actigraph sleep-wake classification is a powerful method, suggesting the necessity to distinguish between subjects with poor and good SE [11].

2) SLEEP STAGE ANALYSIS IN THE MEDICAL DOMAIN

In this section, the focus lies in the automation of sleep stage classification towards home usage with single-channel EEG data. The preprocessing of highly sensitive data is important as artifacts, e.g., from movement, are present, and a large amount of information is available.

a: SIX-STAGE CLASSIFICATION

Attempts to automate sleep stage scoring were performed in 1996 [93], where an NN model with an uncertainty index was presented that was able to classify 6 sleep stages using EEG, EMG, and EOG. Sixty participants were included: 20 suffering from depression, 20 suffering from insomnia and 20 healthy participants. The results showed accuracies of 84.5% for healthy subjects, 81.5% for subjects with depression, and 81% for subjects with insomnia. These approaches have the limitations of collecting data from various sensor sources and are therefore mainly useful in a hospital environment, whereas single-channel EEG approaches could be easier to use at home in the future. Researchers using single-channel EEG data often use the Physionet Sleep EDF database with 8 subjects for validation. With this, an accuracy of 88.62% for empirical mode decomposition with adaptive boosting and DT [99] can be reached, whereas iterative filtering with RF can reach an accuracy of 90.02% [100], complex-valued nonlinear features and complex-valued neural network (CVANN) can reach 91.57% [40], and decomposed two-subband tunable Q-wavelet transform with DT can reach 92.43% [92].

In this setting, NN [40] performs with almost perfect agreement considering Cohen's kappa. We can conclude that preprocessing appears to be a very important aspect to make a single-channel EEG approach perform well. The limitations lie in the size of the data sample with 15k epochs from only 8 subjects, including 4 healthy and 4 participants with mild difficulty falling asleep not diagnosed with sleeping disorders.

b: FIVE-STAGE CLASSIFICATION

The 5-stage classification typically follows the AASM guidelines and is most commonly performed. For single-channel EEG coming from the Sleep-EDF data, (1) 8 subjects with 15k epochs or (2) 20 subjects with 42k epochs are investigated. Approach (1) can reach 87.2% accuracy with Elman-RNN using only 6k epochs [101], 90.11% with DT [99], 91.13% with iterative filtering with RF [100], 93.69% with DT [92], and 93.84% accuracy using CVANN [40]. Approach (2) performs for 20-CV with an accuracy of 82% with CNN-LSTM [102] and 83.5% with CNN on a smartphone [103]. The data sample of 20 subjects only contains healthy individuals; therefore, it is necessary to further investigate patients suffering from sleep disorders. In [41], 41 healthy participants and 42 participants with insomnia were investigated, reaching an overall accuracy of 77% by applying DNN-HMM. Twenty-eight subjects with sleep apnea were considered in [104], reaching 95.88% accuracy for SVM. A total of 5728 patients from the Sleep Heart Health Study (SHHS) were investigated in [42], and applying a CNN resulted in an accuracy of 87% but not performing CV. In [42], better results could be reached for the Sleep-EDF dataset, which is potentially caused by the small number of technicians that participated. This makes the system learn a specific rater's style, which causes difficulties in generalization. Similar accuracy could be reached by Malafeev *et al*. [105] with a CNN-LSTM including 18 healthy patients, 23 patients with narcolepsy and 5 patients with hypersomnia.

For single-channel usage, SVM (16k epochs, 28 participants) [104] and CNN (5k, 5728 participants) [42] tested on a dataset with a larger number of participants affected with sleep disorders presented promising results, and CVANN (15k, 8 participants) [40] performed well on a dataset with a small numbers of participants.

Overall, the N1 stage is often difficult to distinguish [102]. This leads to models that concentrate on this issue, such as in [106]. Filtered single-channel EEG signals from 13 participants of the Sleep-EDF database were investigated by SVM, leading to an accuracy of 92.5% in distinguishing N1 and wake stages. Many approaches for sleep stage classification rely on features and preprocessing data, while others use raw data, such as in Malafeev *et al*. [105]. If no CV is performed, the results must be considered with caution, such as in [42], [99].

c: FOUR-STAGE CLASSIFICATION

To simplify the problem, researchers fuse stages to wake, REM, light (N1, N2), and deep sleep (N3). Approaches with single-lead ECG and CNN achieved 75.4% accuracy when including 16 subjects with sleep issues, 65.6% for

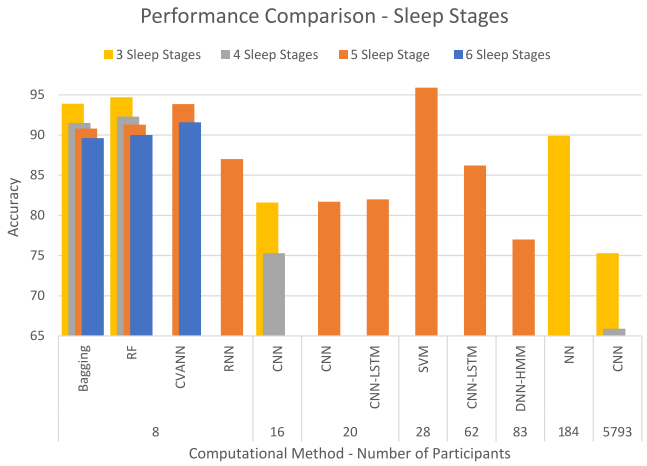


FIGURE 2. Performance comparison for sleep stage approaches in the medical domain that perform CV.

994 subjects with sleep disorders, and 65.9% for 5793 subjects including breathing issues [98]. The investigations in [98] have low Cohen's kappa values and reasonably high accuracy. This result is likely caused by an imbalance in the investigated datasets for different sleep stages, which overestimates the accuracy, whereas Cohen's kappa represents the performance above the baseline of random guessing and is therefore more suitable for imbalanced data; for details, refer to section IV-B.

In contrast, single-channel EEG from the Sleep-EDF database results in 91.2% accuracy with DT [99], 91.5% with DT [92] and 92.29% with iterative filtering with RF [100]. Overall, RF [100] performs the best but has the limitation of involving only 8 participants.

d: THREE-STAGE CLASSIFICATION

Further simplification leads to 3 stages: wake, REM, and NREM (N1, N2, and N3). In this case, a one-channel EEG investigation from Sleep-EDF reached an accuracy of 93.55% with DT [99], 93.9% with DT [92], and 94.6% with RF [100]. Including participants with medical conditions, such as from 184 observations, using NNs reached 89.9% accuracy [38]. Extending the system with EEG, EOG, and Flow reached 89.6% accuracy for healthy individuals and those with restless legs syndrome and sleep apnea [38]. In contrast, ECG data applying a CNN reached an accuracy of 75.3% for SHHS and 81.6% for 16 subjects with sleep issues [98].

Figure 2 summarizes the performance on specific datasets that used comparable CV. We can see an overall trend for deep learning approaches performing worse on smaller datasets such as RNN in Sleep-EDF with 8 participants. Six sleep stages for a higher number of participants are not explored. Hence, two aims are targeted: how well can the guideline-based sleep stages (5 or 6 stages) be detected and which stages can be merged (reducing sleep stages). Fewer investigated sleep stages results in higher accuracy. Overall, newer methods such as deep learning are typically used more often for larger datasets. The trend shows that more data results in worse performance, except for simpler models such as SVM and NN (compared to deep learning approaches).

3) DISCUSSION AND SUGGESTIONS ON SLEEP STAGE ANALYSIS

There are two main areas in sleep stage recognition that target either home-based analysis or sleep stage analysis in the medical domain. Both approaches have individual issues that need to be addressed in the future.

Overall, most methods for home-based assessment have issues with (1) imbalanced data, (2) being unable to correctly classify more complex stages, (3) the limitation of datasets including only healthy participants and only a small number of those and (4) generalizability, caused by non-standardized user-independent classification. It is clear that classifying sleep from wake data is performing well with accelerometers, as movement is the main factor to distinguish these. More complex sleep stages typically require more information, which can be addressed by combining multiple sensors with accelerometers currently tested only on healthy subjects. When reducing the number of sensors, sound is able to distinguish three stages, even for participants with sleep disorders. Wearable devices appear to be promising for healthy subjects and are most commonly applied, but nonwearables are promising for subjects with a medical condition. The overall trend indicates that larger datasets and more diverse datasets degrade the performance, which can be caused by (1) differences in sleep stage detection for people suffering from a medical condition and (2) inconsistency of, e.g., movement during the night for different healthy individuals. The advantage of deep learning methods in larger datasets is compliant with the known necessity for larger training sets, as shown in fig. 2 and table 5. Patterns are more easily distinguished for sleep disorder participants when the datasets are larger. This follows, e.g., from the performance of NNs in fig. 2, where the performance is compared for different dataset sizes for which the number of sleep disorder patients can be found in table 5 (see [38]). The validation method, if mentioned, is rarely CV and is therefore not objective in terms of user-independent classification and representation of influences from the training data. Overall, movement data have limitations when higher granular sleep stages are involved. Simple models currently appear to be more promising than more complex deep learning approaches, which can also be influenced by the quantity of data, such as in [34]. Comparative studies are rarely performed, which is an issue when determining the ideal sensor for home assessment or the ideal machine learning approach.

