Computational Sleep Behaviour Analysis and Application

Sarah Fallmann

Thesis submitted for the degree of Doctor of Philosophy

School of Computer Science and Informatics
De Montfort University

This work has been supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 676157.

July, 2020

Declaration

I declare that the work was solely conducted during registration for the above award with De Montfort University, under University supervision.

I declare that no material contained in the thesis has been used in any other submission for an academic award.

I confirm that the work represented in this submission was undertaken solely by myself except where otherwise attributed. In this thesis editorial assistance was used, in compliance with the regulations outlined in the Code of Practice.

I acknowledge the copyright and other intellectual property rights in relation to my thesis and other work prepared and submitted in the course of my studies shall belong to the University, except where specifically agreed otherwise by the University in writing.

Acknowledgements

I want to thank my supervisors, Prof. Liming Chen and Dr. Feng Chen, for guiding me throughout my PhD programme. Their continual support, feedback, and patience helped me develop my skills, expand my knowledge, and become a better researcher. In addition to my direct supervisors, I would like to thank my advisor, Prof. Li Yang, and assessor, Prof. Raouf Hamzaoui.

My sincere thanks go to all of the individuals who participated in my studies. This includes everyone who took part in my sleep position study in Austria and those who participated in my interview data collection process at Zhongshan hospital in collaboration with Dalian University. I am grateful to Dalian University for hosting me during my time in China; in particular, I would like to thank Prof. Chengsheng Pan for his hospitality. Moreover, I was able to collect data in a smart lab at Great Northern Haven, which was a wonderful experience, thanks to Julie Doyle and her team at the Dundalk Institute of Technology.

None of this would have been possible without the grant from the ACROSSING project, a network funded by the European Union Horizon 2020 Marie Skłodowska-Curie Actions program, that funded my research, training, and networking. I would like to thank everyone involved in organising, lecturing, challenging, encouraging, and discussing ideas throughout my time working on this project. Additionally, I would like to thank the Computer Science and Informatics Department for providing a tuition fee waiver scholarship.

Thanks also go to all of my collaborators involved in either discussions or publications. The exchange of ideas and critical thinking is essential in the development of a thesis—this was only possible because of my involvement and discussions with researchers from a variety of backgrounds. Special thanks also go to Rick van Veen for the continual encouragement, support, and engaging discussions on research.

Last but not least, I would like to express my gratitude for the support from my family during the ups and downs of my entire PhD journey.

Abstract

Sleep affects a person's health and is, therefore, assessed if health problems arise. Sleep behaviour is monitored for abnormalities in order to determine if any treatments, such as medication or behavioural changes (modifications to sleep habits), are necessary. Assessments are typically done using two methods: polysomnography over short periods and four-week retrospective questionnaires. These standard methods, however, cannot measure current sleep status continuously and unsupervised over long periods of time in the same way home-based sleep behaviour assessment can. In this work, we investigate the ability of sleep behaviour assessment using IoT devices in a natural home environment, which potential has not been investigated fully, to enable early abnormality detection and facilitate self-management.

We developed a framework that incorporates different facets and perspectives to introduce focus and support in sleep behaviour assessment. The framework considers users' needs, various available technologies, and factors that influence sleep behaviours. Sleep analysis approaches are incorporated to increase the reliability of the system. This assessment is strengthened by utilising sleep stage detection and sleep position recognition. This includes, first, the extraction and integration of influence factors of sleep stage recognition methods to create a fine-grained personalised approach and, second, the detection of common but more complex sleep positions, including leg positions. The relations between medical conditions and sleep are assessed through interviews with doctors and users on various topics, including treatment satisfaction and technology acceptance. The findings from these interviews led to the investigation of sleep behaviour as a diagnostic indicator. Changes in sleep behaviour are assessed alongside medical knowledge using data mining techniques to extract information about disease development; the following diseases were of interest: sleep apnoea, hypertension, diabetes, and chronic kidney disease. The proposed framework is designed in a way that allows it to be integrated into existing smart home environments. We believe that our framework provides promising building blocks for reliable sleep behaviour assessment by incorporating newly developed sleep analysis approaches. These approaches include a

modular layered sleep behaviour assessment framework, a sleep regularity algorithm, a user-dependent visualisation concept, a higher-granularity sleep position analysis approach, a fine-grained sleep stage detection approach, a personalised sleep parameter extraction process, in-depth understanding on sleep and chronic disease relations, and a sleep-wake behaviour-based chronic disease detection method.

Acronyms

AASM	American Academy of Sleep Medicine.
AHI	apnoea-hypopnoea index.
AUC	area under the curve.
BMI	body mass index.
CKD	chronic kidney disease.
CNN	convolutional neural network.
CV	cross-validation.
DNN	deep neural network.
DT	decision tree.
ECG	electrocardiogram.
EEG	electroencephalography.
EMG	electromyogram.
EOG	electrooculogram.
GMLVQ	generalised matrix learning vector quantisation.
HCHS	Hispanic Community Health Study.
HMM	hidden Markov model.
HRV	heart rate variability.
IoT	Internet of Things.
κ	Cohen's kappa.
KNN	k-nearest neighbour.
LDA	linear discriminant analysis.
LOOCV	leave-one-person-out cross-validation.
LR	logistic regression.
LSTM	long short-term memory network.
LVQ	learning vector quantisation.
MESA	multi-ethnic study of atherosclerosis.
MLP	multilayer perceptron.
NN	neural network.
NNI	normal-to-normal interval.
NSRR	national sleep research resource.
OMR	overnight metabolic rate.
OSAS	obstructive sleep apnoea syndrome.
PCA	principal component analysis.
PPG	photoplethysmography.
PSG	polysomnography.
PSQI	Pittsburgh sleep quality index.
REM	rapid eye movement.
RF	random forest.
R-K	Rechtschaffen and Kales method.
RLS	restless leg syndrome.

RNN recurrent neural network.

ROC receiver operating characteristic curve.

SD standard deviation.

SE sleep efficiency.

SOL sleep-onset latency.

SVM support vector machine.

TST total sleep time.

WASO wake after sleep onset.

Glossary

HCHS	is a dataset provided by the national sleep research resourche (NSRR) containing sleep-related data from polysomnograhpy (PSG) and actigraphy from individuals with different backgrounds.
HRV	is a measure which indicates changes in the time interval between consecutive heartbeats.
LSTM	is a neural network for time series data analysis.
MESA	is a dataset provided by the NSRR containing sleep-related data from PSG and actigraphy from individuals with different backgrounds.
PPG	is an optical technique to measure blood volume changes from the skin surface.
PSG	is the gold standard of sleep assessment usually undertaken in a hospital, including multiple sensor sources attached to the human body.
Shimmer3 sensor	is an intertial measurement unit, providing the possibility to collect data with 9 degrees of freedom from integrated accelerometer, gyroscope and magnetometer sensors.
SOL	is the duration of sedentary time, i.e., the time from light-outs till the time of falling asleep.

Contents

List of Figures

List of Tables

List of Publications

Chapter 2 is based on the published work in the following paper:

[**A**] Fallmann S. and Chen L., *Computational Sleep Behavior Analysis: A Survey*, IEEE Access, vol. 7, pp. 142421–142440, 2019. DOI: 10.1109/ACCESS.2019.2944801

Chapter 3 is based on the published work in the following papers:

[**B**] Fallmann S., Chen L., and Chen F., *A Home-based IoT-enabled Framework for Sleep Behaviour Assessment*, 2019 IEEE Cyber Science and Technology Congress (CyberSciTech), Fukuoka, Japan, 2019, pp. 7–15. DOI: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00018

[**C**] Fallmann S., Psychoula I., Chen L., Chen F., Doyle J., and Triboan D., *Reality and Perception: Activity monitoring and data collection within a real-world smart home*, 2017 IEEE Ubiquitous Intelligence and Computing (UIC), San Francisco, CA, 2017, pp. 1–6. DOI: 10.1109/UIC-ATC.2017.8397533

Chapter 4 is based on the published work in the following paper:

[**D**] Fallmann S., van Veen R., Chen L., Walker D., Chen F., and Pan C., *Wearable Accelerometer Based Extended Sleep Position Recognition*, Proceedings of the 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), Dalian, 2017, pp. 1–6. DOI: 10.1109/HealthCom.2017.8210806

Chapter 5 is based on the following published and submitted work:

[**E**] Fallmann S., Chen L., and Chen F., *Fine-Grained Sleep-Wake Behaviour Analysis*, 2019 IEEE Ubiquitous Intelligence and Computing (UIC), Leicester, UK, 2019, pp. 667-674, DOI: 10.1109/SmartWorld-UIC-ATC-SCALCOM-IOP-SCI.2019.00150

[**F**] Fallmann S., Chen L., and Chen F., *Learning Personalized Sleep-Wake Pat-*

terns from Multi-Source Data Applied to Personalized Sleep Parameter Extraction, Personal and Ubiquitous Computing, Special Issue on Social Networks and Social Ubiquitous Computing, in press

Chapter 7 is based on the published work in the following paper:

[G] Fallmann S. and Chen L., *Detecting Chronic Diseases from Sleep-Wake Behaviour and Clinical Features*, Proceedings of the 5th International Conference on Systems and Informatics (ICSAI), Nanjing, 2018, pp. 1076–1084. DOI: 10.1109/ICSAI.2018.8599388

Further publications:

[H] Kapidis G., Teriús-Padrón J., Fallmann S., Merdivan E., Hanke S., García-Betances R., and Cabrera-Umpierrez M., *Towards Self-management of Chronic Diseases in Smart Homes*, 2018 IEEE International Conference on Pervasive Computing and Communications (PerCom) Workshops, Athens, 2018, pp. 776–781. DOI: 10.1109/PERCOMW.2018.8480163

[I] Triboan D., Chen L., Chen F., Fallmann S., and Psychoula I., *Real-time Sensor Observation Segmentation for Complex Activity Recognition within Smart Environments*, 2017 IEEE Ubiquitous Intelligence and Computing (UIC), San Francisco, CA, 2017, pp. 1–8. DOI: 10.1109/UIC-ATC.2017.8397487

[J] Chen L., Fallmann S., López-de-Ipiña D., Pan C., and Ning H., *Context, Intelligence and Interactions for Personalized Systems*, Journal of Ambient Intelligence and Humanized Computing, vol. 9, pp. 1557–1559, 2019. DOI: 10.1007/s12652-018-0985-y

Chapter 1

Introduction

1.1 Background

Sleep behaviour is a key factor in health and influences the daily lives of all people. Sleep is a recovery mechanism during which heart rate and breathing are slowed, the body approaches a state of paralysis, and the brain processes experiences from throughout the day and relaxes. Sleep is necessary to sustain life, though the evolutionary reasons behind this process have yet to be fully explored. Sleep status can be assessed using physical or physiological measurements, such as respiration rate, heart rate, temperature, and body movement [19]. Based on these measurements, sleep behaviour can be determined in terms of sleep duration, sleep-onset latency (SOL), arousal, wake after sleep onset (WASO), and sleep efficiency (SE) [6]. Good rest benefits overall well-being and enhances the ability to cope with pressure and stress [20].

Everyone has fluctuations between good and bad rest due to several different factors, including lifestyle changes, stress, illness, and environmental conditions. Persistently poor sleep is a severe problem and is likely indicative of chronic disorders and sleep issues. Such a problem should be monitored to avoid these consequences. The importance of sleep analysis must be addressed, as poor sleep is an issue that affects everyone, including those living with a chronic disease, such as diabetes [21], chronic kidney disease (CKD) [22], cardiovascular disease [23], and depression [24,25]. Sleep issues initially emerge as difficulty falling or staying asleep, restless legs, and daytime sleepiness [26]. The extent of sleep issues among those with chronic diseases is substantial: arthritis affects one in five adults, 80% of whom are living with sleep issues [26–28]. In 2013, 16% of all adults were affected by CKD [29, 30], many suffer from conditions such as restless leg syndrome (RLS) and sleep apnoea [31,32]; between 30 and 50% of people diagnosed with cancer deal with sleep issues during

treatment—for several years after treatment, insomnia symptoms persist for 23 to 44% of all patients [33].

Insufficient sleep leads to a loss in productivity and a higher risk of mortality, which result in economic costs of about 2.92% of the gross domestic product (GDP) for Japan, 2.28% for the US, 1.35% for Canada, 1.86% for Germany, and 1.56% for the UK [34]. It is a pervasive issue that can be targeted by maintaining a healthy sleep routine and avoiding sleep deprivation. Sleep disorders are usually related to inadequate sleep quality. They are quite common and result in significant costs to the health care and economic system [19, 35]. They are a seriously under-recognised issue, and this lack of recognition leads to direct and indirect healthcare costs affecting a variety of factors, such as inpatient hospital costs, pharmaceuticals, and work-related injuries [35].

One social consequence of sleep issues is fatigue, which results in a reduction in quality of life. Affected individuals feel unable to actively participate in daily life; they are exhausted and unable to adequately perform tasks. Their mental health is affected by memory loss, concentration issues, drowsiness, and mood changes. Their physical health is affected by a worsened immune system, balance issues, weight gain, high blood pressure, and a higher risk of diabetes and heart disease [36].

Essentially, many problems arise from insufficient or inadequate sleep. The daily lives of individuals are considerably impacted both mentally and physically. The economic costs stemming from sleep issues are a burden for society. As a result, it is necessary to automatically monitor and assess sleep, ideally in a natural sleep environment, without intervening in people's sleep routines. To conduct this home-based sleep assessment, we must employ sensor technology and data analytics.

Sleep behaviour has traditionally been studied in clinical environments. Polysomnography (PSG) is considered the medical gold standard for sleep disorder classification [6]; it monitors brain waves, respiration, heart rate, sleep positions, movement, and sounds. These data are used to assess SE, SOL, arousal index, sleep stages, and other sleep disorder-related parameters. An alternative to PSG is non-invasive actigraphy, which monitors acceleration of movements [19]. Clinicians use these technologies despite their restrictions. The restrictions for PSG include exclusively short-term sleep monitoring, expensive equipment, and the necessary presence of an expert at all times. Additionally, a first-night effect results from sensors influencing sleep behaviour, meaning two consecutive recording nights are necessary, especially for patients with insomnia [37]. For actigraphy, the available information is limited to movement data. Automated processes would provide faster diagnoses

of sleep disorders and insight into new aspects of night behaviour. The development of Internet of Things (IoT) technology would make this possible: sleep trackers for home-usage are reaching the market as both wearables (e.g., smartwatches [38]) and non-wearables (e.g., pressure mats [39]). These trackers are currently limited by the accuracy when compared to PSG.

Current technologies used for sleep assessment (1) interfere with sleep due to the number of necessary sensors, (2) are often not comparable with the gold standard, (3) are misinterpreted by users (e.g., the amount of deep sleep), and (4) are generally costly. For individuals with medical conditions, managing their own health is essential [40, 41], and sleep management is an important part of health monitoring. Individuals suffering from or likely to develop a sleep disorder could seriously benefit from sleep-monitoring systems. These systems allow for self-assessment and self-management on a daily basis and, in turn, the ability to detect anomalies earlier. Medical experts believe that there is a need for reliable home-based systems to monitor natural sleep behaviour [42]. The comprehensive information that would come from such innovations would benefit (1) doctors by helping them facilitate personalised advice and treatment to their patients as well as (2) researchers by enabling them to investigate connections and correlations between various relevant factors (e.g., day and night behaviours) [43]. Evidently, an adaptable system that incorporates doctors' recommendations and users' changing needs is necessary; sleep research seeks to build a reliable sleep-tracking system that enables people to gain insights into their sleep behaviour at home. Computational methods could likely address more complex problems such as the diagnosis of sleep disorders, recurrent sleep patterns, and how sleep influences other areas. Data are analysed and assessed through data mining and machine learning techniques to extract key sleep parameters and indicators. An IoT solution would save costs relative to standard sleep assessment using a PSG [44].

Many studies have investigated the use of various technologies in sleep monitoring and assessment as well as the ways other factors influence sleep. Although substantial progress has been made in this research field, challenges and gaps remain. An overall framework is missing—the lack of a clear guideline or basis for future research and assessment that accurately and understandably represents the many aspects of sleep that is necessary to avoid misinterpretation—and more investigations on sleep-related disease detection during early and advanced stages are necessary. Limitations persist in terms of the accuracy and validity of the proposed methods towards the gold standard—the lack of objective and reliable sleep assessment— (1) reliability being limited to laboratories; (2) investigations on single parameters; (3)

one-model-fits-all approaches; (4) only considering healthy individuals; (5) focusing on one type of domain knowledge; and (6) the lack of adequate comparative outcomes and standards. The field must work towards a defined and reliable system that presents sleep behaviour in a straightforward manner to avoid misinterpretation and clearly highlight ways to improve.

1.2 Research Problem and Issues

In recent years, sleep behaviour analysis has advanced considerably through the introduction of new devices and computational methods; nevertheless, there are still limitations and challenges that must be addressed. These are (1) current reliable assessment and diagnostics being limited to laboratory settings; (2) investigations focusing on single parameters; (3) shortcomings of current sleep behaviour analysis methods because of the use of one-model-fits-all approaches; (4) computational sleep assessment methods considering only healthy individuals; (5) the limitation to one type of domain knowledge either computational or medical; and (6) the lack of adequate comparative outcomes and standards.

The first problem is that current reliable assessment and diagnostics are limited to laboratory settings. Usually, individuals who inform their doctors of their sleep issues are either provided medication or referred to specialists to undergo tests. The usually applied tests are either subjective questionnaires or an expensive and time-limited PSG. Sensor technology is increasingly receiving attention as a way to objectively assess natural sleep behaviour at home. Additionally, home-based sleep assessment are done to make people aware of their sleep status, monitor trends over an extended period of time, detect anomalies, and provide a basis for self-management. Even though several approaches have been investigated for home-based assessment, they are not designed to be used as diagnostic tools and consequently, not actively used by medical people. The main reasons are the reliability and the lack of real-world tested methods. Home-based assessment faces variations in environments and sensors, therefore, reaching a good performance compared to the gold standard is challenging. Other issues in home-based assessment are (1) the users' adherence whom potentially attach sensors incorrectly or prematurely stop the collection, and (2) environmental difficulties due to sensor failures and loss of data in transmission. As sleep is highly influenced by medical conditions, there is an application for proper sleep assessment to be used as a diagnostic tool. Advances in data analysis tools provide the foundation to conduct complex time-series analysis which enable sleep

behaviour-based diagnostics.

The second problem is that current research focuses on single sleep parameters. The main evaluated factors are sleep stages and sleep positions. These investigations do not account for the bigger picture of sleep assessment that limits the detection of changes to one parameter. It is evident that, first, methods needed to assess sleep parameters reliably, but with the increasing interest and methods available, it is a huge challenge to develop a system which can bring different aspects together in one place and present them understandably. These aspects should not only relate to sleep assessment but also to environmental influence factors. This can be achieved with a sleep behaviour assessment framework which can act as a guideline. The main element of such an overall framework is to be still applicable to the specific scenarios and context researchers and users want to investigate. The requirements for such a framework are immense as it is to be adaptable to existing environments, and flexible to users' needs. The benefit of such a framework supersedes the drawback of its challenges, because when available, behavioural interventions can easier be developed and assessed.

The third problem is shortcomings of current sleep behaviour analysis methods because of the use of one-model-fits-all approaches that exclude differences in subject groups. As sleep can be interpreted very subjectively [45] and influence people differently, individuality should be addressed and considered by assessment systems. The currently available systems do not take into account different influence factors, which may have an effect on the performance compared to the gold standard (PSG). Including new features in the training process to develop tailor-made training models for different groups can address this issue. These new features should either include knowledge of the user's condition or more accurately represent behavioural differences.

The fourth problem is that computational sleep assessment methods consider only healthy individuals. In sleep behaviour analysis, there is a need to distinguish healthy individuals from those with a medical condition. However, current research mainly targets healthy individuals. The performance on individuals with a medical condition have shown to be lower which shows that current methods do not generalise well. The investigation of diverse datasets could overcome this issues and reveal currently unknown problems in sleep research. Improving home-use technology for healthy individuals and those with a medical condition can enable the consideration of individual factors and the investigation of more complex problems, such as higher-granularity sleep positions.

The fifth problem is the limitation to one type of domain knowledge either computational or medical. While medical research on sleep behaviour has established significant knowledge of the relations and structure of sleep over the last few decades, it has yet to be sufficiently incorporated in computational analysis methods. Computational investigations lack the fusion with medical knowledge in terms of sleep structure, relations, and influence factors. Medical knowledge can be advantageous when establishing computational sleep behaviour methods. Current sleep research in the medical field concentrates on exploring features that provide potential hints about specific diseases through qualitative analysis. Today, these features can easily be measured continuously at home using sensor technology. Combining quantitative measurements with known relations could help diagnose diseases in the early stages. Early disease detection is important, as people often learn about sleep issues quite late, which prevents preventive approaches. Therefore, to constantly update users on their daily sleep habits [45] and anomalies, data analytics could provide in-depth and important insights by using the available knowledge from the computational and medical field.

The sixth problem is the lack of adequate comparative outcomes and standards. It is not yet standard in sleep research to provide the validation method used; repeated cross-validation (CV) and user-independent validation are critical in providing a reliable outcome for the investigated population. Usually, the target is a generalised approach. As a result, users who are used during training cannot be used for testing; otherwise, the performance ability of the method is invalid. No performance measure has been standardised, but overall accuracy is commonly used. In cases of imbalance in the data, however, further measurements, such as recall and precision, are required. Therefore, we recommend standardising the performance metrics used in sleep behaviour analysis to allow for accurate comparative studies. Comparative studies must be performed to investigate methods using the same datasets to benchmark available approaches and determine the most promising methods. A dynamic framework that incorporates technological advances from various fields could prove helpful for effective comparative assessment.

The outlined problems must be addressed to attain a reliable sleep assessment—data-driven approaches can provide valuable input. The primary motivation for this thesis is to develop solutions for all of the aforementioned problems.

1.3 Aims and Objectives

The overarching aim of the thesis is to develop a home-based modular sleep behaviour assessment framework that is adaptable to individuals' needs and doctors' recommendations. The framework is designed for those who are interested in learning about their sleep behaviour, individuals affected by a medical condition, and medical professionals. The sleep assessment within our standardised framework will allow (1) individuals to actively improve their daily lives and well-being and (2) doctors to provide accurate assessments and recommendations. This framework is built upon adaptive data analysis models and domain knowledge to support a flexible sleep assessment. A layered structure is used to ensure adaptability.

To support sleep parameter extraction, the framework uses a sleep stage recognition process that incorporates influence factors of sleep and a sleep position detection approach for higher-granularity positions. Additionally, an individual regularity assessment per parameter is used to provide in-depth information on consistent sleep routines. Chronic disease detection is also integrated to adapt to changes in users' health status. A visualisation concept provides users and doctors with appropriate information. Finally, experiments and evaluations are performed for all developed technologies.

The key objectives are:

1. To develop a framework that can be used as a guideline for future research by integrating sleep behaviour parameters that are relevant and appropriate for home-based monitoring.
2. To develop a sleep regularity algorithm that reflects the regular sleep habit for individual sleep parameters by measuring differences within multiple days.
3. To design a visualisation concept to ensure information is presented to users accurately and understandably by providing different levels of abstraction.
4. To develop a higher-granularity sleep position analysis that can detect complex positions accurately.
5. To develop a sleep stage recognition approach that incorporates influence factors from clinical history and genetic information to detect personal sleep/wake patterns.
6. To develop a sleep parameter extraction concept that facilitates personalised assessment by introducing dynamic thresholds.

7. To develop an in-depth understanding of relevant stakeholders on the relations between chronic diseases and sleep.

8. To develop a chronic disease detection approach using sleep-wake behaviour and clinical history data while exploiting the temporal aspects of the data.

1.4 Scope and Methodology

The main goal of sleep behaviour analysis is to reliably extract information about sleep parameters and present them easy interpretable to users. Despite the number of opportunities for sleep research, the scope of this thesis is limited to the analysis in a natural home environment in order to address the topics presented in Section 1.3. We restrict our work to analytical data investigations combining sensor data with clinical history data.

To achieve this goal, we adopted the following methodology:

First, we performed a literature review on sleep behaviour analysis related to topics such as sleep characteristics, sleep monitoring, and data collection. Moreover, we conducted an in-depth survey on state-of-the-art computational analysis for sleep stage classification, sleep position recognition, and medical condition investigation; and identified some of the open issues.

Second, we performed several sleep behaviour analysis studies wherein wearable devices were attached to individuals throughout the night and their sleep behaviour was monitored. The studies entail the development of a framework for sleep behaviour assessment at home, relying on domain knowledge. Domain knowledge is represented by a layered structure incorporating data analysis techniques to support the detection of sleep behaviour parameters.

The framework consists of eight layers and four main components: circadian rhythm, regularity, sleep quality, and potential for improvement. The investigation pursues the following steps:

- Based on the literature, extract the most important parameters for objective sleep behaviour assessment.

- Include knowledge from medical researchers about the importance of each parameter and the relevance of specific influence factors on sleep, such as age and gender.

- Determine appropriate parameters for in-home use and test, using national sleep research resource (NSRR) datasets, if the extracted parameters are significantly different within groups (health status, sleep disorders, and chronic diseases).
- Develop a method to calculate regularity for individual sleep parameters
- Establish a sleep behaviour assessment framework that improves upon current approaches and provides an overall picture of important aspects within a modular structure that adapts to technological possibilities, available data sources, and user preference.
- Propose a method to present data to users in an accessible way to avoid misinterpretation and to present data to doctors with a comprehensive picture of users' sleep behaviour status.

The studies we conducted focus on individual aspects of the framework components. A higher-granularity sleep position detection was developed to recognise more complex positions:

- Investigate (1) common positions, especially those that are related to specific diseases; (2) appropriate signal sources; and (3) optimal sensor positions.
- Collect and pre-process the dataset from a simulated and real-world setting on the common positions using three Shimmer3 sensor (accelerometer data used).
- Establish a process of detecting higher-granularity positions by applying generalised matrix learning vector quantisation (GMLVQ).

A fine-grained sleep stage recognition process was developed that is capable of providing dynamic, individual-specific sleep-wake detection, which can be used to detect sleep parameters in a personalised manner:

- Assess which influence factors are affecting sleep and wake stages.
- Propose a fine-grained sleep stage detection approach
- Develop an adaptive sleep parameter extraction process with a personalised threshold-based calculation

Moreover, the perceptions of doctors and individuals with a medical condition are gathered to provide the basis for the development of a chronic disease detection approach:

- Assess which individuals with a medical condition are affected by poor sleep

quality by conducting interviews with affected individuals and specialists. This assessment will underline the importance of sleep for chronic disease affected individuals and the variety of views on the matter.

- Analyse sleep-wake behaviour and determine which clinical history factors are related to symptoms and useful for chronic disease diagnosis.
- Propose a method, including relevant information sources from actigraphy and clinical history, to detect chronic diseases.
- Start preliminary work on forecasting chronic disease development by applying early-stage detection mechanisms.
- Extract clinical history factors that are significant for misclassification.

Third, we performed experiments to evaluate individual approaches for sleep position detection, sleep stage recognition, chronic disease detection, and regularity assessment. Performance was compared to the gold standard of PSG where applicable.

Our investigations were restricted to wearable devices in a home environment and did not take into consideration electroencephalography (EEG). We conducted our data analysis using objective measurements and excluded experiments and comparisons on a subjective level. Further details on the different aspects of our data-driven approach are presented in the subsequent chapters.

1.5 Knowledge Contributions

In addressing the needs of and gaps in sleep behaviour analysis, we made the following knowledge contributions:

- **Modular layered sleep behaviour assessment framework**

  We developed an adaptable sleep behaviour assessment framework that considers individuals' needs and doctors' recommendations. The unique framework characteristics permit personalisation, adaptable presentation, flexibility (towards data sources, needs, and layer integration), and the combination of four main sleep aspects: sleep quality, regularity, circadian rhythm, and the potential for improvement with respect to changes in environmental and sleep hygiene factors. The layered structure can serve as a guideline for future research, as building blocks facilitate smooth integration of technological ad-

vances. This knowledge contribution also includes a sleep regularity algorithm and a user-dependent visualisation concept.

  - **Sleep regularity algorithm**

    We developed a sleep parameter-specific regularity algorithm that considers multiple continuous days to represent individuals' regular sleep habits.

  - **User-dependent visualisation concept**

    We designed a visualisation concept that is adaptable to doctors and users. A layered structure providing different levels of abstraction is followed by a set of rules, providing simple visualisation to avoid misinterpretation by daily users while still allowing for an in-depth view for doctors.

- **Higher-granularity sleep position analysis approach**

  We have developed and tested an approach to detect higher-granularity sleep positions. Real-time assessment at a low computational cost is reached by using a GMLVQ. Representative patterns for individual positions are extracted from real-world data, providing a mobile solution for position detection.

- **Fine-grained sleep stage detection approach**

  We have developed and evaluated a fine-grained sleep stage detection approach that combines sensor data with information from clinical history and genetics. Influences on sleep-wake detection can, therefore, be captured in the models configured for specific groups. This approach is adaptable to available data sources and supersedes current one-model-fits-all approaches. This contribution also involves a personalised sleep parameter extraction process.

  - **Personalised sleep parameter extraction process**

    We developed a process to approach a flexible sleep parameter assessment for groups of individuals. Dynamic thresholds are used to reach personalised sleep parameters and provide a better fit towards the gold standard.

- **Sleep-wake behaviour-based chronic disease detection method**

  We developed and tested a method that uses sleep-wake behaviour to detect chronic disease. This method is designed to deal with temporal data and

multi-dimensional features constructed from clinical history and sensor data—this guarantees a robust model. This knowledge contribution is based on our findings on sleep and chronic disease relations.

– **In-depth understanding on sleep and chronic disease relations**

We performed a thematic analysis on sleep and chronic disease relations by considering how doctors and affected individuals perceive sleep problems (e.g., their thoughts on causes, impacts, assessment, treatment, and technology usage). In-depth understanding of this topic was achieved and the findings encourage further investigation of technological advances in sleep assessment and chronic disease diagnosis.

1.6 Outline of the Thesis

The remaining chapters are structured as follows.

Chapter 2 reviews the literature on sleep behaviour, including sleep monitoring, measurements, and computational analysis methods. It examines in detail sleep stage detection, sleep position recognition, and medical condition classification—its strengths, weaknesses, and the state of the art. Furthermore, challenges and issues for current sleep behaviour analysis are presented and addressed in this thesis.

Chapter 3 addresses the current gap for a guideline in sleep assessment that fuses different aspects of sleep and discusses how this is a knowledge contribution by providing a modular layered sleep behaviour assessment framework and proposing novel approaches to address the one-model-fits-all issue. First, Chapter 3 introduces the overall sleep behaviour assessment framework for home use (i.e., its layered structure, components, and characteristics). Second, we present a sleep regularity algorithm and data visualisation concept for both users and doctors. Third, we look at the opportunities and difficulties of sensor integration as well as the challenges it faces in existing environments (e.g., smart homes). Finally, we evaluate specific aspects with tests and discussions on objective measurements, emphasising, sleep regularity, visualisation, integration aspects, limitations, and implications.

Chapter 4 targets stable state analysis by addressing the lack of a more complex sleep position analysis and contributing to knowledge with a higher-granularity sleep position classification approach. It integrates stable state detection and feature extraction for learning vector quantisation (LVQ). The dataset collected for evaluation

is described, and the position-detection approach is applied in a real-world and simulated setting.

Chapter 5 focuses on sleep stage detection and its application on sleep parameter extraction. It closes the accuracy gap with PSG measurements, addresses the issue of imbalanced datasets and one-model-fits-all approaches, and provides adequate comparative outcomes. By doing so, Chapter 5 contributes to knowledge with (1) fine-grained sleep stage detection through the investigation of influence factors and (2) a proposed personalised sleep parameter extraction process. Additionally, the evaluation steps are described and results are presented and discussed.

Chapter 6 establishes user and doctor opinions on sleep issues and their effects from a chronic disease perspective. Thereby, we contribute to knowledge by investigating themes amid a currently underrepresented population and looking at differences among our sample groups. A summary of the correlation between sleep issues and chronic diseases is provided. The data collection process is described and thematic analysis is performed, in which fundamental sleep problems, causes, impacts, assessment, treatment, and differences are discussed.

Chapter 7 addresses the issue of fusing existing medical knowledge about relations between sleep issues and chronic diseases with computational analysis. It discusses chronic disease detection at early and advanced stages using wake and sleep behavioural data. As a result, it contributes to knowledge with a novel deep-learning approach by incorporating the vital component of time and medical knowledge. It presents the architecture of the long short-term memory network (LSTM) for chronic disease detection, evaluates the approach on a diverse dataset, and discusses the results. In addition, it analyses the features relevant for misclassification.

Chapter 8 reflects on the presented work by providing a summary of the research process and activities, a description of the main research contributions, and a revisit to the success criteria. The chapter concludes with opportunities for future research and concluding remarks.

Chapter 2

Literature Review

2.1 Introduction

In the previous chapter, we gave a brief introduction to sleep behaviour analysis and placed the aim of this thesis in context. In this chapter, we establish a profound understanding of the state of the art of sleep behaviour analysis while concentrating on the computational aspect.

IoT devices are steadily advancing, creating new opportunities and attracting increased attention in home-based sleep assessment. Recent developments, in sensing and data analysis have led to new approaches for sleep monitoring and assessment, leading to an emerging role for personalised smart healthcare options. We examine the current state of technology-based sleep research. We first characterise sleep behaviour and key areas of sleep assessment and introduce a general review of the methodologies used in this domain. We then review the major technological methods and trends associated with sleep monitoring, data collection, and sleep behaviour analysis, from which strengths and weaknesses are extracted. Finally, we highlight and discuss existing issues.

This chapter is structured as follows. Section 2.2 characterises common sleep behaviours in terms of movement, stable states, abnormalities, and how they relate to sleep disorders. Section 2.3 discusses sleep monitoring and data collection and introduces key measurements and devices. The main piece of this chapter is Section 2.4, which reviews computational methods for sleep behaviour analysis in the fields of sleep stage classification, sleep position recognition, and disease diagnostics. Figure 2.1 presents a structural overview.

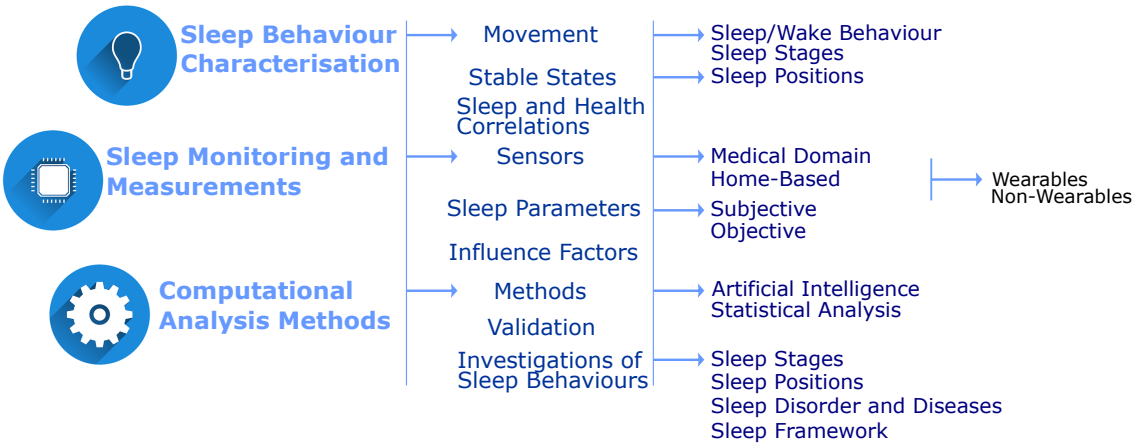


Figure 2.1: Survey structure on sleep behaviour analysis.

2.2 Sleep Behaviour Characterisation

Sleep behaviour can basically be divided into movement and stable states. Movement states contain 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 behaviours can be characterised, possibly leading to the diagnosis of sleep disorders and chronic diseases.

2.2.1 Movement Behaviour

Sleep is a relaxed state that still contains self-induced movements, mainly to prevent pressure ulcers [44]. These self-induced movements are defined as movement states that form behaviour over time. Movement is the primary information source for most sensors. An exception is PSG, which is based on a combination of motion and non-motion information.

Movement behaviour can be used to distinguish sleep from wake episodes [13,46–48]. Based on this information, objective sleep features can be extracted (e.g., sleep continuity [49], SE [48], and time [50]). Combinations of motion and non-motion information constitute sleep motion behaviour, which can be further correlated with sleep stages. Sleep stages describe different levels of sleep, provide hints about patient health, and are one of the most significant parameters 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 with and defined by brain-wave data but are considered difficult to classify and, therefore, can only be distinguished by trained technicians [19]. Abnormal movement during sleep can often be indicative of certain diseases. This includes movement from the eyes, chin, limbs, chest wall, and upper abdomen [51]. Many sleep-related movement disorders, including periodic limb movement disorder, RLS, and sleep-related bruxism [51], can be diagnosed based on irregular sleep movement. These irregular movements can also help to diagnose other disorders, such as rapid eye movement (REM) disorders and sleep apnoea. Sleep apnoea is indicated through respiration effort over abnormal abdomen movement during apnoeas [51]. Abnormal wake-sleep behaviour during the day is used to diagnose circadian rhythm sleep-wake disorders [51]. Additionally, this knowledge can potentially help investigate the severity of insomnia by observing wake and sleep periods during the night. Movement behaviour offers a significant amount of information, but more still can be found by also observing periods with no motion.

2.2.2 Stable State Behaviour

Stable states are the periods in rest with no movement (i.e., the periods while a person does not move); they are mainly related to sleep positions. Sleep positions are entirely independent of sleep stages [52], meaning they provide additional insights into sleep behaviour.

During periods without movement, four basic sleep positions can be measured: supine, prone, right lateral, and left lateral. Sleep position tracking is predominantly motivated by the prevention of pressure ulcers [39, 53] or the influence of position on sleep apnoea [54]. Additionally, certain sleep positions can result in back pain when the hip is rotated [55]. For sleep apnoea, sleeping on the back (i.e., in the supine position), is associated with a higher apnoea-hypopnoea index (AHI) compared to laying on the side [54]. Moreover, numerous sleep parameters, including sleep quality, are influenced by sleep positions [56].

Research mainly concentrates on the four basic sleep positions, but as the influence on different conditions is varied, positions with a higher granularity, including specific arm and leg positioning, are also of interest. Sensors in this field can be (1) wearables [56, 57], (2) imaging devices [58], and (3) sensors applied in or on the bed [39, 53, 59–61]. These approaches will be discussed in more detail later.

2.2.3 Sleep and Health Correlations

Sleep behaviour is manifested in established sleep parameters. These parameters have proven useful in investigating abnormal sleep behaviour. Consequently, abnormal sleep behaviour can determine sleep disorder classifications and can also indicate risk of developing certain chronic diseases.

(a) Sleep Disorders

Certain sleep patterns are used to define several sleep disorders, such as sleep apnoea [62] and insomnia [18]. Currently, the diagnosis of sleep apnoea relies on sensors in a sleep clinic while the diagnosis of insomnia is often based on subjective sleep questionnaires [6]. Sleep apnoea is diagnosed using AHI, which represents the number of apnoea and hypopnoea events per hour [63] (see Table 2.3). An AHI of less than 5 is classified as normal; whereas an AHI between 5 and 15 is classified as mild obstructive sleep apnoea; an AHI between 15 and 30 is classified as moderate sleep apnoea; and an AHI higher than 30 is classified as severe sleep apnoea [64]. For insomnia assessment, experts use the Insomnia Severity Index and Bergen Insomnia Scale [6]. Sleep disorders are generally related to poor sleep quality [64]. Despite being quite common and constituting a significant burden on health care costs and the overall economy, they are underdiagnosed [19, 64]. A chronic lack of sleep can lead to impulsive behaviour, depression, and chronic illnesses. However, when people are able to rest well, their overall mental and physical well-being improves and they are more able to handle pressure and stress [20, 65]. Therefore, monitoring social and sleep behaviours can help with early diagnosis of a wide array of issues, including mental illnesses like major depressive disorder [66].

Insomnia Insomnia can be a cause and a consequence of sleep disorders and is known to seriously affect people's lives. Regarding neurological diseases, such as Parkinson's and Alzheimer's, insomnia can be caused by (1) the disorder itself, (2) secondary symptoms (e.g., pain), and (3) medicinal side effects. There are also associations between insomnia and physical illnesses, such as arthritis. Overall, insomnia is associated with many disorders [67]. Despite its prevalence, however, it is often underdiagnosed and undertreated. A common issue is that of early diagnosis, because of the social stigma attached to sleep. People tend to 'normalise' insomnia; they relate it to ageing or hormonal changes. On the other hand, the perception of people who do report insomnia cannot always be trusted, as they tend to overesti-

mate SOL and underestimate sleep time [68]. The perception of insomnia often differs between patients and healthcare professionals as well. Furthermore, healthcare providers are often unaware of alternatives to non-pharmacological interventions except of sleep hygiene education [69]. Patients concentrate on daytime impairments and self-medication, whereas clinicians concentrate on quantitative descriptors of insomnia instead of the qualitative experiences of patients. This difference leads to the feeling of being misunderstood. As a result, there is a strong need to improve the ability to personalise management of insomnia [69].

(b) Chronic Diseases

Sleep behaviour is related to chronic diseases [26]. Chronic disease is a leading cause of death and disability and, therefore, a significant factor for quality of life of patients and relatives. As populations age, the rate of chronic disease will continue to rise and cause difficulties for society due to the high cost of treatment [70]. Currently, chronic diseases are significantly underdiagnosed, described as the percentage of patients without diagnosis among all patients estimated to have the disease: hypertension up to 40%; diabetes up to 50% [71]; and for CKD up to 74% are undiagnosed [72].

The diagnoses of chronic diseases are currently based on (1) invasive methods, such as blood sugar screening [73]; (2) clinical history, including symptoms and risk factors; or (3) sensor-based data, such as with blood pressure measurements [74]. These methods can also be combined to attain fast and trustworthy results, e.g., by combining clinical history, physical exam, and results from a sleep study to diagnose obstructive sleep apnoea syndrome (OSAS) [64]. For early detection, two main methods are currently used: (1) marker-based clinical analysis and (2) sensor-based behavioural analysis. In method (1), early detection of chronic diseases is based on bio-markers, neuropsychological-markers, or structural-image-markers. Research using this method has been done on many diseases, including Alzheimer's disease [75], diabetes [73], and CKD [76]. Method (2) uses data mining on sensory data, such as actigraphy, to investigate Alzheimer's [77], Parkinson's disease [78], diabetes, and hypertension [18].

It has been clinically observed that people suffering from chronic diseases such as diabetes [21,26], CKD [32], hypertension [79], arthritis, and strokes [26] generally have trouble with their sleep, such as difficulty falling or staying asleep and daytime sleepiness [26]. Patients report that these sleep issues negatively impact their daily lives (e.g., daytime sleepiness) [80]. Research on 2,416 elderly Chinese people indicated

that chronic disease is a primary predictor of poor sleep quality using multiple questionnaires, such as the Pittsburgh sleep quality index (PSQI) [81]. Numerous studies have associated sleep problems with the presence of chronic diseases, including heart disease, hypertension, stroke, and nocturia [81–83]. Hayashino et al. [84] found that the global PSQI score increased as the rate of comorbidity increased. The European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam study investigated 23,620 middle-aged participants and found that people with lower sleep duration (<6h) have a significantly increased risk of stroke, cancer, and chronic diseases generally [85]. Short sleep duration positively correlates with chronic diseases in the Chinese population [86]. In Foley et al. [26], older adults self-reported questionnaires. The results showed that arthritis, obesity, lung disease, diabetes, strokes, and osteoporosis were associated with sleep-related problems, such as breathing pauses, snoring, daytime sleepiness, restless legs, and insufficient sleep [26].

In general, research methods in this field are primarily questionnaires such as PSQI. Semi-structured interviews are seldom applied but have been used to investigate patients' perspectives on sleep [87].

Objective assessment of the relationship between sleep and chronic disease was performed by Herdegen [88]. Medical conditions were associated with sleep architecture changes. Increased awakenings were associated with cancer, while reduced sleep duration, SE, slow-wave, and REM sleep were associated with Parkinson's disease. Several primary sleep disorders affected chronic diseases; OSAS were related with hypertension, while OSAS and RLS were related with diabetes.

Poor health is generally considered to be a risk factor for poor sleep quality. Nevertheless, it is challenging to firmly determine the causal direction. A person's health status is an influence factor for sleep problems that may, in turn, negatively effect an individual's health—the relationship may be bidirectional and complex [81].

Strokes Strokes and sleep are strongly related. There is evidence that sleep issues can increase the stroke risk [89] and that post-stroke sleep disorders impact the rehabilitation and quality of life—if these sleep issues are left untreated, another stroke is more likely to occur [89]. Strokes can influence sleep architecture by reducing the amount of NREM sleep and overall sleep time. Insomnia affects 20–56% of stroke patients; 18% reported that their insomnia began after the stroke. Strokes are also associated with circadian rhythm changes, which results in sleep fragmentation and decreased SE [89].

Cancer Breast cancer is generally associated with significant sleep problems, according to a PSQI assessment [90]. Decreased sleep time and sleep disturbances were associated with pain, nocturia, feeling hot, coughing, and snoring. The treatment of sleep problems in those with cancer is inadequate. Standard treatment involves sleep medication but no cognitive behavioural therapy for sleep, which is known to be an effective approach. Moreover, the study showed that pain-related low quality of life is correlated with poor sleep [90].

Arthritis One study with a sample of 23,134 patients investigated the relationship between arthritis and sleep disturbances. It showed that arthritis and insomnia are connected via joint pain and that both are associated with age. This relationship was especially prevalent in those with depression and anxiety [91].

(c) Discussion on Sleep and Health Correlations

In the field of sleep research, we encountered a lack of (1) semi-structured interviews evaluating influence factors, symptoms, and assessment practices, (2) investigations from the perspective of doctors and patients living with medical conditions, and (3) discussions of varying perspectives. Chapter 6 presents our research, which aims to rectify this issue.

Previous research based on objective and subjective measurements suggests that there are strong relationships between medical conditions and sleep issues, meaning that the use of this knowledge would help to detect chronic disease. In Chapter 7, we investigate this.

2.3 Sleep Monitoring and Measurements

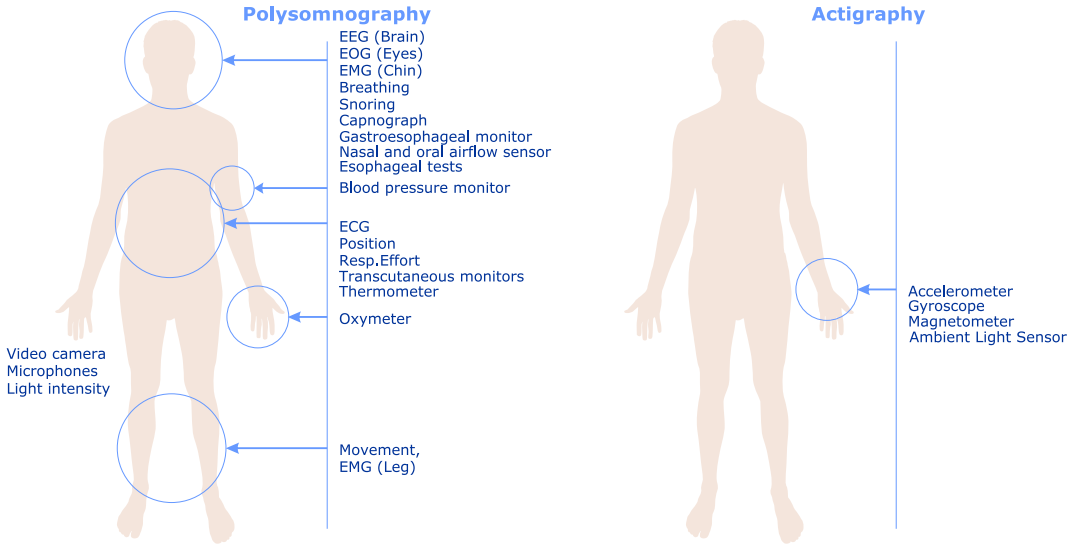
In this section, we discuss home-based technologies, medical devices for sleep assessment, sleep parameters, and influence factors.

2.3.1 Sleep Monitoring in the Medical Domain

The gold standards for assessing sleep disorders and issues are currently PSG and actigraphy.

(a) Polysomnography

PSG is a method that collects sensory data from devices applied to the human body and in the surrounding environment. The wearable devices that can be used in a PSG setting are EEG, electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), pulse oximeters, respiratory monitors, capnogram, transcutaneous monitors, thermometers, oesophageal monitors, nasal and oral airflow sensors, gastroesophageal monitors, and blood pressure monitors [6, 92]. The application areas of these sensors are shown in Figure 2.2(a). The non-wearable sensors applied in the environment are microphones, video cameras, and light intensity sensors [6].



(a) (b)

Figure 2.2: Application area of sensors in the medical field for **(a)** PSG and **(b)** actigraphy. Figures adapted from [1].

PSG monitors brain signals, heart signals, and movement. This method is predominately used for assessing sleep disorders such as sleep apnoea and RLS. Parameters such as sleep stages, SE, SOL, and arousal index can be extracted from the data [45]. To extract knowledge from the data, complex scoring methods are used. The clinical gold standards for sleep scoring with PSG are the Rechtschaffen and Kales method (R-K) method [4, 43] and an alternative method from the American Academy of Sleep Medicine (AASM) [3]. The scoring is generally based on 30-second epochs [45]. The R-K assesses six sleep stages [4]: wake, REM, S1, S2, S3, and S4; AASM [3] assesses five stages: wake, REM, N1, N2, and N3. Results from these two methods are generally compared by simply combining the S3 and S4 (R-K) stages into what AASM considers to be N3. Table 2.1 presents a description of the sleep

stages and the differences between the two methods. This process requires trained technicians who visually inspect data which is time-consuming.

Table 2.1: Sleep stages for the AASM [3] and R-K methods [4].

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 [93]
S3	N3	High Amplitude
S4	N3	Moderate Amount / Large Amount: Slow-Wave-Sleep, Deep-Sleep, Body Relaxes, Rebuilds, and Repairs [93, 94]

High numbers of body-worn sensors often result in misleading sleep behaviour, as it does not represent natural habits. This leads to the advice that at least two consecutive nights of data collection should be performed [37]. PSG is also expensive due to the laboratory setting and the fact that an observer is required to ensure the proper functioning of the applied devices throughout the night [44]. Research on PSG is currently looking to reduce the number of sensors while automatically recognising the main sleep behaviour parameters. This leads to studies with single-channel EEG in sleep stage classification [95–97], which can more easily be applied at home.

(b) Actigraphy

An actigraph unit is a wearable device that is meant to attach to the non-dominant wrist and provide information on sleep-wake cycles [44]. Actigraph units can measure activity over 24 hours, and, therefore, are also used in sleep assessments. The device is widely used for objective sleep quality measurements. The device can be used at home, which means it interferes less with natural sleep behaviour.

Actigraph units are mainly based on accelerometer data but can also include gyroscopes and magnetometers. Ambient light is generally collected to help with wake-sleep recognition. Furthermore, personal inputs can be given that tell the device when the sleeping period starts. The standard assessment involves seven consecutive days to obtain a representative picture of the patient's sleep habits. This can also be

considered the recommended amount of days [98]. This method is an unsupervised wearable approach for sleep analysis. The application area of the sensor is shown in Figure 2.2(b). Current research largely concentrates on extracting knowledge from already-processed data coming from medically-approved devices such as ActiGraphs or Actiwatches. These systems provide activity levels with wake and sleep labels. Activity levels provide the intensity of movement within 30-second intervals.

2.3.2 Home-Based Sleep Monitoring

Sleep monitoring at home is accomplished through wearable or non-wearable devices.

(a) 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 pulse oximeters (performing photoplethysmography (PPG)). Knowledge is generally obtained by using data mining for sleep position detection [57, 99], sleep stage classification [43, 47, 48, 100], heart rate analysis [20], respiration rate [99] analysis, and body temperature monitoring [20, 101].

Accelerometer data are extensively used in the analysis of people's daily activities and behaviour. Accelerometers can measure gravity, thus making it possible to represent the orientation of an object and, therefore, investigate sleep. Sleep position detection is usually investigated using either accelerometers [99] or other wearable wireless devices [57] on the chest or ankles. In general, sleep stage classification uses either one accelerometer on the wrist [47, 48, 100] or a combination of sensors, e.g., two accelerometers (one on the wrist, one on the ankle) and a chest strap [43]; alternatively, such classification can use PPG [46, 102]. Other sensors include thermometers that measure body temperature to extract sleep-wake data [101]. Researchers also sometimes employ wearables that are available on the market, mainly smartwatches, for sleep behaviour analysis under the assumption that automated self-management tools reduce burdens and increase efficiency [103]. Various devices are available and have already been analysed in terms of performance [38,45,50,104–107] and user perception [45, 108]. The devices considered in these 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 [38, 45, 50, 104, 108–112], ResMed S+ [107], EarlySense [106], Smart Eye Masks [14], and Microsoft Band I [92]. A summary of the Bland-Altman mean difference and error percentage is shown in Table 2.2—insignificant values are excluded ($\alpha = 0.05$). This output shows that most of the considered sleep devices overestimate total sleep time (TST). For healthy participants, Actigraph GT3X+ (4 min) provided the most relevant output compared to the Z-machine [50] while the FitBit Charge 2 (-9 min) provided the most relevant output compared to PSG [38]. For SE, Actiwatch (4.8%) performed with the highest accuracy [112] for healthy individuals while the FitBit Alta HR (2%) performed best for individuals with specific diseases. FitBit Charge 2 (24.5 min) represents WASO the most accurate [111] for healthy individuals while Actiwatch (-21.6 min) [112] is the most accurate for participants with medical conditions. FitBit Charge 2 (4 min) represents SOL the best [38] for healthy individuals while FitBit Flex (2.4 min) [112] does so for individuals with medical conditions. The differences between these investigations are also dependent on the participants, meaning that for healthy individuals and those with medical conditions, agreement changes [113]. Both Misfit Shine and Basis Health can distinguish between light and deep sleep, though a comparison reveals that Basis Health performs more accurately [105]. Additionally, accelerometer data extraction is overestimating or underestimating certain sleep parameters relative to PSG [49]. Ravichandran et al. [45] 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 the sleep quality information given to end-users is inaccurate. Design recommendations based on level of automation, understandable visualisation, and emotional influence factors were investigated in Vandenberghe and Geerts [108], which considered Polar Loop, Jawbone Up3, Misfit Shine, and FitBit Flex. Vandenberghe and Geerts recommend that (1) people label their data, which is limited by forgetting as people fail to remember to provide the necessary labels; therefore, it is essential to automatically detect sleep parameters accurately as we propose in Chapter 5 for SOL; (2) visualisations should be understandable and daily comparison available, we target this limitation by user-dependent visualisation and presentation of regularity for the important sleep aspects in Chapter 3; and (3) emotions play a role, therefore, sleep duration should not be referred to as being perfect. This relates to misinterpretation of sleep parameters, which we try to overcome by: providing an overall picture, developing algorithms which are more accurate, and visualising in a layered structure (compare Chapter 3).

Table 2.2: Wearables validated against PSG if not otherwise indicated with Z-machine (Z). Bland-Altman 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		[38]
	-12.3 m		-11.1 m	24.5 m	[111, 113]
Actiwatch	17.8 m	4.8 %B			[112, 113]
	5.8 %*	10.4 %*			[105]
FB* Flex_{S*}	3.0 %*				[105]
Basis Health	7.8 %*				[105]
Pulse O2	6.0 %*				[105]
Misfit Shine	15.3 %*				[45, 105, 108]
GT3X+	4.0 m/Z				[50]
Diseased					
Actiwatch	40.6 m	7.0 %B	-13.5 m	-27.1 m	[104]
	43.9 m	7.5 %B	-12.9 m	-33.9 m	[110, 113]
				-21.6 m	[112, 113]
FB* Flex_{N*}	32.9 m	7.9 %B	-2.4 m	30.5 m	[112, 113]
	46.0 m	8.1 %B		-44.0 m	[104, 108]
FB* Flex_{S*}	-86.3 m	-16.0 %B	11.5 m	74.8 m	[104, 108]
FB* Alta HR	11.6 m	2.0 %B			[109]
Jawbone Up3	39.6 m	6.8 %B	-5.1 m	-34.3 m	[110, 113]

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

(b) Non-Wearable Technology

Non-wearable devices are not attached to the human body, meaning they are the least interfering sleep assessment method because they do not disturb a person's regular sleeping habits. In general, techniques are based on either single devices, such as Kinect sensors, or multiple devices, like those integrated into smart beds. Movement investigations use data collected from sleep trackers. There are already applications on the market that provide insight into user sleep habits to promote sleep self-management with smart devices.

Non-wearable sensors are widely used by researchers. These include load cells, force sensors, air cushions, pressure pads, water-filled vinyl tubes [44], smartphones [66, 114], Shimmer sensors [13], Doppler radar signals [115, 116] also with sound signals [116,117], pressure sensors [39,53,53,59–61,118], and radio signals [119]. Air cushions under the bed, as well as water-filled tubes under a pillow, can be used to collect data

such as respiration rate, heart rate, and body movement to estimate sleep stages. Additionally, pressure pads can be used to evaluate those same three measures as well as snoring and sleep apnoea events [44]. Load cells are often employed to detect body movement and sleep-wake patterns in addition to estimating deep sleep stages. Pressure sensors integrated into mats, beds, or bedsheets typically perform well detecting body locations and positions in the bed [39, 53, 59–61, 118] and can even extract sleep stage information [57]. Imaging devices, such as Kinect sensors [58] and depth cameras [65,120], are also used to detect sleep positions. Shimmer sensors applied on the bed [13], Doppler radar signals [115], and sound signals [116,117] have all been used in sleep-wake detection and sleep stage recognition.

Sleep trackers that are applied within the environment include smartphones, smart mats, and smart beds. Smartphone applications are the most easily accessible for users and, therefore, an inexpensive method for sleep tracking. Such sleep-tracking applications monitor behaviour and environmental influence factors during sleep, including audio, ambient light, and movement [114]; lower mobility and phone usage leads to better sleep [66]. Several applications attempt to provide insights into daily sleep cycles, SE, and duration [121, 122]—some even allow for users to self-report moods and daily habits alongside measuring their sleep behaviour [121].

Bed-mounted sensors are not only good sources of information, but they are easily applied into a user's bed and more accurate than smartphones. There are several devices on the market that use smart mat technology to extract data on body movement, heart rate, and respiration rate [44]. Various sensors can be integrated into smart mats, such as force, piezoelectric, and pressure sensors [44]. Some devices combine this technology with environmental sensors (e.g., Beddit), with which users can see a wide array of statistics, including number of bed exits, sleep quality, body movement, noise levels, and light levels [44]; this information could be useful for various reasons (e.g., pain management and fall prevention) [44]. The locations of non-wearable devices in a sleep environment are depicted in Figure 2.3.

2.3.3 Discussion and Suggestions on Sleep Monitoring

Every approach has advantages and disadvantages that must be carefully considered; the ultimate decision on which sensors to use should vary based on the area of exploration.

Actigraphy and PSG are used for different areas of investigation. PSG offers the highest level of accuracy but is also expensive, restricted to a supervised laboratory

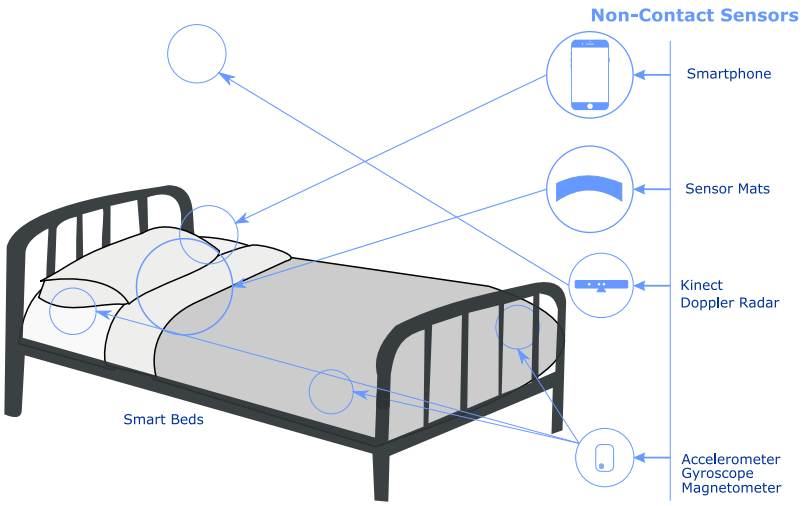


Figure 2.3: Placement of non-wearable sensors in home environments. The figure of the bed is adapted from [2].

setting, and interrupts natural sleep behaviour. PSG requires two days of data collection using more than three sensors while home-based actigraphy requires at least seven days of data collection but only one wearable sensor. Actigraphy is far lower in cost, interferes less with natural sleeping habits, and provides intermediate accuracy. Non-wearable devices applied in the home environment involve the least interference but are less accurate and typically immobile compared to wearables.

Regarding sleep positions recognition, non-wearable devices are generally unable to distinguish between supine and prone positions. While imaging can effectively recognise sleep positions, it leads to privacy concerns and cannot accurately recognise movements through blankets [58]. In contrast, wearable devices offer a mobile solution with few privacy issues that is able to distinguish between multiple people in one bed. However, wearable sensors can lead to discomfort and potentially disrupt natural sleep behaviours. Sleep stage recognition depends on more advanced technology measuring EEG data. Currently, researchers are working to further develop wearable technologies, such as actigraph units, to make home-based self-management more accessible. Details can be found Section 2.4. Which approach to take and devices to use depends on the sleep parameters of interest; for details, see Table 2.2. The easiest devices to start with are smartwatches, as they are reasonably accurate relative to smartphone applications and are more accessible than smart beds. While numerous ready-to-use smartphone applications are available with different collection processes and outputs, issues such as accuracy and validity remain. When choosing the ideal device for investigation, researchers must consider the target group, as this can influence accuracy of wearable devices [113].

2.3.4 Sleep Parameters

The main measurements describing sleep are respiration rate, heart rate, temperature, body movement [19], and brain waves [6]. Considering various hypotheses, different important parameters can be extracted from these measurements and vary in importance per individual. The two ways of measuring sleep are subjective and objective, which present different insights into the sleep of individuals, depends on the aim of research; for details, refer to Table 2.3.

(a) Sleep Parameters from Objective Measurements

Objective sleep parameters are drawn from sensor data. These parameters include sleep stages, disturbances, sleep regularity, SE, duration, SOL, arousals, and spindles, among others. Sleep stages, for example, are determined by heart or respiration rate using the knowledge of existing relations between them [123] while skin temperature is used for estimating disturbances [101]. Sleep continuity is based on SE, arousal index, and the percentage of TST in each sleep stage [49]. Sleep quality, regularity, sleepiness level, and chronotype are considered new insight indicators investigated with wearable devices as opposed to traditional parameters, such as time falling asleep, number of awakenings, and sleep duration [5]. Sleep regularity measures the affinity between sleeping periods from consecutive days [5, 7]. Wake-sleep episodes are the fundamental basis when objectively measuring sleep parameters at home with actigraphy. This means that, with sleep duration, for example, the extracted sleep periods are analysed in relation to wake episodes. Generally, static thresholds are used to define certain parameters. For example, for SOL, the 15-minutes rule is commonly applied [124], despite research having shown that age-based variable thresholds improve the agreement with PSG [125]. These extracted sleep parameters do not compare well with established gold standards. We will propose an approach with dynamic thresholds in Chapter 5.

(b) Sleep Parameters from Subjective Measurements

Subjective sleep parameters are typically assessed by sleep questionnaires that extract information from users by asking questions to assess sleep issues [6]. These sleep parameters include nightmares, bedtime, and rise time; see Table 2.3 for details [5]. The extracted parameters can be assessed with different techniques, such as the SATED assessment [45], Consensus Sleep Diary, PSQI, Mini Sleep Ques-

tionnaire, Epworth Sleepiness Scale, Insomnia Severity Test, and Sleep Disorders Questionnaire [51]. We refer to the work of Ibáñez et al. [51] for a thorough review of subjective methods. Subjective sleep investigation has the advantage that experts are not necessary, but the disadvantage of missing rigor in reported answers [5].

(c) Relevant Sleep Behaviour Assessment Parameters

In this section, we present currently explored parameters concerning sleep quality, regularity, and the circadian rhythm.

Sleep Quality Important roles are played by SE, percentage of TST per sleep stage, arousals [49], and time spent awake [126]; new indicators have also become prominent, however, such as sleep quality, sleepiness level, and chronotype [5]. Self-induced movements are also seen as relevant for good sleep [44] and are investigated through body position measurements, which relate to sleep quality like waking episode frequency [56] as well as sleep disorders, such as sleep apnoea [19] and pressure ulcers [39]. Consequently, sleep positions should be monitored alongside sleep stage and sleep time [127]. Most approaches use generalised methods, but sleep behaviour differs among individuals based on age, biological factors, and lifestyle [128]. Furthermore, irregularity between weekend and weekdays is common for sleep onset time, bedtime, rise time, and sleep period [129].

Circadian Rhythm The circadian rhythm—sleep-wake cycle—responds to external impact factors, such as light exposure [130, 131]. Consequently, sleep-wake cycle alterations imply changes in quality of life and health status [131]; abnormalities help diagnose sleep-wake rhythm disorders [6]. Circadian rhythm analysis suggests that there is a need to integrate circadian measures into medical and care decision processes [132]. Sensors (e.g., actigraphy) monitor circadian rhythm visually or through computational methods, such as the hidden Markov model (HMM) [132]. Several non-parametric measures represent the circadian rhythm: intradaily variability, interdaily stability, and relative amplitude [132]. Interdaily stability represents synchronisation to the 24-hour cycle, intradaily variability represents rhythm fragmentation, and relative amplitude represents activeness during the day [132]. They can be calculated as follows,

$$\text{Intradaily variability} = \frac{n\sum_{j=1}^{m}(\overline{x}_j - \overline{x})^2}{m\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\text{Interdaily stability} = \frac{n\sum_{i=2}^{n}(x_i - x_{i-1})^2}{(n-1)\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

$$\text{Relative amplitude} = \frac{M10 - L5}{M10 + L5}$$

where n is the number of samples; m is the number of samples per day, $\overline{x}_j$ are the hourly means; $\overline{x}$ is the overall average; x_i describes the activity values for each sampling point; M10 is the average activity over the ten consecutive hours with the highest activity level; and L5 is the average activity over the five consecutive hours with the lowest activity level [133].

Sleep Regularity Long-time visualisation (i.e., the regularity of a person's sleep) interests sleep experts [45]. Regular bed and rise times should be preserved throughout the whole week [45,134]. In reality, however, sleep parameters are often irregular between weekend and weekdays such as those for sleep onset time, bedtime, sleep offset, and sleep period [129]. Currently, only day-to-day sleep timing is assessed with a sleep regularity index [130]. Nevertheless, regularity should be assessed for different parameters and included in new frameworks for several reasons, one being, e.g., that regularly high sleep fragmentation suggests underlying issues, such as insomnia. Therefore, a method that can represent regularity for individual sleep parameters over longer periods of time is essential to provide reliable feedback and recommendations to users—this is investigated in Chapter 3. Sleep regularity is independent of sleep duration and relates to performance during the day and the timing of sleep [130].

(d) Discussion on Sleep Parameters

Objective methods monitor and measure individual sleep behaviour in a specific setting, often a hospital or laboratory. In contrast, subjective sleep analysis typically does not need the presence of experts or a specific location. However, subjective methods are less accurate than objective ones [5]. These two methods can be difficult to compare, as definitions are not consistent in objective and subjective feature calculations. In Table 2.3, features and their formulas are presented. Some of the formulas are for PSG but can still be translated to actigraphy; others are developed

specifically for actigraphy data.

Table 2.3: Subjective and objective sleep parameters adapted from [5–12].

Sleep Parameters		
Subjective		
Pains in the Night	Time to Fall Asleep	Rise Time
Real Sleep Duration	Feel Cold	Awakenings
Enthusiasm to Get Things Done	Bad Breath	Snoring
User-Perceived Sleep Quality	Feeling Hot	Nightmares
Somnolence During Activities	Drug Ingestion	Bedtime
Awakenings for the Toilet	# of Co-Sleepers	
Objective		
Time in Each Stage (TiES)	Spindles	Slow Waves
Turning Light-Off Time	K-Complex	Arousals
TST	Turning Light-On Time	
Parameters	**Description**	
Sleep Stages	Wake, REM, S1-S4/N1-N3	
Total Recording Time (TRT)	Lights out to Lights on.	
SE	$100\frac{\text{TST}}{\text{TRT}} = 100\frac{\text{len(Sleep Period)} - \text{WASO}}{\text{len(Sleep Period)} + \text{SL}}$	
SOL	Lights out to first sleep stage.	
REM Latency	Sleep onset to first epoch of REM.	
WASO	WASO = TRT - SOL - TST	
Arousal (AR)	Wake period <10s	
Awakening (AW)	Wake period >10s	
Arousal Index	#AR × 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)	≥4 consecutive LM events	
PLMS Index	PLMS × 60/TST	
Apnoea (A)	Cessation of breathing	
Hypopnoea (H)	Shallow/low-frequency breathing	
AHI	(#A + #H) × 60 / TST	
Respiratory-Effort Related Arousals (RERAs)	Arousals from sleep do not meet the definitions of apnoeas or hypopnoeas but do disrupt sleep.	
Respiratory Disturbance Index (RDI)	RERAs + As +Hs × 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 recording days

2.3.5 Behavioural, Environmental, Genetic, and Health-Related Influence Factors on Sleep

Over the last decade, various studies have tried to find relevant behavioural influence factors relating to sleep; some significant insights were uncovered for specific correlations. In this section, we present currently explored influence factors concerning environmental aspects, sleep hygiene, and clinical history. Furthermore, influences coming from personalisation and presentation are discussed.

Clinical History Influence Factors Clinical history factors have been shown to influence sleep. Such factors include body mass index (BMI), gender, age, sleep disorders, and chronic diseases. Subjective measurement techniques found several risk factors for poor sleep quality: female gender, high number of comorbid conditions, obesity and overweightness, short sleep duration, high alcohol consumption, depression, fatigue, and anxiety [135, 136]. Access to healthcare, good family economic status, and rural residence were found to be associated with high sleep quality [137]. Sleep duration correlates with age, BMI, depression, and fatigue [136] whereas age influences SOL, time spent in bed [138], overnight metabolic rate (OMR) (and, therefore, sleep quality [126]), and the environmental effect on sleep quality [139]. Generally, females have lower-quality sleep, higher SOL, wake up more often, and take sleeping drugs more often but take less frequent naps, whereas males often suffer from daytime sleepiness [138]. Based on changes in circumstances, parameters should be adjusted, and relevant measurements should be monitored to realise a personalised sleep behaviour assessment framework.

Behavioural Influence Factors Objective measurement techniques are used to investigate sleep quality and provide evidence that sleep positions affect quality of sleep [56]. Their results indicate that the increase in time spent in the supine position for younger individuals results in enhancement of sleep. Furthermore, positive correlations were found for younger individuals between number of position shifts and waking episodes and activity index; in the senior group, positive correlations were found between number of position shifts and total minutes scored awake, WASO, activity index, sleep fragmentation index, and waking episodes, as well as negative correlation for SE [56]. Comparing objective actigraphy with subjective questionnaires in children showed that sleep onset time, bedtime, sleep period, and sleep offset time were correlated for these methods during the week but not over

the weekend [129]. Moreover, Valenti et al. [126] showed that time spent awake and age plays a role in OMR and that OMR is directly related to sleep quality. An essential aspect of night-time sleep is self-induced movements; if patients lack these, pressure ulcers can appear [44]. If body movements increase, SE decreases; i.e., ageing reduces SE and boosts the time awake. This implies that awake time is tied to increased body movement and, therefore, to increased OMR [126]. From this, we can draw the conclusion that more time in slow-wave sleep (having the lowest OMR) results in reduced OMR which, in turn, leads to better sleep quality. Further influence factors come from the sleep environment, such as oxygen levels, light, sound, and humidity [20, 44].

Sleep Hygiene Sleep hygiene is comprised of behavioural and environmental factors that can potentially improve sleep quality: regular sleep time; comfortable sleep environment; regular exercise; and caffeine, alcohol, and nap avoidance [134]. Caffeine avoidance and regular bedtime affect sleep positively [134]; naps increase WASO and sleep fragmentation and decrease SE [134].

Environmental Influence Factors The home environment plays a role in sleep and should control temperature, humidity, CO_2 levels, noise, light, or aroma [20,139]; if not ideal deep sleep can be limited [20]. Generally, bedroom air quality is divided into two groups: thermal comfort, including temperature, airspeed, and relative humidity, and air contaminants from physical, chemical, and biological agents [140]. Improvements to the sleep environment can lead to improvements in sleep quality. CO_2 concentration is currently used to indicate indoor air quality, but other contaminants also require consideration and monitoring [140].

Personalisation Research suggests that different sleep parameters hold distinct values for individuals. These distinctions are important for a personalised system [45]. Most notably, the distinctions between healthy individuals and those with a medical condition must be addressed [141] (this is addressed for sleep stages in Chapter 5). The trend in medicine is towards personalised approaches [142], but most approaches still use generalised methods and ignore personal factors such as age, biological factors, and lifestyle [128]. One-size-fits-all methods are not ideal for all individuals—certain diseases are sometimes more common in certain demographic groups [142, 143]. Attempts to personalise treatment based on this kind of subcategorisation, known as clinical phenotyping, have already proven successful for

sleep apnoea patients [144], though not in the assessment itself. Chen et al. has investigated approaches for personalised sleep stage detection for PSG data with individual thresholds [145]. We propose an approach for home-based assessment in Chapter 5.

Presentation One-solution-fits-all approaches are not flexible towards the needs and preferences of users and doctors. Yang et al. propose an intelligent visualisation for sleep apnoea datasets to help analyse the results. The need to display comorbidities and symptoms in a way that is both accessible for patients and comprehensive for doctors was addressed [146]. Current designs typically present daily assessments but lack the ability to show trends over time [45]. Misinterpretation of sleep assessment factors is common [45]. This can and must be addressed by simplifying the presentation for users. Sleep changes as ageing progresses, so appropriate thresholds for age groups [147] must be incorporated. Overall, the preservation of privacy is essential; users must be able to decide what information can be collected and be shared with healthcare providers. We propose a visualisation concept which tackles these issues in Chapter 3.

2.4 Computational Analysis Methods

The main methods for sleep behaviour analysis are data mining techniques, such as statistical analysis and artificial intelligence. Statistical analysis is a well-developed method, whereas the use of artificial intelligence has just recently become popular in the medical field. Despite this, it has quickly proven to be a well-performing method for analysing more difficult scientific problems, such as sleep behaviour and disease detection. Various techniques are available; effectiveness varies based on the investigated sleep problem. Prominent methods are random forests (RFs), decision trees (DTs), support vector machines (SVMs), k-nearest neighbours (KNNs), HMMs, Bayesian classifiers, neural networks (NNs), and deep learning. Methods such as recurrent neural networks (RNNs) (e.g., LSTMs), and convolutional neural networks (CNNs), are the most commonly employed forms of deep learning in sleep research.

We will discuss computational sleep behaviour analysis with a focus on machine learning approaches for the main problems: sleep stage, sleep position, and sleep disorder detection.

2.4.1 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 [148]. RFs are a collection of multiple DTs that can overcome the issue of DTs easily overfitting training datasets due to the law of large numbers [149]. SVMs represent data in a higher-dimensional feature space to separate classes by hyperplanes, which makes them slow to train on large datasets since they must solve a quadratic problem where the number of variables is equivalent to the quantity of training data [150,151]. The performance is sensitive to the choice of the kernel and parameters [150]. KNNs use the k-nearest neighbours within a metric space to decide to which class new data points belong. The model calculates the distance (commonly the Euclidean distance) to every neighbour for each prediction step and performs slowly when many predictions must be made [152]. HMMs are based on Markov chains, in which the current state depends on the previous states to represent transitions and observations [153]. Naive BCs, which assume that features are independent of one another following Bayes' theorem, constitute a very simple approach that requires little training data [16,154]. 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 due to the large number of parameters that are best determined empirically and are difficult to interpret for humans [154]. NNs with more than three layers are considered deep learning methods and are extensively used currently due to their reliable performance and the advantage that, even without pre-processing, discriminant features can be generated from raw data. However, it is not possible to see which features are important or how the outcome is produced; additionally, since a large quantity of training data is necessary, computational costs are quite high. Time-series data are often collected in sleep research; therefore, specific methods are designed to incorporate the time aspect. A prominent deep learning method in the field of time-series data classification is an RNN known as LSTM. The method was first introduced by Hochreiter and Schmidhuber [155] in 1997, and has been extended over time, primarily by additions of forget gates [156], and peep-hole connections [157]. Its use in diagnostics using multiple time-series data has already demonstrated success [158]. 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) [159], draws conclusions based on the data itself, whereas inferential statistics, e.g., logistic regression (LR), draws conclusions based on samples from the population. Classic LR is often used to analyse accelerometer data because it is easy to interpret, though it is subject to limitations. LR is generally unsuitable for learning complex patterns from noisy accelerometer data; therefore, higher-level features are often extracted; this 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 [10].

2.4.2 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 must be aware that the interscorer agreement of human-scored parameters does not have a perfect agreement but rather, one of 82.6% [160]. This can result in a bias towards a rater's style if only one person scores the sleep data. It also means that discussions must take this into account. Data from actigraph units are seldom used for validation [13]. The most common 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 behaviour analysis. Occasionally, the receiver operating characteristic curve (ROC), area under the curve (AUC), or F_1 score is given.

Accuracy is the percentage of predictions that a classifier makes correctly. Sensitivity describes the capability of the classifier to recognise true positives, and specificity indicates the capability that it does not generate a false negative [62]. Specificity is defined as the number of correctly classified negative samples (TNs) divided by the TNs plus the number of incorrectly classified positive samples (FPs), whereas precision is defined as the positive predictive value [97]; calculated using the number of correctly classified positive samples (TPs) divided by the TPs plus the FPs. The F_1 score based on precision and recall can be calculated with macro- and micro-averaging methods. The multiclass F_1 score is based on weighted individual class scores. The macro F_1 score is uniformly weighted, whereas the micro F_1 score is measured by calculating the overall number of false negatives (FNs), true positives (TPs), and FPs [97]. Accuracy and precision are not able to completely describe the situation in multiclass classification. Additionally, for imbalanced classes, accuracy can be misleading; therefore, precision and recall, represented by the F_1 score, are

of importance. Regardless, the drawback is that the F_1 score has no sufficient intuitive explanation [161]. Therefore, the κ statistic was introduced for imbalanced and multiclass classifications. The κ statistic compares the classifier performance to random guessing [161], statistically measuring the agreement between annotators for categorical items [162]. An ROC represents a classifier's performance at different classification thresholds in a graph, providing a global estimation of the classification ability [62]; it is based on the precision and FP rate (1-specificity). The larger the AUC is, the better the classification performs [62]. For multiclass problems, multiple graphs are required.

Note that validation is performed with (1) k-fold CV (k-CV); (2) leave-one-person-out cross-validation (LOOCV), which is favourable; 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 validation is ideal; users who are trained on should not be tested on—otherwise, the generalisability of the method cannot be guaranteed. This can be realised in methods (1) and (2) but needs to be addressed to ensure that the results are trustworthy.

2.4.3 Sleep Stage Classification

The main goals in sleep stage classification are to (1) automate the process that is normally performed by trained technicians and (2) make home-based assessment possible. Validation is usually performed against trained human classification, which includes variability and is thus not optimal [163, 164]. It is especially important to consider differences in performance between measuring healthy subjects and measuring subjects with medical conditions. In the following paragraphs, approaches using sensory monitoring at home and in the medical domain are discussed. The different technologies, their performances, and details can be found in Table 2.4 for home-based sensory data and in Table 2.5 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 outcomes.

Table 2.4: Home-based sleep stage classification in 30-second intervals following AASM [3]; all validated against PSG, except [13] and [14] which are validated against Actiwatch, and a wrist-worn device, respectively. Certain performances were recalculated from confusion matrices.