For sleep stage analysis in the medical domain, challenges remain for (1) the generalization of the model to the general population affected by diseases and (2) dealing with imbalanced classes. This effect can be seen as methods have lower performance on larger and more diverse datasets. This effect is also influenced by the imbalance of the Sleep-EDF dataset containing 8 participants with a majority class of wake, which introduces a model bias and therefore positively influences

TABLE 5. Sleep stage classification comparison for single-channel EEG and one from single ECG [98]. Stage classification in 30-second epochs except for [87]. Performances were recalculated from confusion matrices.

	P*	Dataset	Sample	Epochs (hpP*)	St*	Analysis	Val.*(%•)	Prec.	Rec.	Acc.	κ
6 Stages - Wake, REM, S1, S2, S3, S4											
[99]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	6	AdaBoost	60/ 5/35	79.9	71.8	88.6	0.82
[92]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	6	Bagging	LOOCV	80.4	71.6	89.6	0.84
[100]	8	Sleep-EDF	4H*-4 mild Ins.*	15.136 (~16)	6	RF	10-CV	81	71.7	90	0.84
[40]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	6	CVANN	LOOCV	98.2(S*)	85.8	91.57	0.89
5 Stages - Wake, REM, N1, N2, N3											
[101]	8	Sleep-EDF	4H*-4 mild Ins.*	5.760 (~ 6)	5	RNN	LOOCV	72.4	75.6	87.2	0.76
[99]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	5	AdaBoost	60/ 5/35	84.4	77	90.11	0.84
[92]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	5	Bagging	LOOCV	84.2	77.6	90.8	0.85
[100]	8	Sleep-EDF	4H*-4 mild Ins.*	15.136 (~16)	5	RF	10-CV	86.6	75.4	91.3	0.86
[40]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	5	CVANN	LOOCV	98.4(S*)	91.8	93.84	0.92
[102]	20	Sleep-EDF	H*	41.950 (~17)	5	CNN-LSTM	20-CV	75.6	78.7	82	0.76
[103]	20	Sleep-EDF	H*	41.950 (~17)	5	CNN	20-CV			83.5	0.72
[41]	83	41.3Y*avg*	41H*-42 Ins.*	~79.680 (~ 8)	5	DNN-HMM	5-CV			77	
[102]	62	MASS	H*	58.600 (~ 8)	5	CNN-LSTM	31-CV	82	81.3	86.2	0.80
[42]	5728	SHHS	divers(breathing)	5.384.401 (~ 8)	5	CNN	50/20/30	80.6	77.3	86.8	0.81
[104]	28	35-56Y*	divers(breathing)	15.541 (~ 5)	5	SVM	4-CV	97.42(S*)	88.32	95.88	0.86
4 Stages - Wake, REM, Light Sleep, Deep Sleep											
[99]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	4	AdaBoost	60/ 5/35			91.2	
[92]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	4	Bagging	LOOCV	89	85.6	91.5	0.86
[100]	8	Sleep-EDF	4H*-4 mild Ins.*	15.136 (~16)	4	RF	10-CV			92.29	
[98]	994	CinC2018DB	divers(breathing)	261.946 (~ 2)	4	CNN	10-CV			65.6	0.31
[98]	5793	SHHS	divers(breathing)	400.547 (~ .5)	4	CNN	10-CV	52	64.1	65.9	0.47
[98]	16	SLPD	divers(breathing)	2.829 (~ 2)	4	CNN	10-CV	46.2	53.7	75.4	0.54
3 Stages - Wake, REM, NREM											
[99]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	3	AdaBoost	60/ 5/35			93.55	
[92]	8	Sleep-EDF	4H*-4 mild Ins.*	15.188 (~16)	3	Bagging	LOOCV	91.3	88.8	93.9	0.89
[100]	8	Sleep-EDF	4H*-4 mild Ins.*	15.136 (~16)	3	RF	10-CV			94.6	
[38]	184	M*	divers(RLS*+Apn*)	176.640 (~ 8)	3	NN	10-CV	87.2	85.35	89.9	
[98]	5793	SHHS	divers(breathing)	400.547 (~ .5)	3	CNN	10-CV			75.3	0.57
[98]	994	CinC2018DB	divers(breathing)	261.946 (~ 2)	3	CNN	10-CV			76.5	0.42
[98]	16	SLPDB	divers(breathing)	2.829 (~ 2)	3	CNN	10-CV			81.6	0.63

P-Participants; hpP-Hours per Person; St-Number of Sleep Stages; Val.-Validation; Y-Year; avg-Average; M-Males; S-Specificity; H-Healthy; Ins.-Insomnia; Apn-Apnea
†Sleep stages: wake, (REM, S1), S2, S3, S4; •(training, validation, test) split

the performance. Overall, it appears that single-channel EEG is sufficient to obtain sleep stages; however, the application at home is a future step. In general, online detection is desirable because this can be useful for boosting slow-wave sleep, transcranial stimulation, and acoustic stimulation [94]. In fig. 2, deep learning approaches perform less well for larger datasets, but comparative studies towards simpler models on the same datasets are lacking. The trend of using deep learning methods on larger datasets is in accordance with the known influence of small datasets on the methods; see RNN, which uses only 6k epochs. Therefore, the performance advantages and disadvantages of deep learning cannot be observed and discussed in detail. For smaller datasets, deep learning approaches perform less well, most likely influenced by the need for more training data compared to approaches such as RF and bagging. Six sleep stages are investigated only for smaller datasets, likely because of the newer guidelines following five sleep stages [37]. It is easier to classify fewer sleep stages, as sleep stages that are difficult to distinguish from each other are usually combined. However, the imbalance of different classes positively influences the outcomes as wake is the majority class. Overall, CV and LOOCV are the most commonly used validation methods, which guarantees an objective performance representation in the medical domain. When the only performance metric in an imbalanced dataset is accuracy, the outcomes must be interpreted with caution, as the imbalance often present in these studies influences the accuracy.

D. SLEEP POSITION RECOGNITION

Sleep position recognition generally detects the basic four sleep positions: supine, prone, right, and left lateral [15]. Recently, higher granularity positions have been increasingly investigated. For some approaches, e.g., image-based approaches, the number of detected positions must be reduced based on the limitations of the method. This normally involves excluding the prone position or combining the prone position with the supine position [16], [17], [21], [22], [30], [113]–[116]. Others consider more complex positions, such as right fetus [17], [67], right yearner [17], right log [67], left fetus [17], [67], left log [67], left yearner [17], supine [17], [67] and prone [67]. Different arm [16], [111] and leg [15]

TABLE 6. Sleep position monitoring systems. Ground truth is, if mentioned, based on videos, except for [107], which is based on a smartwatch.

	Data	I*	h* (seP*)	P*	Add to Basic P*	Analysis	Val.*(%•)	Pre.	Rec.	Acc.	κ
Wearables - Simulated Setting											
[108]	ECG-Lead II	12	10 (250)	4		k-means		93(S*)	79		
[109]	Respiration Impedance	16	6.4 (180)	4		SVM	10-CV	99.7	99.7	99.7	0.99
[110]	Koala (wrist+chest)	uc*	- (180)	5	sit	RF		86.5	88.6	88.5	0.86
[23]	W*(chest) + WSN*	2	1.4 (10)	5	30°-incline lateral	kNN				100	
[15]	3 ShS*(ankles+chest)	6	3.8 (25)	8	leg P*	LVQ(Gen.)	20xLOOCV	85.9	82.4	83.6	0.82
[111]	3 W*(chest+arms)	10	0.4 (10)	8	arm P*	kNN		92.2	92.5	92.2	0.91
[15]	3 ShS*(ankles+chest)	6	3.8 (25)	8	leg P*	LVQ(Ind.)	20x10-CV			99.8	
Wearables - Real-World Setting											
[107]	Huawei SW*	16	~560	4		RF	LOOCV	TPR"=91.8; FPR"=0.03			
[46]	Accelerometer(chest)	7	~56	4		LDA		99.5	99.5	99.5	0.99
[112]	Accelerometer(chest)	13	88.4	5	sit/stand	Rules		99.1	95.6	99.2	0.98
[15]	3 ShS*(ankles+chest)	2	14	6	leg P*	LVQ(Gen.)				97.7	
Nonwearables - Simulated Setting											
[22]	Pressure sens.(uBP*)	2	uc*	3	no prone	BC	10-CV	81.4	78.7	78.7	0.71
[113]	Pressure mat	13	uc*	3	no prone	RBM-DNN	10-CV	82.6	79.4	82.7	
[114]	Pressure mat(uBP*)	1	120 (uc*)	3	no prone	NN-BN	5-CV			89.9	
[115]	Pressure mat	uc*	312 S*	3	fuse prone & supine	PCA-SVM	10-CV	94.2	94.1	94.1	0.91
[21]	Pressure mat	9	uc*	3	no prone	GMM-kNN	60/40	98.4	98.4	98.4	0.98
[116]	Pressure sen.(uBP*)	2	0.3 (60)	3	no prone	RF	33/67			98.4	
[117]	HBT*	58	3.9 (60)	4		NN	10-CV			72	
[20]	Pressure distr.	3	6 (300)	4		LR	33/67	90.5	90.1	90.2	0.88
[17]	Pressure mat	6	uc*	5	leg P*, no prone	PCA-kNN	70/30	97.7	97.6	97.7	0.97
[67]	Pressure bedsheet	14	3360 S*	6	leg P*	MCR	LOOCV	83.3	83	83	0.81
[69]	3D-Artec/Kinect	3	uc*	6	leg P*, no prone	SVM	67/33	92.7	92.5	92.5	0.91
[16]	Pressure mat	20	uc*	8	leg, arm P*, no prone	kNN	10-CV	97.1	97.1	97.1	0.97
[21]	Pressure mat	9	uc*	13	body, limb P*, no prone	GMM-kNN	60/40			91.6	
Nonwearables - Real-World Setting											
[30]	3D-Asus Xtion cam	78	1880S*	4	sit/stand, no prone	CNN	5-CV	95	94.9	94.9	0.93
[67]	Pressure bedsheet	3	7800S*	6	leg P*	MCR-HMM	LOOCV	86.5	84.7	85.6(F*)	