	Data	P*	Characteristics	Epochs (hpP*)	Analysis	Val.*(%•)	Prec.	Rec.	Acc.	κ
5 Stages - Wake, REM, N1, N2, and N3										
[92]	Microsoft Band I	39	H*/30M*/19−64Y*	37,000 (~ 8)	RNN	LOOCV	64.5	65	60.5(F*)	
[43]	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), and Deep Sleep (N3)										
[106]	EarlySense	63	SD*/45M*/17−72Y*	70,965 (~ 9)	Commercial		63.8	60.6	64.5	0.46
[165]	Accelerometer, PWS*	100	H*	108,000 (~ 9)	Thresholds		64.4	59.7	68.5	
[107]	ResMed S+	38	H*/21M*	~41,040 (~ 9)	Commercial	53/47/			70	0.53
[166]	CPAP flow	400	SD*	360,000 (7.5)	C/RNN-CRF	60/20/20	65	71.8	74.1	0.57
[119]	Radio Signals	25	H*/15M*	~90,000 (~30)	CNN-RNN	75/25/	79.6	75.8	79.8	0.70
[100]	Accelerometer	36		~34,560 (~ 8)	Kstar, Bagging, RC, RSS, RF	10-CV	85(S*)	70	87.5	
3 Stages - Wake, REM, and NREM (N1, N2, N3)										
[116]	DR*, Microphone	24	SD*/21M*/43.4±13Y*	17,280 (6)	RF	46/54/			64.4	
[117]	Microphone	250	SD*/162M*/19−84Y*	~180,000 (~ 6)	NN	60/40/	82.6	80.9	87.3	0.72
2 Stages - REM and NREM										
[14]	Smart Eye Mask	7	H*/4M*/20−24Y*	~7,920 (~ 6)	RF	11-CV			80	
[102]	PPG*, Accelerometer	15	H*	~16,200 (~ 9)	DT		81.4(S*)	86.3	81.4	
2 Stages - Sleep and Wake										
[46]	PPG*, HRV*	10	SD*/5M*/56±8.79Y*	8,452 (~ 7)	KNN, SVM	50/50/	79(S*)	79	79.4	0.59
[47]	Accelerometer	22	7M*/85.7±3.7Y*	~23,760 (~ 9)	CNN		68(S*)	80		
[13]	5 Shimmer sen.	3	H*/1M*/21−30Y*	24,062 (~67)	RF		86(S*)	93	90	
[167]	Video	10	avg. 22.6 Y*	~6,000 (~ 5)	Thresholds				92.13	
[48]	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-F_1 score; •(training/validation/test) split

(a) Home-Based Sleep Stage Analysis

Wearable and non-wearable devices are investigated in terms of their sleep stage classification abilities. Sleep stages monitored at home with sensor technology provide a less-biased data source, as the subjects are in their natural sleep environment. If the number of body-worn sensors is held to a minimum, sleep behaviour is influenced less than it is during PSG [37].

Five-Stage Classification An ideal approach would be able to distinguish between the six or five sleep stages, depending on the chosen guideline. Currently, research is mainly focused on the five AASM stages. In Reimer et al. [43], MSR accelerometer data from the wrist and ankle, and Zephyr BioHarness 3^2 data from the chest were collected from 26 individuals in a sleep lab to investigate sleep stages and compare results to those obtained using PSG. The data were analysed using RF and deep learning. Deep learning was used for unsupervised feature learning, followed by a deep belief network (DBN) built from stacked restricted Boltzmann machines (RBM). The DBN approach achieved an average 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 Zhang et al. [92] to collect heart rate and actigraphy 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 F_1 score of 60.5% in the comprehensive group, in which resting and non-resting sleep were included. The performance of the RNN approach is affected by the dataset size, which, at just 37,000 epochs, is likely too small.

Using sensors at home, which interferes less with sleep habits, is showing solid results for using accelerometer data and a chest strap with an RF method. The limitations in five-stage classification are that (1) only healthy participants are considered and (2) the validation datasets are relatively small, especially for deep learning. Overall, reliable outcomes across all classes can be seen when comparing κ.

Four-Stage Classification Because it is difficult to distinguish between N1 and N2, the two stages are occasionally fused together as light sleep and compared to N3 (deep sleep) [166]. This results in fours stages: wake, REM, light sleep, and deep sleep. In Aggarwal et al. [166], continuous positive air pressure (CPAP) flow signals from 400 subjects were analysed 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 with a weak κ of only 0.57. We can conclude that the dataset is highly imbalanced and cannot sufficiently detect minority classes based on only one split for training and testing. Similarly, radio waves can be analysed by combining CNN and RNN [119]. This approach reached a moderate κ of 0.70 with up to 79.8% accuracy for the 25 healthy subjects that participated, but investigated only one data split. Ultimately, the advantages of both main deep learning methods were incorporated. Specifically, the CNN was able to separate wake and REM stages, whereas RNN could separate deep and light sleep [119]. Accelerometer data were collected in Yeo et al. [100] from the non-dominant hand for 36 individuals. They analysed different classifiers and performed feature selection, concluding that Random Committee 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 discussed, but the study is most likely based on healthy participants. In contrast to machine learning approaches, equation threshold-based approaches are also investigated. In Fujimoto et al. [165], this is done by recording data for 100 participants from a wrist-worn device featuring a 3-axis accelerometer and a reflective photoelectric volume pulse sensor. The system reached an accuracy of approximately 68.5%, which is fairly low relative to other accelerometer approaches; however, it is validated on a larger database. The commercial ResMed S+ device bearing an ultra-low-power radio-frequency sensor was analysed in Zaffaroni et al. [107]. Respiration and body-movement signals are gathered. The validation was performed with three technicians using majority voting for an overall score. 38 adults were assessed with an accuracy of 70% compared to the 82% arrived at by the standard scoring system. In contrast, the Early Sense sensor, which uses piezoelectric sensors, reached only 64.5% accuracy with a weak κ of 0.54 [106] but included 63 subjects with medical conditions.

An RC approach for accelerometers appears to be promising for four-stage classification with an average accuracy of 87.5% [100]; while a non-wearable device and deep learning method such as those used in Zhao et al. [119] comes with certain advantages, but their accuracy is also significantly lower at just 80%.

Three-Stage Classification Researchers have occasionally further narrowed down the problem to wake, REM and NREM stages. In Chung et al. [116], a sound sensor and a Doppler radar sensor were combined to detect sleep-wake episodes prior to an NREM-REM classification. Different features for each problem were used, reaching an accuracy of 64.4% using RF for 24 patients with sleep disorders.

Each step included a personal-adjustment structure based on a threshold derived from the likelihood ratios from the RF classification. In Eliran et al. [117], audio signals from microphones were analysed for 250 participants with a one-layer NN; the process reached 87.3% accuracy. Moreover, Smart Eye Masks were applied to determine REM and NREM sleep in seven healthy subjects using photo reflectors and accelerometers [14]. The researchers used a RF and obtained 80% accuracy for 10-CV. In contrast, Renevey et al. [102] utilised optical wrist-worn devices by applying a trained DT and reached an overall accuracy of 81.35%. This study considered PPG and a 3-axis accelerometer data from 15 participants.

Although it is difficult to compare different data sources, it appears that with only three stages considered, audio signals combined with NN perform quite well even with subjects with a medical condition. However, this is based on the outcomes of only one training-testing case.

Two-Stage Classification Wake-sleep classification is often the first step towards fine-grained sleep stage classification. This classification can be performed with PPG [46], accelerometers [47], actigraph units [48], and Shimmer sensors on the bed [13]. In Uçar et al. [46], sleep-wake stages for ten patients with sleep apnoea were classified by KNN and SVM using PPG, from which heart rate variability (HRV) 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 8.000, which included participants with sleep disorders. Twenty-two elderly individuals participated in a study aimed at collecting accelerometer data. These data were analysed using a CNN approach, which was compared to a standard sleep-wake classification approach called ESS [168], which is based on the estimation of stationary sleep-segments; it increased the specificity from 54% to 68% and decreased the sensitivity from 82% to 80% [47]. In McDowell et al. [13], five Shimmer sensors were applied on the bed of three individuals and validated against a Philips Actiwatch. Undersampling and oversampling were used to prepare the data for RF methods. In this way, the overfitting issue was addressed and the researchers achieved results with a sensitivity of 93% and specificity of 86%. Alternatively, a rule-based approach using an actigraph unit was proposed by Kuo et al. [48]; it reached an accuracy of 92.16%, specificity of 71.3%, and sensitivity of 95.02%. They tested the system with 81 subjects split by poor and good SE. Four different rules were introduced from movement density and density thresholds. A descriptive analysis was performed by Liao and Yang [167]. For this purpose, camera recordings were analysed and validated against actigraphy and PSG. Frame difference and motion history were

used to classify body motion, which is an indicator of wake episodes. Data from ten subjects reached 92.13% accuracy for the video-based system compared to 91.24% for Actiwatch with 91.24% accuracy.

Threshold-based sleep-wake classification using actigraph units is a powerful method. This suggests that it is necessary to distinguish between subjects with poor and good SE [48].

(b) Sleep Stage Analysis in the Medical Domain

In this section, the focus lies in the automation of home-based sleep stage classification with single-channel EEG data. The pre-processing of highly sensitive data is important as artefacts (e.g., from movement) are present and a large amount of information is available.

Six-Stage Classification One attempt to automate sleep stage scoring was made in 1996 [163] when an NN model with an uncertainty index was presented that was able to classify six sleep stages using EEG, EMG, and EOG. There were 60 subjects participating—20 suffered from depression, 20 suffered from insomnia, and 20 were healthy. The results showed accuracies of 81.5% for subjects with depression, 81% for subjects with insomnia, and 84.5% for healthy subjects. These approaches are limited in that they collect data with numerous sensors, meaning that they are mainly useful in a hospital environment. Single-channel EEG approaches, in contrast, could be easier used at home in the future. Researchers working with single-channel EEG data often use the Physionet Sleep EDF database with eight subjects for validation. In this way, an accuracy of 88.62% for empirical mode decomposition with adaptive boosting (AdaBoost) and DT [169] can be reached; iterative filtering with RF can reach 90.02% [170]; complex-valued non-linear features and complex-valued neural networks (CVANN) can reach 91.57% [95]; and tunable Q-wavelet transformations with Bagging (DTs) can reach 92.43% [162].

In this setting, NN [95] performs with almost perfect agreement considering κ. We can conclude that pre-processing appears to be essential to successful single-channel EEG approaches. Its limitations lie in the size of the data sample with 15.000 epochs from only eight subjects—four healthy subjects and four who have not been diagnosed with sleep disorders but have mild difficulty falling asleep.

Table 2.5: Sleep stage classification comparison for single-channel EEG and one from single-lead ECG [15]. Stage classification in 30-second epochs except for [16]. 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, and S4											
[169]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	6	AdaBoost	60/ 5/35	79.9	71.8	88.6	0.82
[162]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	6	Bagging	LOOCV	80.4	71.6	89.6	0.84
[170]	8	Sleep-EDF	4H*-4 mild Ins.*	15,136 (~16)	6	RF	10-CV	81	71.7	90	0.84
[95]	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, and N3											
[171]	8	Sleep-EDF	4H*-4 mild Ins.*	5,760 (~ 6)	5	RNN	LOOCV	72.4	75.6	87.2	0.76
[169]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	5	AdaBoost	60/ 5/35	84.4	77	90.11	0.84
[162]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	5	Bagging	LOOCV	84.2	77.6	90.8	0.85
[170]	8	Sleep-EDF	4H*-4 mild Ins.*	15,136 (~16)	5	RF	10-CV	86.6	75.4	91.3	0.86
[95]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	5	CVANN	LOOCV	98.4(S*)	91.8	93.84	0.92
[172]	20	Sleep-EDF	H*	41,950 (~17)	5	CNN-LSTM	20-CV	75.6	78.7	82	0.76
[173]	20	Sleep-EDF	H*	41,950 (~17)	5	CNN	20-CV			83.5	0.72
[96]	83	41.3Y*avg*	41H*-42 Ins.*	~79,680 (~ 8)	5	DNN-HMM	5-CV			77	
[172]	62	MASS	H*	58,600 (~ 8)	5	CNN-LSTM	31-CV	82	81.3	86.2	0.80
[97]	5728	SHHS	divers(breathing)	5,384,401 (~ 8)	5	CNN	50/20/30	80.6	77.3	86.8	0.81
[174]	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, and Deep Sleep											
[169]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	4	AdaBoost	60/ 5/35			91.2	
[162]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	4	Bagging	LOOCV	89	85.6	91.5	0.86
[170]	8	Sleep-EDF	4H*-4 mild Ins.*	15,136 (~16)	4	RF	10-CV			92.29	
[15]	994	CinC2018DB	divers(breathing)	261,946 (~ 2)	4	CNN	10-CV			65.6	0.31
[15]	5,793	SHHS	divers(breathing)	400,547 (~ .5)	4	CNN	10-CV	52	64.1	65.9	0.47
[15]	16	SLPD	divers(breathing)	2,829 (~ 2)	4	CNN	10-CV	46.2	53.7	75.4	0.54
3 Stages - Wake, REM, and NREM											
[169]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	3	AdaBoost	60/ 5/35			93.55	
[162]	8	Sleep-EDF	4H*-4 mild Ins.*	15,188 (~16)	3	Bagging	LOOCV	91.3	88.8	93.9	0.89
[170]	8	Sleep-EDF	4H*-4 mild Ins.*	15,136 (~16)	3	RF	10-CV			94.6	
[93]	184	M*	divers(RLS*+Apn*)	176,640 (~ 8)	3	NN	10-CV	87.2	85.35	89.9	
[15]	5,793	SHHS	divers(breathing)	400,547 (~ .5)	3	CNN	10-CV			75.3	0.57
[15]	9,94	CinC2018DB	divers(breathing)	261,946 (~ 2)	3	CNN	10-CV			76.5	0.42
[15]	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-Apnoea; †Sleep stages: wake, (REM, S1), S2, S3, S4; •(training/validation/test) split

Five-Stage Classification Five-stage classification typically follows the AASM guidelines and is commonly performed in the medical domain. For single-channel EEG coming from the Sleep-EDF data, either (1) 8 subjects with 15,000 epochs or (2) 20 subjects with 42,000 epochs are investigated. Approach (1) can reach 87.2% accuracy with Elman-RNN (using only 6,000 epochs) [171], 90.11% with DT [169], 91.13% with iterative filtering and RF [170], 93.69% with tunable Q-wavelet transformations and Bagging (DTs) [162], and 93.84% with CVANN [95]. Approach (2) performs for 20-CV with an accuracy of 82% using CNN-LSTM [172] and 83.5% with CNN on a smartphone [173]. The data sample of 20 subjects contains only healthy individuals. Evidently, it is necessary to further investigate patients suffering from sleep disorders. In Shahin et al. [96], 41 healthy participants and 42 participants with insomnia were investigated, reaching an overall accuracy of 77% by applying deep neural network (DNN)-HMM. Twenty-eight subjects with sleep apnoea were considered by Koley and Dey [174], which reached 95.88% accuracy for SVM. A total of 5,728 patients from the Sleep Heart Health Study (SHHS) were investigated in Sors et al. [97], which applied a CNN and reached an accuracy of 87% but did not perform CV. Malafeev et al. [141] could reach a similar accuracy with a CNN-LSTM including 18 healthy patients, 23 patients with narcolepsy and five patients with hypersomnia. Sors et al. [97] reported that they reach better results for the smaller Sleep-EDF dataset (no exact results reported); this is potentially caused by the small number of technicians that participated. A lower number of technicians make the system learn a specific rater's style, which causes generalisation problems.

For single-channel usage, SVM (16,000 epochs, 28 participants) [174] and CNN (5,384,401 epochs, 5,728 participants) [97] tested on a dataset with a larger number of participants with sleep disorders, both of which presented promising results. CVANN (15,000 epochs, 8 participants) [95] performed well on a dataset with a small number of participants.

Overall, the N1 stage is often difficult to distinguish [172]. This leads to models that focus on this issue, such as the one in Aboalayon et al. [175]. Filtered single-channel EEG signals from 13 participants of the Sleep-EDF database were investigated with SVM; this led to an accuracy of 92.5% in distinguishing N1 from wake stages. Many approaches for sleep stage classification rely on features and pre-processed data while others use raw data, such as Malafeev et al. [141]. If no CV is performed, the results must be considered with caution, such as for Sors et al. [97] and Hassan and Bhuiyan [169].

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 have achieved 75.4% accuracy when including 16 subjects with sleep issues, 65.6% for 994 subjects with sleep disorders, and 65.9% for 5,793 subjects with breathing issues [15]. The investigations by Li et al. [15] have low κ values and reasonably high accuracy. This result is likely caused by an imbalance in the investigated datasets for different sleep stages. With imbalanced datasets accuracy is potentially overestimated, whereas κ represents the performance above the baseline of random guessing and, therefore, is more suitable for imbalanced data; for details, see Section 2.4.2.

In contrast, single-channel EEG from the Sleep-EDF database results in 91.2% accuracy with DT [169], 91.5% with tunable Q-wavelet transformations and Bagging (DTs) [162], and 92.29% with iterative filtering and RF [170]. Overall, RF [170] performs the best but has the limitation of involving only eight participants.

Three-Stage Classification Further simplification leads to three 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 [169], 93.9% with tunable Q-wavelet transformations and Bagging (DTs) [162], and 94.6% with RF [170]. Including participants with medical conditions, those using NNs reached 89.9% accuracy [93]. Extending the system with EEG, EOG, and flow allowed for 89.6% accuracy for healthy individuals and those with RLS and sleep apnoea [93]. In contrast, ECG data with a CNN reached an accuracy of 75.3% for SHHS and 81.6% for 16 subjects with sleep issues [15].

Figure 2.4 summarises 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 eight participants. Six sleep stages with 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). Investigating fewer sleep stages results in higher accuracy. Newer methods, such as deep learning, are generally used more often for larger datasets. Trends indicate that more data typically results in worse performance, some exceptions being simpler models, such as SVM and NN (compared to deep learning approaches).

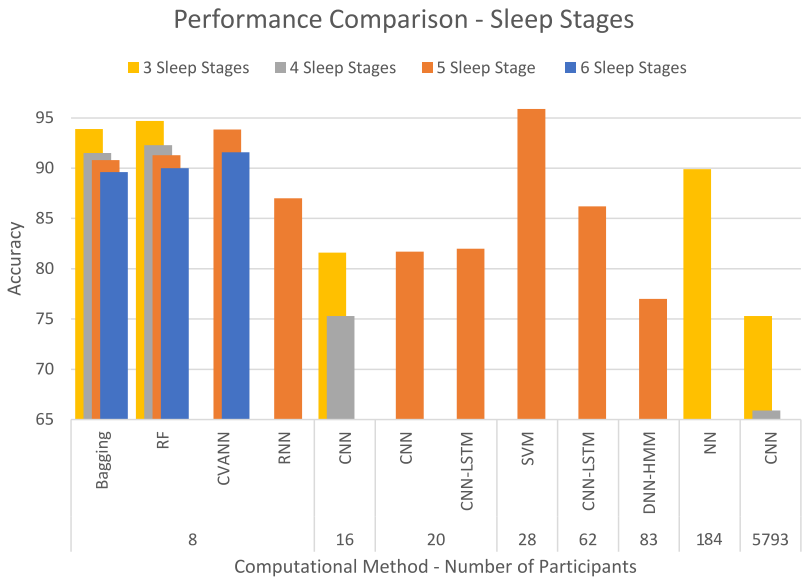


Figure 2.4: Performance comparison for sleep stage approaches that perform CV in the medical domain.

(c) 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 must be addressed in future research.

Most methods for home-based assessment have several issues: (1) training on imbalanced data; (2) inability to correctly classify complex stages; (3) exclusion of unhealthy participants in most datasets and a limited number of participants (non-diverse datasets); and (4) low generalisability, caused by non-standardised user-independent validation. It is clear that accelerometers are effective at classifying sleep from wake data, as movement is the main measurement used to distinguish sleep from wake stages. More complex sleep stages typically require additional information, which can be addressed by combining multiple sensors with accelerometers—this is currently tested only on healthy subjects. When reducing the number of sensors, audio sensors can distinguish between three stages, even for participants with sleep disorders. Wearable devices are promising for healthy subjects and are most commonly applied, whereas non-wearables are promising for subjects with medical conditions. Overall trends indicate that larger datasets and more diverse datasets worsen performance, which could be a result of (1) differences in sleep stage detection for people suffering from a medical condition and (2) inconsistency in movement or other elements throughout the night among different healthy individuals. The advantage of deep learning methods in larger datasets is in

line with the known necessity for more extensive training sets, as shown in Figure 2.4 and Table 2.5. Patterns are more easily distinguishable for sleep disorder participants when datasets are larger. This follows, e.g., from the performance of NNs shown in Figure 2.4, where the performance is compared for different dataset sizes for which the number of sleep disorder patients can be found in Table 2.5 (see [93]). The validation method, if mentioned at all, is rarely CV, and, is resultantly not objective in terms of user-independent validation and representation of influences from the training data. Overall, movement data have limitations when higher-granularity sleep stages are involved. Simple models are currently more promising than complex deep learning approaches. Comparative studies are rarely performed, which makes it difficult to determine the ideal sensor for home assessment or the ideal machine learning approach. Generally, one-model-fits-all approaches are investigated, so no personal or other influence factors are considered. Mainly, testing against one technician's scoring is performed but is restricted, as technicians have an individual component. As a result, agreement is not always present [141]. Applied machine learning algorithms enhance the automated practices but have then the restriction to learn one technician's style. Typically, one data source is accessed for information and a fusion of multiple sources is not considered. We address these issues in Chapter 5, where we present a fine-grained sleep stage detection approach.

For sleep stage analysis in the medical domain, challenges remain for (1) the generalisation of the model to people affected by diseases and (2) dealing with imbalanced classes. These challenges are evident in the fact that methods perform worse on larger and more diverse datasets. This effect is also influenced by the imbalance of the Sleep-EDF dataset containing eight participants with a majority class of wake, which introduces a model bias, and, therefore, positively influences performance. Overall, it appears that single-channel EEG is sufficient to obtain sleep stages—however, home use is still a future step. The low number of features and signal sources required for single-channel EEG could make it possible to easily extract in-depth information at home. Despite this potential, the gold standard remains PSG, which is not feasible in a home setting. In general, online detection is desirable because it can be useful for boosting slow-wave sleep, transcranial stimulation, and acoustic stimulation [164]. In Figure 2.4, deep learning approaches perform worse with larger datasets, but comparative studies of simple models with the same datasets are lacking. The trend of using deep learning methods on larger datasets is in line with the known influence of small datasets on the methods—see RNN, which uses only 6,000 epochs. Therefore, the performance advantages and disadvantages of deep learning cannot be observed and discussed in detail. For smaller datasets,

deep learning approaches perform worse, most likely influenced by the need for more training data compared to approaches like RF and bagging. Six sleep stages are investigated only with smaller datasets, likely because of the newer guidelines that allow for the consideration of fewer (five) stages [3]. It is easier to classify fewer sleep stages, as the ones that are difficult to differentiate between are simply combined. Overall, CV and LOOCV are the most commonly used validation methods, which guarantees objective performance 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, which is often present in these studies, influences the accuracy.

2.4.4 Sleep Position Recognition

Sleep position recognition generally detects the four basic sleep positions: supine, prone, left lateral, and right lateral. Higher-granularity positions are increasingly being investigated as well. For some approaches (e.g., image-based), the number of detected positions must be reduced due to methodological limitations. This normally involves excluding the prone position or combining it with the supine position [39,53, 60, 61, 65, 181–184]. Other approaches detect more complex positions on top of the four basic ones, including right fetus [53, 118], right yearner [53], right log [118], left fetus [53, 118], left yearner [53], and left log [118]. Varying arm positions [39, 179], leg positions (see Chapter 4), and angles [57, 60], are considered higher-granularity positions. These positions can be associated with certain issues, such as back pain [55]. Furthermore, the basic set of positions can be extended with an unknown state, normally the sitting/standing position [65, 178, 180], explaining, e.g., bed exits.

In this section, outcomes are distinguished by wearable and non-wearable devices, as sensors in the medical field are rarely used (only in [176, 177]). A summary can be found in Table 2.6, where different methodologies and their results are listed.

(a) Sleep Position Analysis Using Wearables

Wearables are often investigated because they are easily applied and provide high accuracy for position detection. Some systems have almost perfect performance for detecting sleep positions, though they normally have issues, such as the limited number of participants investigated and the lack of user-independent validation.

In general, single sensors can effectively detect basic sleep positions. Shinar et

Table 2.6: Sleep position monitoring systems. Mentions of ground truth are based on video in all instances except [17], which is based on a smartwatch.

	Data	I*	h* (seP*)	P*	Add to Basic P*	Analysis	Val.*(%•)	Pre.	Rec.	Acc.	κ
Wearables - Simulated Setting											
[176]	ECG-Lead II	12	10 (250)	4		k-means		93(S*)	79		
[177]	Respiration Impedance	16	6.4 (180)	4		SVM	10-CV	99.7	99.7	99.7	0.99
[178]	Koala (wrist+chest)	uc*	- (180)	5	sit	RF		86.5	88.6	88.5	0.86
[57]	W*(chest) + WSN*	2	1.4 (10)	5	30°-incline lateral	KNN				100	
[179]	3 W*(chest+arms)	10	0.4 (10)	8	arm P*	KNN		92.2	92.5	92.2	0.91
Wearables - Real-World Setting											
[17]	Huawei SW*	16	~560	4		RF	LOOCV	TPR"=91.8; FPR"=0.03			
[99]	Accelerometer(chest)	7	~56	4		LDA		99.5	99.5	99.5	0.99
[180]	Accelerometer(chest)	13	88.4	5	sit/stand	Rules		99.1	95.6	99.2	0.98
Non-Wearables - Simulated Setting											
[61]	Pressure sens.(uBP*)	2	uc*	3	no prone	BC	10-CV	81.4	78.7	78.7	0.71
[181]	Pressure mat	13	uc*	3	no prone	RBM-DNN	10-CV	82.6	79.4	82.7	
[182]	Pressure mat(uBP*)	1	120 (uc*)	3	no prone	NN-BN	5-CV			89.9	
[183]	Pressure mat	uc*	312 S*	3	fuse prone & supine	PCA-SVM	10-CV	94.2	94.1	94.1	0.91
[60]	Pressure mat	9	uc*	3	no prone	GMM-KNN	60/40	98.4	98.4	98.4	0.98
[184]	Pressure sen.(uBP*)	2	0.3 (60)	3	no prone	RF	33/67			98.4	
[185]	HBT*	58	3.9 (60)	4		NN	10-CV			72	
[59]	Pressure distr.	3	6 (300)	4		LR	33/67	90.5	90.1	90.2	0.88
[53]	Pressure mat	6	uc*	5	leg P*, no prone	PCA-KNN	70/30	97.7	97.6	97.7	0.97
[118]	Pressure bedsheet	14	3360 S*	6	leg P*	MCR	LOOCV	83.3	83	83	0.81
[120]	3D-Artec/Kinect	3	uc*	6	leg P*, no prone	SVM	67/33	92.7	92.5	92.5	0.91
[39]	Pressure mat	20	uc*	8	leg, arm P*, no prone	KNN	10-CV	97.1	97.1	97.1	0.97
[60]	Pressure mat	9	uc*	13	body, limb P*, no prone	GMM-KNN	60/40			91.6	
Non-Wearables - Real-World Setting											
[65]	3D-Asus Xtion cam	78	1880S*	4	sit/stand, no prone	CNN	5-CV	95	94.9	94.9	0.93
[118]	Pressure bedsheet	3	7800S*	6	leg P*	MCR-HMM	LOOCV	86.5	84.7	85.6(F*)	

*I-Individuals; h-hours; seP-seconds per position; P-Positions; Val.-Validation; uBP-upper bed part; HBT-hydraulic bed transducers; W-Wearables; WSN-wireless sensor network; ShS-Shimmer Sensors; SW-Smartwatch; uc-unclear; S-Samples; S-Specificity; F-F_1 Score; •(training/test) split

al. [176] used an ECG device and achieved a specificity of 93% and a sensitivity of 79% with a k-means iterative algorithm. In Liu et al. [177], respiration impedance signals are measured among 16 individuals; the researchers achieved 99.7% accuracy for 10-CV.

Extending approaches with additional wearable devices reached an accuracy of 88.5% with RF [178] and 92.2% with KNN for ten subjects [179]. Barsocchi [57] studied the feasibility of using a transmitter and receivers in the sleep environment to distinguish four main positions and lateral incline of 30%. They were able to reach 100% accuracy with just two sensors and KNN.

The best matching rates, however, were achieved by Barsocchi [57], which was limited by a small sample of just two participants. Higher-granularity positions including variations in hand movement were considered by Kwasnicki et al. [179].

The next step is the real-world application, which has rarely been considered. Smartwatches can be applied, e.g., as they were by Sun et al. [17], where RF performed the best with a TP rate of 91.8% for 16 participants and objective LOOCV. The most promising models, however, involve an accelerometer placed on the chest and reach 99.5% accuracy with linear discriminant analysis (LDA) for seven subjects [99]; even rule-based approaches can reach 99.2% accuracy for 13 subjects [180] including an unknown state.

Overall, many reviewed approaches had a low number of participants and, moreover, often failed to state the used validation method.

(b) Sleep Position Analysis Using Non-Wearables

Non-wearables usually require more complex analysis methods because they often produce images that need to be classified.

The use of three sleep position classifications is very common, as distinguishing between prone and supine is quite difficult for most sensors. Consequently, Hsia et al. [61] used pressure sensors at the upper part of the bed to investigate three positions by focusing on the influences coming from hand positioning and laying angle with a Bayesian classifier (BC). The results showed a low accuracy of 78.7% for two individuals and were disproportionately influenced by the laying angle. Pressure mats are commonly applied for position detection; they have reached up to 98.4% accuracy with the Gaussian Mixture Model (GMM)-KNN for nine individuals [60] as well as with RF for two subjects and no CV [184]. Other investigations have reached

94.1% accuracy with PCA-SVM [183], 89.9% with NN-Bayesian Network (BN) for one person [182], and 82.7% with RBM-DNN for 13 individuals [181]. Evidently, simple data analysis models already appear promising. More complex approaches, such as hydraulic bed transducers, were placed under a mattress to distinguish between the four basic sleep positions among 58 participants with an NN [185]. This led to an accuracy of 72%—that is fairly low compared to the 90.2% achieved by the approach with pressure distribution and LR, which only included three participants and no CV [59]. Yousefi et al. [53] utilised a pressure mat to detect five different positions, including supine, left/right yearner, and left/right fetus, from six subjects. The position detection was based on a three-step algorithm using normalisation, eigenspace projection, and a KNN classifier. The average accuracy of this detection reached 97.7%. Liu et al. [118] investigated a mobile, easy applicable solution. In this case, a pressure-sensitive bedsheet was used to monitor six sleep positions, including supine, prone, log, and fetus, by sparse classifiers with minimum class residual (MCR); the process reached an accuracy of 83% for 14 subjects [118]. To obtain a more detailed picture, cameras capable of 3D measurements have been used and have reached 92.5% accuracy using SVM for three individuals [120]. Pressure mats can be used to detect higher-granularity positions—Pouyan et al. [39], e.g., classified eight different positions in the bed, excluding prone, using pressure mats. Their algorithm creates a pressure image that is processed using size- and shift-invariant image processing. Positions were classified by computing the Hamming distance between the signature images and the presented sample. Results showed an accuracy of 97.1% for 20 subjects. Multiple angles of three positions were detected by Ostadabbas et al. [60]; they detected 13 sleep positions from nine individuals with a GMM-based clustering approach and reached an accuracy of 91.6%.

Studies with whole nights of data from the real-world are rarely conducted. However, some that have been conducted have involved Kinect devices [58], 3D-Asus Xtion cameras [65], and pressure bedsheets [118]. Kinect sensor data were collected to distinguish five sleep positions among 20 students [58]—no blankets were used. In one study, a single-depth camera was able to distinguish supine, left lateral, right lateral, and the unknown position (sitting/standing). The researchers included 78 patients and reached 94.9% accuracy with a CNN [65]. Liu et al. detected six positions by using pressure-sensitive data from a bedsheet. Three nights from three people were investigated using a MCR-HMM, resulting in 86.5% precision and 84.7% recall [118].

To obtain a fast and easy visualisation of the current state of sleep position investigations, we incorporated Figure 2.5, which compares all CV investigations.

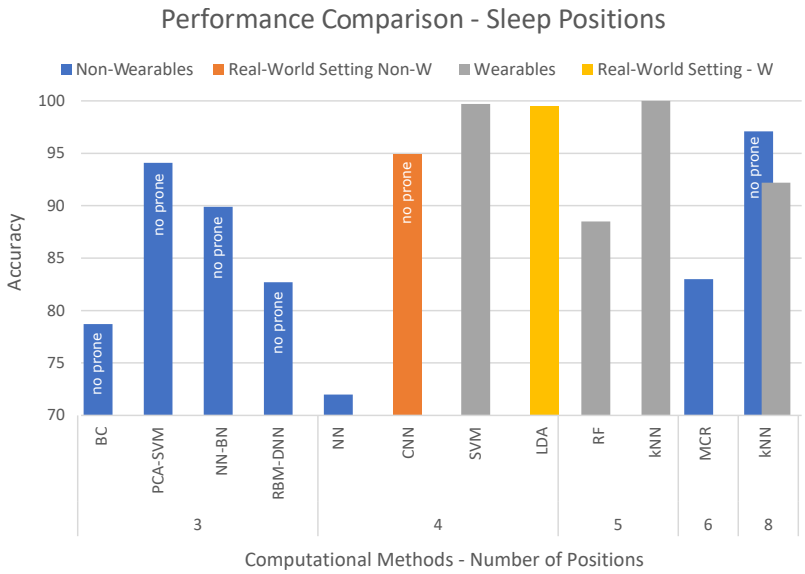


Figure 2.5: Performance comparison for sleep position approaches using CV.

Three-position investigations primarily use non-wearable devices, whereas the basic four-position and higher-granularity-position investigations typically use wearables. Overall, simple models with no temporal component, such as KNN and SVM, appear to be appropriate for positions investigations. Little research has been done under real-world conditions.

(c) Discussion and Suggestions on Sleep Position Analysis

Sleep position analysis can be performed with wearable and non-wearable devices in simulated or real-world settings. Wearable accelerometers, particularly when placed on the chest, can reach high accuracy with simple machine learning models. Alternatively, respiration monitoring is also a promising approach. While research has evaluated various data sources and technologies, a computing-efficient and robust solution that is adaptable to individual needs and can accurately detect higher-granularity positions is still lacking. Wearable devices are commonly used when tracking sleep positions, but as these are worn on the body, discomfort can be a problematic factor. Therefore, improving usability and comfort is of high interest.

Non-wearable devices, however, are also good options. Pressure mats are commonly used and allow for high accuracy. Recent investigations have attempted to produce 3D images, however, this methods are limited because of privacy concerns. Figure 2.5 shows that non-wearable devices cannot accurately distinguish the prone from the supine position (partially overcome by Liu et al. [118] and Barsocchi et al. [59]) 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, applications in the real world are rarely reported but are nonetheless important. Simple models, rather than more complex approaches, appear to be appropriate for sleep position investigations. These complex approaches use NN or deep learning and are more notably influenced by the quantity of training data. A small number of participants is usually investigated, which comes with a drawback in that consistency in sleep positions 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 to provide reproducible results that are able to enhance recognition rates for higher-granularity positions. Furthermore, user-independent validation needs to become a standard to guarantee generalisability. In the future, other sensor sources will become available and could lead to more advanced investigations.

2.4.5 Investigation of Sleep Disorders and Diseases

Abnormal sleep behaviour was originally classified by sleep experts and further used to diagnose individuals with sleep disorders. Recently, however, behaviour has been investigated in automated decision-making for (1) sleep disorders, such as sleep apnoea [18, 62, 186, 187] and insomnia [18], and (2) certain chronic diseases, such as diabetes, hypertension [18], Alzheimer's disease [77], and Parkinson's disease [78]. Established sleep parameters can be used to investigate disorders and abnormal sleep behaviour. The potential for the field of computational behaviour analysis to help in diagnostics has not yet been sufficiently explored.

(a) Sleep Apnoea Investigation

Sleep Apnoea Detection Javaid et al. [187] investigated the use of a non-wearable Impulse Radio Ultra-Wide Band Radar panel under the mattress to detect sleep apnoea events. The analysed dataset consists of 25 hours of data from 4 subjects with apnoea. The overall match against PSG reached an accuracy of 70% for 5-CV with an LDA on extracted statistical features. In Aggarwal et al. [18], sleep

apnoea diagnostics was investigated using actigraphy. The researchers investigated the Hispanic Community Health Study (HCHS) [188] with 1887 individuals; the results for two classes (affected and unaffected by sleep apnoea) showed an accuracy of 68% with a CNN [18]. An alternative for sleep apnoea-hypopnoea syndrome diagnosis, a home-based oximetry sensor, was investigated by Gutiérrez-Tobal et al. [189]. A dataset of 320 subjects (one night each) was analysed using different machine learning algorithms, such as LDA, LR, Bayesian multilayer perceptron, and AdaBoost. A split of 60% was used for training while a split with the remaining 40% was used 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. [190] reached the diagnostic ability closest to the machine learning approach by applying statistical methods; they achieved a better accuracy of 93.7% for an AHI of 30, but the sample was made up of only surgical patients (475 patients, one night each). Khandoker et al. [62] analysed two sleep apnoea types among 83 patients by applying SVM to 125 sets of ECG records and extracting 24 features. This approach reached an accuracy of 92.85% for LOOCV. An optimisation problem on kernels led to the conclusion that the polynomial kernel with a degree of three provides the best results. The research was limited to patients with no history of cardiovascular disease or central sleep apnoea. The model could prove useful for determining the CPAP therapy by analysing the change in probabilities in the outcome [62].

It is difficult to compare the approaches because the data sources used and classes investigated are different. Nevertheless, while non-wearable devices can recognise sleep apnoea events with an accuracy of 70%, they are outperformed by wearable sensors such as ECG, oximetry, and actigraphy sensors.

Sleep Apnoea Treatment CPAP devices are utilised to treat patients suffering from sleep apnoea by providing pressure to help them through apnoea episodes. Araujo et al. [186] used CPAP and electronic health records of 3.588 patients to investigate an approach meant to detect patients who were likely to discontinue the therapy. The researchers performed feature selection and used oversampling to counteract the dataset's imbalanced distribution. Different techniques were applied, such as linear regression, LR, DT, and SVM—RF and extreme gradient boosting, were the only ones to significantly improve the results. Deep learning for time series classification did not perform well, largely due to the lack of data. The results showed that extreme gradient boosting produced the best F_1 score including health records with an accuracy of 85% for 10-CV, which is an improvement relative to the current state of the art.

(b) Chronic Disease Investigation

As discussed in Section 2.2.3, chronic diseases affect not only a person's health, but also introduce sleep behaviour changes. These changes are investigated to obtain reliable indicators for early and advanced stages of diseases.

Aggarwal et al. [18] investigated sleep behaviour changes by considering actigraphy data from a broad population to detect sleep apnoea, insomnia, diabetes, and hypertension. A CNN was applied to the HCHS dataset, introducing a method for embedding activities. However, the study has one major limitation: An imbalanced dataset is used to train a supervised machine learning approach, where the majority of subjects are not suffering from the disorders. Consequently, its models and baselines are highly biased towards predicting the majority class [18]. The accuracy or related F_1 (micro), when including individuals with seven nights of data, was 69% for hypertension, 44% for diabetes, 68% for apnoea, and 41% for insomnia.

Early disease detection is especially important for the diagnoses of both Alzheimer's and Parkinson's disease, which relate to sleep [77, 78]. Early stages of Alzheimer's disease already affect sleep behaviour based on the relationship between β-amyloid (Aβ) and sleep quantity and quality, as measured through actigraphy [77]. Sleep behaviour also shows relationships for patients with Parkinson's disease. Data from actigraph units have been explored to investigate these correlations [78]. These newly elucidated relationships could seriously enhance early disease detection.

(c) Suggestions for Disorder and Disease Investigations

In general, it is necessary to use the existing knowledge about the relationships between sleep and certain diseases to enhance and promote self-management and early detection. Automating sleep disorder diagnoses (such as sleep apnoea with daily technologies) could enable easy and inexpensive assessments.

With the increasing prominence of deep learning technology capable of processing large amounts of data, chronic diseases could be more effectively investigated using sleep data. Based on current sensor accuracy and accessibility, investigations on disease diagnostics is certainly a promising prospect with the potential to make early diagnoses at home.

In the investigations presented here, a large number of individuals are usually involved. When the validation method is stated, it is CV, which guarantees an objec-

tive process. We decided to present the accuracy metric in this section for consistency, as most failed to present other measurements.

2.4.6 Sleep Behaviour Assessment Frameworks

From the three primary assessment methods (PSG, sleep questionnaires, and actigraphy), various sleep parameters can be extracted and used to classify sleep disorders; e.g., insomnia patients are diagnosed through SOL and sleep time [6]. These subjective and objective methods each have individual characteristics that affect the validity and precision of the extracted information [6]; details are provided in Section 2.3.