*I-Individuals; h-hours; seP-seconds per adopted position; P-Positions; Val.-Validation; uBP-upper bed part; HBD-hydraulic bed transducers; W-Wearables; WSN-wireless sensor network; ShS-Shimmer Sensors; SW-Smartwatch; uc-unclear; S-Samples;S-Specificity; F-F1 Score; •(training,test) split

positions, as well as angles [21], [23], are considered higher granularity positions. These positions can relate to certain issues, such as inducing back pain [15]. Furthermore, positions can be extended by an unknown state, normally the sitting/standing position, such as in [30], [110], [112], explaining, e.g., bed exits.

In this section, outcomes are distinguished by wearable and nonwearable devices, as sensors in the medical field are rarely used, such as in [108], [109]. A summary can be found in table 6, where different methodologies and their results are listed.

1) SLEEP POSITION ANALYSIS USING WEARABLES

Wearables are often investigated because these devices are easily applicable and provide high accuracy for posture detection. Some systems have almost perfect performance for detecting sleep positions but normally show issues related to the number of participants involved and user-independent detection.

In general, single sensors can detect basic sleep positions well. Shinar *et al*. [108] used an ECG device, achieving a specificity of 93% and sensitivity of 79% with a k-means iterative algorithm. In [109], respiration impedance signals are measured from 16 individuals and achieved a 99.7% accuracy for 10-CV.

Extending approaches with additional wearable devices can achieve similar outcomes for an even higher number of positions, such as in [15], with 99.8% for LVQ personalized per individual with repeated CV on 6 individuals. Other multiple wearable approaches reached an accuracy of 88.5% with RF [110], 83.6% for a general LVQ approach using LOOCV for 6 individuals [15], and 92.2% with kNN for 10 subjects [111]. Barsocchi [23] studied the feasibility of a transmitter and receivers in the sleeping environment to distinguish four main positions, also considering a lateral incline of 30% [15]. They could reach 100% accuracy by using at least two sensors and kNN.

The best matching rates could be achieved in [23], which has the limitation of a very small sample of two participants. In [15], similar high performance is achieved with individual trained models, which makes generalizing the system more difficult but can distinguish higher granularity positions including leg movement, whereas [111] includes hand movement in the higher granularity position investigation.

The next step is the real-world application, which has been considered in different research studies. Smartwatches, for example, can be applied, such as in [107], where RF performed the best with a TP rate of 91.8% for 16 participants and objective LOOCV. The most promising models include an accelerometer placed on the chest, reaching an

accuracy of 99.5% with LDA for 7 individuals [46], and even rule-based approaches can reach 99.2% accuracy for 13 subjects [112] including an unknown state. Six positions could be recognized with an accuracy of 97.74% with LVQ, having the drawback of a small data sample of 2 participants and multiple sensors attached.

Overall, participant numbers are low when validating the approaches, and many approaches fail to state the used validation method.

2) SLEEP POSITION ANALYSIS USING NONWEARABLES

Nonwearables usually need more complex methods for analysis because they often produce images that need to be classified.

Three sleep posture classifications are very common because distinguishing the prone posture from the supine posture is difficult. Consequently, Hsia *et al*. [22] exploited pressure at the upper part of the bed to investigate three postures, focusing on the influences coming from hand postures and laying angle with a BC. The result showed a low accuracy of 78.7% for 2 individuals and was influenced by the laying angle. Pressure mats are commonly applied for detection, reaching up to 98.4% accuracy with GMM-kNN for 9 individuals [21] and RF for 2 subjects and no CV [116]. Other investigations have used PCA-SVM, resulting in 94.1% accuracy [115], 89.9% with NN-BN for one person [114] and 82.7% with RBM-DNN for 13 individuals [113]. Hence, simple data analysis models already appear promising. More complex approaches, e.g., hydraulic bed transducers, were used under a mattress to distinguish the four main sleep postures from 58 participants by using an NN [117], also including the prone position. This led to an accuracy of 72%, which is low compared to the approach with pressure distribution and logistic regression (LG) with an accuracy of 90.2%, which only included 3 participants and no CV [20]. Yousefi *et al*. [17] utilized a pressure mat to detect five different positions, including supine, yearner, and fetus, from six subjects. The posture detection was based on a three-step algorithm using normalization, eigenspace projection, and a kNN classifier. The average accuracy of this detection reached 97.7%. A mobile, easy applicable solution was investigated in [67]. In this case, a pressure-sensitive bedsheet was used to monitor six sleeping positions, including supine, prone, log, and fetus, by sparse classifiers with Minimum Class Residual (MCR), reaching an accuracy of 83% for 14 subjects [67]. To obtain a more detailed picture, camera devices for 3-D measurements can be used and have reached an accuracy of 92.5% using SVM on 3 individuals [69]. Pressure mats can be used to detect higher granularity positions, e.g., Pouyan *et al*. [16] classified eight different bed postures excluding prone. The proposed algorithm creates a pressure image that is processed using size and shift-invariant images. Classification was performed by computing the Hamming distance between the signature images and the presented sample. The results showed an accuracy of 97.1% for 20 subjects. Multiple angles of the three positions could be detected by

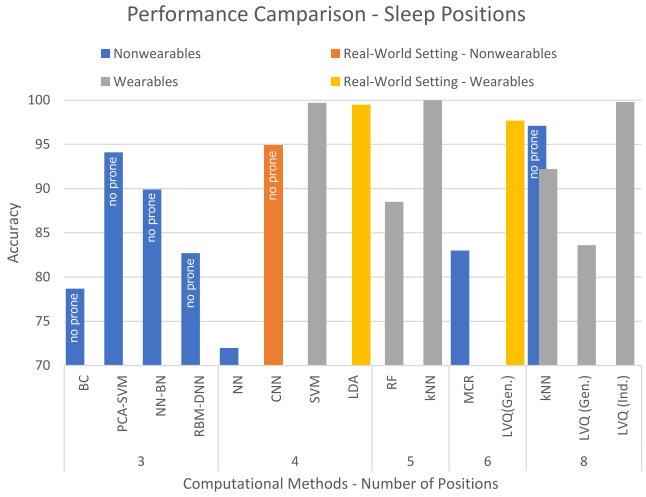


FIGURE 3. Performance comparison for sleep position approaches that perform CV.

Ostadabbas *et al*. [21]. They detected 13 sleeping positions from 9 individuals with a GMM-based clustering approach, reaching an accuracy of 91.6%.

Real-world applications considering whole nights of data are rarely investigated. However, existing investigations include Kinect devices [24], 3D-Asus Xtion cameras [30], and pressure bedsheets [67]. Kinect sensor data were collected to distinguish five sleep positions from 20 students [24], and no blankets were used [15]. A single depth camera was able to distinguish supine, left, right, and empty positions. They included 78 patients and obtained results of 94.9% accuracy with a CNN [30]. For multiple positions, collecting pressure-sensitive data from a bedsheet over three nights from three people were used, resulting in 86.5% precision and 84.7% recall [67].

To obtain a fast and easy visualization of the current sleep position investigation state, we incorporated fig. 3, which compares all CV investigations. Mainly nonwearable devices are used for three position investigations, whereas wearables are used to target the basic four or more granularity positions. Overall simple models such as kNN, SVM and LVQ with no temporal component appear to be appropriate for the investigation in comparison to more complex approaches such as NN and DNN. Little research has been done on real-world applications.