A home-based approach enables the objective assessment of sleep through long-term data collection in a natural environment. This objective assessment intrigued sleep experts [45] and led to the initiation of IoT-based approaches. Consequently, self-management is increasingly receiving attention in sleep assessment, as daily IoT devices become more accessible through smartphones and watches (providing insight into sleep behaviour) [45]. IoT-enabled self-management tools for sleep reduce burdens, are more efficient than self-reports [103], and allow experts to more easily detect problems with sleep behaviour [45]. However, commercially available devices often focus on sleep parameters that cannot be independently changed, such as sleep stages. Users tend to misinterpret these parameters and what they mean for sleep quality [45]. To provide appropriate feedback, connections and correlations must be measured and adequately presented to users [43].

Current research on IoT-based assessment investigates various data analysis methods and new sensors to try to close the performance gap with PSG. These approaches normally concentrate on improving one specific aspect of sleep with sensors, be they wearable or non-wearable (compare Section 2.3.2). Nevertheless, multiple factors influence and define sleep behaviour; hence, multiple aspects should be integrated into assessment methods. Nam et al. [127] proposed a sleep quality monitoring flow with two sensors: an accelerometer and a pressure sensor. Sleep measurements (activity, heart rate, and respiration) are used to extract various sleep parameters: sleep positions, sleep stages, sleep time, and apnoea events. These sleep parameters describe sleep quality through a proposed equation [127]. Contrarily, Doppler radar signals gauge respiratory rate, movement, and bed exits using SleepSense, a cost-effective, unobtrusive, and privacy-preserving system [94]. Multiple sleep parameters can also be analysed by two accelerometers, including heartbeat, respiratory motion,

sleep position, and movement [159].

Research suggests that every parameter has a different value to different individuals (e.g., sleep duration). Current approaches fail to incorporate individuality despite it being an essential aspect of any personalised system [45]. Other important aspects are missing as well: regularity, circadian rhythm [130], flexibility, influence factors from clinical history, and environmental factors. Some attempts have been made to incorporate environmental factors, such as light, humidity, temperature, sound, and air quality [20]. Deen [20] proposed a smart sleep environment for elderly people that allows for customisation, feedback, and recommendations. Other proposals focus on smartphones to monitor important factors of sleep: noise, location, light intensity, and movement. Doryab et al. [66] found that less mobility and phone usage relate to better sleep, and more social interaction increased sleep quality. While there is some evidence to back these relationships, generalisability is still lacking. Numerous ready-to-use smartphone and smartwatch applications exist on the market, but most have accuracy issues [6,19,105], or provide limited information compared to PSG and actigraphy [105] (for details, see Section 2.3.2). Sleep experts are not fully convinced by smart devices, as performance among them differ widely. They do, however, see the benefit of these trackers in combination with feedback on regularity, sleep hygiene, and environmental factors [45].

2.5 Critical Analysis

In this section, we highlight the uniqueness of our approach in comparison with relevant work from literature and the remaining open issues.

2.5.1 Uniqueness of our Approach

The uniqueness of our approach lies in the combination of different relevant sleep behaviour aspects to establish a reliable sleep behaviour assessment framework which accounts for influence factors, personalisation, regularity, accurate visualisation, and the potential from the environment (for details, see Chapter 3). We propose a higher-granularity sleep position approach which extracts representative patterns from eight individual positions including leg positioning in Chapter 4. A comparative study is performed for guided and real-world settings. The proposed fine-grained sleep stage detection handles multi-source data and incorporates influence factors from health status, gender, and race. It supersedes current one-model-fits-all approaches,

Table 2.7: Summary of the uniqueness of our research compared to the most similar approaches in literature.

Lit.	Purpose	Research Elements	Comparison to our Approach
		Sleep Framework	
[127]	Extract sleep quality from collected accelerometer and pressure sensor data	Sleep positions, sleep stages, sleep time, and apnoea events, 3 participants	We incorporate currently missing aspects: regularity, circadian rhythm, influence factors, environmental factors, sleep hygiene, and personalisation of the assessment. We use an adaptable approach which can deal with different sensor data and combine multiple elements. Validation on a diverse and large dataset is performed.
[94]	Noncontact and cost-effective sleep monitoring with Doppler radar signals	Respiratory rate, movement, and bed exits, 3 participants, 1 in a real-world setting	
[159]	Monitoring sleep behaviour with two accelerometers	Heartbeat, respiratory motion, sleep positions, and movement	
		Sleep Positions	
[179]	Detect 8 positions with arm postures	Wearables; guided setting; 3 accelerometers	Higher-granular positions with different leg postures are not extensively investigated with wearables. In [179] arm positions where investigated with KNN, we propose LVQ with the advantage of easy feature selection and perform a comparative study between guided and real-world settings including the prone position.
[39]	Detect 8 position with limb postures, excluding prone	Non-Wearables; guided setting; pressure mat	
[60]	Detect 13 positions with limb postures, excluding prone	Non-Wearables; guided setting; pressure mat	
[118]	Detect 6 positions including leg postures	Non-Wearables; real-world setting; pressure bedsheet	
		Sleep Stages	
[46]	Sleep-Wake stage detection	PPG and HRV from 10 participants	We designed a fine-grained sleep stage detection approach which can handle multi-source data and incorporates influence factors from health status, gender, and race for diseased and healthy individuals.
[47]	Sleep-Wake stage detection	Accelerometer data from 22 participants	
[167]	Sleep-Wake stage detection	Video recordings from 10 participants	
[48]	Sleep-Wake stage detection	Accelerometer data from 81 participants	
		Chronic Disease Detection	
[18]	Classify existing chronic diseases	Actigraphy data from individuals affected by apnoea, diabetes, hypertension are analysed with CNNs	We designed LSTMs to deal with temporal aspects while fusing clinical history and actigraphy data to detect apnoea, diabetes, hypertension, and CKD at early and later stages.

and is tested on a large and diverse dataset (compare Chapter 5). Chronic disease detection is performed from sleep-wake behaviour by introducing a LSTM to deal with the temporal data and multi-dimensional features coming from clinical history and sensor data (for details, see Chapter 7). In-depth understanding on sleep and chronic disease relations was established in our study (compare Chapter 6) on the perception from doctors and effected individuals. A summary of the uniqueness of our approach in relation to relevant literature is provided in Table 2.7.

2.5.2 Open Issues

In recent decades, sleep behaviour analysis has advanced considerably through the introduction of new devices and computational methods; nevertheless, there are still limitations and challenges that must be addressed by the research community. Common challenges that must be met include (1) incorporating sleep behaviour differences from both healthy and unhealthy populations into methodologies; (2) including medical knowledge in terms of sleep structure, relationships, 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; (6) providing adequate comparative outcomes and standards; and (7) establishing a framework to present different aspects of sleep at home using sensor technology.

There is a strong need in sleep behaviour analysis to distinguish between healthy subjects and subjects with medical conditions. Currently, research mainly targets healthy individuals for sleep analysis. When subjects with medical conditions are included, overall performance is typically lower, but this can be addressed by either training models separately for different groups to establish personalised approaches; or by including new features in the training process that would either include knowledge of users' health or represent behavioural differences more accurately. We address this aspect, in chapters 3 and 5, by investigating personalised approaches and include new features to increase performance and, in turn, reliability.

Medical researchers established knowledge of the structure, i.e., the time aspect of sleep changes, and relations of, e.g., sleep and health factors, over the last several decades. This knowledge can be an advantage when conducting a computational sleep behaviour analysis. In Chapter 4, we establish an approach to investigate sleep position changes throughout the night that advances the currently used subjective questionnaires. Including temporal aspects while assessing sleep can likely improve performance but has not yet been investigated extensively. We include

the time aspect of sleep in our chronic disease detection approach in Chapter 7. Current research mainly concentrates on automating human scoring tasks on well-known problems such as sleep stage classification. A drawback of this computational scoring is the usage of inconsistent human labelling with an inter-scorer agreement of only 82.6% [141, 160]. One way to bring new insight into sleep behaviour analysis that is not based on human labelling 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 [7]. The regularity of sleep [7] represents trends over consecutive nights, which is a step towards long-term visualisation and helps to draw more specific conclusions. In Chapter 3, we develop an algorithm that represents the sleep regularity. This sleep regularity provides more insight than the regularity index by representing the regularity of individual sleep parameters. Furthermore, the automatic detection of shorter underlying structures using machine learning techniques is rapidly emerging, such as for K-complexes [191] and sleep spindle detection [191, 192]. Eventually, research will extract this information from sensor sources other than EEG to make it easier to gather at home.

Improving technologies for home use entails addressing potential discomfort and integrating them into existing smart devices. Advances will likely move in the direction of wearable sensors that can be integrated into clothing and single-channel EEG devices that are easily applicable. Advances could also be made in non-wearables, which could be seamlessly integrated into the environment; existing methodologies could be further explored, as non-wearables are currently less accurate. In certain areas, such as sleep stage detection, research must be conducted to develop technology that can accurately detect higher levels of sleep stages within a natural home environment. In Chapter 5, we targeted this issue by bringing context to the current differences in sleep-wake detection performance for different groups of people.

In the future, the issue of small datasets could be overcome through user contributions (i.e., by crowdsourcing sleep data to contribute to sleep research). Consumer wearables will progressively adapt to facilitate the sharing of sleep tracking data with researchers and could be contextualised by users' personal information, such as gender, age, and health status—studies could possibly be large enough to overcome any inherent noise or inconsistencies. More diverse datasets including these factors would enable researchers to address differences in healthy and diseased individuals. Still, a balance needs to be persevered between diverse aspects of the dataset during training to counteract inconsistency, i.e., the bias towards the majority class. It is not yet standard for current research to provide the validation

method used, but repeated CV and user-independent validation are critical for reliability. A generalised approach is usually targeted, so users who are trained on cannot be tested—otherwise the performance ability of the method is unreliable. No standardised performance measure exists but overall accuracy is fairly common. In the case of imbalanced data, however, further measurements, such as recall and precision, are necessary. Therefore, we recommend that the performance metrics used in sleep behaviour analysis be standardised to easily allow for comparative studies. Comparative studies are needed to investigate different methods used on the same datasets to benchmark available approaches and assess the most promising methods. Throughout this thesis, we address balance issues in datasets and provide comparative results by using publicly available data and repetitive CV to assure reproducibility.

It is well known that there are many correlations between sleep and certain chronic diseases, and while some researchers work on them, these correlations do not receive sufficient attention. The research being conducted is typically aimed at exploring features that might provide hints about specific diseases. These features can be measured continuously with sensor technology at home. Combining this data with known relationships could increase the potential for early-stage diagnosis of some diseases. In Chapter 7, we were driven by the current knowledge on relationships between sleep and chronic diseases (see Chapter 6) to investigate a computational approach for chronic disease detection through wake-sleep behaviour. People often learn about medical issues too late for preventive treatment. Computational analysis could enable us to uncover the patterns in sleep data that are unobtainable using current methods. Consequently, with these patterns users could be shown their sleep habits [45] and learn about medical issues earlier.

There is currently no framework able to combine all of the individual investigated sleep aspects. A framework offers several benefits: (1) integration of different facets on which researchers can build upon, (2) efficiency, (3) scalability towards user's focus, and (4) robustness. Thus, a framework acts as a guideline for future researchers and the first step towards a reliable system. Overall, many aspects of sleep are very subjective [45]; therefore, individuality should be addressed and considered in this framework. This would allow for the fusion of different sleep aspects from objective, subjective, and environmental [20] perspectives into an automatic approach for sleep assessment and self-management. This can be extended by known relationships, e.g., between physical activity and sleep, which could actually improve outcomes, such as in [10]. Consideration of routines and behaviours, including environmental aspects, sleep hygiene, regularity, and circadian rhythm [45, 132, 134],

would serve to enhance sleep behaviour assessment frameworks. Models could also benefit from including specifications that vary based on moon phase, season, day of the week, and periodic personal sleep behaviour changes. We develop a sleep behaviour assessment framework with the following aspects in mind: (1) a personalised solution to allow for self-management, (2) flexibility towards doctors' and users' needs, preferences, and recommendations, (3) adaptability to available data sources, (4) inclusion of regularity and circadian rhythm, (5) integration of the potential effects of environmental factors, (6) validation on larger datasets to extract more reliable relationships, (7) sensor considerations to improve sleep disorder diagnoses, and (8) personalised recommendations for users. On top of a framework providing this overall picture, researchers should determine the best combination of sensors and the effect of the output on the next day's activities. An overall framework would lead to new investigation areas that could provide new findings and potentially solve the mystery around sleep. Investigating information provided to users (see Chapter 3) and validating individual parameters will improve the current sleep assessment process.

2.6 Summary

This chapter provided a systematic, comprehensive review on the state of the art in computational sleep behaviour analysis. We specifically focused on the latest developments in sleep monitoring, modelling, and computational analysis methods for sleep assessments using sensor technologies, which have recently been made available—as well as easier, quicker, and inexpensive—to the general public at home. This is a rapidly growing and dynamically changing research area. While previous research has mainly sought to exploit and automate human expert knowledge, one apparent trend has been to apply data-driven techniques to investigate data from various perspectives, rather than through human labelling alone, to discover new insights. In this context, we identified challenges and issues faced in sleep behaviour assessment, some of which we will continue to address in the following chapters.

Chapter 3

Sleep Behaviour Assessment Framework

3.1 Introduction

Sleep impacts health and well-being [20]. Thus, by assessing sleep behaviour, one can gain insight into general health. The assessment of sleep has a long history and has proven useful in diagnosing diseases and offering lifestyle indications [6]. In recent years, self-assessment has received growing attention as being revealing not only for users but also for doctors, as they are able to extract additional and important factors from a natural sleep environment. The modern healthcare system as a whole is troubled in that many of its individual systems are unable to communicate with one another and, thus, fail to provide optimal care [193]. Current sleep behaviour assessment is restricted to controlled hospital environments limited by (1) the duration of monitoring (usually two days) and (2) medical sleep parameters to diagnose diseases. As health care shifts from reactive to preventive care through digital health and IoT technologies, the demand for home-based sleep assessment continues to grow.

Current research largely concentrates on one specific aspect of sleep. In contrast, we aim to bring various aspects together to form an overall framework for sleep behaviour assessment. We believe that integrating key behavioural sleep aspects can enhance current assessment if we incorporate existing medical and computational knowledge. This means that, within computational reason, the perception from sleep experts on important aspects and thresholds should be integrated. We address the following lacks that were presented in Chapter 2: (1) an overall framework; (2) personalised solutions; (3) flexibility towards doctors' and users' needs, preferences, and recommendations; (4) flexibility to available data sources; (5) inclusion of sleep regularity and circadian rhythm; (6) integration of the potential to improve sleep quality; and (7) incorporation of sleep disorder diagnosis. To gain

more insight into people's sleep behaviour at home and subsequently address objective 1, we propose a sleep behaviour assessment framework that considers sleep quality, regularity, circadian rhythm, environmental conditions, and sleep hygiene. This framework considers personal preferences, influence factors, doctors' recommendations, and clinical history to allow for personalised medical and behavioural assessment. To address objective 2, we develop a sleep regularity algorithm which is embedded in the design of the framework. The visualisation of the sleep behaviour assessment is key, therefore, to address objective 3, we design a visualisation concept to adequately present the sleep parameters to users. Additionally, the framework follows a modular, service-oriented design that adapts to personal needs and available technologies. We propose this with the knowledge that frameworks can integrate different facets, are scalable, can be used efficiently, and are robust; they can act as guidelines for future research and can be adapted. With this framework, we hope to provide an objective, trustworthy assessment that minimises subjectivity. The framework's structure will be followed throughout this thesis. In the following chapters, we discuss how we integrated various methodologies for specific sleep aspects to provide general sleep assessment for users to monitor, analyse, fuse, and present information. Our personalised, home-based framework can adapt to users' and doctors' needs, represent regularity, and present information in a way that does not contribute to misinterpretation.

The structure, components, and characteristics of our framework are presented in Section 3.2; this section also details the technologies, devices, parameters, and influence factors used in our assessment process. Section 3.3 provides an overview of the challenges facing integration into existing smart homes. The analysis on specific sleep framework aspects is explained in Section 3.4. In Section 3.5, we explain the outcomes and discuss limitations. A summary is provided in Section 3.6 alongside future research directions.

3.2 The Sleep Behaviour Assessment Framework

Introducing basic structure in sleep assessment would provide focus and support by combining different facets and perspectives; thus increases accessibility. Our generally applicable framework for sleep behaviour assessment provides this structure by addressing the current gaps. To build this sleep behaviour assessment framework, we consolidated various sources of medical knowledge and technological possibilities. Our framework combines these perspectives to support a full picture of assessment

inside a person's home and remain flexible. The home-based framework enables users to self-assess their sleep behaviour with IoT devices in the long term; the same information can be provided to doctors so that they can diagnose problems faster and more effectively. The assessment is divided into two main parts: biological—circadian rhythm, regularity, and sleep quality—and potential—environmental aspects and sleep hygiene.

We present the framework as follows: (1) the layered-structure of the framework, (2) the sleep assessment components, and (3) the modular-design and flexibility characteristics. The design and elements of the framework can be seen in Figure 3.1, which is the basis for the following descriptions.

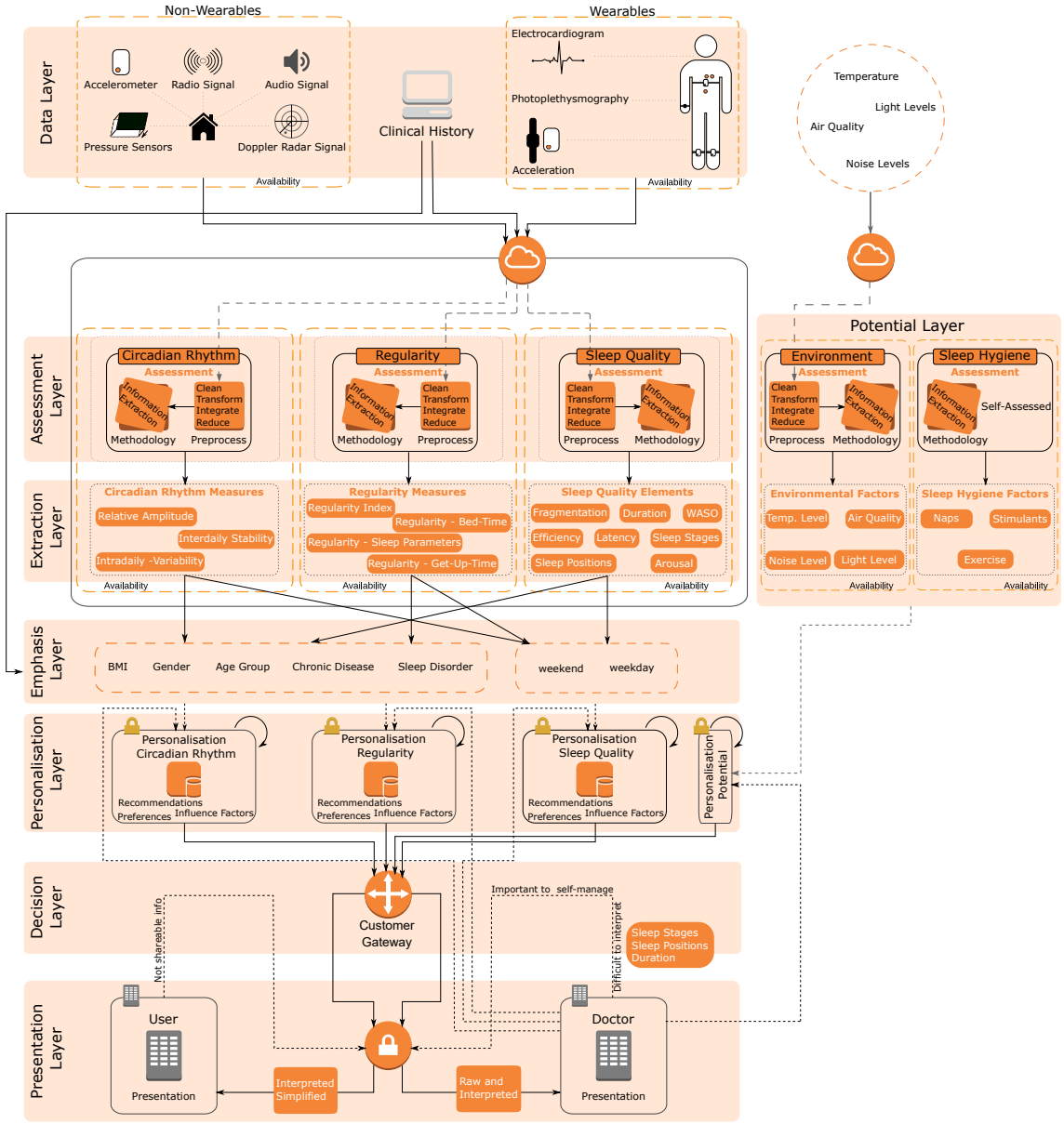


Figure 3.1: Sleep behaviour assessment framework.

3.2.1 The Layered Structure

The framework is built on layers: data, assessment, extraction, emphasis, personalisation, potential, decision, and presentation. These are connected and interact with each other. These layers serve as building blocks for a standardised sleep assessment process. Some of them—emphasis, personalisation, and potential—can even be removed if necessary for particular circumstances; this will be discussed more in Section 3.2.3.

(a) Data Layer

The data layer gathers information from wearable and non-wearable devices, which is then analysed using the appropriate methodologies. Specific IoT sensors must be chosen based on accuracy trade-off (balance between accuracy and other factors, such as user comfort and monitoring period) and parameters of interest. Furthermore, a full clinical history assessment is included to incorporate individual influence factors on sleep, such as BMI, age, gender, diagnosed sleep disorders, and chronic diseases. Environmental data include temperature, air quality, light levels, and noise levels. Sleep hygiene usually relies on self-reporting, which should ideally be gathered through a digital diary.

(b) Assessment and Extraction Layers

The assessment and extraction layers are strongly connected; the assessment layer covers pre-processing (i.e., preparing the data for information extraction) while the extraction layer deals with the extraction of relevant elements for specific components. The information extraction can be realised through a combination of machine learning, equation-based extraction, and statistical methods. The method used is based on the parameters of interest (e.g., machine learning for sleep stages and equation-based extraction for SOL), which are obtained in the extraction layer; details are provided in Section 3.2.2.

(c) Potential Layer

The potential layer provides information on aspects that can be interpreted and actively changed by users. Based on individual environmental and behavioural (sleep

hygiene) aspects, users can aim for lifestyle changes and doctors can provide recommendations. Component details are outlined in Section 3.2.2.

(d) Emphasis Layer

As discussed in Chapter 2, various factors affect sleep and should be integrated into a sleep behaviour assessment framework, such as clinical history and the current day (weekend/weekday). The emphasis layer highlights parameters that are particularly important to certain cohorts. Specific cohorts, such as individuals affected by sleep disorders or chronic diseases, need a framework that can adapt to their needs. For example, sleep apnoea worsens when lying on the back [19] and BMI groups vary significantly for specific parameters, such as SOL and SE (see Section 3.5). This layer allows for insights that can help doctors judge sleep parameters based on medical knowledge and provide outcomes that account for influence factors. We propose to incorporate information about gender, BMI, age, chronic diseases, sleep disorders, and the current day.

(e) Personalisation Layer

The personalisation layer considers the different extracted and emphasised parameters and the potential to improve sleep. These parameters are fused with doctors' recommendations and personal influence factors to present personalised feedback [45]. This layer visualises trends over long periods while considering the individual-specific impacts that sleep parameters can have.

(f) Decision Layer

This decision-making support mechanism transports information based on individual needs and preferences, and encourages active interaction between users and doctors. Here, presentation decisions are made based on users' communication choices (e.g., the restriction on sharing information with doctors) and doctors' recommendations (geared towards personalised self-management and parameters challenging to interpret). The preferences and recommendations are then fused with the extracted personalised parameters.

(g) Presentation Layer

The presentation layer presents sleep behaviour parameters to the users. For the standard user, an interpreted and simplified presentation is suggested. Certain parameters, such as sleep stages, positions, and duration, will only be monitored by their doctors, as they are difficult to interpret. Doctors should be able to access raw measurements and interpreted parameters to provide new recommendations for presentation and personalisation. The user can overrule any of these presentation recommendations by restricting the information that is shared.

3.2.2 Sleep Assessment Components

The framework focuses on four sleep assessment components: circadian rhythm, regularity, sleep quality, and potential from the environment and sleep hygiene. IoT sensors enable continuous home-based monitoring, meaning influence factors can be measured and set into context with sleep behaviour. Hence, it is essential to include specific elements for each sleep assessment component; these elements are described in the following paragraphs.

(a) Sleep Quality

The main purpose of assessing sleep quality is to help patients self-manage their sleep and, in turn, their health. To provide adequate feedback, connections and correlations must be measured [43]. Sleep quality elements concentrate on the timing of sleep, fragmentation, and movement, among other parameters, which are monitored using various devices, including actigraph units, pressure sensors, and video cameras [6,19]. Other parameters considered include SE, WASO, SOL, arousal, and number of awakenings [5, 49, 105] (see Table 2.3), as well as more complex parameters like sleep stages [43] and sleep positions [179]. These elements are relevant in determining a patient's sleep status, and each parameter can be integral to the diagnosis of a certain disease [6].

Data on sleep stages are mainly relevant for doctors to help detect anomalies and diagnose diseases; for users, the data are difficult to interpret and largely irrelevant [45]. Sleep stage assessment at home is generally based on HRV [46] or actigraphy data [19], but these both continue to lack in comparison to the gold standard of PSG. Therefore, we present a fine-grained sleep-wake detection approach in Chapter 5.

Sleep positions affect sleep quality [56] and are associated with sleep apnoea [19] and pressure ulcers [39]. Additionally, as sleep positions have been found to be unrelated to sleep stages [52]; new and distinct insight can be obtained from measuring them. Basic sleep positions are monitored by sleep trackers; in Chapter 4, however, we propose a standardised means of considering common, higher-granularity positions.

The combination of all of these elements is likely to result in a more reliable system that can provide personalised insight into a person's sleep behaviour.

(b) Circadian Rhythm

Our proposed framework considers three main circadian rhythm measures: relative amplitude, intra-stability, and intra-variability. For details, see Section 2.3.4.

(c) The Potential for Improvement

Users need a possibility to act on low assessment values, which the potential for improvement facilitates. This potential includes both environmental and sleep hygiene potential.

As stated in Chapter 2, environmental factors affect sleep; therefore, altering the sleep environment has the potential to improve sleep quality. Smart devices are able to monitor temperature, noise levels, light levels, and air quality. As individuals are unique in their responses to environmental factors, our framework considers this individuality.

Sleep hygiene factors should be assessed using a digital diary. We propose keeping subjective factors to a minimum and only using them when necessary due to the many drawbacks of subjective measurements. We decided on the most important factors: naps, substance use (primarily caffeine and alcohol), and exercise.

(d) Sleep Regularity

The regularity of sleep includes individual regularity measures, such as standard rise times and bedtimes, and overall regularity. Overall regularity is assessed using the sleep regularity index [130]; individual regularity measures must be assessed individually.

We recommend considering the regularity of sleep parameters and sleep timing, such as fragmentation, duration, SE, SOL, bedtime, and rise time. Regularity measures are extracted by summarising the changes between days over long periods of time (and can be adapted based on the available amount of data). To represent the regularity of specific sleep parameters, we propose the following sleep regularity algorithm applied to each sleep parameter. It involves the subtraction of each value x per day from a set of parameter values m (see Equation (3.1)). After subtracting, the norm is applied ($||: \mathbb{R} \to \mathbb{R}^+$), which results in $(n^2 - n)/2$ values represented as a vector s. The standard deviation (SD) is applied to the vector s, which results in the sleep regularity r (see Equation (3.2)). We propose that this approach be applied to each sleep parameter.

$$
\begin{aligned}
&m = (x_1, x_2, \ldots, x_n) \\
&s_k = |m_i - m_j| \qquad\qquad i = (0, n],\ j = (0, i] \\
&\qquad\qquad\qquad\qquad k = (0, \frac{n^2 - n}{2}] \\
&s = (s_1, \ldots, s_{\frac{n^2-n}{2}})
\end{aligned}
\tag{3.1}
$$

$$
r = \sigma_{|s|} = \sqrt{\frac{\sum_{k=0}^{n}(s_k - \overline{s})^2}{n-1}} \tag{3.2}
$$

The lower the resultant values, the more regular the sleep parameters. We include an option to incorporate weekend and weekday differences, as weekends usually include sleep compensation for the weekdays. Therefore, the values in m which refer to weekends can be excluded, if deemed necessary. The regularity assessment is flexible to the number of days available, as shown later.

(e) Personalisation

Influence factors include BMI, gender, age, medical conditions, and days of the week (weekday/weekend). A personalised approach fosters an appropriate integration of these factors; our framework's emphasis layer targets the integration of medical status and assessment schedule (weekday/weekend).

It is essential to emphasise specific parameters based on individual characteristics, as they significantly differ (see Section 3.5). In the emphasis layer, we incorporate

Table 3.1: μ-function: Thresholds for age groups.

Category	Age	SOL(min)				
		0–15	16–30	31–45	46–60	61+
Teens	14–17					
Young Adults	18–25					
Adults	26–64					
Older Adults	≥ 65					

Category	Age	WASO (min)						
		≤ 10	11–20	21–30	31–40	41–50	51–60	60+
Teens	14–17							
Young Adults	18–25							
Adults	26–64							
Older Adults	≥ 65							

Category	Age	SE (%)					Sleep Duration (h)				
		≥ 95	85–94	75-84	65–74	≤ 64					
Teens	14–17						<7	7	8–10	11	>11
Young Adults	18–25						<6	6	7–9	10–11	>11
Adults	26–64						<6	6	7–9	10	>10
Older Adults	≥ 65						<5	5–6	7–8	9	>9

this aspect. Listed characteristics of a user should be controlled and kept up to date, as a person could develop a medical condition. We investigate this aspect in Chapter 7, where we develop a deep-learning approach that continuously monitors the current status of the user—if a significant change is detected, a doctor's visit is recommended and the personalised parameters can be adjusted accordingly. People tend to catch up the lost sleep time during the week. Therefore, weekdays and weekends should consciously be separated, as they cannot be considered equal. The focus lies on weekdays, as a regular sleep habit should be preserved.

Influence factors affect recommendations on important things to monitor for. For an asthma patient, sleep position on one's back are likely more relevant, than time spent in bed; for insomnia patients the opposite would likely be true. Medical recommendations will result in changes to how information is presented to the user. These personalised recommendations can come directly from doctors or from the system itself, which will be prepared with basic recommendations for several scenarios.

(f) Presentation

We suggest that presentation vary based on the target group: individuals or their doctors. We propose a visualisation concept providing simplified and interpreted presentation for users and an in-depth presentation for doctors.

User's Presentation This visualisation consists of three modules: sleep quality, sleep regularity, and circadian rhythm. The visualisation is designed in three layers, which each provide a different level of information (see Figure 3.3). The unprocessed parameters are personalised through doctors' recommendations, indi-

vidual characteristics, and users' preferences. The visualisation of the parameters follows a rating matrix in a five-level colour-based structure. For age groups different thresholds were determined by Ohayon et al. [147] and Hirshkowitz et al. [194] (see Table 3.1), and applied to the sleep parameters (see Equation (3.3)). The matrix is coloured from green to red and correspondingly reaches from appropriate levels to inappropriate levels (refer to Figure 3.2). For sleep regularity, no thresholds are available; therefore, we suggest extracting thresholds empirically on a 5-level basis dependent on the data investigated, and applying them similarly to Equation (3.3) (see Equations (3.4) and (3.5)). The individual ratings are combined using the proposed Equation (3.6), to step up one layer of visualisation (i.e., to present a more abstract visualisation); inappropriate values are weighted higher (function f) while weekends are weighted lower (function m).

$$vis_1^x = \mu(r^x) \quad r^x \in \{\text{Sleep Quality Elements}\} \tag{3.3}$$

$$vis_1^y = \nu(r^y) \quad r^y \in \{\text{Circadian Rhythm Measures}\} \tag{3.4}$$

$$vis_1^z = \lambda(r^z) \quad r^z \in \{\text{Sleep Regularity Measures}\} \tag{3.5}$$

$$vis_k = \frac{1}{n} \sum_{i \in [1,n]} m(f(vis_{k-1}^i)) \quad k = 2, 3 \tag{3.6}$$

$$f(x) = \begin{cases} x & \text{if } x \in [0, 3] \\ 2\ x & \text{if } x \in (3, 5] \end{cases} \tag{3.7}$$

$$m(x) = \begin{cases} x & \text{if day is weekday} \\ \frac{3}{5}\ x & \text{if day is weekend} \end{cases} \tag{3.8}$$

Figure 3.2: Colour scale based on threshold levels.

Doctor's Presentation The visualisation concept for doctors focuses on the aspects relevant for individual patients (see Figure 3.4). Parameters that are important for diagnoses and interpretations are presented most clearly to doctors. For example, if someone suffers from sleep apnoea, sleep positions are relevant, so the presentation will be altered to reflect that, i.e., the presented circles in Figure 3.4

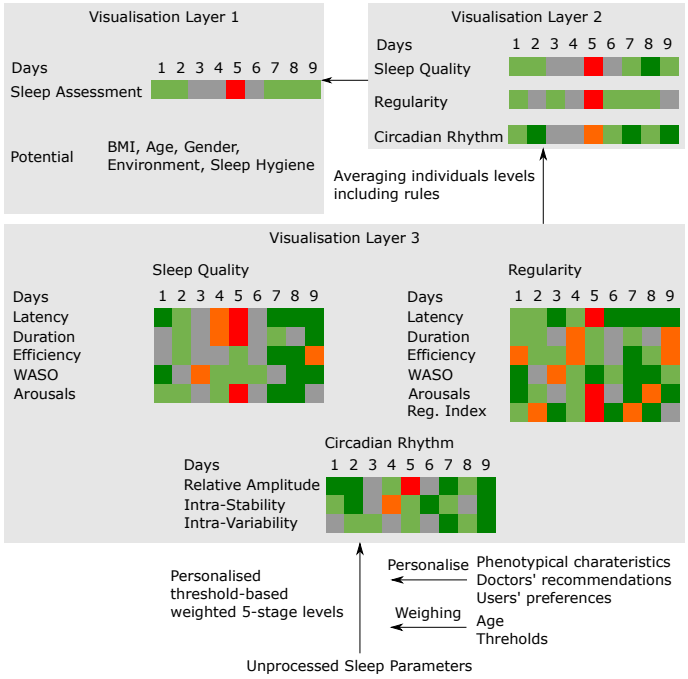


Figure 3.3: Visualisation concept for users.

will be bigger. The information displayed to doctors is often that which would lead to misinterpretation among users, such as sleep duration and sleep stages. Furthermore, the raw parameter values can be accessed directly, though availability depends on user privacy settings.

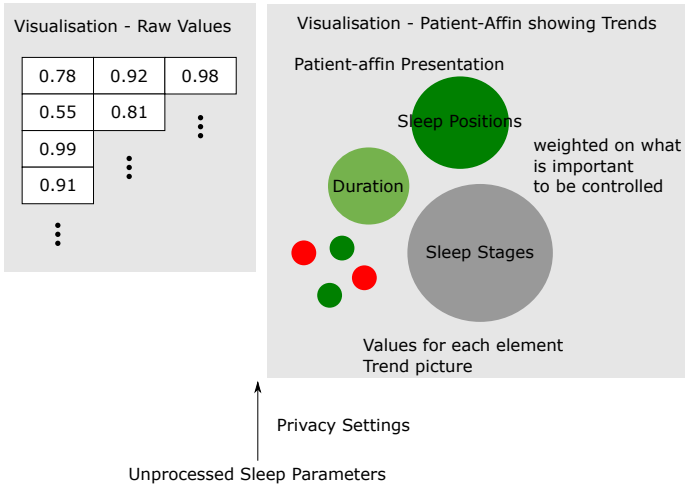


Figure 3.4: Visualisation concept for doctors.

3.2.3 Modular Design and Flexibility Characteristics

Our framework has unique characteristics and a design that encourages self-assessment and flexibility.

Self-Assessment Self-assessment can lead to the misinterpretation of certain parameters [45]. To change this, doctors offer recommendations on what the users should pay attention to. The focus on long-term monitoring provides users with insight into changes and regularity instead of isolated daily measurements.

Flexibility to Available Sensors The framework is fully adaptable to varying availability of sensors, whether it is restricted by the user or by accuracy considerations (certain sensors are more accurate for specific parameters [6, 19]). Different sensors can be used in the data layer, as long as the data structure is compatible, the information can be fused and used in the other layers.

Flexibility Towards Customers Doctors and self-managing users have different needs. Doctors need a comprehensive picture of the available information to arrive at conclusions, interpretations, and potential diagnoses. Users need pre-interpreted outcomes to avoid misinterpretation and promote effective self-management, e.g., by providing information on behaviour which can be changed actively.

Modular Design, Preferences, and Recommendations The fusion of users' preferences and professional knowledge makes our system reliable and understandable. Knowledge is integrated by advising users on what information should be important to them. Doctors' recommendations help emphasise the parameters that are relevant to the users and their goals, as defined by their established preferences. Personalisation is a key factor here, as sleep parameters affect individuals differently. The framework has a modular design, i.e., variable layers that adapt to individuals. The potential, emphasis, and personalisation layers can be excluded entirely, if necessary (e.g., if no influence factors are present or a general population approach is preferred). The potential layer is automatically dismantled if no environmental devices or digital diary entries are present. The exclusion of these layers has no negative impact on the functioning of the overall framework. The benefit of a modular design is that (1) computational solutions can be developed independently from the framework and later incorporated; (2) the framework is customisable; (3) the

existing blocks can be used by researchers which leads to greater efficiency as they do not need to invent knew structures; and (4) consistency is assured.

3.3 Sleep Assessment Experiments in Smart Homes

Due to advances in technology, homes are getting smarter and sensors are becoming better at monitor sleep behaviour in a natural environment. However, integrating these new technologies and data sources into existing environments remains a challenge.

Combining ambient, wearable, and binary sensors for multi-purpose analyses is of significant interest. Such a combination can improve a home's ability to facilitate independent living for individuals purposes, e.g., self-management of a medical condition. This can improve a system in private households providing independent living for individual purposes, e.g., self-management of a medical condition. To be adaptable to new rapidly growing technologies [195], technological infrastructure in users' homes must be highly flexible to various devices and applicable to all types of homes. We performed a study presented in the next sections to investigate current challenges.

3.3.1 Sleep Assessment Environment Design

To provide an effective infrastructure for multi-purpose analysis seeking to, for example, monitor sleep, fluid intake, and daily activities, two sources of data need to be fused: (1) embedded sensors in the existing environment and (2) ambient and wearable sensors. This kind of arrangement is shown in figures 3.5 and 3.6. This example was actually an experiment conducted for two weeks by researchers in a real-world smart home. The environment at Great Northern Haven has an infrastructure of 16 smart homes equipped with ambient sensors and actuators. Older adults have permanently occupied fifteen of the apartments since 2010; the remaining apartment continues to be used for research and demonstration purposes [196].

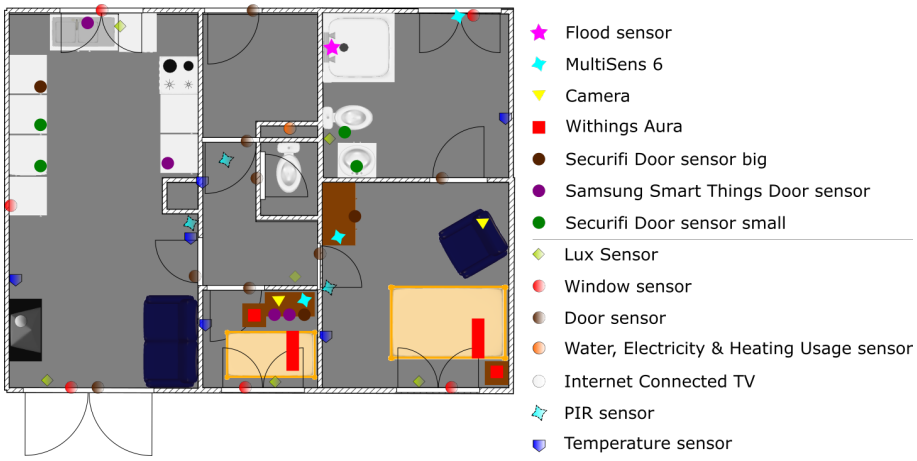


Figure 3.5: Floor plan of DKIT smart environment: permanent (semi-filled, framed) and additional (filled) smart sensors.

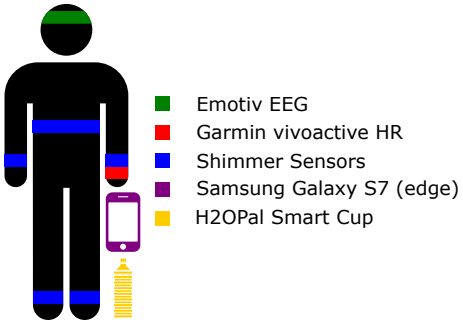


Figure 3.6: Used wearable devices and their body location.

(a) Use Cases of Interest

The challenges in activity monitoring in smart homes by fusing additional sensors into the environment are explored for some use cases of interest. Sensor data are collected for different user scenarios, incluing monitoring sleep and daily activities which affect sleep.

Table 3.2: Use case: Sleep assessment.

Description	The use case describes sleep quality monitoring in a smart environment using environmental and wearable sensors.
Actors:	Adults [healthy (**HEA**), depressed suffering from insomnia (**DEP**), and affected by **CKD**], Smart home
Conditions:	Sleep behaviour: **HEA** normal **DEP** problem related to sleep deprivation: falling asleep (onset), staying asleep (maintenance), and waking up early **CKD** problem related to sleep disturbances: wake-ups, nocturia caused by medication and fluid consumption.
Action sequence:	1. Adjust required sensors for sleep: **HEA,DEP** Withings Aura, smartphone, camera **HEA,CKD** Shimmer3 sensors, smartwatch, Withings Aura, smartphone 2. **HEA** falls asleep **DEP,CKD** has difficulty falling asleep 3. **DEP** wakes up during the night and has difficulty falling asleep again **CKD** wakes up to go to the bathroom

Table 3.3: Use case: Napping.

Description	The use case is related to napping during the day.
Actors	Adult, Smart Home
Conditions:	Adult feels tired during the day
Action sequence:	1. Adult adjusts sensors including smartwatch, Shimmer3 sensors, Withings Aura, smartphone 2. Adult naps in bedroom

As sleep quality is an important factor of overall well-being [20], it is considered in our use cases for night-time (Table 3.2) and daytime (Table 3.3) sleep monitoring. Daytime sleep periods are somewhat controversial in the literature [134]; generally,

Table 3.4: Use case: Fluid consumption.

Description	The use case illustrates the fluid consumption with a Smart Cup.
Actors **Condition:** **Action sequence:**	Adult, Smart Home Person is thirsty 1. The person feels thirsty 2. The person goes to kitchen to fill the Smart Cup 3. The person drinks 4. Sensors in the Smart Cup measure quantity of fluid drunk

Table 3.5: Use case: Physical activity.

Description	The use case is related to the physical activity of an adult.
Actors **Conditions:** **Action sequence:**	Adult, Smart Home Adult is inactive 1. The person adjusts wearable device (smartwatch) 2. Physical activity inside the house is captured by the smart home environment and a smartwatch, while outside the house it is captured only by the smartwatch

Table 3.6: Use case: Bed time and morning routine.

Description	The use case describes bedtime and morning routine of an adult using environmental and wearable sensors.
Actors **Conditions:** **Action sequence:**	Adult, Smart Home Adult wants to go to bed or get up 1. The person adjusts required sensors on the body during bathroom routine (smartwatch, Shimmer3 sensors) 2. The environment measures presence in the bathroom: use of the sink, toilet, shower, and the hand acceleration from Shimmer3 sensors 3. Person starts the bathroom routine: comb hair, wash face, wash hands, brush teeth, use the toilet, and shower, if necessary

however, it's recommended to get no more than 30 minutes [197]. Relevant conditions are simulated: healthy adults without problems, depressed patients with insomnia, and patients with CKD (see Table 3.2). Furthermore, influences on sleep in relation to sleep hygiene and physical activity are considered.

Sleep hygiene is used as a treatment for insomnia [197, 198]; this includes:

light exposure measured via light sensors in the house and time spent outside
exercise measured with environmental sensors and smartwatch (Table 3.5)
fluid consumption (Smart Cup) including caffeine consumption (Table 3.4)
environment noise level, humidity, and temperature inside the room
nap duration (Table 3.3)
routine bedtime procedure measured with environmental and wearable sensors (Table 3.6)

(b) Sensor Deployment and Data Collection

The smart home at Great Northern Haven consists of various wired environmental sensors, including window, door, consumption (energy, water, heating), passive infrared sensor (PIR), Lux (illuminance), and temperature sensors (see Table 3.7).

Table 3.7: Existing smart environment sensors at Great Northern Haven.

Sensor type	Count	Location	Detection
Door	7	doors	open/close
Window	5	windows	open/close
Electricity, water, & heating usage	1	boiler	usage
PIR	3	ceiling	movement
Lux	6	ceiling	light
Temperature	6	walls	temperature

The smart home's utility was extended with wearable and environmental devices: door, movement, flood sensors, Shimmer3 sensors, and Withings Aura (see Table 3.8).

Two setups were designed for sleep assessment; the first involves a Withings Aura and a smartphone under the pillow, while the second expands upon the first with smartwatch and Shimmer3 sensors applied to the chest and extremities. The first setting collects data throughout the night, while the second setting expands upon the first by gathering additional data during 30–45 minute-long naps during the day.

Table 3.8: Additional sensors for the smart environment. R1, R2 stand for resident bedrooms, K for kitchen, and B for bathroom.

Sensor type	**Count**	**Location**	**Collected for**
Shimmer3 Sensors	5	extremities, chest	daily activities, sleep
Camera	2	bedroom (R1,R2)	annotation
Smartwatch	1	wrist (R1)	sleep, active
Withings Aura	2	bedroom (R1,R2)	sleep
Samsung Galaxy	2	bed (R1,R2)	sleep
iPhone	2	smart home	annotation
H2OPal Smart Cup	2	smart home	fluid intake
EEG Emotiv	1	head (R1)	daily activities
Securifi Door large	3	cupboard (K), wardrobe (R1,R2)	open/close
Securifi Door small	4	fridge (K), cupboard (K), toilet (B), tab (B)	open/close
Samsung Smart Things Door Sensors	4	tab (K), cupboard (K), wardrobe (R1)	open/close
MultiSensor6	3	bedrooms (R1,R2), bathroom	movement, temperature, light, humidity, vibration
Flood Sensor	1	bathroom	water

The Shimmer3 sensors collect accelerometer, gyroscope, magnetometer, pressure, and temperature data. The Garmin vivoactive HR smartwatch gathers data about sleep stages, heart rate, and movement. The Withings Aura uses an air pillow under the mattress to calculate sleep stages, heart rate, temperature, luminosity, and noise. The integrated smartphone sensors collect data at 0.1- to 1-second intervals. Fluid consumption is measured with the H2O Pal Smart Cup as millilitres drunk throughout each day. Additionally, binary sensors in the flat continuously record various daily activities (refer to Section 3.3.1(a)).

3.3.2 Experiment Design Considerations and Lessons

Here we present challenges that must be faced when devices are added and discuss how they are addressed throughout the process.

Hardware Setup We established an ideal setting that integrated binary and wearable sensors. At the start of the experiment, we encountered difficulty applying and pairing some of the additional movement and door sensors, so they had to be excluded in further investigations. Some sensors were not easy to install, as we sought

to use them for purposes they were not designed for. For example, door sensors were used as tap sensors to gather data about temperature and water flow; these were difficult to apply in the bathroom and kitchen.

Connection Difficulties We faced failures of Bluetooth connection and synchronisation between wearable devices and smartphones throughout the study. Other difficulties included internet connection problems with some devices, including the Aura Withings and the SmartCam, which was used to collect the ground truth.

Malfunctioning Device software issues presented another challenge; some devices stopped recording data prematurely or failed to connect properly. These difficulties led to a lack of data during the experiment, which have to be met by an algorithmic methodology that can handle incomplete and noisy datasets.

Data Extraction Data collection was occasionally interrupted by problems with data transmission and raw data extraction from third-party device providers. Some providers did not provide access to the raw data from their devices. The output from some other devices is provided in a convoluted manner that has to be studied in detail to understand.

Data Cleaning Data pre-processing includes data cleaning to remove erroneous and unwanted values, data interpolation to deal with potential missing values, and data transformation to correctly format the data.

As the collected data are from various types of sensors in a natural environment, it is possible for noisy and redundant information to be generated (e.g., by dead batteries, failures in sensor readings, losses in wireless transmission, and other unforeseen events that occur in real homes). To perform data cleaning, specific filters must be applied based on sensor characteristics and the number of people present. For example, Bayes and particle filters clean data from binary sensors in scenarios with multiple residents; median filters remove noise and avoid abnormal measurements. The noisy data collected by Shimmer3 devices can be reduced by signal processing; for the accelerometer, low-pass filters remove the gravity component from the values and high-pass filters reveal the real acceleration.

Data Processing Data from different devices are fused and related to activities within the smart environment using timestamps from sensors and annotations from

cameras and the smartphone. The smartphone contains a customised application that allows for the annotation of daily activities and their duration. Camera data are used to obtain a ground truth for sleep movement. Smartphone annotation is mainly used for daily activities in the bathroom and bedroom (refer to Section 3.3.1(a)).

Feature selection is performed for Shimmer3 sensors to extract relevant features for sleep positions (see Chapter 4). Collection frequency among sensors varies. For instance, Shimmer3 sensors on the extremities collect data at a frequency of 50Hz while on the chest they do so at one of 520Hz; smartphone sensors gather data at a frequency of 1–10Hz. Therefore, we suggest to either downsample or process them independently and fuse the outcomes. For example, a high frequency is necessary when a Shimmer3 chest sensor is used for respiration detection; when used for sleep position detection, however, the lowest frequency suffices.

For sleep assessment, different sensors must be used and combined. Physical exercise affects sleep positively [134]; this can be monitored by a smartwatch counting steps and heart rate both outside and inside the home; the activeness is calculated based on the movement. Fluid intake can be analysed using Smart Cup data. Subsequent changes in sleep behaviour can be considered through the correlation of outputs from these single-data sensors. We face some difficulty in matching timestamps and data sourced from devices at different frequencies. The annotated information must be considered carefully, as the annotation time of some activities (ground truth) start earlier than the corresponding data collection time. The formatting of the various extracted datasets from providers must be adjusted to enable integration and comparability.

Occupancy The occupancy of the smart home in terms of the number of persons varied throughout the study. The main challenge during data processing is the extraction of behaviour patterns considering the number of inhabitants living in the smart home. Additionally, visits to the smart home from guests and external researchers took place on numerous occasions, and these visits must be considered during data processing because occupancy dramatically affects data collection from environmental devices.

3.4 Sleep Behaviour Analysis on Specific Framework Aspects

In this section, we present the datasets and investigation settings used for our analyses to investigate (1) the necessity and importance of specific layers and elements of the framework structure, (2) the focus on objective measurements, and (3) the reliability of the developed sleep regularity algorithm.

3.4.1 Dataset

To validate the individual aspects, a real-world dataset was explored. The NSRR provided us with the HCHS [199, 200] and multi-ethnic study of atherosclerosis (MESA) datasets [201], which contain data collected at home on PSG, actigraphy, individual characteristics, and sleep questionnaire information. The participants were instructed to wear a wrist-worn actigraph unit, the Philips Respironics Actiwatch Spectrum, for a week. The data were collected in an unsupervised setting [201]. Records were scored by a trained technician at the Boston Sleep Reading Center [202] using current guidelines and specific algorithms to derive elements like wake/sleep patterns and activity levels. Actigraphy data are available in 30-second intervals. The datasets provide features that are necessary for our investigation, such as ethnic background, gender, and health status.

Additionally, we simulate data to show the functionality of the regularity approach.

HCHS The HCHS [188] is a multi-centre epidemiologic study investigating acculturation in disease development and risk factors. Actigraphy data were collected from 2.252 participants between 2010 and 2013. From those, data from 1,887 are publicly available. Individuals suffering from various diseases, including hypertension, apnoea, diabetes, and CKD, participated in the study. Provided clinical history includes individual medical information, such as BMI and family history.

MESA The initial objective of the MESA study was to examine disparity of sleep disorders and behaviour across ethnic groups and genders as well as determine the association between sleep and sub-clinical atherosclerosis [201, 203]. 2,237 individuals participated between 2010 and 2013; data are available from 2,159 individuals.

This dataset is quite diverse, as it includes data from individuals of different races, genders, age groups, and degrees of health [201].

3.4.2 Sleep Behaviour Analysis Protocols

It is essential to provide evidence of the emphasis layer's importance and the need to exclude certain sleep parameters in user-specific presentation. The investigation is based on two elements: (1) one-way ANOVA to extract significant differences ($\alpha = 0.05$) between sleep parameters among groups of people distinguished by genetic information and health status considering weekend and weekday influences; (2) pairwise Pearson correlation between sleep parameters in order to either include or exclude them from the assessment, where we consider a correlation of 0.80 the threshold for exclusion. Validation is performed using the HCHS dataset; data fulfilled the assumptions for both tests. Investigation (1) establishes the understanding of what influence factors are relevant for the individual sleep parameters incorporating the weekday information. Based on this, the need for the emphasis layer will become clearer, and influence factors can be revealed that should be considered to achieve personalised parameters. Investigation (2) facilitates which parameters can be excluded to guarantee a resource-efficient assessment. We use the actigraphy data from HCHS to extract the following sleep parameters and circadian rhythm measures: sleep fragmentation, SE, rise time, bedtime, relative amplitude, intradaily variability, and interdaily stability.

As already stated, the motivation for this framework is to objectively represent sleep behaviour. To justify this choice, we investigated the relationships between sleep parameters from objective measurements and subjective ratings, which are provided in the HCHS dataset [199, 200]. We tested two machine learning models, an LSTM and a GMLVQ, for the prediction of subjective sleep quality and used the sleep parameters mentioned in the preceding paragraph as features. These features are calculated from the available days for each individual, with highly correlated features excluded. The dataset contains five sleep quality classes which are provided by a subjective question-based assessment.

The proposed sleep regularity algorithm presented in Section 3.2.2(d) is designed to provide a regularity measure for each sleep parameter individually. To show the impact of our regularity algorithm, we investigate a simulated data scenario. In this simulated scenario, the functioning of the algorithm is investigated in a four-day and seven-day assessment, to show the influences of weekend inclusion and exclusion.

Table 3.9: ANOVA tests: Significant sleep parameter differences among groups (e.g., apnoea groups differ based on fragmentation).

labels	all	WD*	WE*	all	WD	WE	all	WD	WE
	Gender			Apnoea			CKD		
SF*	✓	✓	✓	✓	✓	✓			
SE*	✓	✓	✓	✓	✓	✓			
SOL*	✓	✓	✓			✓	✓	✓	
TS*	✓	✓	✓				✓	✓	
RA*	✓	✓	✓	✓	✓	✓	✓	✓	
VAR*	✓	✓	✓	✓	✓	✓	✓	✓	
ST*	✓	✓	✓	✓	✓	✓	✓	✓	
	BMI			Insomnia			Hypertension		
SF				✓	✓	✓			
SE	✓	✓		✓	✓				
SOL	✓	✓		✓	✓		✓	✓	
TU*	✓			✓	✓	✓			
TS	✓	✓	✓						
RA	✓	✓		✓	✓		✓	✓	
VAR	✓	✓					✓	✓	
ST	✓	✓	✓	✓	✓		✓	✓	
	Age						Diabetes		
TU	✓	✓	✓				✓	✓	
TS	✓	✓	✓				✓	✓	
VAR	✓	✓	✓						

*WD-weekday; WE-weekend; RA-relative amplitude; SF-Fragmentation; SE-Efficiency; SOL-Sleep Onset Latency; TU-Rise time; TS-Bedime; VAR-Intradaily Variability; ST-Interdaily Stability

3.5 Results and Discussion

In this section, we present the results of the analysis and discuss the individual elements of the framework including the visualisation concept. Furthermore, the gaps, integration aspects, and limitations are provided.

3.5.1 Emphasising

Results indicate that it is reasonable to include an emphasis layer, as significant differences can be found between groups. The groups are divided into multiple categories for gender (female, male), apnoea (no apnoea, mild, moderate-severe apnoea),

CKD (no CKD, stage 1, stage 2), BMI (<25, <30, >30), insomnia (no clinically significant insomnia, subthreshold insomnia, moderate severity, severe), hypertension (no hypertension, hypertension), age (18–24, 25–34, 35–44, 45–54, 55–64), and diabetes (no diabetes, prediabetes, diabetes). As can be seen in Table 3.9, certain parameters are correlated while others are not. For example, SOL is significantly different for BMI groups on weekdays but not during weekends. These relations are integrated into the emphasis layer. The results provide preliminary evidence of the importance of emphasis as a means of representing individual influence factors.

The pairwise correlation shows that sleep fragmentation and SE are highly negatively correlated (-0.81). This suggests that only one of these values needs to be assessed and presented. One way to determine which one should be included is to use doctors' recommendations or users' preferences. Furthermore, all of the sleep parameters are significantly different with p-values smaller than 0.05 for age and genders. This is the case for combined data, weekend, and weekdays. Other parameters show differences depending on the day (e.g., BMI groups are affected by the weekday for SE, SOL, relative amplitude, and variability).

3.5.2 Sleep Parameters from Objective Measurements

The multi-class problem, even with optimisation techniques, like standardisation, did not show promising results. The GMLVQ results in an averaged 10-fold cross-validated accuracy of 35% for weekdays, 39% for weekends, and 32% overall for the multi-class problem; LSTM results in even lower accuracies.

These negative results in our data-driven approaches suggest that the subjective and objective parameters are not related or cannot properly represent each other. These outcomes can be interpreted in one of two ways: (1) subjective measures do not bring the optimal outcome and objective measures are preferable or (2) both are independent and bring new information into a sleep assessment. In this chapter, we followed the second interpretation, but following chapters focus on objective measurements.

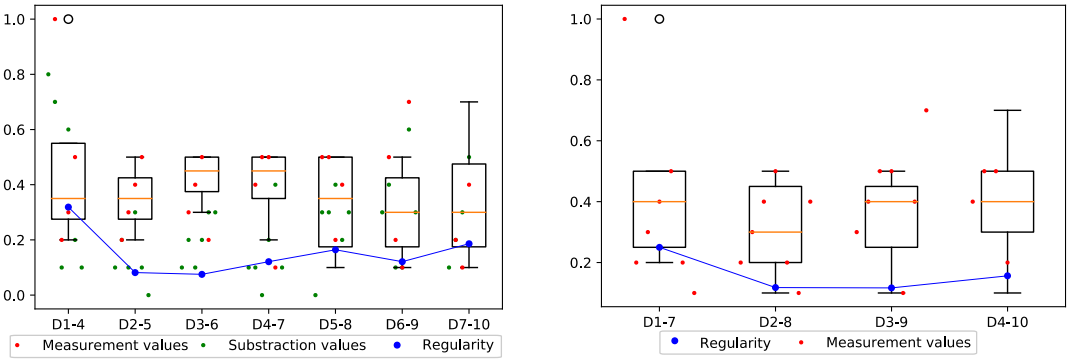
3.5.3 Sleep Quality and Personalisation

Reliable measurements are the basis for accurate home-based sleep assessment. The importance of adaptable approaches was mentioned before, and personalised approaches can represent this flexibility. We will present the outcomes for our pro-

posed approaches to sleep positions in Chapter 4, sleep stages in Chapter 5, and detection of medical conditions in Chapter 7.

3.5.4 Regularity

We investigated a specific routine for sleep regularity assessment. Therefore, a behaviour routine is simulated: [1, 0.2, 0.3, 0.4, 0.5, 0.5, 0.2, 0.1, 0.4, 0.7], describing the values of one sleep parameter for 10 consecutive days. We tested how regularity changed over these days, when using a window of four or seven days of values, as can be seen in Figure 3.7. The target could be either to include weekends or exclude them. The x-axis describes the days used; D1–4, for example, indicates the assessment of regularity for day 1, 2, 3, and 4. The plot displays our regularity approach by showing the measured values (red dots) and their distribution (boxplot), the subtraction values (green dots), and the regularity assessment (blue). The resultant changes in regularity are shown in the blue curve. We can see that the individual values of regularity are different for the four-day and seven-day assessments. With four-day assessments, we see two drops and increases in regularity; when seven days are considered, we lose the information of the second drop in the four days method (D6–9).



(a) (b)

Figure 3.7: Sleep regularity for simulated data using **(a)** four-day assessment **(b)** seven-day assessment.

These results suggest that days of the week influence sleep regularity and that, for various situations, weekend and weekday results may need to be separately represented, which is possible with our flexible approach. However, this flexibility comes with the drawback of sensitivity to individual values. Overall, our regularity approach can represent more information than the often-assessed regularity index,

which fuses all information. The advantage of our approach is that we can extract information about which parameters have strong or weak regularities. Many conclusions could be drawn from this information—strong regularity in high SOL, for example, could suggest insomnia.

3.5.5 Visualisation Concept

The proposed visualisation concept would likely motivate people to monitor and self-manage their current sleep habits and subsequently develop healthier behaviours to prevent and detect conditions at an early stage. The concept aims to avoid misinterpretation, which is currently a prominent issue.

The proposed concept makes users' data accessible to them. The different layers allow for varying levels of abstractions—data needs to be assessed more closely in a lower layer if one of the parameters is marked as inappropriate. The concept also enables doctors to closely monitor their patients from afar in an objective manner by fusing individual-specific sensor data to extract the essential information. This information-sharing process can help detect underlying conditions and prompt the necessary treatments earlier. These treatments can then be adapted to a personalised plan to increase efficiency.

3.5.6 Integration Aspects

There were various sleep assessment aspects pointed out in Chapter 2 that need to be investigated; some of them are addressed in our proposed framework. The proposed assessment considers sleep regularity and circadian rhythm, which are important sleep aspects that have largely been ignored by previous assessments. Our framework enable behaviour change approaches, because it integrates the environmental and sleep hygiene potential. The inclusion of IoT sensors and long-term monitoring targets underdiagnosis and aims to increase the accessibility of self-management strategies. The validation tasks described in Section 3.4.2 constitute a step towards closing the gap left by small, single-subject-type samples and providing results that better represent the general population. The fusing of various influence factors (age, gender, and environmental condition) is addressed in the emphasis and personalisation layers. The assessed elements largely come from objective measurements taken using technology; we only include subjective assessments to effectively incorporate clinical history and sleep hygiene. The combination of objective and subjective

methods can potentially (1) improve the accuracy of subjective tests with objective measurements and (2) bring in additional subjective information that is sometimes not sufficiently reflected by objective properties. Additionally, our objective sleep assessment avoids the common limitation of experts needing to be present by integrating methods into our IoT-enabled framework that minimise this need. The consolidation of doctors' and users' perspectives creates a more powerful approach, as it can adapt to different needs and take different preferences and recommendations into account. Our framework can also adapt to advances in technology and data analytics because it is built to be flexible regarding available data sources. The minimum requirement in the proposed investigation is an actigraph unit, which is inexpensive and can be used to extract sleep parameters.

3.5.7 Limitations

This study assumes that clinical history data is available or can be assessed with IoT technology. The availability of gender is a notably strong assumption. In the real world, however, not all clinical features are available, as they are either not shared by users or not assessed.

Furthermore, we assume that doctors make recommendations about what should be monitored for individual users. The potentially lack of these recommendations, is only partly overcome by preparing the system with basic recommendations for several scenarios.

Our use of personalisation can result in difficulty when comparing users. Our approach moves away from the one-model-fits-all approach but ends up moving towards a subgroup-based comparison. Therefore, it may not be possible to directly compare, for example, a male with sleep apnoea with a healthy female.

3.6 Summary and Future Directions

This chapter contributes to knowledge by developing a layered sleep behaviour assessment framework that supports IoT-enabled home-based monitoring; the key aspects considered are sleep quality, regularity, circadian rhythm, environmental conditions, and sleep hygiene. We have described the measurements and techniques used for monitoring and analysis; especially, the developed sleep regularity algorithm and user-dependent visualisation concept contribute to knowledge signifi-

cantly. The framework can handle data from multi-source channels and incorporate clinical history and preferences; this enables it to provide a personalised medical and behavioural assessment. The framework has been designed as a modular, service-oriented paradigm that facilitates extendibility and flexibility to both doctors' and users' needs. While large-scale implementation and evaluation are planned, initial analysis have shown positive results, proving the framework's viability. The underlying methods will be discussed in further detail in the subsequent chapters.

This framework can act as a straightforward guideline due to its block-based structure, ability to adapt to technological advances, and standardised process (from monitoring to presentation) incorporating current relevant sleep behaviour aspects. As our proposed framework is designed for home-based assessment, testing in a natural environment is critical for all incorporated technologies. The inclusion of doctors' recommendations means the framework can have tangible effects on users' lives. This makes it a useful tool, creating the baseline for future investigation and testing. The inclusion of IoT sensors and long-term monitoring procedures may very likely enhance sleep disorder diagnosis by increasing accessibility and allowing for faster detection at home. The integration of various influence factors is addressed in the emphasis and personalisation layer. This process could be further improved by developing concrete methods for analysing and investigating relations with sleep parameters. Our framework is, besides diagnostics and self-management, intended to provide adequate feedback through detected connections and correlations, e.g., between day behaviour and sleep architecture [43]. Future research could focus on ways to enhance feedback and assist users with improving their habits. Our work on subjective and objective assessment suggests that both are independent. This outcome will encourage researchers to investigate them in more detail.

In the following chapters, we concentrate on core aspects of the framework that address sleep position recognition, sleep stage classification, and chronic disease detection; the structure of our framework eases the integration of these technologies.

Chapter 4

Higher-Granular Sleep Position Detection

4.1 Introduction

The previous chapter has presented a framework designed to improve sleep behaviour assessment. The framework incorporates non-motion aspects, also known as stable states. Predominantly, sleep positions explain these non-motion periods. Sleep positions are a significant influence factor for sleep quality and, therefore, must be investigated. Current research mainly investigates four sleep positions in guided settings—for details, see Section 2.4.4—but, the variation of impacts extends to more complex positions (see Section 2.2). Hence, higher-granularity positions need to be considered. Current approaches face several difficulties: (1) non-wearables are unable to distinguish the supine position from the prone position and they are immobile as sensors need to be installed in the environment; (2) imaging techniques lead to privacy concerns and have difficulty with detection through blankets; and (3) wearables can cause potential discomfort. We decided to pursue wearable devices, as they provide mobility and the ability to gather accurate data with more than one person in the same bed. They can also be more easily integrated into existing smart devices.

In this chapter, we target objective 4 by investigating higher-granularity positions among the general population. In addition to the four basic positions, the proposed approach looks at four classes that consider common leg position variation. To do this, we target the following gaps extracted from literature (see Section 2.4.4): (1) lack of a robust and computing-efficient approach; (2) no approach to track higher-granularity positions; and (3) lack of comparative studies between natural environments and guided settings. To address issue (1), a robust and computing-efficient method is applied that has low computational costs and can reliably classify the positions of interest. The literature suggests that distance-based approaches perform

well for sleep position detection. Therefore, we decided on GMLVQ, a low-cost, distance-based method that supports real-time assessment, because it can make a decision based on prototypes representing each of the sleep positions. To investigate a computing-efficient system, we conducted experiments to explore the necessity of three sensors and the potential to reduce the number of features. To address issue (2), we extract and investigate the most common positions among the wider population [204]. The primary sensor is attached to the chest, as this has been shown to be effective [57]. To further investigate the common positions, the number of sensors is increased to three in order to monitor relevant lower limb movements. This results in data on leg and torso positions, enabling insight into higher-granularity positions. To address issue (3), we compare performance on datasets from both guided and natural settings to assess the feasibility of comparative analysis using mobile solutions. Employing accurate sleep position classification under computing-efficient constraints implies a more reliable and efficient sleep assessment process within our framework. Higher-granularity position detection is a solution that offers more comprehensive information about sleep behaviour. This approach contributes to the framework's sleep quality component as well as the following layers: data, assessment, and extraction.

This chapter is structured in the following way. In Section 4.2, we introduce our approach for sleep position recognition with a focus on a low-computational-cost model. We are especially interested in our approach's performance differences in real-world and guided settings. Section 4.3 describes how we collected data and ran our experiment while investigating the differences between one-model-fits-all and personalised approaches. The results and findings are provided in Section 4.4 and discussed in Section 4.5. Finally, in Section 4.6, we offer a summary of the work, its implications for future studies, and indicate the focus of the next chapter.

4.2 Sleep Position Classification

In this section, we describe our proposed approach for recognising complex sleep positions.

As already stated, we opted for accelerometer-based detection and used a combination of sensors placed on the chest and legs to cover positions of interest. The approach can be divided into three main steps. These steps are shown in Figure 4.1: (1) the stable state (non-movement) detection, followed by (2) the feature calculation, and (3) a classification algorithm that considers the features and decides on

the current sleep position.

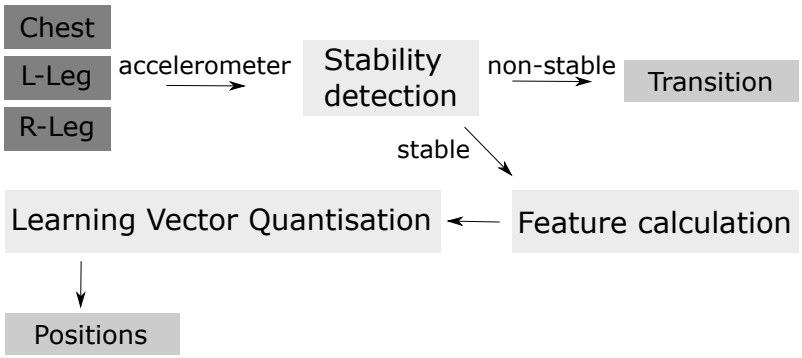


Figure 4.1: Sleep position detection for accelerometers applied on the chest, left leg (L-Leg), and right leg (R-Leg).

Sleep positions appear in the absence of movement; therefore, movement periods, known as transitions, must be excluded from the analysis. To distinguish between stable and non-stable periods, the accelerometer data must be divided. The data collected from chest and leg sensors are classified using the acceleration-moving variance method [205] as either stable or non-stable. This step follows Equation (4.1), where α is empirically set to 0.5.

$$T(\mathbf{z}) = \frac{1}{N}\sum_{t=1}^{N} \|z_t - \overline{z}\|^2 < \alpha \tag{4.1}$$

While other methodologies can detect stable states, the acceleration-moving variance method requires no additional data, such as gyroscope and magnetometer data.

When the stable states are extracted, the features can be calculated. The information on the stable periods is compressed by calculating the mean for each of the three accelerometer axes over a sliding window, resulting in a nine-dimensional data vector. The sliding-window concept is depicted in Figure 4.2. A window with size w slides over each axis with datapoints $(d_1, \ldots, d_n)$. For each iteration, such as i_1, the mean is calculated over the data within this window $(d_1, \ldots, d_w)$. Afterwards, the window slides further with an overlap (of, e.g., 50%) and, for the next window $(d_{\frac{w}{2}}, \ldots, d_{\frac{3}{2}w})$, the mean is calculated. These steps continue until the end of the data is reached.

The axes are further referred to as X, Y, and Z. The features are described as

$$\mathrm{X_{Left}}, \mathrm{Y_{Left}}, \mathrm{Z_{Left}}, \mathrm{X_{Right}}, \mathrm{Y_{Right}}, \mathrm{Z_{Right}}, \mathrm{X_{Chest}}, \mathrm{Y_{Chest}}, \mathrm{Z_{Chest}},$$

where Left is for the sensor applied to the left leg, Right is for the sensor on the

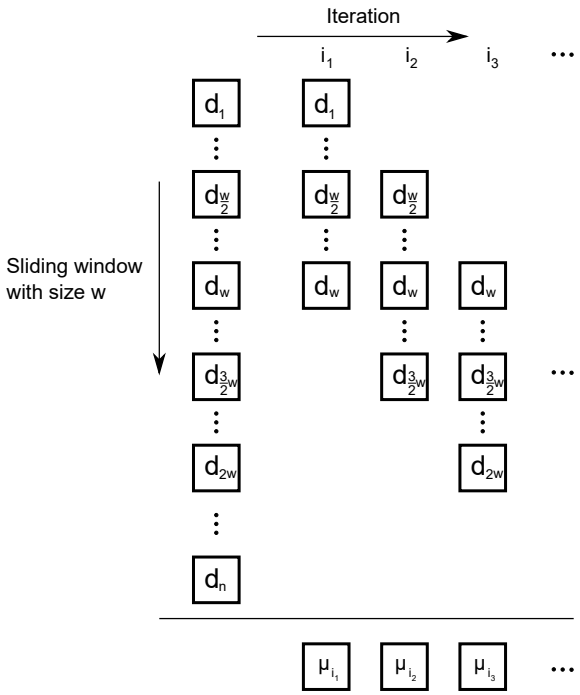


Figure 4.2: Sliding window concept; calculating mean μ for each iteration over one sliding window with an overlap.

right leg, and Chest is for the sensor attached to the chest.

The last aspect of the approach is the classification using GMLVQ.

4.2.1 Classification Algorithm

Environments with limited computational resources require a fast and appropriate method. Therefore, we selected GMLVQ, because it is resource-efficient and has never been applied to this specific problem before. GMLVQ is a distance-based classifier similar to the KNN algorithm. However, it only computes the distance between the model's prototypes and a novel data pattern. Consequently, GMLVQ scales better when new data is introduced, as the number of prototypes does not increase. This makes GMLVQ a fast method which is appropriate for environments with limited computational resources.

In this section, the GMLVQ is described in detail by summarising its original introduction from Schneider et al. [206]. GMLVQ is an adaptive, distance-based supervised-learning method. The basis of the model is the collection of prototypes that are found during the training phase and describe typical representations for each class. We constructed a relevance matrix (Λ) that transforms the data in a way that

makes classes more distinguishable. The distance measure follows Equation (4.2).

$$d^{\Lambda}(\vec{\omega}, \vec{\xi}) = (\vec{\xi} - \vec{\omega})^T \Lambda (\vec{\xi} - \vec{\omega}) \tag{4.2}$$

A prototype $\vec{\omega}$ and an n-dimensional feature vector $\vec{\xi}$ are considered; Λ describes an $n \times n$ matrix. This matrix takes into consideration the correlations between features and rotation of the axes. Every pair of features is weighed, which results in a more robust performance [206, 207]. Overall, GMLVQ provides an intuitive model that makes the data easier to understand. It enables us to directly interpret prototypes, as they now exist in the same space. Furthermore, the diagonal values of Λ can be interpreted as the relevance of a single feature, whereas the off-diagonal values can be interpreted as the relevance of every combination of two features [206].

4.3 Experiment Design

The application of our proposed method is critical to provide insight into its performance. As already stated, it is important to consider data collected in a person's natural environment. As the scope is rather unique, we needed to gather our own data to investigate the problem of interest. This investigation is structured in the form of different experiments in normal and computing-efficient settings. These experiments explore individual and generalised models. Moreover, the effects of reduced sampling rate and features are explored.

4.3.1 Data Collection

The main goal of this data collection is to investigate higher-granularity sleep positions than are typically considered. Information on ten different sleep positions is considered: supine, prone, left lateral, and right lateral positions, with additional positions being classified based on leg positioning—eight positions are distinguished in the analysis process, as the difference between the others cannot be captured, resulting in merging four of the positions into two. Data were gathered from three sensors attached to chest and both ankles. The dataset includes data from real-world and guided settings. Ethical approval was received by De Montfort University and all participants permitted us to collect and use their data by signing a consent form.

(a) Planning and Sampling

The research was conducted in the home environment of the individual participants to reach the most natural circumstances.

Healthy participants from both genders and two age groups (28–33 years and 55–58) were recruited to guarantee that the results can apply to a wide population. However, gender differences and age-specific performance are not major factors in this analysis. For one person from each age group and gender, real-world data were collected.

Six individuals participated, four between 25 and 33 with three males and one female and two between 55 and 58 with one male and one female. For the real-world setting, two individuals were representatively selected—one per age group and gender was selected (for a detailed overview of the sample distribution, refer to Table 4.1).

Table 4.1: Sample distribution in terms of gender and age group.

Setting	**Age Group**		**Gender**	
Guided	28–33	4	Female	1
			Male	3
	55–58	2	Female	1
			Male	1
Real-World	28–33	1	Male	1
	55–58	1	Female	1

(b) Data Collection Method

For each participant, accelerometer data were gathered with a sampling frequency of 52Hz. Additionally, two participants were continuously monitored in a real-world setting over one night.

The guided scenario was performed under the most natural circumstances possible. The participants were in a bed, with a pillow, and covered by a blanket. They were asked to wear three Shimmer3 sensors, one on each of their ankles and one on the chest (placement illustrated in Figure 4.3(a)). The sensors were applied in the same orientation for all participants. The participants were instructed to assume each of the ten most common sleep positions [52] for at least 25 seconds each; this process was repeated for nine cycles to capture any potential variance. One cycle consisted of ten positions, and no transition was performed twice. These cycles were designed

by the researcher and communicated to the participant. Figure 4.3(b) illustrates the investigated positions, described as follows:

Supine position laying on the back with straight legs (Sup); two variations were monitored, those being right leg bent (S_{RLB}), and left leg bent (S_{LLB}).

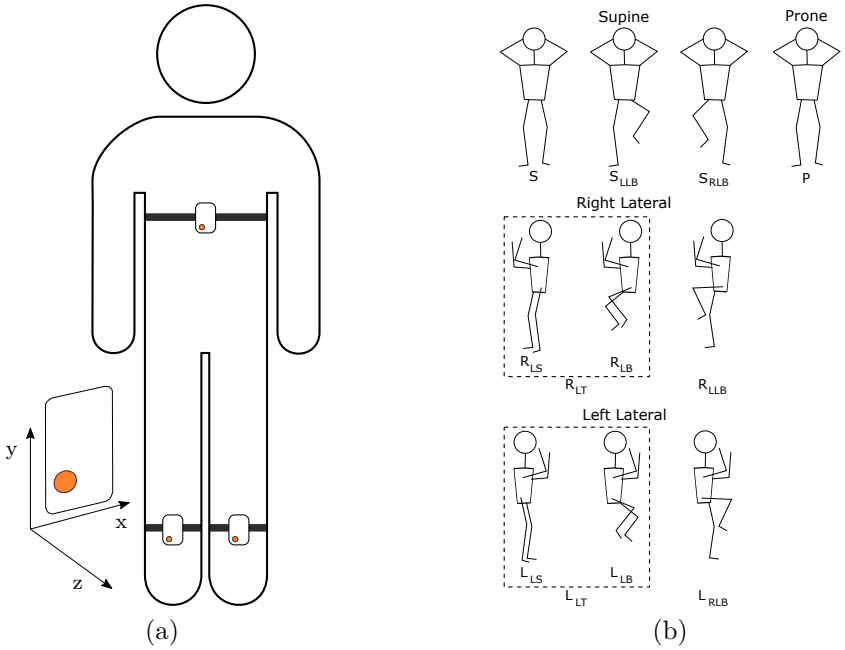
Prone position laying on the stomach (P).

Left lateral position laying on the left side with either both legs together (L_{LT}) or right leg bent (L_{RLB}).

Right lateral position laying on the right side with either both legs together (R_{LT}) or left leg bent (R_{LLB}).

Arm sensors were not required because the arms are always bent in a similar way for all of the positions shown in Figure 4.3(b). A researcher marked the start and end time of each position on a mobile device and later annotated relative to the available designed cycles.

Sensor placement in the real-world scenario was the same as that in the guided scenario. The two individuals were continuously monitored for about seven hours. Camera recordings were made to allow for later annotation of stable states.



(a) (b)

Figure 4.3: **(a)** Body placement of the sensors. **(b)** Monitored higher-granularity positions; dotted lines indicate two positions that are merged into one, such as R_{LS} (legs straight) and R_{LB} (legs bent).

4.3.2 Pre-Processing

First, the acceleration-moving variance method was applied to all acceleration data [205] to classify the non-movement periods as positions of interest. The cut-off threshold for stable states was empirically determined as $\alpha < 0.5$. The positions were determined using camera recordings for the real-world setting and instruction sheets for the guided settings. The camera recordings were destroyed following the annotation.

Furthermore, the raw data were processed into windows with the size of the sampling rate (52Hz) and an overlap of 50%. To obtain the feature vector for every window, the values in each window had to be averaged. For certain experiments, we applied downsampling to reach about 6Hz. This was done to address computing efficiency in relation to sampling rates.