3) DISCUSSION AND SUGGESTIONS ON SLEEP POSITION ANALYSIS

Sleep position analysis can be performed for wearable and nonwearable devices in a laboratory or real-world setting.

Overall, wearable accelerometers, particularly on the chest, can reach high accuracy with already simple machine learning models. Alternatively, respiration monitoring also appears to be a promising approach. Although research has investigated different data sources and technologies, an energy-efficient, robust solution that is able to accurately detect finer granularity positions and that is adaptable to

individual needs is still lacking. Furthermore, validation on larger data samples is necessary. Wearable devices are commonly used when tracking sleep positions, but as these are worn on the body, discomfort can be present; therefore, improving usability and comfort is of interest [15].

One alternative is the usage of nonwearables; in this context, pressure mats are commonly applied and reach high accuracy with the limitation of distinguishing prone from supine, which was partly overcome in [20], [67]. Most recent investigations have attempted to investigate 3D images, which raises privacy concerns.

Figure 3 shows that nonwearable devices cannot accurately distinguish the prone from supine position, as shown by the NN approach, which reaches 72% accuracy. Therefore, three-position approaches are investigated. By contrast, wearables perform better for more complex positions. Overall, real-world applications are rarely reported but are indeed important, as a real-world application in [15] showed that fewer positions are captured from the participants. Overall simple models appear to be appropriate for the investigation, in comparison to more complex approaches. These complex approaches use NN or deep learning and are more influenced by the quantity of training data. A small number of participants is usually investigated, which comes with a drawback as consistency in sleep postures between individuals cannot be assumed. Many research studies fail to report the validation method used, which makes comparison difficult. Furthermore, many studies report performance from one split of the data, which is not an objective representation of the investigated population, as mentioned previously.

Research needs to target real-world applications, providing reproducible results that are able to enhance recognition rates for finer-grained positions. Furthermore, user-independent validation needs to become a standard to guarantee generalizability. In the future, other sensor sources will become available and can potentially target more advanced investigations.

E. INVESTIGATION OF SLEEP DISORDERS AND DISEASES

Abnormal sleep behavior was originally classified by sleep experts and further used to diagnose individuals' health status in terms of sleep disorders. Recently, behavior has been investigated in automatic decision making for (1) sleep disorders such as sleep apnea [25]–[27], [118], [119] and insomnia [27] and (2) specific chronic diseases such as diabetes, hypertension, CKD [25], [27], Alzheimer's [32], and Parkinson's disease [33]. Established sleep parameters can be used to help investigate disorders and abnormal sleep behavior. The field of using computational behavior analysis in helping diagnose certain diseases is not extensively explored. In this survey, we present approaches that are interesting in our opinion.

1) SLEEP APNEA INVESTIGATION

a: SLEEP APNEA DETECTION

Javaid *et al*. [119] investigated a nonwearable Impulse Radio Ultra-Wide Band Radar panel under the mattress for the detection of sleep apnea events from radar signals. The analyzed dataset consists of 25 hours of data from 4 subjects with apnea. The overall match against PSG reached an accuracy of 70% for 5-CV with a linear discriminant classifier (LDA) on extracted statistical features. In [27] and [25], sleep apnea stages were investigated by comparing sleep apnea towards unaffected participants with a wearable actigraph, separating mild stages in [25]. Both studies investigated the HCHS study [120] with 1887 individuals. In [25], 170 individuals are included for the investigation of sleep apnea. The results for two classes showed an accuracy of 68% with a CNN [27] and 81% (six times 10-CV) [25] with an LSTM on a balanced dataset. An alternative for sleep apnea-hypopnea syndrome diagnosis was investigated with a home-based oximetry sensor in [121]. A dataset of 320 subjects (one night each) was analyzed with different machine learning algorithms, such as LDA, LG, Bayesian multilayer perceptron, and AdaBoost with one split of 60% for training and 40% for testing. AdaBoost with LDA performed the best depending on the AHI: for 5 (92.9%), 10 (87.4%), and 30 (78.7%). Chung *et al*. [122] reached the closest diagnostic ability to the machine learning approach, achieving better accuracy of 93.7% for an AHI of 30, but including only surgical (one night of 475 patients) and not regular patients. In [26], two sleep apnea types were analyzed for 83 patients by applying SVM on 125 sets of ECG records, extracting 24 features. This approach reached an accuracy of 92.85% for LOOCV. An optimization problem on kernels led to the conclusion that the polynomial kernel with degree of 3 provides the best results. The research was limited to patients with no history of cardiovascular disease and central sleep apnea. The model could be useful for determining the CPAP therapy by analyzing the change in probabilities in the outcome [26].

It is difficult to compare the approaches as the used data sources and investigated classes are different. Nevertheless, nonwearable devices can recognize sleep apnea events with an accuracy of 70%, but they are outperformed by wearable sensors such as ECG, oximetry and actigraph sensors.

b: SLEEP APNEA TREATMENT

CPAP devices are utilized for treating patients suffering from sleep apnea by providing pressure to help them through apnea episodes. In [118], an approach was investigated to detect patients who were likely to discontinue the therapy. CPAP and electronic health records from 3588 patients were used. Feature selection was performed, and oversampling the dataset counteracted the imbalanced distribution. Different techniques were applied, such as linear regression, LR, DT, and SVM, whereby RF and boosting trees such as XGBoost considerably improved the results. Deep learning for time series classification did not present good performance, mainly due to the lack of data. The results showed that XGBoost reaches the best F1 score including health records with 85% for 10-CV, which is an improvement compared to the current state-of-the-art.

2) CHRONIC DISEASE INVESTIGATIONS

Chronic diseases affect individuals' lives but also introduce sleep behavior changes [25]. These changes are investigated to obtain indications for early and later stages of diseases.

Sleep behavior changes were investigated in [27] and [25], where actigraph data were considered from a broad population to detect insomnia [27], diabetes, hypertension, and CKD [25]. In [27], a CNN was applied, introducing a method for embedding activities. In contrast, in [25], an LSTM was investigated, considering the major limitation of an imbalanced dataset in [27]. Therefore, a balanced dataset was adopted for training, reaching accuracies for 6 times 10-CV of 72.5%, 62%, and 76.7% for diabetes, hypertension, and CKD, respectively [25]. In [25], 475 individuals affected by hypertension, 325 by diabetes and 60 by CKD are investigated, all with the same number of healthy subjects and 6 nights of data for each person. In [27], the outcomes were 69% for hypertension and 44% for diabetes, including individuals with 7 nights of data. The accuracy for hypertension extraction was higher in [27], but precision and recall values could be improved in [25]. This is an important step in obtaining reproducible outcomes.

Early disease detection is relevant for the diagnosis of Alzheimer's and Parkinson's diseases, which relate to sleep [32], [33]. Early stages of Alzheimer's disease already affect sleep behavior based on the relation of β-amyloid (Aβ) with sleep quantity and quality manifested in actigraph data [25], [32]. Sleep behavior also shows relationships for patients with Parkinson's disease investigated using actigraph data [25], [33]. Early disease detection could thrive from these newly elucidated relationships.

3) SUGGESTIONS FOR DISORDER AND DISEASE INVESTIGATIONS

In general, it is necessary to investigate and use the existing knowledge of relationships between sleep and certain diseases because this knowledge can enhance and promote self-management and help diagnose diseases at an early stage. Automating sleep disorder diagnoses such as sleep apnea with daily technologies could provide an easy and inexpensive assessment.

With the increasing usage of deep learning technologies that are able to handle larger amounts of data for individual classifications, chronic diseases can be investigated from sleep data. Based on the existing basis of sensor accuracy and availability, the diagnosis of diseases and disorders is a promising future investigation area with the potential to make early diagnoses possible and accessible at home.

In the investigations presented here, a greater number of individuals are usually involved. When the validation method is stated, CV is applied, which provides an objective method. We decided to present the accuracy metric in this section for consistency, as most investigations used balanced data or failed to present other measurements.

V. CHALLENGES AND FUTURE TRENDS

In recent decades, sleep behavior analysis has advanced considerably by introducing new devices and computational methods; nevertheless, there are still certain limitations and challenges that must be addressed by the research community. Common challenges are (1) the inclusion of sleep behavior differences coming from healthy and sick populations into methodologies; (2) including medical knowledge in terms of sleep structure, relations and influence factors; (3) improving current technologies for home usage; (4) validating on larger and more diverse datasets; (5) addressing imbalanced datasets and their issues; and (6) providing adequate comparative outcomes and standards.

In sleep behaviour analysis, there is a need to distinguish healthy subjects from subjects with a medical condition. However, current research targets mainly healthy individuals when analyzing sleep. When subjects with a medical condition are included and investigated, the overall performance is usually lower. This aspect can be addressed by either training models separately for different user groups or including new features in the training process. These new features should either include knowledge of the user's condition or represent behavioral differences more accurately.