This approach is unable to distinguish straight legs from both legs bent in the lateral positions. Consequently, the two classes are fused for right lateral and left lateral into one class, from $(R/L)_{LB}$ and $(R/L)_{LS}$ to $(R/L)_{LT}$. This merging is illustrated in Figure 4.3(b) by the dotted boxes.

4.3.3 Experimental Settings

Primary interest with this experiment lies in the performance differences between guided and real-world settings. Therefore, experiments are performed to investigate, (1) personalised models, (2) general models, (3) effects of a reduced sampling rate, (4) effects of a reduced number of sensors, and (5) effects of a reduced number of features.

The experiments were divided into general and personalised model groups. The personalised GMLVQ is trained using 80% of the data from one person; the other 20% is used for testing. The experiment repeated a 10-fold CV 20 times. The general model is trained using the overall dataset with a LOOCV which is averaged over 20 runs. This means that the system tests on one person and trains on all other participants. The experiments were conducted to provide insight into the potential generalisation of the approach. Moreover, we conducted experiments with a reduced sampling rate and a lower number of sensors to illustrate computing-efficient performance changes.

4.4 Results

In this section, we provide the outcomes from personalised and general models in guided and real-world settings.

4.4.1 Application in a Guided Setting

In this section, we provide the results from the guided scenario and offer insight into the overall performance and functionality of the presented approach.

(a) Personalised Models

The personalised models are trained on one person's data with separate CVs. The test results in 99.8% averaged accuracy over all validation runs.

(b) General Model

In Table 4.2, the averaged confusion matrix is shown for the general model using LOOCV. This model achieves an overall accuracy of 83.62% for our scope of detect-

Table 4.2: Averaged confusion matrix for the general model.

	Sup	Sup_{LLB}	Sup_{RLB}	R_{LT}	R_{LLB}	L_{LT}	L_{RLB}	P	Recall
Sup	94.3	2.04	3.64	0	0	0	0	0	91.35
Sup_{LLB}	3.94	96.06	0	0	0	0	0	0	96.24
Sup_{RLB}	1.83	0	98.17	0	0	0	0	0	98.20
R_{LT}	0	0	0	81.8	18.2	0	0	0	83.30
R_{LLB}	0	0	0	53.41	44.65	0	0	1.93	42.32
L_{LT}	0	0.85	0	0	0	91.7	7.45	0	91.75
L_{RLB}	0	0	0	0	0	30.87	67.35	1.78	67.37
P	0	0	0	0	7.86	0	0	92.14	88.75
Prec.	95.87	96.17	96.33	79.30	63.04	88.71	71.13	96.75	83.62

ing eight different sleep positions. However, the model has difficulty distinguishing between lateral positions. Furthermore, the results include some misclassifications, e.g., between the lateral and prone position. Based on these outcomes, we trimmed the number of potential positions to six and reran the experiment to find the source of the issue. By fusing the lateral sub-classes into the more straightforward right lateral and left lateral, the results indicate an increase in performance With only six positions, results show an accuracy of 98.31%.

(c) Mechanisms to Improve Computing Efficiency

The effects of computing-efficient design on performance are explored to gain insight into what setting is appropriate in terms of (1) sampling rate, (2) number of sensors, and (3) number of features.

Sampling Rate We downsampled the original data to a sampling rate of 6Hz for each participant and reran the experiments. Results indicated that the reduction from 52Hz to 6Hz did not affect the performance of the different models.

Number of Sensors We excluded the chest sensor, as the ankles are relevant for most of the positions, leaving us with just the two ankle sensors. For the general model, this results in an accuracy of 79.71% for eight positions and 96.08% for six positions. In contrast, the personalised models reach an accuracy of around 99.8%, meaning the sensor reduction seemingly had no impact.

Feature Selection One of the advantages of the GMLVQ model is the high availability of the relevance matrix Λ, from which important features can be extracted. The feature rankings based on the averaged relevant values are shown in Figure 4.4(a) alongside their SD; the three most relevant features—X_{Chest}, Z_{Left}, and Z_{Right}—are shown in Figure 4.4(b). For each personalised model and the general model, the order of relevance is unique (see Table 4.3). When applying the reduced data, including the combination of the three most relevant features, an approximate accuracy of 78% can be reached for the general model.

Table 4.3: Feature ranking for the personalised and general models.

Model	X_{Chest}	Y_{Chest}	Z_{Chest}	X_{Left}	Y_{Left}	Z_{Left}	X_{Right}	Y_{Right}	Z_{Right}
Personalised 1	1	9	6	5	4	8	3	2	7
Personalised 2	3	9	8	4	7	1	6	5	2
Personalised 3	2	9	4	3	5	7	8	6	1
Personalised 4	2	9	5	7	6	8	1	4	3
Personalised 5	1	8	6	5	2	4	9	3	7
Personalised 6	1	5	7	6	2	4	8	3	9
General	1	9	6	5	7	2	4	8	3

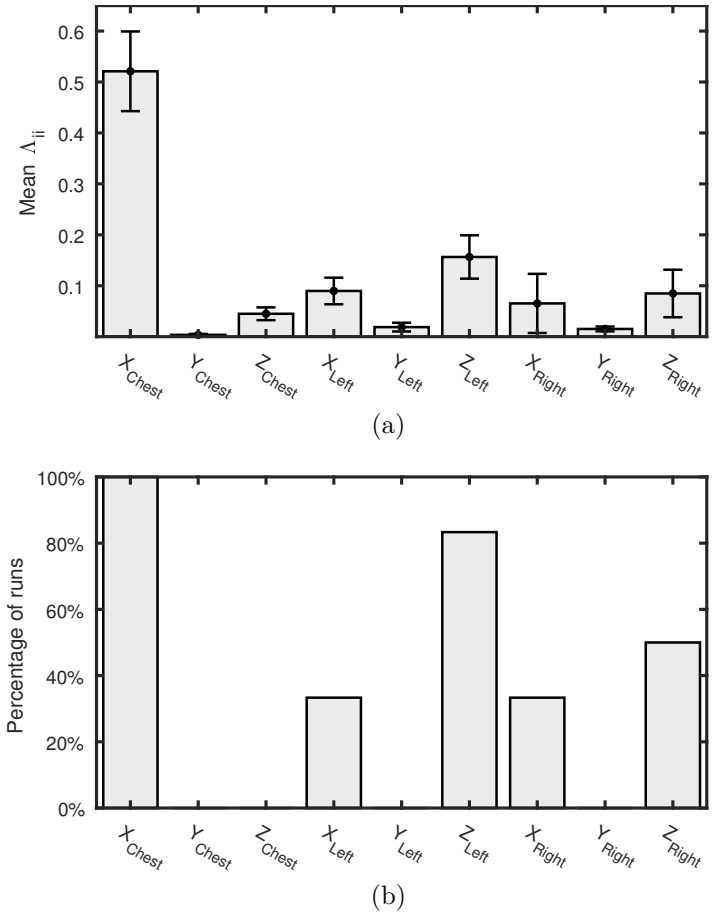


Figure 4.4: **(a)** Averaged relevance matrix diagonal and SD for each feature. **(b)** Percentage a feature was in the top 3 relevant features across 20 10-fold CVs.

4.4.2 Real-World Application

In this section, the general model is applied to data from two participants that were continuously monitored over one night.

In a real world setting, scope sometimes changes. Only six positions could be monitored; as we cannot control the sleep positions a person takes, only these six are considered. The positions for both participants are:

$$S, S_{LLB}, S_{RLB}, R_{LT}, R_{LLB}, L_{LT}$$

For participant one, applying the general model reaches an averaged accuracy of 98%. However, when we use only the three most relevant features, the accuracy decreases to 89.5%. In contrast, the personalised model reaches an averaged accuracy of only 58.39%.

When applying the general model to participant two, an averaged accuracy of 97.48% is reached. This decreases to 95.53% when only the three most relevant features are used. The personalised model fails in this experiment.

4.5 Discussion

In this section, we discuss the findings we obtained by applying the sleep position approach in guided and real world settings.

4.5.1 Application in a Guided Setting

Personalised Model Applying the personalised model to our collected data reaches nearly perfect performance. This suggests that the individuals are consistent in their positioning during the guided experiment. The lack of a perfect score is likely due to the inclusion of some small parts of transitions in the training set. Personal models come with a privacy advantage; no personal information needs to be shared or used in a more widespread model. However, the personalised approach comes with a limitation, as the model has to be retrained for every individual. This retraining requires an annotated dataset, which is time-consuming and a challenge to collect. One potential solution to this issue is an unsupervised machine learning approach, which could handle unlabelled data.

General Model The general model has an advantage over the personalised model in that it can be provided to any individual with no need for additional personal data to be provided or annotated; an off-the-shelf prototype capable of reasonable outcomes can be provided.

Misclassification generally results from the absence of orientation changes between positions, such as between the lateral positions $(R/L)_{LT}$ and $(R/L)_{(L/R)LB}$. An accelerometer can capture the orientation of the sensor, but not the relative position in space.

The general model also has certain restrictions, as it represents only general information on positions. Consequently, individual components of the positions cannot be represented, which leads to misclassification. However, there are ways to potentially improve the model. First, the classification task could be reduced to six positions resulting in orientation-based distinction. Second, small, personal datasets could be collected and annotated to retrain the model, and, therefore, incorporate individual components. Thirdly, a different method of analysis could improve detection performance between lateral and prone positions. For example, a rule-based approach to distinguish between the basic positions can be combined with individual-trained models to accurately detect sub-classes. This individual-trained models target to differentiate between higher-granularity positions, e.g., L_T and L_{RLB}.

This experiment goes beyond most research on the topic by expanding upon the current list of commonly investigated positions (see Table 2.6). Following the completion of our research, other researchers investigated sleep positions with a focus on arm positions using KNN and three devices on the torso and arms [179]. While it is not possible to directly compare these approaches, our use of GMLVQ comes with an advantage over that of KNN: easy interpretation through the provision of prototypes.

4.5.2 Mechanisms to Improve Computing Efficiency

Feature selection The general model can differentiate between positions even when fewer features are considered. It's important to note that this comes with a decrease in accuracy but requires less computational power. Our main finding is the individuality of different positions for each personalised model, which is displayed in Table 4.3. As personal information from individuals is essential when training and testing this model, the possibility exists to differentiate people on their personalised models. Comparing the feature ranking from the general model to that from the

personalised models suggests that the general model is less accurate for those that do not share the same feature ranking; for them, the personalised models offer the optimal results.

Reduced Sampling Rate When the sampling rate is reduced, a more computing-efficient setting is achieved. Despite the fairly large reduction, from 52Hz to 6Hz, performance remains similar. This suggests that no information is lost in the stable state data to distinguish between the positions. We conclude that a computing-efficient solution can be both viable and successful in sleep behaviour analysis, e.g., with a dynamic sampling rates monitoring stable states with a low frequency and movement periods with a high frequency.

Reduced Number of Sensors We investigated performance with a reduced number of sensors because of the potential discomfort and high cost that can result from the use of multiple sensors. While the use of only leg sensors with the general model resulted in a worse performance, results stayed largely the same with the personalised model. This is because each individual likely adopted positions consistently during the guided setting.

4.5.3 Real-World Application

The general model performs accurately in a real-world setting. Even though the model is not trained on real-world data, positions can be effectively detected. In contrast, the personalised model fails. We conclude that the guided-setting data does not accurately represent the real-world scenario for personalised models. A general model based on guided-setting data is still feasible, however, because it is trained on a subgroup of individuals; as a result, it captures position variability and, therefore, can provide more effective position detection. A more computing-efficient approach with only the three most relevant features reaches similar outcomes, resulting in an approach with only three 1-axis accelerometers.

Relative to current research, we can detect more different positions in the real-world application, because we consider limb movements (refer to Table 2.6). However, this does introduce some potential discomfort through the application of three sensors to the body.

4.5.4 Limitations

The findings of this study must be seen in light of some limitations within our scope. First, only a few people were included in our data collection process, which, of course, limits the potential generalisation of the outcomes. We counteract this effect to some degree by including age and gender factors in our analysis. Second, only eight of the ten most common positions can be detected by our approach. Third, annotation is likely not 100% accurate. Automatic detection, along with research notes and recordings, was used to keep label annotation as reliable as possible.

4.6 Summary and Future Directions

This chapter contributes to knowledge by successfully implementing an approach to detect higher-granularity sleep positions with wearable devices. This is an improvement upon current sleep position detection, as eight of the most common positions were considered using acceleration data gathered by Shimmer3 devices. Our approach works well with low sampling rates and in a real-world setting. Even three one-axis accelerometers are sufficient and reducing the number of sensors only slightly weakens performance. We provided a comparative study that enhanced current analysis in real-world applications and could accurately determine higher-granularity positions in guided settings. Overall, we conclude that the general model is the most appropriate option for a real-world application, as it more effectively captures the variability of positions. Compared to related work (depicted in Figure 2.5), our general model can reliably detect more positions (six instead of four) in the real-world; additionally, we improve the performance on eight positions in guided settings.

Monitoring these positions provides additional insight into a person's sleep behaviour. This can be a good starting point for medical assessments and behavioural interventions, such as position treatment for sleep apnoea patients [208, 209]. Sleep-related medical assessments are usually based on subjective questionnaires, which often do not accurately depict the positions taken while falling asleep. Our approach makes it possible to recognise, e.g., positions that prompt someone to fall asleep or wake up. Hence, it could inspire research and discussion on this topic to investigate the real correspondence of subjective position reporting and objective recognition. The sleep position recognition approach should be integrated into more broad sleep quality representations (as we are doing with our sleep behaviour assessment frame-

work). The insight provided through more complex sleep positions alongside professional medical advice can help in better sleep self-management. Bed exit detection could also be targeted with this approach with a few small modifications to the trainings data. In this study, participants were instructed to wear three Shimmer3 devices and did not report discomfort, however, the potential for discomfort and usability issues must still be considered by future studies. Clothing-based sensors are one way to avoid the possibility of discomfort [210]; certain integration solutions are already available on the market, such as Enflux [211] and Xsens MVN [212]. Another approach is the application of compact sensors directly on the skin [213]. Wearable technologies are becoming more practical every day, meaning the study of even more complex positions will be possible in time. Future studies could look at, for example, the now-difficult fused positions. As already stated, these are quite hard to distinguish with only orientation information. The positioning of the feet could soon be included, which could help detect positions that are not differentiated by orientation alone. With a higher sampling rate, the accelerometer data can be integrated twice, theoretically resulting in the true positioning. Currently, however, this comes with a limitation that numerical integration increases small errors drastically. Therefore, integration should be conducted over a short time, which could be realised as the transition periods in real life are measured to be around three seconds.

Sleep positions describe non-movement periods. To comprehensively describe sleep behaviour, however, movements must be considered as well. Movement during sleep is heavily related to sleep stages, which are further explored in the context of health, gender, and race in the next chapter.

Chapter 5

Fine-Grained Sleep Stage Detection

5.1 Introduction

In the previous chapter, we explored the non-movement periods during sleep through higher-granularity sleep positions. Additionally, we discussed the insights they offer into sleep behaviour. Sleep stages constitute another important element of sleep behaviour; in contrast to sleep positions, these are extracted from movement (see Section 2.3). Current technologies used to assess sleep and wake stages have various performance limitations; we try to overcome these by fusing and learning from multiple data sources. Sleep parameters, which indicate how well a person sleeps, are usually calculated using wake and sleep stages. Sleep parameters extracted from automatically detected sleep-wake stages still lack in comparison to the gold standard, in which stages are annotated manually by experts. Therefore, we develop an adaptive sleep parameter extraction process that is capable of interpreting sleep-wake stages.

Traditionally, human experts classify sleep stages using PSG measurements, which is a time-consuming task. Advancements in IoT technology have enabled common devices to assess sleep behaviour at home; this ability could potentially eliminate expensive in-hospital monitoring. We target several limitations described in the literature (discussed in Section 2.4.3). Commonly, one-model-fits-all approaches are employed despite evidence showing that individual sleep behaviour varies due to many factors, including biological traits, age, medication, health, and lifestyle [128, 214]. We address these limitations by presenting an adaptive learning approach that uses multiple data sources in order to achieve objective 5. This provides us with a fine-grained sleep-wake recognition process capable of capturing and characterising the factors that influence sleep stage recognition at different levels of granularity. We go beyond the commonly used physiological data and develop a diverse dataset

by integrating clinical history (health status) and genetic information (gender and race). We employ a multilayer perceptron (MLP) model that is trained and tested on this diverse dataset. Diversity plays an important role in training machine learning methods [141]; especially crucial is the inclusion of both healthy and diseased individuals. We avoid the common mistake of testing against just one technician's scoring by using a dataset scored by multiple technicians. The problem with one technician is, that they have individual styles and machine learning algorithms learn them. This issue can be reduced by including scoring by multiple technicians and, therefore, potentially reduce variation in the trained models. A direct method would be to train one model per technician and combine the answers of the different models. However, the dataset provides an averaged scoring and, therefore, indirectly deals with the issue. Generally, only one data source (e.g., actigraphy) is used; research has yet to employ a combination of different data sources. Actigraphy has proven successful in home-based classification; HRV features have emerged fairly recently. For this reason, we explore these sources and their combination, as the sensors capable of extracting this information could help to achieve a more accurate classification. We emphasise the importance of balanced datasets, as imbalanced data can lead to unreliable classification with a bias towards the majority class. To ensure comparability with other approaches, we use a publicly available dataset. Measuring sleep parameters at home with actigraphy is usually done by interpreting wake-sleep episodes using static thresholds without regard for various individual influence factors, such as age [125]. To improve upon standard sleep parameter extraction from sleep-wake behaviour, we develop a process that adapts to users' needs and subsequently addresses objective 6.

The ability to accurately detect sleep stages and extract sleep parameters provides us with the following advantages in terms of the framework introduced in Chapter 3. First, this information can help experts diagnose diseases and detect abnormalities in natural-environment sleep behaviour. Second, it constitutes a better basis for measuring sleep parameters, which allows for a more accurate and objective assessment of sleep quality. This approach contributes to our framework's assessment and extraction layers as well as its sleep quality, regularity, and circadian rhythm components.

The rest of this chapter is organised as follows: In Section 5.2, we detail our sleep stage recognition approach and describe how we extract sleep parameters through our sleep-wake recognition process. In Section 5.3, we explore the influence factors of sleep stage recognition and perform tests to validate our approach. The outcomes are provided in Section 5.4 and discussed further in Section 5.5. A summary of the

work and future research directions are provided in Section 5.6.

5.2 Sleep Stage Recognition and Application

In this section, we describe the application of multi-source data learning, feature extraction, and pre-processing for sleep stage recognition. We also discuss how we combine our recognition approach with a personalised sleep parameter extraction process.

5.2.1 Multi-Source Data Learning Approach for Sleep-Wake Recognition

We propose a multi-source data learning approach with a fine-grained structure to detect sleep and wake stages. Our aim is to move away from a one-model-fits-all approach towards a personalised approach. We consider a two-stage classification problem following the gold-standard of 30-second scoring intervals; these intervals are scored as sleep or wake stage.

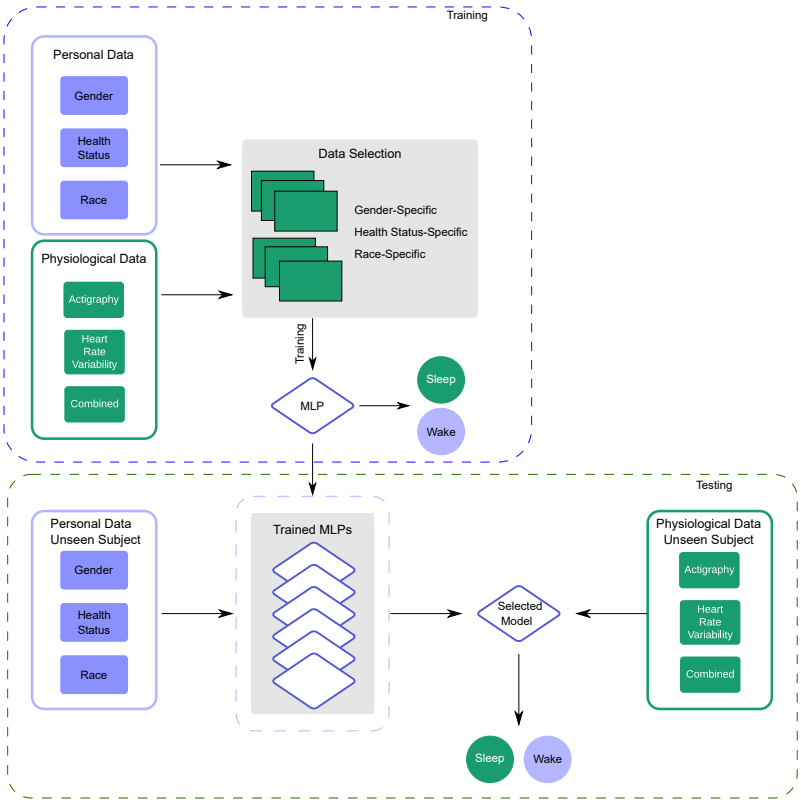


Figure 5.1: Multi-source data learning approach for sleep-wake detection.

A visual representation of our multi-source data learning approach is presented in Figure 5.1. We incorporate two types of data: personal data, including gender, health status, and race; and physiological data, including actigraphy, HRV, or a combination of the two. The physiological information is fused with clinical history and genetic information (influence factors). Depending on the specifics of the personal data, a portion of this fused information is selected and a final dataset is assembled. The selected training datasets are pre-processed and features are extracted (extraction explained in more detail in Section 5.2.3). The features are used to train MLPs with one hidden layer that has the same size as the list of features. The structure is depicted in Figure 5.2, which shows the number of input features $(F_1, F_2, ..., F_n)$ and nodes in the hidden layer $(N_1, N_2, ..., N_n)$. The overall model consists of three layers: input layer, hidden layer, and output layer. The output layer describes the classification targets, which are wake (W) and sleep (S) stages. The number of nodes is adjusted to the accessible sensory data input features, either being 18 for HRV, 7 for actigraphy, or 25 for both (details in Section 5.2.3).

We decided on the MLP empirically based on initial experiments comparing methods considered appropriate in the literature. These methods show similar outcomes with respect to accuracy; the best outcomes across several experiments are presented here: GMLVQ, 78.2%; KNN, 79.0%; MLP, 79.1%; AdaBoost, 79.2%; and RF, 80%. A model with a simple structure that is still able to learn from the data can counteract the possibility of overfitting. MLP contains three layers and still performs well; the tested AdaBoost and RF have a more complex structure—AdaBoost is tested on 100 estimators and RF is tested with 50 estimators and a depth of 30. The results from AdaBoost and MLP are not significantly different (t-test on two related samples). Furthermore, the SD over the repeated experiments was lower for MLP compared to AdaBoost, therefore, was chosen, as it suggests a more stable performance. We investigated the mean squared error (MSE) for the training and test sets and found that RF was overfitting (MSE was high for testing and low for training). MLP did not show overfitting, as a consequence was chosen as being an appropriate model for further investigations.

For testing on unseen individuals, a model is selected based on subject's personal information. Features from the physiological data are then calculated and given to the selected model to assess sleep and wake stages.

We illustrate the idea of a fine-grained adaptive structure against the traditional general approach in Figure 5.3. The information used to reach a granular adaptive model details gender, race, and clinical history. In our tests, we include

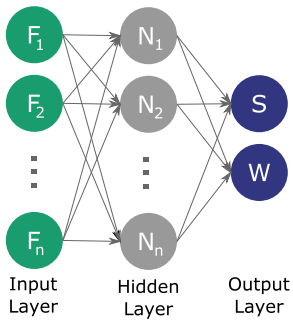


Figure 5.2: Multilayer perceptron design.

four races—White, Asian-American, African-American, and Hispanic—and three diseases—RLS, insomnia, and sleep apnoea. The standard data for general sleep-wake approaches are used to generate specific models. This data typically includes physiological data from a single data source (e.g., actigraphy or HRV). We expand upon this with multiple data sources by fusing actigraphy and HRV. Figure 5.3 shows the steps to achieve fine-grained models. First, gender information determines gender-specific models. Second, the health status of an individual—included if available—determines a health-status-specific classification. Third, race identification, if available, determines a race-specific model. This information can also be combined to select a model with two or all three influence factors. Each of these steps is flexible; if at least one is included, a fine-grained sleep-wake recognition approach is reached. The advantage of a fine-grained approach is that influence factors can be incorporated into individual models.

The proposed approach is designed to be adaptable to (1) available personal information and (2) available sensory data. First, the availability of gender, health status, and race is not guaranteed. Therefore, only one is necessary to reach a somewhat personalised approach. Second, the approach can incorporate multiple data sources by combining calculated features through a 30-second raw data structure. This is feasible by adapting the size and overlap of the windows when calculating individual features. Generally, this approach can be applied to any data source that provides information on sleep in a similar structure. The design of our approach allows for varying levels of granularity, ranging from gender-specific to race-health-gender-specific.

5.2.2 Personal Sleep Parameter Extraction

Sleep parameters can be extracted from sleep-wake stages assessed by accelerometers, but this approach still does not achieve comparable performance to the gold

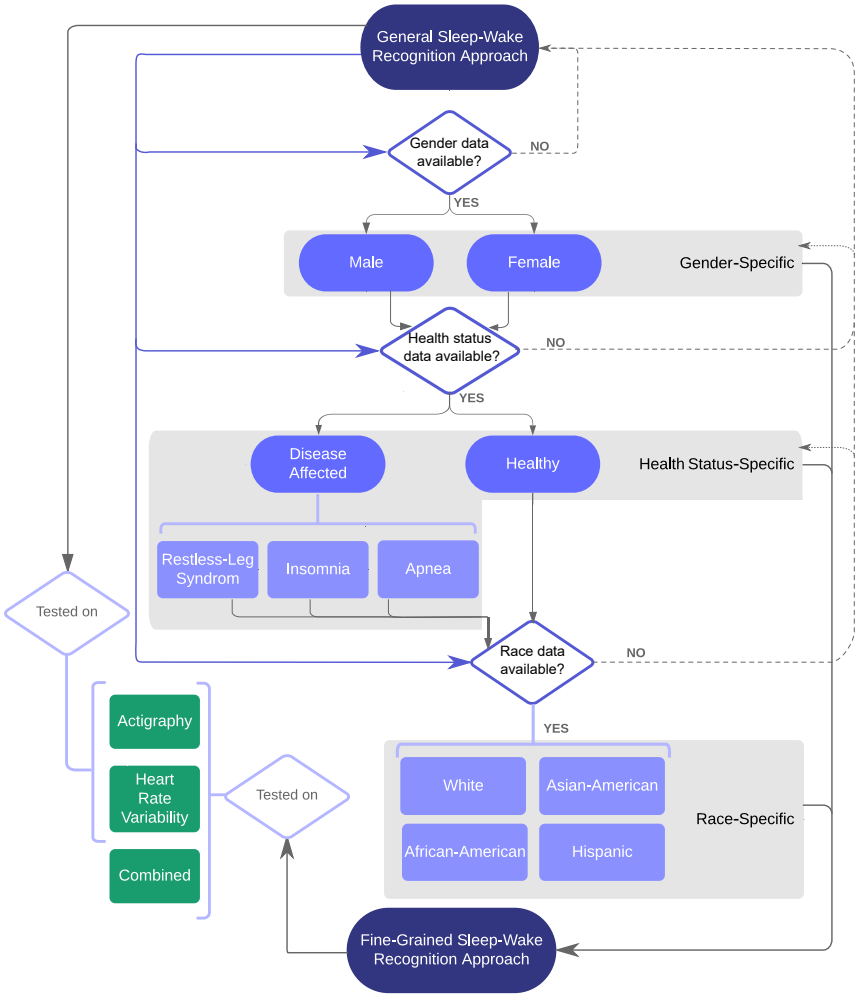


Figure 5.3: Illustration of the steps from a general to a fine-grained approach.

standard (PSG). Therefore, we developed a personalised sleep parameter extraction process that incorporates our fine-grained sleep-wake detection approach (Section 5.2).

The wake-sleep episodes are the basis for sleep parameter extraction. This means, e.g., for sleep duration that the extracted sleep episodes are summarised with the knowledge of sleep onset and sleep offset. Currently, static thresholds are used to define sleep parameters such as SOL [124]. These approaches are not personalised for individual groups despite it being well known that group factors influences accuracy such as age [125]. Therefore, we incorporate dynamic thresholds that change depending on personal factors. We propose to personalise the thresholds for sleep parameters based on subgroups (i.e., health status, race, and gender). In this investigation, we follow an empirical threshold determination (details in Section 5.3.2).

5.2.3 Feature Extraction and Pre-Processing

Feature extraction and pre-processing are essential, especially since we want to validate an approach adaptable to different data sources. Since the used data sources provide different sampling rates, we adopt a common 30-second interval structure (the standard used in PSG). Furthermore, we consider a balanced training and test set. We do not use the raw actigraphy and HRV data directly; we extract specific features.

For actigraph data, the provided data include activity counts and light data in 30-second intervals. Furthermore, we calculate the mean and SD features over 5-minute overlapping windows. Activity counts summarise the intensity of activeness at a time. While light gives information about the ambient light divided into white-, red-, green- and blue-light. These light features can help to judge the day-night-rhythm. Light exposure at night has shown effects on sleep, blue light increased alertness and red light increased cortisol levels [215]. Furthermore, blue light is emitted by smartphones and can help to judge a by movement unrelated wakefulness with the blue light level. Both can cause awakenings or difficulties falling asleep, therefore, they are important aspects to assess sleep and wake stages. When using light as features patterns can be learned by the MLP which reflect the influences and help better judge sleep and wake stages. Overall seven features are extracted: activity counts, mean, SD and four light features.

To extract the HRV features, an analysis based on ECG is performed by the NSRR [202] to extract QRS complexes and, therefore, R-points. The QRS com-

plex describes three of the graphic deflections in a typical ECG (see Figure 5.4). From these R-points, normal-to-normal intervals (NNIs) and cardiac inter-beat intervals (RRs)—the intervals between adjacent QRS complexes—are determined (for details compare [216]). Certain intervals are excluded based on the exclusion criteria presented in [217] to remove artifacts: Excluding individuals (1) with sleep stages of <2 hours, (2) with normal-to-normal-intervals (NN) <1.000, (3) NN $<$ 0.35 sec or $>$ 2.5 sec, (4) <180 NN within 5-minute windows.

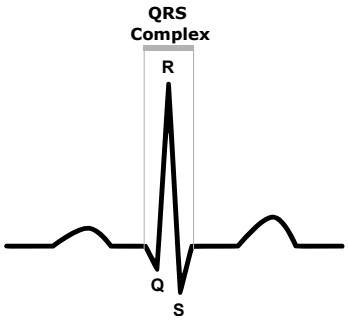


Figure 5.4: QRS complex for ECG.

To receive the HRV features the information is fused into features getting values for 30-seconds each coming from 256Hz sampling rate. The idea is based on the amount of extracted beats within epochs of 30 seconds. In Table 5.1, the extracted features are provided, divided into time- and frequency-domain features.

The included time-domain features were: (1) NNI averages; (2) ratio of successive NNI beats to RR intervals (cardiac inter-beat intervals); (3) instantaneous heart rate average; (4) NNI SD; (5) 5-minute-interval-based SD of NNI averages; (6) 5-minute-interval-based average of NNI SDs; (7) percentage of disparity of contiguous NNIs bigger than 10, 20, 30, 40, and 50 ms; and (8) square root extraction from the mean of squared difference between contiguous NNIs [217]. The frequency-domain measurements we used are based on the spectral power of the NNIs: (1) up to 0.4 Hz; (2) between 0 and 0.003 Hz (ultra-low-frequency power); (3) between 0.003 and 0.04 Hz (very low-frequency power); (4) between 0.04 and 0.15 Hz (low-frequency power); (5) between 0.15 and 0.4 Hz (high-frequency power); (6) low-frequency power to high-frequency power ratio; (7) normalised high-frequency power; and (8) normalised low-frequency power [217]. The HRV measurements are calculated either over the whole sleeping period or over 5-minute intervals according to the NSRR formulation in [217]. Whole night values are treated as a constant feature for data from a specific subject. As traditional resampling can cause a reduction in high-frequency components and the Fast Fourier transformation needs balanced distributed samples, Lomb periodograms are used for frequency-domain spectra calculation of irregular

Table 5.1: HRV features which are calculated and excluded (x) based on their pre-assessed pairwise correlation.

Feature	Description	x	Feature	x
	Time-Domain			
	30-sec Intervals		Whole Sleep Period	
NN_RR	ratio of consecutive NNs* over all RRs*	x	Tot_NN_RR	x
AVNN	average of all NNs*		Tot_AVNN	
IHR	average instantaneous heart rate		Tot_IHR	
SDNN	standard deviation of all NNs*	x	Tot_SDNN	
rMSSD	square root of the mean of the squares of difference between adjacent NNs*		Tot_rMSSD	
	percentage of differences between adjacent NNs* that are			
pNN10	>10 ms		Tot_pNN10	x
pNN20	>20 ms		Tot_pNN20	x
pNN30	>30 ms	x	Tot_pNN30	x
pNN40	>40 ms	x	Tot_pNN40	
pNN50	>50 ms	x	Tot_pNN50	
	5-min Intervals			
SDANN	standard deviation of the averages of NNs*			
SDANNINDX	mean of the standard deviations of NNs*			
	Frequency-Domain			
	5-min Intervals		Whole Sleep Period	
TOTPWR	total NNs* spectral power up to 0.4 Hz	x	Tot_TOTPWR	x
ULF	ultra-low FP*: [0, 0.003 Hz]		Tot_ULF	
VLF	very low FP*: [0.003, 0.04 Hz]		Tot_VLF	x
LF	low FP*: [0.04, 0.15 Hz]	x	Tot_LF	
HF	high FP*: [0.15, 0.4 Hz]		Tot_HF	x
LFn	normalized LF	x	Tot_LFn	x
HFn	normalized HF	x	Tot_HFn	x
LF-HF-ratio	the ratio of low to high FP*	x	Tot_LFHF	

*NNs-normal sinus beat intervals; RRs-cardiac inter-beat intervals; FP-frequency power

data samples [200,217]. Furthermore, correlated features are excluded from the HRV measurements, reducing the number of features from 38 to 18.

In order to train and test in a balanced manner, we apply downsampling and ensure balance on gender (equal number of male and females) and sleep stages (equal number of wake and sleep events). As the MLPs classify per event and do not look at the time series, we can delete different values from the majority class randomly per repeat of CV. Moreover, the data is standardized per feature with the z-score: $z_i = (x_i - \overline{x})/\sigma$, where x_i is the observed value, $\overline{x}$ is the mean of the sample, and σ is the standard deviation. For the case of non-gender-specific training, we keep a ratio of 1/1 for male/female and sleep-wake stages, creating a gender- and sleep stage-balanced dataset. Furthermore, we reject all individuals with less than 400 examples of sleep-wake stages.

5.3 Evaluation of the Proposed Approach

The following scenarios are investigated to validate our approach and test the usefulness of the personalised models fused with granular-adaptive data analysis. In this section, we present the considered experimental data and the experimental settings to investigate the performance and influence factors of our approach.

5.3.1 Dataset

We use the MESA dataset described in Section 3.4.1 with the exclusion criteria detailed in Section 5.2.3 to obtain a diverse dataset. The diversity comes from the inclusion of participants (1) from different ethnic backgrounds, including African-American, White, Hispanic, and Asian-American; (2) of both genders; (3) aged between 45 and 84; and (4) with different health statuses, either healthy or diagnosed with a sleep disorder (sleep aponea, insomnia, or RLS) [201]. The segmentation of the dataset, corresponding to the data presented in Table 5.2, shows the number of people and average age with SD per data source as well as the considered diversity factors. The diversity factors include race, health status, and gender. We can clearly see an unequal distribution, which should be taken into account. For the case of females and males together, we provide the number of available data used in the two different test settings (i.e., gender-balanced and gender-imbalanced cases; details in Section 5.3.2).

In this investigation, we use the first night of actigraphy, which was collected simultaneously with PSG [200,201,203] to validate our fine-grained approach. The scoring of the recordings from actigraphy has a 76.71% match with the PSG scores while downsampling and averaging show only 73.45% coherence, suggesting a majority class bias (see Section 5.4). This shows the necessity of improving current recognition rates and establishing a fine-grained approach. The ground truth of PSG for sleep stages is given by the R-K, which we translated to wake and sleep stages. This means wake is considered normally but S1, S2, S3, S4, and REM are fused into one stage called sleep. We employed sensors that could be easily applied at home. These sensors are actigraphy and ECG measuring HRV. To use the dataset, we need to extract features and pre-process in line with the process described in Section 5.2.3.

Table 5.2: Diversity-based data segmentation incorporating gender, race, and health status.

		ACITGRAPHY						HRV					
		Together		Male		Female		Together		Male		Female	
Race	HS	No.	No.G.*	No.	Age	No.	Age	No.	No.G.*	No.	Age	No.	Age
All	All - HS	1,641	1,458	896	69.1±9.0	745	69.1±9.0	1,576	1,236	862	68.9±8.9	714	69±8.9
All	He	1,367	1,226	740	69.3±9.1	627	69.2±9.0	1,313	1,030	712	69.2±9.0	601	69.1±8.9
All	Apn	120	96	50	67.8±7.7	70	66.6±7.3	116	82	50	67.8±7.7	66	66.5±7.3
All	Ins	99	64	66	66.6±8.7	33	67.6±9.5	94	56	63	66.7±8.8	31	67.7±9.6
All	RLS	73	54	46	67.9±9.2	27	68.1±9.2	69	52	43	67.1±8.7	26	68.5±9.2
W	All - HS	612	566	325	69.2±9.1	287	69.3±9.0	591	492	315	68.9±9.0	276	69.3±9.0
W	He	518	474	277	69.6±9.3	241	69.6±9.1	500	410	268	69.3±9.2	232	69.5±9.1
W	Apn	46	32	17	66.5±9.0	29	65.9±6.9	44	24	17	66.5±9.0	27	66.1±7.0
W	Ins	24	16	16	64.5±5.9	8	64.9±9.4	23	16	15	63.8±5.4	8	64.9±9.4
W	RLS	26	24	14	67.8±7.9	12	69.7±10.2	26	24	14	67.8±7.9	12	69.8±10.2
A	All - HS	176	156	97	69.2±8.4	79	68.0±9.2	166	132	91	69.7±8.5	75	67.6±8.8
A	He	129	144	71	69.5±8.1	58	67.9±9.0	122	92	67	69.9±8.2	55	67.4±8.2
A	Apn	8	2	1	77.0±0.0	7	64.1±10.0	7	2	1	77.0±0.0	6	62.7±10.2
A	Ins	28	22	17	66.8±7.9	11	69.4±9.5	25	20	15	67.5±8.1	10	69.1±9.9
A	RLS	8	8	4	65.8±12.0	4	62.8±8.02	8	8	4	65.8±12.1	4	62.8±8.0
B	All - HS	453	382	256	69.0±8.8	197	69.5±8.7	435	312	249	69.1±8.7	186	69.3±8.5
B	He	382	332	211	69.2±8.8	171	69.6±8.7	367	272	205	69.3±8.8	162	69.4±8.6
B	Apn	40	38	20	69.6±7.0	20	68.2±7.4	39	32	20	69.5±7.0	19	67.8±7.4
B	Ins	20	6	17	66.8±8.9	3	66.7±14.2	19	2	17	66.8±8.9	2	71.0±17.0
B	RLS	16	10	11	68.9±12.2	5	69.0±11.3	14	8	10	67.1±11.2	4	71.8±10.9
H	All - HS	400	354	218	68.9±9.5	182	68.7±9.2	384	300	207	68.5±9.4	177	68.7±9.1
H	He	338	306	181	68.9±9.5	157	68.7±9.4	324	256	172	68.4±9.3	152	68.6±9.2
H	Apn	26	22	12	65.8±6.3	14	67.0±6.8	26	22	12	65.8±6.3	14	67.0±6.8
H	Ins	27	20	16	68.4±11.7	11	67.9±9.4	27	18	16	68.4±11.7	11	67.9±9.4
H	RLS	23	12	17	67.8±8.1	6	67.8±6.5	21	12	15	66.7±7.6	6	67.8±6.5

* No.G. - Number of individuals in gender balanced case; No. - Number overall; All - All individuals; W - White; A - Asian;
B - African-American, H - Hispanic; HS - Health status; He - Healthy; Apn - Apnoea; Ins - Insomnia

5.3.2 Sleep-Wake Analysis Scenarios

The validation of our proposed personalised sleep-wake recognition approach is performed in various scenarios that are described in detail in this section. As already stated, we conducted health-status-specific and race-specific experiments. Health status categories include: all health statuses (shown as All-HS); healthy (He); diagnosed with apnoea (Apn); diagnosed with insomnia (Ins); and diagnosed with RLS (RLS). Race categories include: all races (All-Races); African-American (B); White (W); Hispanic (H); and Asian-American (A). All investigations are validated with six times 10-fold CV to ensure representative outcomes and we report averaged accuracies with SD in Section 5.4. Accuracy is considered sufficient, as data are always balanced for sleep stages. Based on changes of SD across different repeats of CV, we find the influences that the used training data has; if the SD is higher, the approach works better for certain training-/testing-splits than for others (a low SD is preferable). The scenarios are investigated for different accessible data sources: actigraphy, HRV, and the two fused together (fusion is done on the feature level).

The structures of the sleep-wake analysis scenarios include the following steps:

First, we compare our fine-grained approach on actigraphy data with the original sleep-wake classification, which is part of the used dataset [202]. The original sleep-wake classification is denoted as 'Original Sample' and is based on the imbalanced dataset with respect to sleep-wake stages. In the original approach, activity counts are analysed by comparing the activity levels within $\pm$ two minutes time period ($<$ 40 is sleep, $>=$40 is wake), details can be found in [218]. Additionally, we explore the performance in the downsampled case, referred to as the 'Under-sampled'. This scenario is further referred to as scenario (1).

Second, we analyse data for females (Female) and males (Male); all data from the specific gender groups are used in the model. The average of these outcomes is also given (Avg. Gender) in order to be comparable to the overall performance when personalisation is not integrated. Furthermore, the race-specific and health status-specific analyses are performed. We will refer to 'Avg. Races' as the average of all individual race-specific trained models and 'Avg. Health Status' as the averaged performance of all health status-specific models. Combinations of, e.g., gender and race information in a model are referred to ase'. These scenarios are further referred to as scenarios (2).

Third, we compare the outcomes with the case in which the data from both genders are taken together to train a more general model. This general approach is explored

with a 1/1 ratio for female/male and sleep-wake stages per individual (All; (3a)) or imbalanced towards gender using all data from all individuals (All-biased; (3b)). These scenarios are further referred to as scenarios (3a) and (3b).

Fourth, we consider gender as an extra feature in the data, either balanced sleep stages and gender (Incl. Gender; (4a)) or imbalanced towards gender (Incl. Gender-Biased; (4b)). These scenarios are further referred to as scenarios (4a) and (4b).

Fifth, we determine the best approach in terms of training data. Models are trained on individuals with a medical condition and tested on both healthy individuals and individuals with a medical condition. Additionally, models are trained on healthy individuals and tested on both healthy individuals and individuals with a medical condition. In the results section, we refer to this analysis as 'Train Disease - Test Healthy' (i.e., train model on diseased individuals and test on healthy individuals).

Lastly, we investigate the performance of the variable-threshold-based parameter extraction process built upon our wake-sleep assessment approach. We use the lights-off time from PSG provided in the MESA dataset to make a direct comparison with the actigraphy. This means that both data sources are used from the lights-off time onwards. The scope of the functionality investigation lies in SOL, because SOL is important to show abnormalities of sleep behaviour and to indicate insomnia.

5.4 Results

In this section, the results of the investigated scenarios are presented for both actigraphy and HRV data.

5.4.1 Comparison of Approaches

In Figure 5.5, the outcomes of the original sleep-wake classification [202] are compared with the outcomes of our fine-grained approach for actigraphy data. The 'Original Sample' case reached an accuracy of around 76%. For the 'Under-Sampled' case, the repeated-averaged accuracy decreases by around 3%. For example, for 'All-HS', a decrease from 76.71% to 73.45% can be observed.

Furthermore, the results are given for: 'All' (compare to Section 5.3.2, 3a), 'Gender' (4a), 'Avg. Gender', 'Avg. Races', 'Avg. Gender and Race', and 'Avg. Health Status'.

In Figure 5.5, actigraphy-based recognition rates are improved by around 5% when using MLPs, compared to the downsampled recognition methodology that is currently used. For example, 'All-HS' reached 78.98% for 'Avg. Gender' compared to the downsampled 73.45%.

The 'Avg. Health Status' models reach an accuracy of 77.88%, while gender-specific personalised models on top of the health-specific classification can achieve an averaged accuracy of 78.47% (79.16, 79.38, 78.06, 77.29) and even 78.75% (78.87, 78.93, 79.05, 78.15) if race is also specified. Overall, the gender-specific approach reaches the best outcome, with 78.98% in the 'All-HS' case.

The effect of including race- and gender-specific data is especially prominent in the diseased cases (e.g., 79.05% accuracy for insomnia 'Avg. Gender and Race' compared to 78.06% for 'Avg. Gender'). The results show that a fine-grained model is superior to standard one-model-fits-all approaches, especially for diseased individuals (e.g., 79.05% for 'Avg. Gender and Race' compared to 77.01% for 'All').

Furthermore, it is evident that including gender (Incl. Gender) as a feature during training does not mentionably change the recognition rate towards 'All'. This can be seen, e.g., in 'Ins' which reaches an accuracy of 77.01% for 'All', which is very similar to 77.05% for 'Incl. Gender'.

5.4.2 The Impact of Personal and Physiological Data

In figures 5.7, 5.8, and 5.9, the results from (2)–(4) are presented for 'Male', 'Female', 'Avg. Gender' (see Section 5.3.2 2), 'All' (3a), and 'Incl. Gender' (4a).

The SD is not given in 'All' and 'Incl. Gender' to allow for visual clarity; overall, however, the SDs for 'All' and 'Incl. Gender' are similar and 1–5% higher than in the individual-trained models. For example, 'White, RLS, All' has an SD of 7.8% and 'White, RLS, Incl. Gender' has one of 7.6% while the gender-specific 'Avg. Gender' SD is only 0.95%.

Furthermore, it should be noted that the SD is lower in scenarios (3b) and (4b), gender-imbalanced analysis, relative to scenarios (3a) and (4a), balanced case (see Figure 5.6). Put more concretely, for (3b) 'All-Races, Apn, All' the SD is 0.5%; for (4b) 'All-Races, Apn, Incl. Gender' SD is 0.4%; for (3a) SD is 3.4%; and for (4a) SD is 3.7%. There is an accuracy difference between (3a) and (3b) as well as between (4a) and (4b) for HRV; this difference is not as obviously present for actigraphy. For example, with HRV, (3a) 'A, He, All' reaches an accuracy of 79.1% while (3b)

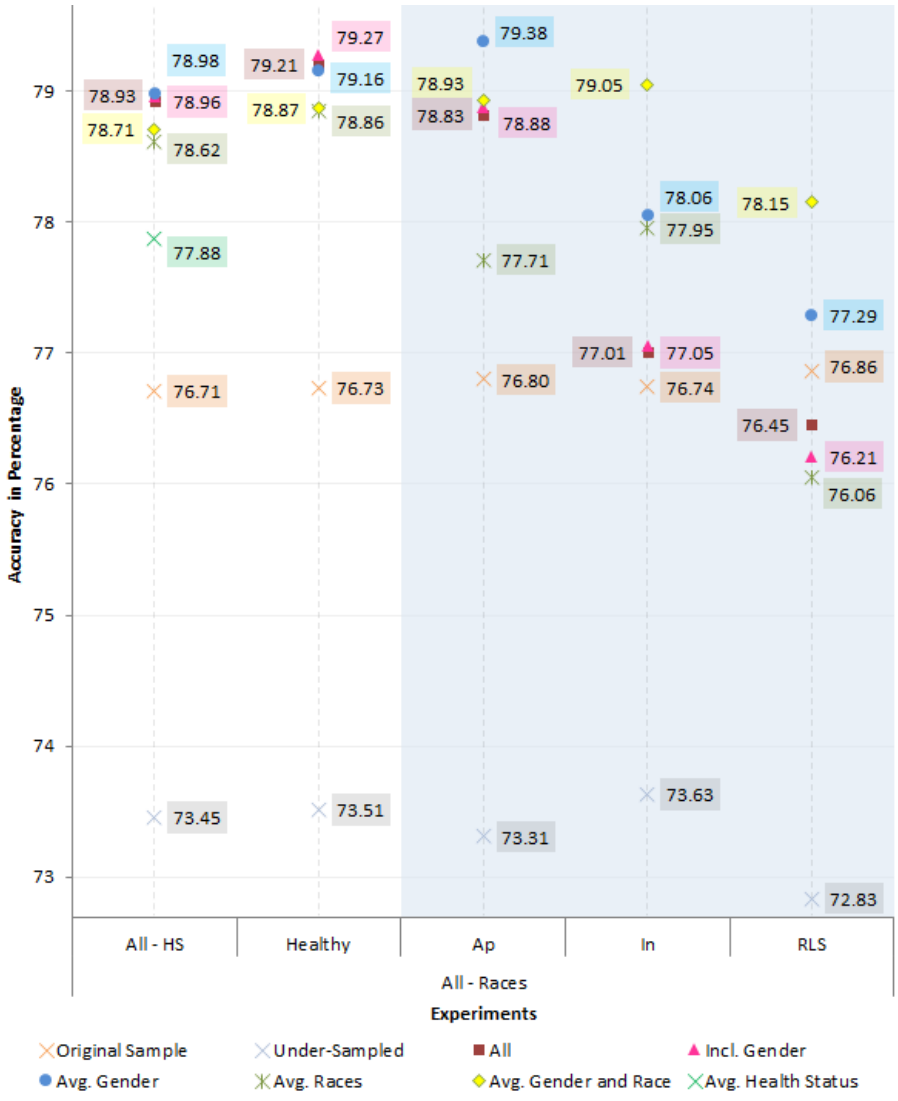


Figure 5.5: Performance overview of original sleep-wake classification and the proposed approach.

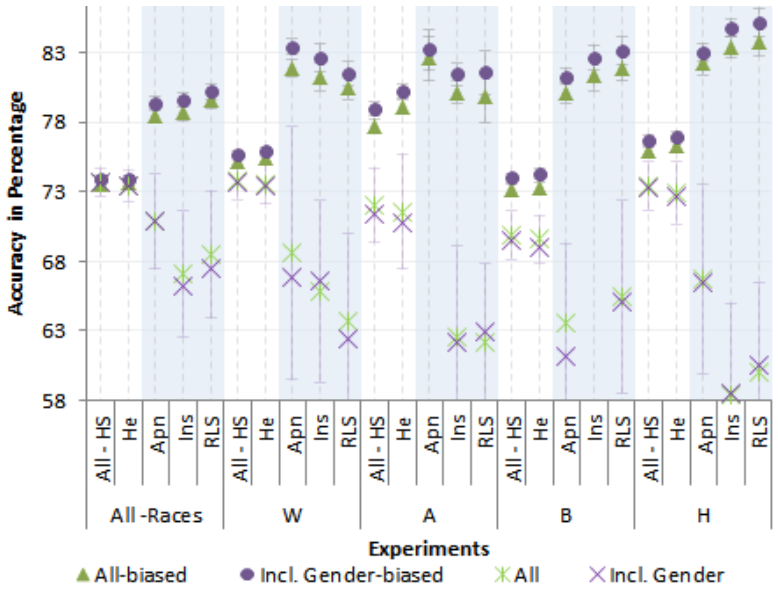


Figure 5.6: Gender-balanced and gender-imbalanced HRV training results.

reaches an accuracy of 71.6%.

In Figure 5.7, accuracy is generally higher for male participants compared to females. For example, 'All-HS' in the 'All-Races' case can reach 80.04% accuracy for males and 77.92% for females. In contrast, for HRV in Figure 5.8, the outcomes are more similar; the individual female models perform better overall with an 80.60% averaged accuracy compared to one of 79.64% for the male models. In the combined case in Figure 5.9, however, the males (85.77%) outperform the female models (84.65%) again.

The effect of fewer individuals is represented in higher SDs (see Table 5.2). For example, 'HRV, Female, A, Apn' with six individuals has an SD of 1.6% while 'HRV, Female, W, He' with 601 individuals has an SD of 0.2%. It should be noted that the 'Male, Asian, Apnoea' case contains just one participant, meaning it is unrepresentative. It is excluded from the presented results, as no conclusions can be drawn from it.

For 'He' ('White, He, Male' with SD of 0.2%) and 'All-HS' data (SD of 0.2%), the SD is usually lower than in the disease models ('Apn': SD of 1.1%, 'Ins': SD of 0.8%, and 'RLS': SD of 0.8%; see Figure 5.7).

The average SD for HRV is generally higher, with 1.08% for 'Avg. Gender', 4.27% for 'All', and 0.68% for 'All-biased', while, for actigraphy, it is 0.79% for 'Avg. Gender', 3.42% for 'All', and 0.59% for 'All-biased', and, for the combined case, 0.86% for 'Avg. Gender', 4.09% for 'All', and 0.6% for 'All-biased'.

Overall, the accuracy for actigraphy data has less variation in races and health statuses (ranging from 76.3% to 82.2% for 'Avg. Gender') compared to that for HRV (ranging from 74.2% to 86.2%); this is evident in figures 5.7 and 5.8. However, it is more challenging to distinguish between wake and sleep stages in the Hispanic group with actigraphy (averaged 'Avg. Gender' accuracy of 76.65%), than with HRV (81.22%).

HRV in Figure 5.8 indicates that, if we look at both diseased and healthy individuals, our method performs better for diseased individuals than for healthy ones. For example, 'White' averaged diseased results reach 82.05% accuracy; 'White' averaged healthy results reach just 76.64% accuracy. This trend is also visible when the data sources are combined (see Figure 5.9). Actigraphy shows less variable accuracy rates in this respect (in 'B' averaged diseased outcomes reach 78.46%, 'He', 80.08%; see Figure 5.7), though between male and female models, actigraphy shows more variable accuracy rates.

Based on these trends, we investigate the combined cases (see Figure 5.9). The trends are consistent with the general trend reported for the HRV case. The combination of sources improved the accuracy from the HRV case by around 5% (from 74.5% to 80.6% for 'All-Races, He'). For actigraphy, only a 1% increase is observed for the case of the healthy individuals (from 79.21% to 80.87% for 'All-Races, He'). However, it also shows a 5–8% increase in the diseased individuals (from 77.29% to 85.47% for 'All-Races, RLS').

5.4.3 Models Trained on Healthy and Diseased Individuals

In figures 5.10, 5.11, and 5.12, the analysis outcomes from setting (5) are given for actigraphy data, HRV data, and them both together. Results are provided for the models trained on individuals with a medical condition and tested on healthy individuals ('Train Disease - Test Healthy') and as well as for the models trained on individuals with a medical condition and tested on individuals with a medical condition ('Trained Disease - Test Disease'). Additionally, the average of the outcomes ('Avg. Disease Trained') is given. The fields are coloured in shades of red for the models trained on individuals with a medical condition and coloured in shades of blue for models trained on healthy individuals. Furthermore, we provide the average of the individual models trained on individuals with specific medical conditions ('Avg. Ind. Disease Models') and the average of the individual models trained on healthy and diseased individuals ('Avg. Ind. Models').

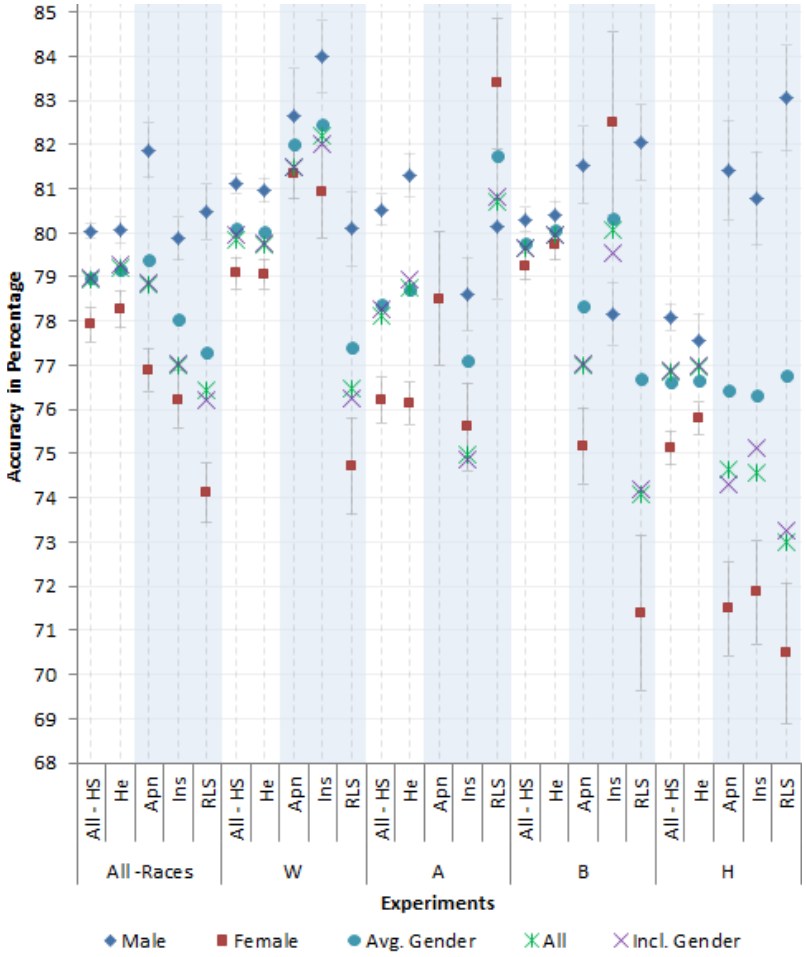


Figure 5.7: Outcomes for actigraphy data: Accuracy with SD for female (Female), male (Male), and both together (All); averaged female and male (Avg. Gender); and both together including gender as a feature (Incl. Gender).

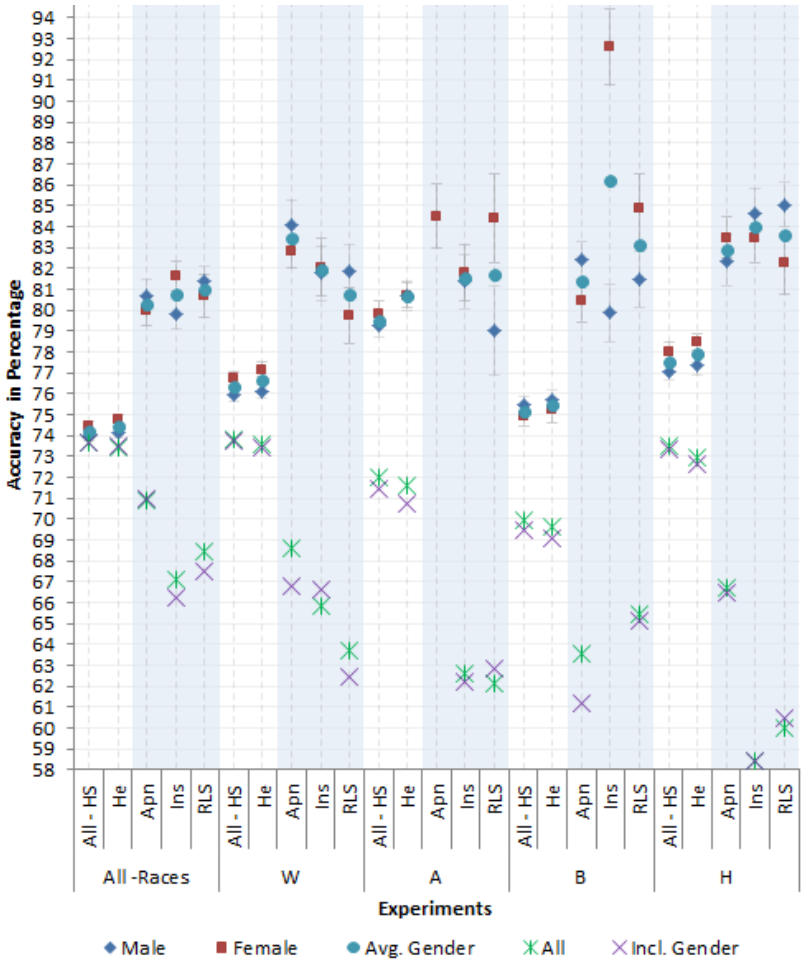


Figure 5.8: Outcomes for HRV data: Accuracy with SD for female (Female), male (Male), and both together (All); averaged female and male (Avg. Gender); and both together including gender as a feature (Incl. Gender).

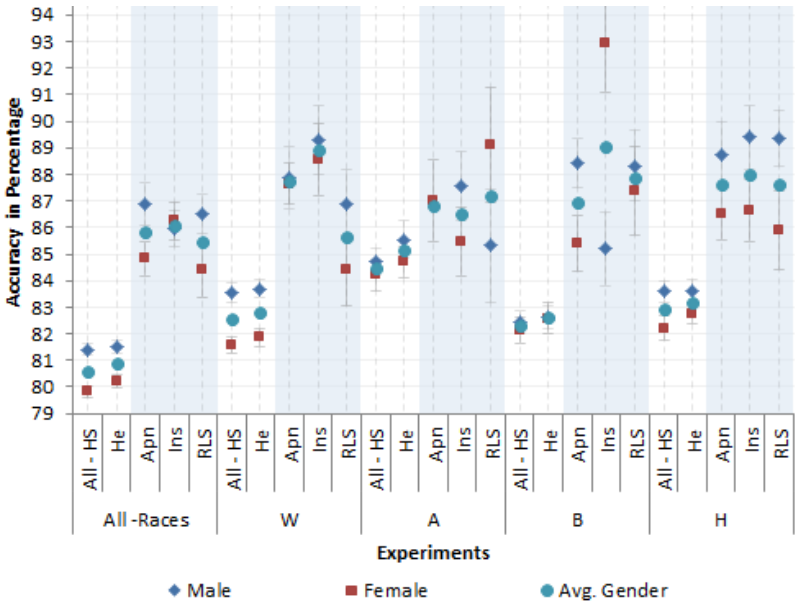


Figure 5.9: Outcomes for the combined data of actigraphy and HRV: Accuracy with SD for female (Female), male (Male), and both together (All); averaged female and male (Avg. Gender); and both together including gender as a feature (Incl. Gender).

From Figure 5.10, an apparent performance difference between male and female groups is visible in all investigations (e.g., 'All-Races, Males' reaches 80.4% while 'All-Races, Females' reaches 77.01%). Additionally, it is difficult to detect the sleep-wake stages for Hispanic females—this problem does not exist with Hispanic men.

When comparing figures 5.11 and 5.12, we can see a similar trend for HRV and the combined case; the trained models are most distinctive in the HRV and combined case (Figure 5.12), meaning models trained on diseased and healthy individuals are preferable. For example, for white males, 85.85% is reached with the fine-grained model compared to 82.73% and 81.43% for 'Avg. Disease Trained' and 'Avg. Healthy Trained', respectively. This trend is also present in Figure 5.10 but only when race is considered (e.g., for Asian males: 82.28% is reached for 'Avg. Ind. Models', 80.17% for 'Avg. Healthy Trained', and 79.87% for 'Avg. Disease Trained').

5.4.4 Comparison to Existing Methods

In Table 5.3, we compare our proposed model to standard sleep-wake recognition (presented in Section 2.4.3). Note that we calculated the macro-average for recall on a balanced dataset, meaning that recall and accuracy are the same. It is not possible to accurately compare our outcomes with Wolz et al. [47] and McDowell et al. [13], as their ground truth is provided by an actigraph unit while ours is based

Figure 5.10: Diseased-individual- and healthy-individual-trained models tested on diseased and healthy individuals for actigraphy data.

Figure 5.11: Diseased-individual- and healthy-individual-trained models tested on diseased and healthy individuals for HRV data.

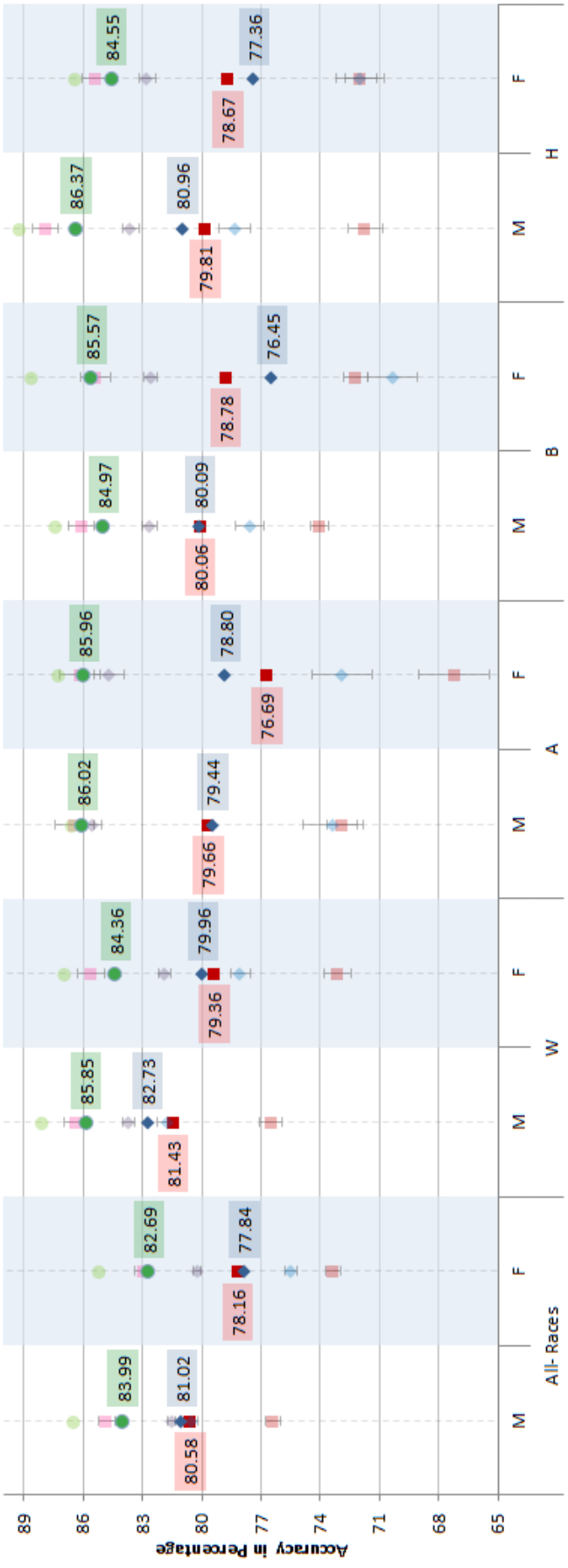


Figure 5.12: Diseased-individual- and healthy-individual-trained models tested on diseased and healthy individuals for combined actigraphy and HRV data.

on PSG. Only Uçar et al. [46] provide a 10-fold CV. However, we also repeat this analysis six times and on a bigger dataset to provide results that more accurately represent the general population. Our model is trained on the highest number of individuals and with the highest level of diversity of any study done before on this topic. Overall, we can reach the highest κ of 0.69 using our fine-grained approach on gender and health statuses by combining two data sources.

Evidence shows that camera recordings lead to the highest levels of accuracy [167]. The studies, conducted by Kuo et al. [48] and Khademi et al. [128], are very similar in accuracy, though Khademi et al. [128] reach a recall of 38%, which is considered low. Our accuracy is lower than that of Kuo et al. [48], but we use a balanced dataset in terms of gender and sleep-wake stages with more participants. We also use processed activity counts and light levels instead of raw accelerometer data. Additionally, our approach reaches higher κ values than Parro and Valdo [219] despite us including both males and females while they only included males.

Table 5.3: Comparison with other methods.

	No	HS*	Sensor	Feature No	Acc.	Rec.	κ	Method
[46]	10	Ap	PPG	28	77.4	79.0	0.59	KNN
[167]	10	-	Camera	-	92.1	-	-	-
[48]	81	poor/ good SE	Actigraphy	2	89.7	92.9	0.62	-
[128]	54	diverse	Actigraphy	39	87.0	38.0	-	XGB*
[219]	43	male	Actigraphy	-	85.3	-	0.53	RQA*
Ours	1,641	diverse	ECG	18	79.1	79.1	0.58	MLP
	1,576	diverse	Actigraphy	1+6	79.0	79.0	0.58	MLP
	1,576	diverse	Combined	25	84.6	84.6	0.69	MLP

* HS- Health Status; RQA - recurrence quantification analysis; XGB - extreme gradient boosting

5.4.5 Personalised Sleep Parameter Extraction

We have shown that our sleep-wake recognition approach is an improvement in accuracy over the standard approach ('Original Sample'). We use this approach as the basis for our personalised sleep parameter extraction, which we investigated here as a means to obtain SOL. We compare all investigated approaches to the gold standard of sleep assessment, PSG.

In Table 5.4, we present an optimal threshold level for the original—a one-threshold-fits-all approach—and our approach. Furthermore, we investigate whether personalising the thresholds results in improvements to the approach. The original SOL

calculation (Org.) uses the 15-minutes rule on the original sleep-wake stages obtained from actigraphy (provided by NSRR [202]). Our variable-threshold process incorporates personalised sleep stage detection. In searching for one optimal threshold for all groups of individuals, we reach improvement over the original in mean difference with a threshold level of 25 minutes. The mean difference is lower for our sleep-wake analysis approach with 20.16, compared to 20.74 in the original. When we use the full possibility of our proposed approach, we explore dynamic thresholds (variable-minute approach) that are adapted to various subgroups. The personalisation of these thresholds is relevant; therefore, we explored threshold values between 2 and 50 for categories based on gender, health status, and race. This produces a better correlation with PSG at 17.78 mean difference, superior to 20.74 for the original and 20.16 for our optimal static threshold. Furthermore, we investigate the Pearson and Spearman correlation with PSG; Pearson correlation represents linear relations and Spearman correlation depicts monotonic relations. It should be noted that there could still be a relationship even if both are found to be zero, as some cases are not represented (e.g., a parabola). Our investigations reach a lower Spearman correlation, however, our approach reaches a lower mean difference, which we consider the leading indicator. Thresholds must be adjusted alongside changes to personal factors to reach the optimal match with PSG. A comparison between the original 15-minute rule and our variable personalised threshold (variable-minute approach) is depicted in Figure 5.13. For male participants (shown in green), improvement is visible. Overall, our approach tends to underestimate SOL while the original tends to overestimate it.

5.5 Discussion

5.5.1 Comparison of Approaches

The difference between the original and the downsampled case indicates that sleep states, which are downsampled, more often match with the sleep stages of the ground truth (PSG) than it is the case with the wake stages. This demonstrates the better fit between the majority class (sleep states) and the ground truth from PSG.

The lower outcome for ‘Avg. Health Status’ demonstrates the influence of the training data on the recognition rate and the importance of fine-grained approaches, as we can increase the performance by personalising the model in terms of gender and race.

Table 5.4: SOL: Correlations and sum of differences.

Approach	Threshold	Correlation		Mean Difference
		Pearson	Spearman	
Original	15	0.68	0.73	20.74
Our	25	0.63	0.57	20.16
Original Male	15	0.63	0.73	26.50
Our Male	15	0.63	0.62	24.08
Personalised Thresholds				
Gender-Specific				
Female	25	0.60	0.50	20.29
Male	25	0.72	0.69	19.94
			Mean	20.16
Health Status-Specific				
Healthy	25	0.67	0.60	19.98
Apnoea	20	0.61	0.56	14.20
Insomnia	20	0.62	0.50	27.43
RLS	25	0.65	0.64	18.09
			Mean	19.78
	HS*	Gender	Mean	19.67
Race-Specific				
White	25	0.60	0.54	17.84
Asian	20	0.77	0.62	17.28
African	20	0.66	0.58	22.87
Hispanic	25	0.58	0.60	23.56
			Mean	20.05
	Race	Gender	Mean	19.69
HS*	Race	Gender	Mean	17.78

HS - Health Status

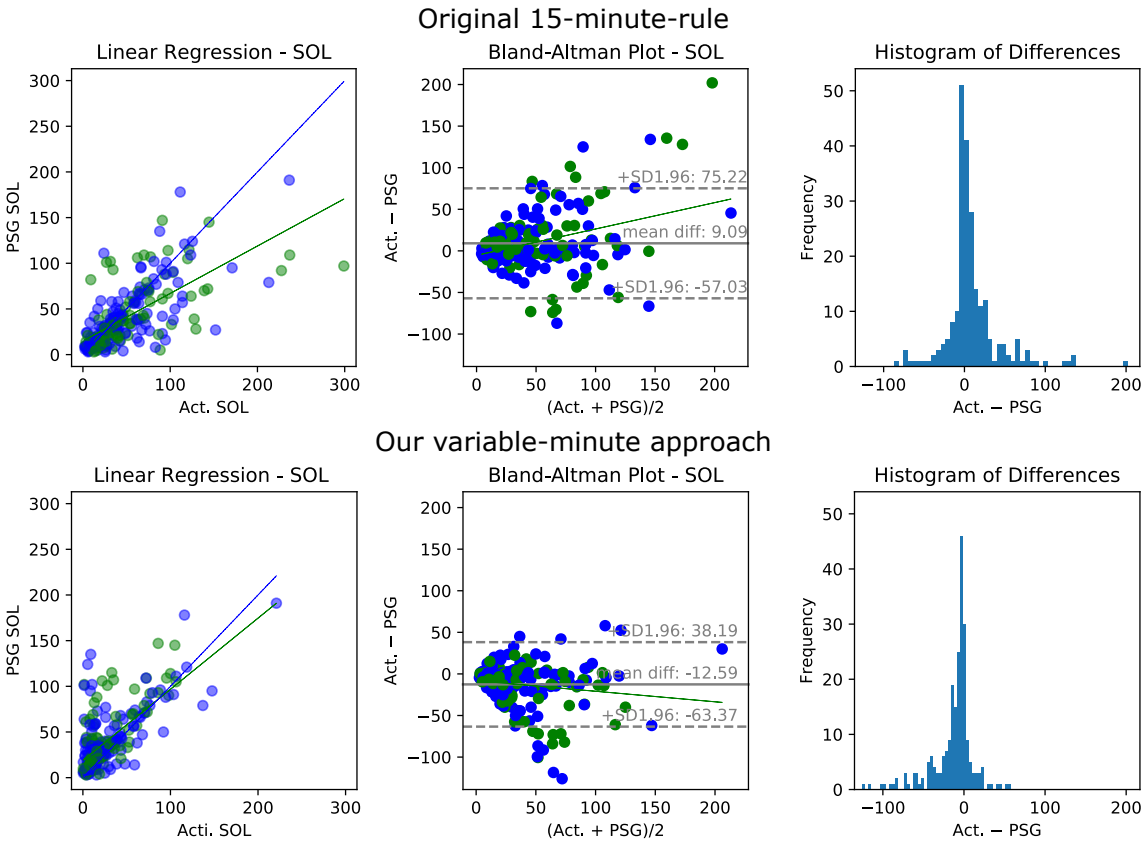


Figure 5.13: Comparison of SOL approaches; the original sleep-wake algorithm using a 15-minute-rule and our fine-grained model using a variable-minute approach with PSG. Linear regression, Bland-Altman plot, and histogram of differences are presented.

The findings show that fine-grained models have an advantage over standard one-model-fits-all approaches. Notably, individuals with a medical condition show an improvement in performance (e.g., 'Avg. Gender and Race' (79.05%) and 'All' (77.01%)). This is a positive result, as diseased individuals are usually the target group for sleep behaviour analysis. This effect likely comes from the fact that gender, race, and health status are influence factors for sleep.

Another finding is the importance of properly including the information during the training process; we tested the inclusion of gender as a feature 'Inc. Gender' and gender represented as two separate models. The gender-specific models can represent the differences in gender better.

5.5.2 The Impact of Personal and Physiological Data

The differences in SD between the general approach and the personalised gender-specific case show the stability and reliability of our fine-grained perspective in sleep-wake recognition, as the SD is lower for repeated analysis.

The finding that SD for HRV data is lower in the imbalanced case than it is in the balanced case suggests that a bias is introduced and the majority class is represented better. Additionally, it suggests that patterns are different between individuals and different timepoints. This can be explained by changes in HRV between different sleep stages; therefore, HRV was generally used, in literature, to investigate higher-granularity stages. In contrast, actigraphy is based on movement being more consistent over time, so this effect is not as prominent.

The finding revealing that models for male participants tend to perform better than models for female participants suggests that males and their movements are more consistent while females vary more by individual components, meaning that individual-trained models would have an advantage.

The number of individuals affects performance and must be considered when reviewing the outcomes.

The result that SD is lower for models on healthy individuals and all health statuses together relative to disease-specific models could either be caused by the influence of the individual numbers or the classification task at hand.

The average SD for HRV is higher than for actigraphy or the combined case. These findings indicate that actigraphy has a positive influence on SD in the combined case.

Which data source is ideal varies by case and desired outcome. With actigraphy, outcomes are more stable over all fine-grained models, while HRV shows higher variability (ranging from 74.2% to 86.2%; see Figure 5.8). However, certain subgroups face difficulties and low reliability with actigraphy but good performance with HRV, such as Hispanics. These findings demonstrate that the choice of data source is quite relevant when detecting sleep and wake stages.

We found that HRV performs better when considering diseased individuals rather than healthy individuals. This suggests that the HRV differences between sleep and wake stages are more significant in diseased cases. Actigraphy, however, shows more differences between male- and female-individual models while disease-specific models perform slightly worse than those with healthy individuals.

Overall, a combination of HRV and actigraphy performs best. Presumably, actigraphy pushes the recognition rate in healthy cases while HRV pushes it in diseased cases. It is likely that the actigraphy and HRV data are independent of each other and are thus able to capture different behaviour. As a result, a combination of the data sources features is a promising approach. This is an important finding which, to our knowledge, has not been investigated before.

5.5.3 Models Trained on Healthy and Diseased Individuals

In our investigations, male and female groups show differences in performance for actigraphy data. This relates to the idea mentioned earlier of males having less variability in body movements. For specific groups, sleep and wake episodes are more difficult to separate for females (e.g., Hispanics). This could be caused by the smaller individual set of 157 females compared to the 181 in the male case (see Table 5.2). More likely, however, is that the gap is caused by underlying differences in sleep-wake patterns between males and females. This, once again, makes the personalised fine-grained models a great solution for more reliable sleep-wake stage recognition.

Our results demonstrate that individual-trained models on healthy and diseased individuals are preferable. Our fine-grained sleep-wake behaviour analysis approach is the optimal method of investigation. Furthermore, disease-trained models ('Avg. Ind. Disease Models') are preferable for HRV, whereas for actigraphy, it is not so clear; best models need to be determined based on individual use cases.

5.5.4 Comparison to Existing Methods

Our investigation is the only one provided in Table 5.3 which considers a balanced dataset for training (see figures 5.5 and 5.6); however, downsampling can influence the recognition rate. We are also the researchers to investigate, in a study on this subject, the combination of multiple data sources and the relevance of clinical history and personal information. We incorporate the highest number of individuals and the highest level of diversity. We use sensors that are easy to apply at home and do not interfere much with the privacy of individuals; video recordings, used in studies like Liao and Yang [167], continue to raise privacy concerns. Characteristics of our validation data—fewer features and higher accuracy—make our HRV model a better choice than other methods, such as Uçar et al. [46]. Overall, we reach the highest κ value, 0.69, by using our fine-grained approach on gender and health statuses and combining two data sources. Many approaches test against just one technician, making that technician's style too influential [141]. Our dataset is scored by multiple technicians, so it does not encounter this issue. In addition, we investigate the combination of actigraphy and HRV as a means to enhance model accuracy, which was achieved.

Compared to other approaches, we simply use five features, one from actigraphy and four from light levels, and can reach good κ results for a balanced dataset, but we do not reach the same accuracy rates. This gives us the impression that the larger and more diverse dataset used in our study influences the performance and, in turn, better represents the general population. Overall, a balanced dataset represents a realistic recognition rate but leads to a decrease in performance and higher variability in SD, such as in the original sleep-wake recognition model (see figures 5.5 and 5.6). However, our multi-source data learning approach for granular sleep-wake behaviour detection adapts to these effects and provides comparable accuracy, improving when multiple data sources are available.

5.5.5 Personalised Sleep Parameter Extraction

The personalised sleep parameter extraction shows a better mean difference and indicates that the thresholds must be adjusted to fit better with varying personal factors. This means, e.g., that sleep apnoea and healthy individuals require different thresholds. Furthermore, results indicate that our multi-source data learning approach for sleep-wake stages applied in a real-world environment can improve sleep parameter extraction. The variable threshold technique shows advantages in

the Bland-Altman plot with a better fit but tends to underestimate the SOL. Particularly for male participants, SOL can be improved by our method, likely due to their more consistent sleep-wake patterns. Overall, we can see that a personalised approach is optimal; however, further analysis is still necessary to determine the best thresholds levels for all groups.