Medical research on sleep behavior has been performed over the past decades, and knowledge of the relations and structure has been established. This knowledge can be an advantage when constructing a computational sleep behavior analysis, e.g., the relation with health factors and the time aspect of sleep changes throughout the night. Including the time aspect while assessing sleep aspects can likely improve the performance but has not been investigated extensively. Current research mainly concentrates on automating human scoring tasks on well-known problems such as sleep stage classification. This has some drawbacks because machines have the potential to classify stages better than humans. Human-labeled data are known to be scored differently between sleep experts and cannot reach a 100% match [105]. Specifically, this interscorer agreement following the AASM rules is only approximately 82.6% [90]. One way to bring new insights into sleep behavior analysis that is not based on classic features is pattern recognition and unsupervised approaches to describe sleep differently. Emerging approaches for sleep assessment are investigating new features, such as the sleep regularity index [75]. The regularity of sleep [75] represents the trend of consecutive nights, which is a step towards long-term visualization and helps to draw more specific conclusions. Furthermore, the automatic detection of shorter underlying structures using machine learning techniques is emerging, such as for K-complexes [123] and sleep spindle detection [123], [124]. Eventually, research will extract this information from sensor sources other than EEG, which are easier to apply at home.

Improving technologies for home usage include addressing potential discomfort and integration into existing smart devices. Sensors and technologies will develop towards

wearable sensors that are integrated into clothing and single-channel EEG devices that are easily applicable. Another direction includes nonwearables, which can be seamlessly integrated into the environment, and the used methodologies will be further explored as nonwearables are currently less accurate. In certain areas such as sleep stage detection, research must be conducted to accurately detect higher levels of sleep stages within a natural home environment.

In the future, the issue of small datasets could be overcome by user contributions, i.e., crowdsourcing sleep data to contribute to sleep research. Consumer wearables will progressively adopt to facilitate users sharing their sleep tracking data with researchers and could be enhanced by users' personal information, e.g., gender, age, and medical conditions. Hence, larger studies can be conducted, and inconsistencies and inherent noise can be overcome to a certain extent. More divers datasets allow to address the issue of imbalance by considering the inconsistency during training. Therefore, a balanced dataset should be created for training purposes to counteract the bias towards the majority class. It is not yet standard for papers to provide the validation method used, but it is made clear throughout this paper that repeated cross-validation and user-independent validation are critical. Repeated cross-validation provides a reliable outcome for the population investigated. Usually, a generalized approach is targeted; therefore, users who are trained on cannot be tested on, else the performance ability of the method is falsified. No standard performance measure exists, but overall accuracy is commonly used. However, in the case of imbalanced data, further measurements, such as recall and precision, are needed. Therefore, we recommend standardization of the performance metrics used in sleep behavior analysis to guarantee comparative studies. Comparative studies must be performed to investigate methods of the same datasets to benchmark available approaches and to be able to assess the most promising methods.

It is already known that there are many correlations between sleep and daily life, as well as specific chronic diseases, which are not used extensively. Research is being performed to explore features that provide hints about specific diseases. These features can be measured continuously with sensor technology at home. Combining this with the known relations could help diagnose diseases in an early stage [25], [27]. The reason for this interest is that people normally learn about issues quite late, which does not allow preventative approaches. Computational analyses could provide insights into sleep data that are not obtainable by currently used methods and therefore target educating and showing users their sleep habits [35], preferably at home. Overall, sleep is very subjective and individual [35]; therefore, individuality should be addressed and transported to developed systems. This will lead to fusing different sleep elements from objective, subjective, and environmental [29] perspectives into an automatic approach for sleep assessment and self-management. Another attempt is to fuse known relations, such as the knowledge from physical activity into sleep research, which can actually improve results, such as in [78].

VI. CONCLUSION

In this paper, we provide a systematic, comprehensive review on the state-of-the-art in research and practices in computational sleep behavior analyses. We specifically focus on the latest developments in sleep monitoring, modeling, and computational analysis methods for sleep assessments using sensor technologies, which can be used by the general public at home and are easier, quicker, and inexpensive. This survey offers in-depth knowledge and insights into this increasingly important research field to effectively guide the reader through vast amounts of literature. We have also highlighted the challenges and future research trends that will inform, inspire, and guide researchers, technology developers, and healthcare practitioners in research, innovation, and service provision. This is a rapidly growing, dynamically changing research area. Whereas previous research has mainly been undertaken to exploit and automate human expert knowledge, one apparent trend is to apply data-driven techniques to investigate data from various perspectives rather than human labeling only, inferring and discovering new insights directly. It is also expected that, with the prevalence and maturity of daily technologies and the availability of cloud-based computational power, the gap between clinic-based and home-based sleep assessments will vanish in the very near future. Whereas this will require close collaborations and knowledge sharing among healthcare professionals, research experts and users, it opens up opportunities that will potentially lead to transformations in future healthcare.

REFERENCES

[1] A. Roebuck, V. Monasterio, E. Gederi, M. Osipov, J. Behar, A. Malhotra, T. Penzel, and G. D. Clifford, "A review of signals used in sleep analysis," *Physiol. Meas.*, vol. 35, no. 1, pp. R1–R57, 2013.

[2] V. Ibáñez, J. Silva, and O. Cauli, "A survey on sleep assessment methods," *PeerJ*, vol. 6, p. e4849, May 2018.

[3] J. Newell, O. Mairesse, P. Verbanck, and D. Neu, "Is a one-night stay in the lab really enough to conclude? First-night effect and night-to-night variability in polysomnographic recordings among different clinical population samples," *Psychiatry Res.*, vol. 200, nos. 2–3, pp. 795–801, 2012.

[4] U. Reimer, S. Emmenegger, E. Maier, Z. Zhang, and R. Khatami, "Recognizing sleep stages with wearable sensors in everyday settings," in *Proc. 3rd Int. Conf. Inform. Commun. Technol. Ageing Well e-Health*, 2017, pp. 172–179.

[5] V. Ibáñez, J. Silva, and O. Cauli, "A survey on sleep questionnaires and diaries," *Sleep Med.*, vol. 42, pp. 90–96, Feb. 2018.

[6] T. Penzel and R. Conradt, "Computer based sleep recording and analysis," *Sleep Med. Rev.*, vol. 4, no. 2, pp. 131–148, 2000.

[7] D. Waltisberg, B. Arnrich, and G. Tröster, "Sleep quality monitoring with the smart bed," in *Pervasive Health*. London, U.K.: Springer, 2014, pp. 211–227.

[8] M. K. Uçar, M. R. Bozkurt, C. Bilgin, and K. Polat, "Automatic sleep staging in obstructive sleep apnea patients using photoplethysmography, heart rate variability signal and machine learning techniques," *Neural Comput. Appl.*, vol. 29, no. 8, pp. 1–16, 2018.

[9] R. Wolz, J. Munro, R. Guerrero, D. L. Hill, and Y. Dauvilliers, "Predicting sleep/wake patterns from 3-axis accelerometry using deep learning," *Alzheimer's Dementia*, vol. 13, no. 7, p. P1012, 2017.

[10] A. McDowell, M. P. Donnelly, C. D. Nugent, L. Galway, and M. J. McGrath, "Addressing the challenges of sleep/wake class imbalance in bed based non-contact actigraphic recordings of sleep," in *Proc. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2013, pp. 4654–4657.
[11] C.-E. Kuo, Y.-C. Liu, D.-W. Chang, C.-P. Young, F.-Z. Shaw, and S.-F. Liang, "Development and evaluation of a wearable device for sleep quality assessment," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1547–1557, Jul. 2017.
[12] P. Celka, K. Martinmaki, T. Korhonen, N. Santaniemi, and J. Virkkala, "Sleep-wake detection and computation of sleep continuity from a wrist unit in children, adolescents and adults," in *Proc. EMBEC NBC*, 2017, pp. 1049–1052.
[13] M. E. Rosenberger, M. P. Buman, W. L. Haskell, M. V. McConnell, and L. L. Carstensen, "Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices," *Med. Sci. Sports Exerc.*, vol. 48, no. 3, pp. 457–465, 2016.
[14] D. Lorrain and J. De Koninck, "Sleep position and sleep stages: Evidence of their independence," *Sleep*, vol. 21, no. 4, pp. 335–340, 1998.
[15] S. Fallmann, R. van Veen, L. Chen, D. Walker, F. Chen, and C. Pan, "Wearable accelerometer based extended sleep position recognition," in *Proc. IEEE 19th Int. Conf. e-Health Netw., Appl. Services*, Oct. 2017, pp. 1–6.
[16] M. B. Pouyan, S. Ostadabbas, M. Farshbaf, R. Yousefi, M. Nourani, and M. D. M. Pompeo, "Continuous eight-posture classification for bed-bound patients," in *Proc. 6th Int. Conf. Biomed. Eng. Inform.*, Dec. 2013, pp. 121–126.
[17] R. Yousefi, S. Ostadabbas, M. Faezipour, M. Farshbaf, M. Nourani, L. Tamil, and M. Pompeo, "Bed posture classification for pressure ulcer prevention," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug./Sep. 2011, pp. 7175–7178.
[18] G. D. Pinna, E. Robbi, M. T. La Rovere, A. E. Taurino, C. Bruschi, G. Guazzotti, and R. Maestri, "Differential impact of body position on the severity of disordered breathing in heart failure patients with obstructive vs. Central sleep apnoea," *Eur. J. Heart Failure*, vol. 17, no. 12, pp. 1302–1309, 2015.
[19] A. Nojiri, C. Okumura, and Y. Ito, "Sleep posture affects sleep parameters differently in young and senior japanese as assessed by actigraphy," *Health*, vol. 6, no. 21, pp. 2934–2944, 2014.
[20] P. Barsocchi, M. Bianchini, A. Crivello, D. La Rosa, F. Palumbo, and F. Scarselli, "An unobtrusive sleep monitoring system for the human sleep behaviour understanding," in *Proc. IEEE 7th Int. Conf. Cognit. Infocommun.*, Oct. 2016, pp. 91–96.
[21] S. Ostadabbas, M. B. Pouyan, M. Nourani, and N. Kehtarnavaz, "In-bed posture classification and limb identification," in *Proc. IEEE Biomed. Circuits Syst. Conf.*, Oct. 2014, pp. 133–136.
[22] C.-C. Hsia, Y.-W. Hung, Y.-H. Chiu, and C.-H. Kang, "Bayesian classification for bed posture detection based on kurtosis and skewness estimation," in *Proc. IEEE 10th Int. Conf. e-health Netw., Appl. Services*, Jul. 2008, pp. 165–168.
[23] P. Barsocchi, "Position recognition to support bedsores prevention," *IEEE J. Biomed. Health Inform.*, vol. 17, no. 1, pp. 53–59, Jan. 2013.
[24] J. Lee, M. Hong, and S. Ryu, "Sleep monitoring system using kinect sensor," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 10, 2015, Art. no. 875371.
[25] S. Fallmann and L. Chen, "Detecting chronic diseases from sleep-wake behaviour and clinical features," in *Proc. IEEE 5th Int. Conf. Syst. Inform.*, Nov. 2018, pp. 1076–1084.
[26] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings," *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 1, pp. 37–48, Jan. 2009.
[27] K. Aggarwal, S. Joty, L. F. Luque, and J. Srivastava, "Co-morbidity exploration on wearables activity data using unsupervised pre-training and multi-task learning," 2017, *arXiv:1712.09527*. [Online]. Available: https://arxiv.org/abs/1712.09527
[28] T. Van Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, "Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks," *IEEE J. Biomed. Health Inform.*, to be published.
[29] M. J. Deen, "Information and communications technologies for elderly ubiquitous healthcare in a smart home," *Pers. Ubiquitous Comput.*, vol. 19, nos. 3–4, pp. 573–599, 2015.
[30] T. Grimm, M. Martinez, A. Benz, and R. Stiefelhagen, "Sleep position classification from a depth camera using bed aligned maps," in *Proc. IEEE 23rd Int. Conf. Pattern Recognit.*, Dec. 2016, pp. 319–324.
[31] A. Doryab, J. K. Min, J. Wiese, J. Zimmerman, and J. Hong, "Detection of behavior change in people with depression," in *Proc. AAAI Workshops*, 2014, pp. 12–16.
[32] Y.-E. S. Ju, J. S. McLeland, C. D. Toedebusch, C. Xiong, A. M. Fagan, S. P. Duntley, and D. M. Holtzman, "Sleep quality and preclinical alzheimer disease," *JAMA Neurol.*, vol. 70, no. 5, pp. 587–593, 2013.
[33] M. Louter, J. B. Arends, B. R. Bloem, and S. Overeem, "Actigraphy as a diagnostic aid for REM sleep behavior disorder in Parkinson's disease," *BMC Neurol.*, vol. 14, no. 1, 2014, Art. no. 76.
[34] X. Zhang, W. Kou, E. I.-C. Chang, H. Gao, Y. Fan, and Y. Xu, "Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device," *Comput. Biol. Med.*, vol. 103, pp. 71–81, Dec. 2018.
[35] R. Ravichandran, S.-W. Sien, S. N. Patel, J. A. Kientz, and L. R. Pina, "Making sense of sleep sensors: How sleep sensing technologies support and undermine sleep health," in *Proc. Conf. Hum. Factors Comput. Syst.*, 2017, pp. 6864–6875.
[36] A. Kales and A. Rechtschaffen, *A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects* (NIH Publication), no. 204. Washington, DC, USA: United States Government Printing Office, National government publication, 1968.
[37] American Academy of Sleep Medicine, *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications*. Darien, IL, USA: American Academy of Sleep Medicine, 2007.
[38] A. Procházka, J. Kuchynka, O. Vyšata, P. Cejnar, M. Vališ, and V. Marík, "Multi-class sleep stage analysis and adaptive pattern recognition," *Appl. Sci.*, vol. 8, no. 5, p. 697, 2018.
[39] F. Lin, Y. Zhuang, C. Song, A. Wang, Y. Li, C. Gu, C. Li, and W. Xu, "SleepSense: A noncontact and cost-effective sleep monitoring system," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 1, pp. 189–202, Feb. 2017.
[40] M. Peker, "An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms," *Neurocomputing*, vol. 207, pp. 165–177, Sep. 2016.
[41] M. Shahin, B. Ahmed, S. T.-B. Hamida, F. L. Mulaffer, M. Glos, and T. Penzel, "Deep learning and insomnia: Assisting clinicians with their diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 6, pp. 1546–1553, Nov. 2017.
[42] A. Sors, S. Bonnet, S. Mirek, L. Vercueil, and J.-F. Payen, "A convolutional neural network for sleep stage scoring from raw single-channel EEG," *Biomed. Signal Process. Control*, vol. 42, pp. 107–114, Apr. 2018.
[43] S. Briscoe, E. Hardy, M. F. Pengo, C. Kosky, A. J. Williams, N. Hart, and J. Steier, "Comparison of 7 versus 14 days wrist actigraphy monitoring in a sleep disorders clinic population," *Chronobiol. Int.*, vol. 31, no. 3, pp. 356–362, 2014.
[44] M. Häggström and F. Michel. (2013). *Human_Body_Silhouette.Svg, Background Transparent*. Accessed: Nov. 30, 2018. [Online]. Available: https://commons.wikimedia.org/wiki/File:Polysomnografie_Body_en_106.svg
[45] P. Robinson. (2007). *Steel Bed.Svg*. Accessed: Dec. 3, 2018. [Online]. Available: https://commons.wikimedia.org/wiki/File:Steel_bed.svg
[46] Z. Zhang and G.-Z. Yang, "Monitoring cardio-respiratory and posture movements during sleep: What can be achieved by a single motion sensor," in *Proc. IEEE 12th Int. Conf. Wearable Implant. Body Sensor Netw.*, Jun. 2015, pp. 1–6.
[47] M. Yeo, Y. S. Koo, and C. Park, "Automatic detection of sleep stages based on accelerometer signals from a wristband," *IEIE Trans. Smart Process. Comput.*, vol. 6, no. 1, pp. 21–26, 2017.
[48] R. J. E. M. Raymann, D. F. Swaab, and E. J. W. Van Someren, "Skin temperature and sleep-onset latency: Changes with age and insomnia," *Physiol. Behav.*, vol. 90, nos. 2–3, pp. 257–266, 2007.
[49] P. Renevey, R. Delgado-Gonzalo, A. Lemkaddem, M. Proença, M. Lemay, J. Solà, A. Tarniceriu, and M. Bertschi, "Optical wrist-worn device for sleep monitoring," in *Proc. EMBEC NBC*, 2017, pp. 615–618.
[50] M. J. Duncan, C. Vandelanotte, S. G. Trost, A. L. Rebar, N. Rogers, N. W. Burton, B. Murawski, A. Rayward, S. Fenton, and W. J. Brown, "Balanced: A randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults," *BMC Public Health*, vol. 16, no. 1, p. 670, 2016.

[51] J. D. Cook, M. L. Prairie, and D. T. Plante, "Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy," *J. Affect. Disorders*, vol. 217, pp. 299–305, Aug. 2017.

[52] J. Mantua, N. Gravel, and R. M. C. Spencer, "Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography," *Sensors*, vol. 16, no. 5, p. 646, 2016.

[53] A. Tal, Z. Shinar, D. Shaki, S. Codish, and A. Goldbart, "Validation of contact-free sleep monitoring device with comparison to polysomnography," *J. Clin. Sleep Med.*, vol. 13, no. 3, pp. 517–522, 2017.

[54] A. Zaffaroni, E. P. Doheny, L. Gahan, Y. Ivanov, H. Kilroy, N. O'Mahony, and D. O'Rourke, "Non-contact estimation of sleep staging," in *Proc. EMBEC NBC*, 2017, pp. 77–80.

[55] M. de Zambotti, A. Goldstone, S. Claudatos, I. M. Colrain, and F. C. Baker, "A validation study of fitbit charge 2 compared with polysomnography in adults," *Chronobiol. Int.*, vol. 35, no. 4, pp. 465–476, 2018.

[56] B. Vandenberghe and D. Geerts, "Oh, what a night! Effortless sleep monitoring at home," in *Proc. Conf. Hum. Comput. Interact.*, 2015, pp. 417–424.

[57] J. D. Cook, S. C. Eftekari, E. Dallmann, M. Sippy, and D. T. Plante, "Ability of the Fitbit Alta HR to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: A comparison against polysomnography," *J. Sleep Res.*, vol. 28, no. 4, 2018, Art. no. e12789.

[58] J. Cook, M. L. Prairie, and D. T. Plante, "Ability of the multisensory jawbone up3 to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: A comparison against polysomnography and actigraphy," *J. Clin. Sleep Med.*, vol. 14, no. 5, p. 841–848, 2018.

[59] Z. Liang and M. A. C. Martell, "Validity of consumer activity wristbands and wearable EEG for measuring overall sleep parameters and sleep structure in free-living conditions," *J. Healthccare Inform. Res.*, vol. 2, nos. 1–2, pp. 152–178, 2018.

[60] S. Kang, J. M. Kang, K.-P. Ko, S.-C. Park, S. Mariani, and J. Weng, "Validity of a commercial wearable sleep tracker in adult insomnia disorder patients and good sleepers," *J. Psychosomatic Res.*, vol. 97, pp. 38–44, Jun. 2017.

[61] S. Matsui, T. Terada, and M. Tsukamoto, "Smart eye mask: Sleep sensing system using infrared sensors," in *Proc. ACM Int. Symp. Wearable Comput.*, 2017, pp. 58–61.

[62] M. de Zambotti, N. Cellini, A. Goldstone, I. Colrain, and F. Baker, "Wearable sleep technology in clinical and research settings," *Med. Sci. Sports Exerc.*, to be published.

[63] J. Behar, A. Roebuck, J. S. Domingos, E. Gederi, and G. D. Clifford, "A review of current sleep screening applications for smartphones," *Physiol. Meas.*, vol. 34, pp. R29–R46, Jun. 2013.

[64] T. Rahman, A. T. Adams, R. V. Ravichandran, M. Zhang, S. Patel, J. A. Kientz, and T. Choudhury, "Dopplesleep: A contactless unobtrusive sleep sensing system using short-range Doppler radar," in *Proc. ACM Int. Joint Conf. Pervas. Ubiquitous Comput.*, 2015, pp. 39–50.

[65] K.-Y. Chung, K. Song, K. Shin, J. Sohn, S. H. Cho, and J.-H. Chang, "Noncontact sleep study by multi-modal sensor fusion," *Sensors*, vol. 17, no. 7, p. 1685, 2017.

[66] E. Dafna, A. Tarasiuk, and Y. Zigel, "Sleep staging using nocturnal sound analysis," *Sci. Rep.*, vol. 8, Sep. 2018, Art. no. 13474.s

[67] J. J. Liu, Q. Wu, M.-C. Huang, N. Alshurafa, M. Sarrafzadeh, N. Raut, and B. Yadegar, "Sleep posture analysis using a dense pressure sensitive bedsheet," *Pervas. Mobile Comput.*, vol. 10, pp. 34–50, Feb. 2014.

[68] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals: A conditional adversarial architecture," in *Proc. 34th Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 4100–4109.

[69] A. Ren, B. Dong, X. Lv, T. Zhu, F. Hu, and X. Yang, "A non-contact sleep posture sensing strategy considering three dimensional human body models," in *Proc. 2nd IEEE Int. Conf. Comput. Commun.*, Oct. 2016, pp. 414–417.

[70] Y. K. Choi, G. Demiris, S. Y. Lin, S. J. Iribarren, A. C. Landis, H. J. Thompson, and T. M. Ward, "Smartphone applications to support sleep self-management: Review and evaluation," *J. Clin. Sleep Med.*, vol. 14, no. 10, pp. 1783–1790, 2018.

[71] E. Fino and M. Mazzetti, "Monitoring healthy and disturbed sleep through smartphone applications: A review of experimental evidence," *Sleep Breathing*, vol. 23, no. 1, pp. 13–24, 2019.

[72] A. A. Ong and M. B. Gillespie, "Overview of smartphone applications for sleep analysis," *World J. Otorhinolaryngol. Head Neck Surg.*, vol. 2, no. 1, pp. 45–49, 2016.

[73] F. Snyder, J. A. Hobson, D. F. Morrison, and F. Goldfrank, "Changes in respiration, heart rate, and systolic blood pressure in human sleep," *J. Appl. Physiol.*, vol. 19, no. 3, pp. 417–422, 1964.

[74] F. de Arriba-Pérez, M. Caeiro-Rodríguez, and J. M. Santos-Gago, "How do you sleep? Using off the shelf wrist wearables to estimate sleep quality, sleepiness level, chronotype and sleep regularity indicators," *J. Ambient Intell. Humanized Comput.*, vol. 9, no. 4, pp. 897–917, 2018.

[75] J. R. Lunsford-Avery, M. M. Engelhard, A. M. Navar, and S. H. Kollins, "Validation of the sleep regularity index in older adults and associations with cardiometabolic risk," *Sci. Rep.*, vol. 8, Sep. 2018, Art. no. 14158.

[76] K. S. Park and S. H. Choi, "Smart technologies toward sleep monitoring at home," *Biomed. Eng. Lett.*, vol. 9, no. 1, pp. 73–85, 2019.

[77] A. Dias, L. Gorzelniak, J. Rudnik, D. Stojanovic, and A. Horsch, "Detecting periodic limb movements with off-the-shelf accelerometers: A feasibility study," *Stud. Health Technol. Inform.*, vol. 192, nos. 1–2, pp. 152–156, 2013.

[78] A. Sathyanarayana, S. Joty, L. Fernandez-Luque, F. Ofli, J. Srivastava, A. Elmagarmid, T. Arora, and S. Taheri, "Sleep quality prediction from wearable data using deep learning," *JMIR mHealth uHealth*, vol. 4, no. 4, p. e125, 2016.

[79] D. Wang, K. K. Wong, G. C. Dungan, II, P. R. Buchanan, B. J. Yee, and R. R. Grunstein, "The validity of wrist actimetry assessment of sleep with and without sleep apnea," *J. Clin. Sleep Med.*, vol. 4, pp. 450–455, Oct. 2008.

[80] J. Accardo, *Sleep in Children With Neurodevelopmental Disabilities: An Evidence-Based Guide*. Cham, Switzerland: Springer, 2019.

[81] S. Güneş, K. Polat, and Ş. Yosunkaya, "Efficient sleep stage recognition system based on EEG signal using k-means clustering based feature weighting," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 7922–7928, 2010.

[82] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[83] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs modeling for highly imbalanced classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 39, no. 1, pp. 281–288, Feb. 2009.

[84] S. Abe, *Support Vector Machines for Pattern Classification*, 2nd ed. London, U.K.: Springer, 2010.

[85] N. Bhatia and A. Vandana, "Survey of nearest neighbor techniques," *Int. J. Comput. Sci. Inf. Secur.*, vol. 8, no. 2, pp. 302–305, 2010.

[86] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[87] K. A. I. Aboalayon, W. S. Almuhammadi, and M. Faezipour, "A comparison of different machine learning algorithms using single channel EEG signal for classifying human sleep stages," in *Proc. IEEE Long Island Syst., Appl. Technol.*, May 2015, pp. 1–6.

[88] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Burlington, MA, USA: Morgan Kaufmann, 2011.

[89] P. Jiang and R. Zhu, "Dual tri-axis accelerometers for monitoring physiological parameters of human body in sleep," in *Proc. IEEE Sensors*, vol. 6050, Oct./Nov. 2016, pp. 1–3.

[90] R. S. Rosenberg and S. Van Hout, "The american academy of sleep medicine inter-scorer reliability program: Sleep stage scoring," *J. Clin. Sleep Med.*, vol. 9, no. 1, pp. 81–87, Jan. 2013.

[91] S. Kampakis. (2016). *The Data Scientist-Everything About Data Science, Blockchain and AI*. Accessed: Jan. 7, 2018. [Online]. Available: http://thedatascientist.com/performance-measures-cohens-kappa-statistic/

[92] A. R. Hassan and A. Subasi, "A decision support system for automated identification of sleep stages from single-channel EEG signals," *Knowl.-Based Syst.*, vol. 128, pp. 115–124, Jul. 2017.

[93] N. Schaltenbrand, R. Lengelle, M. Toussaint, R. Luthringer, G. Carelli, A. Jacqmin, E. Lainey, A. Muzet, and J.-P. Macher, "Sleep stage scoring using the neural network model: Comparison between visual and automatic analysis in normal subjects and patients," *Sleep*, vol. 19, no. 1, pp. 26–35, 1996.

[94] A. Patanaik, J. L. Ong, J. J. Gooley, S. Ancoli-Israel, and M. W. L. Chee, "An end-to-end framework for real-time automatic sleep stage classification," *Sleep*, vol. 41, no. 5, pp. 1–11, 2018.

[95] K. Fujimoto, Y. Ding, and E. Takahashi, "Sleep stage detection using a wristwatch-type physiological sensing device," *Sleep Biol. Rhythms*, vol. 16, no. 4, pp. 449–456, 2018.

[96] K. Aggarwal, S. Khadanga, S. Joty, L. Kazaglis, and J. Srivastava, "A structured learning approach with neural conditional random fields for sleep staging," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2018, pp. 1318–1327.
[97] W.-H. Liao and C.-M. Yang, "Video-based activity and movement pattern analysis in overnight sleep studies," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.
[98] Q. Li, Q. Li, C. Liu, S. P. Shashikumar, S. Nemati, and G. D. Clifford, "Deep learning in the cross-time frequency domain for sleep staging from a single-lead electrocardiogram," *Physiol. Meas.*, vol. 39, no. 12, 2018, Art. no. 124005.
[99] A. R. Hassan and M. I. H. Bhuiya, "Automatic sleep scoring using statistical features in the EMD domain and ensemble methods," *Biocybernetics Biomed. Eng.*, vol. 36, no. 1, pp. 248–255, 2016.
[100] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Comput. Appl.*, vol. 28, no. 10, pp. 2959–2978, Oct. 2017.
[101] Y.-L. Hsu, Y.-T. Yang, J.-S. Wang, and C.-Y. Hsu, "Automatic sleep stage recurrent neural classifier using energy features of EEG signals," *Neurocomputing*, vol. 104, pp. 105–114, Mar. 2013.
[102] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.
[103] A. Koushik, J. Amores, and P. Maes, "Real-time sleep staging using deep learning on a smartphone for a wearable EEG," 2018, *arXiv:1811.10111*. [Online]. Available: https://arxiv.org/abs/1811.10111
[104] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Comput. Biol. Med.*, vol. 42, no. 12, pp. 1186–1195, 2012.
[105] A. Malafeev, D. Laptev, S. Bauer, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, J. M. Buhmann, and P. Achermann, "Automatic human sleep stage scoring using deep neural networks," *Frontiers Neurosci.*, vol. 12, p. 781, Nov. 2018.
[106] K. A. I. Aboalayon, H. T. Ocbagabir, and M. Faezipour, "Efficient sleep stage classification based on EEG signals," in *Proc. IEEE Long Island Syst., Appl. Techn. Conf.*, May 2014, pp. 1–6.
[107] X. Sun, L. Qiu, Y. Wu, Y. Tang, and G. Cao, "Sleepmonitor: Monitoring respiratory rate and body position during sleep using smartwatch," in *Proc. ACM Interact. Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, p. 104, 2017.
[108] Z. Shinar, A. Baharav, and S. Akselrod, "Detection of different recumbent body positions from the electrocardiogram," *Med. Biol. Eng. Comput.*, vol. 41, no. 2, pp. 206–210, 2003.
[109] G. Liu, K. Li, L. Zheng, W.-H. Chen, G. Zhou, and Q. Jiang, "A respiration-derived posture method based on dual-channel respiration impedance signals," *IEEE Access*, vol. 5, pp. 17514–17524, 2017.
[110] P. Jeng and L.-C. Wang, "Stream data analysis of body sensors for sleep posture monitoring: An automatic labelling approach," in *Proc. 26th Wireless Opt. Commun. Conf.*, Apr. 2017, pp. 1–4.
[111] R. M. Kwasnicki, G. W. V. Cross, L. Geoghegan, Z. Zhang, P. Reilly, A. Darzi, G. Z. Yang, and R. Emery, "A lightweight sensing platform for monitoring sleep quality and posture: A simulated validation study," *Eur. J. Med. Res.*, vol. 23, no. 1, 2018, Art. no. 28.
[112] H. Yoon, S. Hwang, D. Jung, S. Choi, K. Joo, J. Choi, Y. Lee, D.-U. Jeong, and K. Park, "Estimation of sleep posture using a patch-type accelerometer based device," in *Proc. 37th Conf. Proc. IEEE Eng. Med. Biol. Soc.*, Aug. 2015, pp. 4942–4945.
[113] M. B. Pouyan, J. Birjandtalab, M. Heydarzadeh, M. Nourani, and S. Ostadabbas, "A pressure map dataset for posture and subject analytics," in *Proc. IEEE EMBS Int. Conf. Biomed. Health Inform.*, Feb. 2017, pp. 65–68.
[114] W. Viriyavit, V. Sornlertlamvanich, W. Kongprawechnon, P. Pongpaibool, and T. Isshiki, "Neural network based bed posture classification enhanced by Bayesian approach," in *Proc. 8th Int. Conf. Inform. Commun. Technol. Embedded Syst.*, May 2017, pp. 1–5.
[115] G. Matar, J.-M. Lina, J. Carrier, A. Riley, and G. Kaddoum, "Internet of Things in sleep monitoring: An application for posture recognition using supervised learning," in *Proc. IEEE 18th Int. Conf. e-Health Netw., Appl. Services*, Sep. 2016, pp. 1–6.
[116] M. Wang, P. Liu, and W. Gao, "Automatic sleeping posture detection in ballistocardiography," in *Communications, Signal Processing, and Systems*. Singapore: Springer, 2019, pp. 1785–1792.
[117] M. Enayati, M. Skubic, J. M. Keller, M. Popescu, and N. Z. Farahani, "Sleep posture classification using bed sensor data and neural networks," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jun. 2018, pp. 461–465.
[118] M. Araujo, R. Bhojwani, J. Srivastava, L. Kazaglis, and C. Iber, "ML approach for early detection of sleep apnea treatment abandonment: A case study," in *Proc. Int. Conf. Digital Health*, 2018, pp. 75–79.
[119] A. Q. Javaid, C. M. Noble, R. Rosenberg, and M. A. Weitnauer, "Towards sleep apnea screening with an under-the-mattress IR-UWB radar using machine learning," in *Proc. IEEE 14th Int. Conf. Mach. Learn. Appl.*, Dec. 2015, pp. 837–842.
[120] P. D. Sorlie, L. M. Avilés-Santa, S. Wassertheil-Smoller, R. C. Kaplan, M. L. Daviglus, A. L. Giachello, N. Schneiderman, L. Raij, G. Talavera, M. Allison, L. LaVange, L. E. Chambless, and G. Heiss, "Design and implementation of the hispanic community health study/study of latinos," *Ann. Epidemiol.*, vol. 20, no. 8, pp. 629–641, 2010.
[121] G. C. Gutiérrez-Tobal, D. Álvarez, A. Crespo, F. del Campo, and R. Hornero, "Evaluation of machine-learning approaches to estimate sleep apnea severity from at-home oximetry recordings," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 882–892, Mar. 2019.
[122] F. Chung, P. Liao, H. Elsaid, S. Islam, C. Shapiro, M. Colin, and Y. Sun, "Oxygen desaturation index from nocturnal oximetry: A sensitive and specific tool to detect sleep-disordered breathing in surgical patients," *Anesthesia Analgesia*, vol. 114, no. 5, pp. 993–1000, 2012.
[123] S. Chambon, V. Thorey, P. J. Arnal, E. Mignot, and A. Gramfort, "A deep learning architecture to detect events in EEG signals during sleep," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process.*, Sep. 2018, pp. 1–6.
[124] K. Lacourse, J. Delfrate, J. Beaudry, P. Peppard, and S. C. Warby, "A sleep spindle detection algorithm that emulates human expert spindle scoring," *J. Neurosci. Methods*, vol. 316, pp. 3–11, Mar. 2018.

SARAH FALLMANN received the M.Sc. degree in mathematics and biomedical engineering from the Vienna University of Technology, Vienna, Austria, in 2015. She is currently pursuing the Ph.D. degree in computer science with De Montfort University, Leicester, U.K. She was with the Austrian Institute of Technology and is currently employed as a Researcher within the European H2020 ITN Acrossing project. Her current research interests include sleep behavior analysis, and applying machine learning techniques in the field of data analysis for sleep parameters, sleep positions, sleep stages, and disorder investigations.

LIMING CHEN received the B.Eng. and M.Eng. degrees from the Beijing Institute of Technology (BIT), China, and the Ph.D. degree in artificial intelligence from De Montfort University, U.K. He has extensive research expertise. He is currently a Professor of data analytics with the School of Computing, Ulster University, U.K. He has published over 200 peer-reviewed articles in top-tier IEEE and ACM TRANSACTIONS and journals. His current research interests include data analytics, artificial intelligence, cyber-physical systems, and their application in assisted living and smart healthcare. He is an IET Fellow, and has secured multimillions research grants from the UK, EU research funding bodies.