5.5.6 Limitations

Limitations in our analyses are (1) the necessary inclusion of clinical history and personal information; (2) downsampling; (3) lack of comorbidity consideration; (4) varying number of individuals; (5) HRV's extraction from ECG. First, clinical history and personal information must be available for our granular adaptive approach. If none of this information is provided (may not be shared by real-world users), a fine-grained sleep stage detection cannot be provided. Additionally, our approach does not cover for non-binary individuals; to overcome this future researchers need to determine whether their sleep behaviour corresponds to gender identification or sex-at-birth. Second, when downsampling is performed, not all the data from the majority class are necessarily used, even if repeated; to reach a balance in sleep and wake stages, the majority class is reduced. We sampled randomly for each repetition and did not cover for already-used or not-used periods. This means potentially that not all data are used. Third, comorbidity is not considered in our approach; it may have an influence on the recognition rate that we are unaware of. Fourth, as mentioned throughout this chapter, the number of individuals varies within groups and differences in individual counts can influence performance. Fifth, HRV features are extracted from ECG, which has more contact points with the human body and is considered less comfortable than PPG. Monitoring PPG with pulse rate variability is a promising alternative to HRV extraction from ECG with good coherence [220]. Another potential limitation could come from the way data has been collected, as PSG and actigraphy have been gathered simultaneously. Therefore, PSG could potentially influence the natural sleep habit negatively. This issue is difficult to overcome, because a ground truth is necessary.

5.6 Summary and Future Directions

This chapter contributes to knowledge in two folds. First, we proposed a fine-grained sleep-wake behaviour analysis approach that is adaptable to gender, race, and health

status as well as accessible physiological data sources. Second, a variable threshold technique for sleep parameter extraction was developed which incorporated the sleep-wake detection approach. The validated sources include HRV and actigraphy data. Analysis using our fine-grained approach revealed the influence factors on performance for sleep-wake recognition. These proved useful when categorising for variable threshold extraction. The investigations produced results comparable to those found in the sleep-wake detection literature; our use of variable thresholds in sleep parameter extraction constituted an improvement over the currently available ground truth. Our findings are as follows. First, gender-separated models provide higher accuracy, indicating that gender is an influence factor on how wake and sleep periods are differentiated. Second, disease-dependent classification is relevant, as it shows clearly better results compared to the healthy cases for HRV and the combined case. Third, the inclusion of multiple data sources (the combination of HRV and actigraphy) improves the recognition rate considerably and makes our approach more closely matching with the gold standard, PSG. Fourth, the findings demonstrate the potential for variable thresholds for sleep parameters extraction.

The basis for sleep behaviour investigation is a reliable sleep-wake recognition approach. We have shown that (1) personalised models, considering clinical history and genetic information and (2) models trained of multiple data sources, can improve the performance of such approaches. These outcomes serve as a basis for reliable sleep parameter extraction through variable threshold adaptation for our sleep behaviour assessment framework. Incorporation of the knowledge brought about by our investigations could lead future researchers to new insights. Perhaps our personalised approach could eventually find similar success even when sleep is divided by the guideline-specific stages: N1, N2, N3, and REM. We propose an investigation of multiple two-class classification problems that would ultimately be fused together into an overall model. One would distinguish between REM from NREM sleep, one would distinguish between light (N1, N2) and deep sleep (N3); and one would distinguish between N1 and N2. This method includes multiple personalised models for pairs of sleep stages, each optimised for an individual problem. Furthermore, the incorporation of multiple data sources suggests a promising direction for boosting recognition rates. Future research could look at new combinations of already-investigated data and potentially open up new avenues of investigation for sleep-wake classification. The variable-threshold process allows for the possibility of personalised, home-based sleep health monitoring, which is reliable.

Sleep stage behaviour analysis is currently used to detect anomalies and diagnose diseases in a laboratory. Home-based analysis makes a proper investigation of the

relationship between sleep and chronic disease possible. To show the necessity of such an investigation, we assessed the perspectives from individuals affected by medical conditions and doctors in related fields. We targeted specific chronic diseases to investigate how chronic diseases can induce sleep changes.

Chapter 6

Understanding of Sleep and Chronic Disease Relations

6.1 Introduction

In previous chapters, we developed methods to improve sleep behaviour analysis for home-based assessment. Past research and our investigations have demonstrated relations between sleep architecture changes and medical conditions. To effectively implement home-based assessments, the use of daily technology is key—individuals should feel comfortable with it. We performed semi-structured interviews to gather perspectives from individuals and doctors on sleep issues, technology usage, and their daily lives.

This chapter addresses elements that are missing from the literature, including (1) semi-structured interviews to assess influence factors, symptoms, and used sleep assessment methods; (2) investigations into sleep perspectives of doctors and individuals affected by a medical condition; and (3) discussions on differences in those perspectives. We interviewed individuals suffering from a chronic disease about (1) their sleep issues; (2) the effects of sleep quality on their daily life; (3) the influence factors for their sleep issues; and (4) their usage of technology. From doctors, we were interested in their views on: (1) sleep issues; (2) prominence of sleep issues in their specific field of medicine; (3) influence factors for sleep problems; and (4) technology usage at home. Subsequently, we address objective 7, as an in-depth understanding of sleep and chronic disease relations is developed. We use a qualitative approach to obtain the perspective of reported sleep issues in the Chinese population. We looked at the Chinese population because they are under-represented in current studies. Throughout this chapter, we refer to individuals living with a chronic disease as *affected individuals*.

Structurally, this chapter is divided into four sections. First, our research methods are presented in Section 6.2. Second, the results from our interviews are provided in Section 6.3. Third, we discuss the findings on sleep and perspective differences, as well as the limitations of this study, in Section 6.4. Lastly, Section 6.5 includes a summary, implications for future studies, and the focus of the following chapter.

6.2 Methods

For this study, we conducted explorative qualitative analysis with semi-structured interviews. In this section, we explain the methods used for planning, sampling, data collection, and data analysis.

6.2.1 Design and Settings

The interview process was planned to take place in China at the Affiliated Zhongshan Hospital of Dalian University over one week in October 2017. This hospital has 2,200 beds, 2,887 employees, and can accommodate 5,000–7,000 people per day. At our request, the hospital director contacted doctors with experience in chronic diseases. All doctors interviewed were knowledgeable about chronic diseases and sleep problems. Participants living with a chronic disease were contacted by their doctors. Dalian University provided a translator to support our interviews with translations and to take notes. Additionally notes are taken by the researcher during the interviews. The combined notes were made into transcripts and subsequently summarised following the interviews.

We gathered a diverse sample from various departments to gather opinions from individuals with various chronic disease backgrounds. The goal was to include at least one expert's opinion and that of two or three affected individuals per chronic disease. Family members could be present during the interview and offer clarifications if necessary. We followed the inclusion and exclusion criteria, described in the next section, for the process of recruiting interviewees.

(a) Inclusion and Exclusion Criteria

Participants were included in the study if they fulfilled the inclusion criteria. The interviewed participants are either affected individuals or doctors in relevant fields.

Table 6.1: Inclusion and exclusion criteria.

	Inclusion	**Exclusion**
General	Adult Provide consent Volunteer	Individual under 18 years Don't provide consent
Individuals with a chronic disease	Suffering from chronic disease Able to communicate	Healthy individuals Unable to communicate
Doctors	Knowledgeable about sleep issues Specialist in a specific chronic disease Has contact with affected individuals Medical education	Doctors from unrelated fields Has no personal contact with patients

All participants were adults and gave consent to be interviewed and to have their data collected and analysed.

In this study, we recruited doctors from various departments who worked with individuals suffering from a chronic disease, such as insomnia, cancer, stroke, and arthritis. The doctors needed to have a medical degree, be a specialist for a specific chronic disease, and know about sleep issues and common sleep disorders. Furthermore, they needed to regularly be in contact with affected people.

The affected individuals needed to be able to communicate, understand the questions, give their consent, be willing to explain their sleep issues, and be conscious of their difficulties. As participant 10 was unable to communicate effectively due to the state of their disease, we excluded their data from the analysis.

6.2.2 Ethical Considerations

The investigation was performed with ethical approval from De Montfort University. All individuals participated in this study voluntarily. Potential participants were informed of the aim of the study, the questions, and the interview setting; written informed consent was received prior to data collection. The participants have the right to withdraw from the study at any time. All participant identities are kept completely anonymous and their data are stored securely. Participants did not approve the collection of audio data.

6.2.3 Data Collection

Interviews took place in the hospital for about 5–10 minutes with a translator present at all times. The interviews were not recorded; instead, notes were taken. Based on the knowledge and state of the interviewees, interview questions were sometimes adjusted or left out. Interviews with affected individuals were conducted by a researcher with a translator, doctor, and occasionally a family member (or multiple) present. Interviews with doctors were conducted face-to-face by a researcher.

Participants received the study information sheet translated into Chinese (by a colleague at De Montfort University). Details of the study were explained in Chinese by a translator. Participants could ask questions before giving their consent. Those who gave consent were interviewed by a researcher about their experiences with medical conditions. We followed two different protocols in a semi-structured style for doctors and affected individuals. The guideline questions were divided into 'Introduction Questions', 'Sleep Issues', and 'Self-Management'.

We started with introduction questions, such as *'What is your profession'* and *'What is your background'* for doctors, and *'How long have you been suffering from this chronic disease'* for affected individuals.

We asked for the doctors' perspective regarding sleep problems: *'Are chronic disease patients affected by sleep problems'* and *'Which problems do they usually have'*. Furthermore, we asked: *'What method is currently used to assess sleep'* and *'What would you suggest to people with sleep issues'*. For affected individuals, we asked *'What kind of problems do you have during sleep'*, *'Is it affecting your daily life'*, and *'Is it a burden'*. To gather more details on the nature of their problems, we investigated further: *'What difficulties and challenges do you face during sleep and during the day after a bad rest'*. Furthermore, we were interested in what they think affects their sleep: *'Do subjective or objective elements relate to your sleep quality'*.

As there is currently a trend towards home-based assessment and self-management, we were interested to know if doctors see an advantage in self-management tools and information they could extract from home-based devices. First, we asked if they know about options for home-based sleep assessment and self-management: *'Are their options to self-manage sleep issues from your point of view'*. Our interest was also to extract *'What information would you like to get from a patient that would help in the treatment decision process'*. Generally speaking, we sought to know *'Do you think the data collected could help you, as a doctor, check patient's conditions'* and, get more information on sleep problems such as potential sources

(e.g., environmental influences). We were also interested to ask affected individual *'Do you self-manage something in your daily life'* and *'Do you think technology would be helpful'*.

6.2.4 Participants

The survey was conducted in collaboration with Dalian University and the Affiliated Hospital of Zhongshan Hospital in China. Participants included 18 people from five different departments:

- Rehabilitation department
- Ear-Nose-Throat department
- Orthopaedic department
- Oncology department
- Neurology department

The participants were distributed as follows: 33.4% from the Rehabilitation department; 11.1% from the Orthopaedic department; 33.4% from the Oncology department; 16.7% from the Neurology department; and 5.6% from the Ear-Nose-Throat department (see Table 6.2).

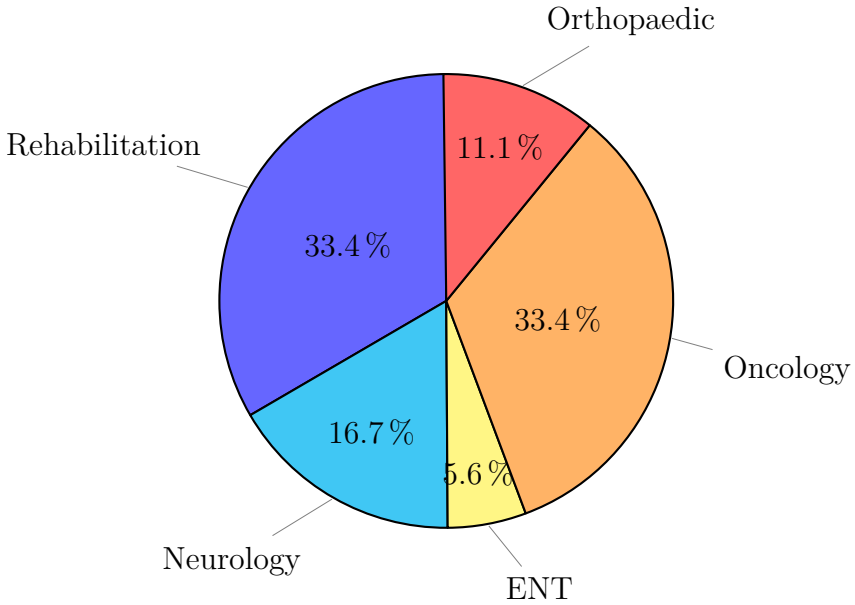


Table 6.2: Department statistics of participants.

We interviewed 6 females and 12 males, of which 7 are doctors and 11 are affected individuals. The gender distribution for doctors is 57.1% males (4 individuals) and 42.9% females (3 individuals); the gender distribution for affected individuals is 18.2% females (2 individuals) and 81.8% males (9 individuals). We recruited participants living with insomnia, cancer, stroke, and arthritis. The gender distribution for those subgroups can be found in Table 6.3.

Table 6.3: Number of participants in terms of gender and affiliation.

Source	Gender		Affiliation	
Affected Individuals	Female	2	Insomnia	1
			Cancer	1
	Male	9	Insomnia	1
			Stroke	4
			Cancer	3
			Arthritis	1
Doctors	Female	3	Neurology	1
			Rehabilitation	2
	Male	4	Oncology	2
			Orthopaedic	1
			ENT	1

6.2.5 Data Analysis

A qualitative content research analysis—an inductive thematic analysis—is conducted. The interviews were semi-structured and followed key questions to define the area of exploration and guide the participant through the session.

The process included the following steps (steps 5–9 are taken from [221]): (1) plan the study; (2) collect data; (3) transcribe and summarise each participant's interview notes; (4) compare notes with the translator; (5) familiarise ourselves with the data; (6) extract initial codes; (7) group the information thematically; (8) revisit the data to review the themes; and (9) name the themes. This includes the coding of the text, sorting of these codes, and categorising the statements of the respondents towards themes. The extracted codes are manifested in the individuals' answers; personal interpretation is kept to a minimum at this stage. Inductive thematic analysis is data-driven, which broadens the possible findings beyond the researcher's specific area of interest [221]. We further grouped the themes by doctors and affected individuals to allow for a discussion about the differences between them. Finally, we

provide an overall discussion about the themes and the relationships among them.

To provide a rigorous research, we included both genders for doctors and affected individuals to provide a rich sample of studied perspectives. During analysis, we revisited codes, categories, and themes to objectively extract our findings [222]. Selected interview notes are presented in our results to provide support for the extracted themes.

6.3 Results

In this section, we present the results of our thematic analysis. The analysis revealed six general themes: sleep problems, causes, impacts, assessments, treatments, and technology usage. Moreover, subthemes related to the overarching theme were identified (for details, refer to Figure 6.1). For readability and anonymity, we introduce an abbreviation—an ID—for each interviewee. The participants are categorised as either a doctor 'D' or an affected individual 'P' and numbered to create an individual ID. For example, D2 relates to a doctor with the ID-number 2. It should be mentioned that the included notes are not direct quotes but rather general sentiments. We indent the notes and include the person's ID, medical condition/expertise, and gender at the end of each note.

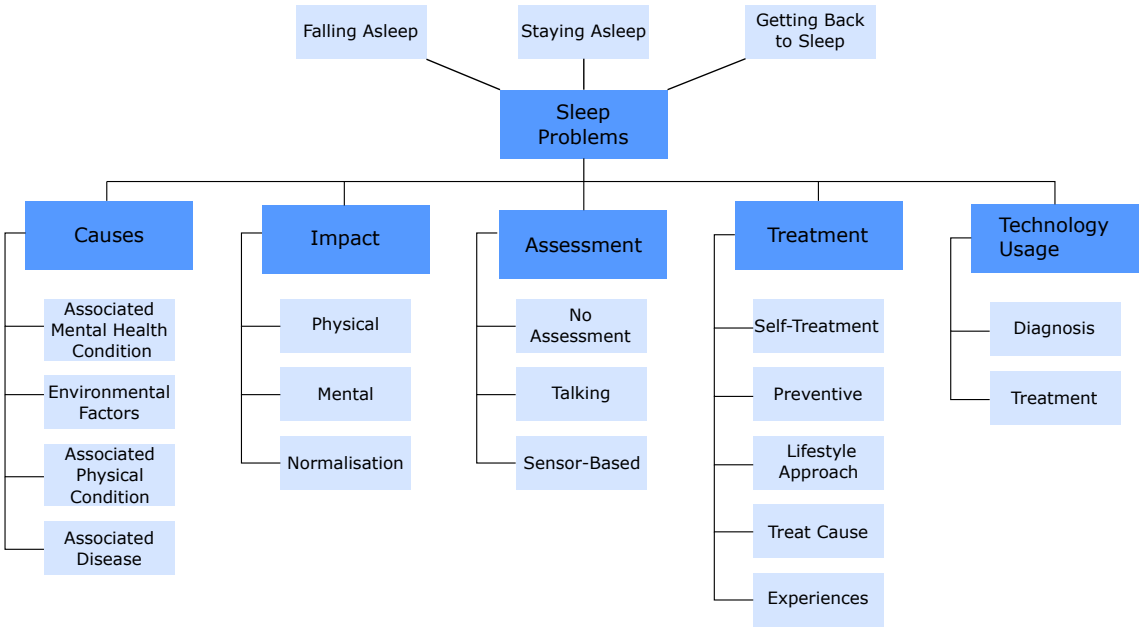


Figure 6.1: Thematic analysis for sleep perspective of doctors and affected individuals.

6.3.1 Sleep Problems

All affected individuals report that their sleep is troublesome.

> ... sleep is very bad (P7, Cancer, Male)
>
> ... sleep is heavily affected (P9, Cancer, Male)
>
> ... do not sleep very well (P5, Stroke, Male)

People generally report similar sleep issues concerning falling asleep, staying asleep, and getting back to sleep. In Figure 6.2, we can see that for affected individuals all three categories are equally common, while for doctors falling asleep issues are reported more frequently. This can likely be attributed to the doctors treating more patients that had a stroke, as stroke patients suffer from insomnia more often [89]. In the following subsection, we will provide examples and details for these three sleep problem classes.

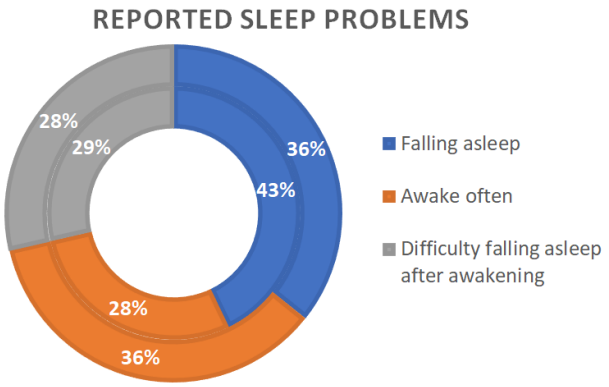


Figure 6.2: Statistics for the reported sleep problems from affected individuals (outer circle) and doctors (inner circle).

(a) Falling Asleep

Doctors and participants agree that falling asleep is an issue. It is perceived by many affected individuals independent from their condition:

> It takes more than 30 minutes to fall asleep without medication (P1, Insomnia, Female)
>
> ... huge problems falling asleep (P8, Cancer, Male)
>
> ... difficult to fall asleep (D1, Neurology, Female)

Interviewees use words such as “huge” and “difficult”, demonstrating the severe impact of their medical condition on their sleep. For some, it is an persistent issue, but appears not every day.

> 1–2 times per week, hindered ability to fall asleep (P11, Arthritis, Male)

(b) Staying Asleep

Sleep quality is strongly connected with sleep fragmentation, i.e., staying asleep throughout the night is essential to provide an individual with a good rest. However, many interviewees reported trouble staying asleep.

> Wakes up every 1–2 hours (P3, Stroke, Male)
>
> Wakes up three times each night ... has shallow sleep (P1, Insomnia, Female)
>
> ... wakes up a lot (P7, Cancer, Male)
>
> Individuals wake up often during the night (D3, Rehabilitation, Female)

(c) Getting Back to Sleep

When a person wakes up, it is often perceived as challenging to fall asleep again. Doctors are aware of this often-reported issue.

> It takes 2–3 hours to fall asleep again (P1, Insomnia, Female)
>
> ... cannot fall asleep again (P5, Stroke, Male)
>
> ... hard to fall asleep again (D4, Rehabilitation, Female)

6.3.2 Causes

The perspective of affected individuals is the most relevant, as they can explain present sleep symptoms and their relation to other chronic-disease-related symptoms they face. Doctors are able to provide an overall view on potential causes. Some interviewees see their sleep issues as being related to their mental health, environment, physical condition, or disease. In Figure 6.2, the main cause for sleep issues is reported to be the disease or comorbidities affected individuals are suffering from, followed by the physical condition and environmental influence aspects. For

doctors the main reported cause is the physical condition, followed by associated diseases. Mental Health and the environment can have an impact as well. Each of the different causes are further discussed in the following subsections.

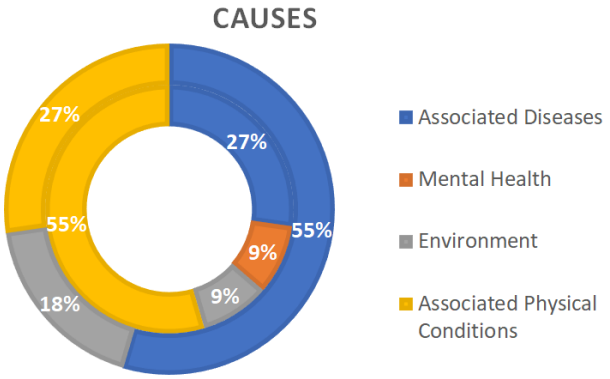


Figure 6.3: Statistics for the causes from affected individuals (outer circle) and doctors (inner circle).

(a) Associated Mental Health Conditions

Doctors mention mental health associations.

> Insomnia is often caused by mood and pressure in young adults (D1, Neurology, Female; comment: mood refers to anxiety and depression)
>
> Post-stroke depression can cause sleep issues (D1, Neurology, Female)

(b) Environmental Factors

Environmental factors affect sleep directly and are reported by many individuals and doctors.

> Light and noise influence sleep immensely (P7, Cancer, Male)
>
> Waking up is caused by the environment, like when my granddaughter is playing (P2, Insomnia, Male)

(c) Associated Physical Conditions

People also mention physical conditions as a cause of sleep issues, such as sleep positions, nocturia, pain, and lifestyle habits, refer to Figure 6.4. The statistics show a quite similar distribution for doctors and affected individuals the only difference is

that positioning is mentioned by doctors as associated physical condition. Overall, pain is the main cause for sleep issues followed by nocturia, and changes in lifestyle habits.

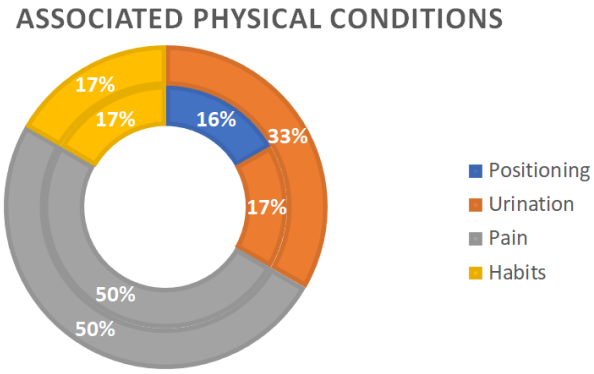


Figure 6.4: Statistics for the associated physical conditions from affected individuals (outer circle) and doctors (inner circle).

Sleep positions are mentioned by one of the doctors in relation to breathing issues.

> Positioning affects breathing while asleep (D2, Ear-Nose-Throat, Male)

Furthermore, nocturia is mentioned mainly by individuals who have suffered a stroke.

> ... need to go to the bathroom (D3, Rehabilitation, Female)
>
> ... need to release during the night is directly connected to sleep issues (P4, Stroke, Male)

Pain as an influence factor is controversial among affected individuals and doctors. Doctors relate sleep issues to specific pain levels but individuals report issues at lower thresholds.

> ... exists a relationship between sleep issues and pain (P11, Arthritis, Male)
>
> ... suffers from sleep issues (P9, Cancer, Male, Pain Level: 4)

Some doctors have a different opinion on pain in relation to sleep.

> Very seldom do patients awake because of pain (D7, Arthritis, Male)
>
> Sleep issues related to pain start at a pain level of seven (D6, Oncology, Male)

Lifestyle habits are another cause, as well as a consequence, of sleep issues.

> ...being bound to the bed is why I developed insomnia (P2, Insomnia, Male)
>
> Getting up and going to sleep very early decreases sleep issues (D1, Neurology, Female)
>
> Sleep quality improves when going to be early (P1, Insomnia, Female)

(d) Associated Diseases

Affected individuals often suffer from comorbidity, which is mentioned as one of the leading causes of sleep issues along with chronic diseases generally.

> There exists a relationship between most chronic diseases and sleep issues (D1, Neurology, Female)
>
> Stroke is strongly related to sleep (D3, Rehabilitation, Female)
>
> Sleep is affected by high blood pressure (P5, Stroke, Male)
>
> Hypertension is related to sleep issues (P6, Stroke, Male)

Some individuals mention that certain conditions do not relate to sleep issues

> Cancer does not influence insomnia; no relationship between diabetes and sleep issues (P1, Insomnia, Female)
>
> Only 10–15% of stroke patients have sleep issues (D4, Rehabilitation, Female)

6.3.3 Impact

In general, bad sleep affects individuals' daily lives and their perceived health status.

(a) Influences on the Next Day

Most of the affected individuals reported influences on the next day.

> Next day is considerably influenced by bad rest (P4, Stroke, Male)

(b) Mental Condition

Sleep issues can cause mental health issues, such as hallucinations, anxiety, and feeling annoyed, refer to Figure 6.5. In the statistics, we can see that sleep is often reported by doctors and affected individuals to impact anxiety, while feeling annoyed is only reported by affected individuals. Hallucinations can be caused by sleep issues as well and are reported by doctors and affected individuals.

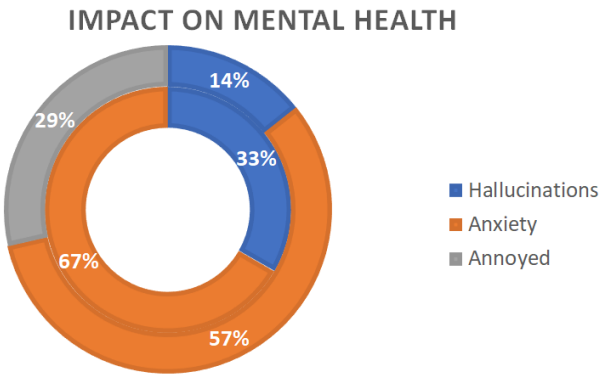


Figure 6.5: Statistics for the impacts on mental health from affected individuals (outer circle) and doctors (inner circle).

> Sleep issues can cause mental health disorders, such as visual hallucinations and anxiety (D1, Neurology, Female)
>
> ... dreams often ... sometimes affected by anxiety (P1, Insomnia, Female)
>
> Sensitive nerves (P7, Cancer, Male)
>
> Individuals feel anxiety and often feel worried (D3, Rehabilitation, Female)
>
> ... feels annoyed and depressed the next day (P2, Insomnia, Male)
>
> ... feels miserable (P9, Cancer, Male)

Furthermore, people can feel hopelessness and despair.

> ... sickening feeling during the day (P8, Cancer, Male)
>
> Daily life considerably affected (P4, Stroke, Male)
>
> The situation is uncomfortable (P6, Stroke, Male)
>
> Sleep is heavily affected (P9, Cancer, Male)

Affected individuals are concerned with their family's perspective of their state.

> Family members would describe me as not anxious and not depressed, while my perception is the opposite (P1, Insomnia, Female)
>
> My family would describe me as being moody (P2, Insomnia, Male)

(c) Physical Condition

Physical condition is reported as affected the next day. The impact on the physical condition is shown in Figure 6.6. We can see that affected individuals only report feeling tired the next day and dizziness likely caused by blood pressure issues. Doctors further report the influence on the reaction time, which is reasonable as after a bad rest, affected individuals are often exhausted and have concentration issues [36].

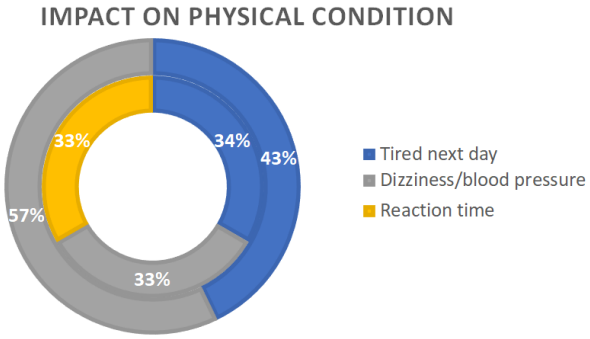


Figure 6.6: Statistics for the impacts on physical conditions from affected individuals (outer circle) and doctors (inner circle).

> Feeling tired after getting up (D1, Neurology, Female)
>
> Exhausted and no strength the next day (P3, Stroke, Male)
>
> Difficulty getting up (P5, Stroke, Male)
>
> Dizziness has already led to multiple falls (P1, Insomnia, Female)
>
> Slow reaction time (D1, Neurology, Female)

(d) Normalisation

Affected individuals tend to normalise their condition [69].

> P1 explains that she was a pig farmer. Therefore, she is used to being tired. She does not care a lot about sleep and states that her life is not affected too much. It is a phase that started to get worse in the last few months (P1, Insomnia, Female)

> Insomnia has phases (P2, Insomnia, Male)

Responses suggest that some individuals get used to insomnia. For example, P1 mentioned that her daily life is not affected, as she was used to heavy work and, as a consequence, being tired. Doctors who are not directly involved in assessing sleep, but who specialise in a chronic disease, often normalise the sleep condition of their patients.

> People are often sick and tired, so they most often do not have sleep issues (D4, Rehabilitation, Female)

6.3.4 Assessment

In general, three types of assessment are conducted by doctors: no assessment, talking, and sensor-based assessment, where no assessment is the most common, for details, see Figure 6.7.

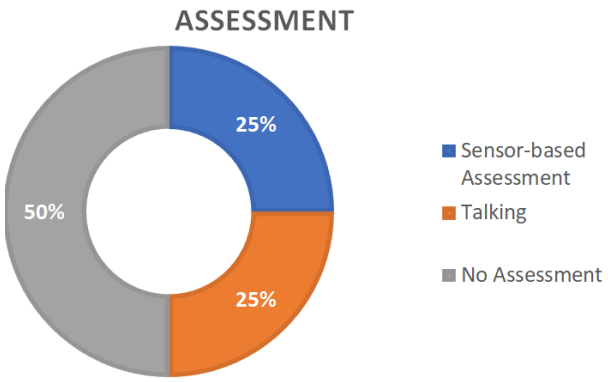


Figure 6.7: Statistics for the used methods for assessment from doctors.

(a) No Assessment

Most commonly reported was that no sleep problem assessments take place.

> In general, no assessment is used (D3, Rehabilitation, Female)

Some doctors are also aware that current methods are not promising and a change is necessary, others make aware of the difficulty of assessment for certain patient groups.

> No questions are asked about sleep but would be necessary (D4, Rehabilitation, Female)

> Difficult to communicate with patients (D3, Rehabilitation, Female)

(b) Talking

Talking with an individual can provide insight into their issues and symptoms. Doctors typically talk with their patients but do not make scaled assessments.

> Talking face-to-face with the patient as part of a psychological assessment is the method used to assess sleep issues (D3, Rehabilitation, Female)

(c) Sensor-Based Assessment

Sensor-based assessment involves various sensors applied to the body. This usually takes place in a laboratory under the observation of a specialist.

> For patients with breathing issues, a two-part assessment is done during a sleep-lab assessment. First, sleep quality is assessed based on sleep stages, awakenings, and oxygen levels. Second, a trend picture over time is created from the collected data (D2, Ear-Nose-Throat, Male)

> No sensor-based assessment is done (P1, Insomnia, Female)

6.3.5 Treatment

Treatment is essential for affected individuals as their life is severely impacted by their issues. We discuss in this section the different treatment methods shown in Figure 6.8. Especially common for affected individuals is self-treatment, followed by treating the cause usually with medication, and changing the lifestyle. Rarely preventative medication is reported by affected individuals, even though, it is commonly reported by doctors. This could come from people not being aware of their preventative treatment.

(a) Self-Treatment

A common method among our interviewees is self-treatment.

> ... uses traditional Chinese medicine; moving to another province improved sleep quality (P2, Insomnia, Male)

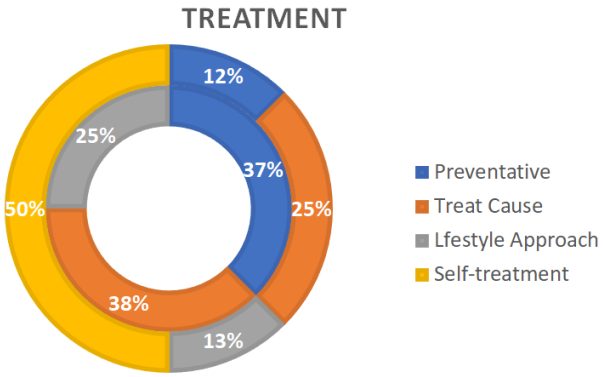


Figure 6.8: Statistics for applied treatment methods from affected individuals (outer circle) and doctors (inner circle).

> Going to bed early improves sleep quality (P1, Insomnia, Female)
>
> Change in the environment would help (P6, Stroke, Male)

(b) Preventive

Preventive methods are used to treat an issue before it appears.

> Preventive medicine is given to patients (D3, Rehabilitation, Female)
>
> Patients receive medication easily to sleep better, to keep them fit for the next day (D4, Rehabilitation, Female)
>
> Traditional Chinese medicine is used for prevention (P1, Neurology, Female)

(c) Lifestyle Approach

Changes in sleep hygiene can improve a person's sleep quality and is used as a treatment.

> Suggestions are given to improve health, such as exercising, listening to music, communicating with friends, bathing feet in hot water, drinking milk, and going to bed at the same time (D1, Neurology, Female)
>
> Lose weight, stop smoking, and stop consuming alcohol are suggestions to improve sleep issues (D2, Ear-Nose-Throat, Male)

(d) Treat the Cause

The cause is usually treated when pain occurs and is seen as the reason for sleep issues.

> Pain killers are the treatment; no pain means no sleep issues (D5, Oncology, Male)

Medications, such as sedatives, are given to patients with medical conditions.

> A combination of antidepressants and sedatives are given to patients (D1, Neurology, Female)
>
> ... receives medication (P4, Stroke, Male)

(e) Experience

Doctors point out that there is a lack of assessment and an issue with preventive medication being too easily prescribed.

> Patients are easily medicated to keep them fit for the next day of exercises (D4, Rehabilitation, Female)

Affected individuals have positive and negative experiences with treatment. Some see treatment as effective.

> Treatment helps (P9, Cancer, Male)
>
> Treatment and medication help (P6, Stroke, Male)

Others do not see an improvement

> Sleep issues arise even when on medication (P4, Stroke, Male)
>
> Medication does not change anything (P3, Stroke, Male)

6.3.6 Technology Usage

Most individuals have never used technology—only two have.

(a) Diagnosis

For assessment and diagnosis, technology could be very useful. Doctors with patients with sleep issues are keen to use technology and get insight into sleep behaviour.

> As many details as possible would be helpful in the assessment process (D1, Neurology, Female)

However, they also see disadvantages in current assessment methods and believe there is a need for improvement. Figure 6.9 shows that doctors are in general positive towards technology but see the need for change to make technology more reliable almost as equally important.

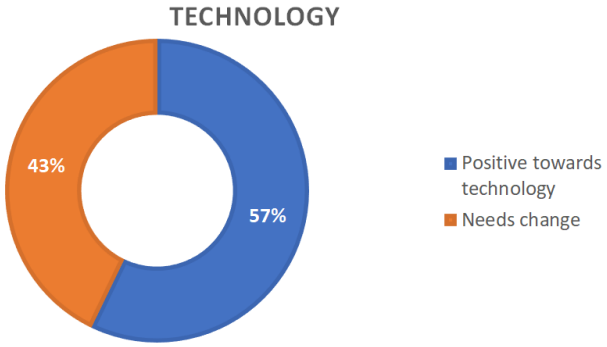


Figure 6.9: Statistics for the opinion towards technology from doctors.

> There are many elements in PSG. Significant education is necessary to become a sleep expert who can analyse the data ... the automated systems that are currently used are not reliable in detecting different sleep stages and need to be controlled by human experts. Therefore, automation is not as accurate as human assessment (D2, Ear-Nose-Throat, Male)

(b) Treatment

While most of the interviewees did not have experience with devices, some did report the use of technology.

> Blood sugar screening testing performed at home (P1, Insomnia, Female; comment: P1 did not view this as using technology)

The effectiveness of certain devices and new potential use cases were pointed out.

> Technological device to stimulate the body described positively as being useful (P5, Stroke, Male)
>
> No real issue if the pain level is measured, controlled, and treated; Technology could help in this aspect (D5, Oncology, Male; comment: D5 refers to tools which assess pain levels as technology)

6.4 Discussion

In this section, the findings are discussed and their relation to literature is examined.

6.4.1 Sleep Problems

Poor sleep quality is an issue for people living with a chronic disease [81] and has been reported by all of our interviewees. Our findings indicate that sleep quality affects people living with specific chronic diseases differently. There are three major sleep issues that are reported by affected individuals and doctors: problems falling asleep, problems staying asleep, and issues falling asleep after waking up.

6.4.2 Causes and Impacts

We focus our scope to analyse affected individuals' perspectives of sleep and their condition. Note that there are still other influence factors for sleep issues.

Arthritis and joint pain are directly related to sleep issues [91]. This relation is in line with our findings. We found that cancer-affected individuals report pain as a cause for awakening; this is supported by literature [67, 90]. Strokes have also been shown to influence sleep [89], which is supported by our findings (especially by participants with nocturia).

Insomnia-affected individuals have a very subjective perspective on their sleep. They misjudge specific objective symptoms as being more severe by, for example, overestimating SOL and underestimating total sleep time [68]. Nevertheless, insomnia symptoms are often trivialised [69], which is likely a cause behind affected individuals normalising their condition. For example, P1 denied that she was affected by insomnia but mentions symptoms including dizziness and fatigue. Interestingly,

there is an apparent gap in perspective between family members and affected individuals. For example, P1 describes herself as having anxiety but says that family members would describe her as neither depressed nor anxious.

In our interviews, environmental factors were often mentioned as a cause for sleep issues. This is likely due to the fact that people are more aware of their environment because they suffer from shallow sleep.

Our findings suggest that sleep issues are related to specific mental and physical conditions, including hallucinations, sensitive nerves, and feeling annoyed. The feeling of being annoyed is likely caused by feeling tired and weak. Doctors mentioned hallucinations, though affected individuals mainly referred to them as dreams.

Chronic diseases clearly influence sleep behaviour; on the other side, poor sleep affects peoples' daily lives and increases the risk of developing a chronic disease [80, 85]. In this study, all participants with various chronic diseases had sleep issues; most reported effects in their daily life. In conclusion, knowledge about the relationship between sleep and chronic diseases could enable data analytics to effectively diagnose medical conditions using sleep behaviour data (refer to Chapter 7).

There are many reasons why sleep issues are worth considering. For example, people suffering from cancer and arthritis often claim that they wake up or have difficulty falling asleep due to pain. Difficulty falling asleep could also be a sign of insomnia.

6.4.3 Assessment and Treatment

Treatment of sleep issues is not always adequate and medication is provided easily; behavioural change therapy is rarely suggested [90]. The main problems here are (1) failure to employ assessment methods and (2) preference of preventive medication. For some doctors, especially those who specialise in stroke and rehabilitation, it is challenging to communicate with their patients, as their diseases often make it difficult for them to communicate effectively. Many other fields simply view sleep assessment as unnecessary, likely because doctors receive no instruction on it. Some doctors see the importance of changing the current approach with stroke patients.

Many affected individuals treat themselves with traditional Chinese medicine and lifestyle changes. They try out various things to help them improve their sleep. Doctors also suggest lifestyle changes to their patients. Our interviews suggest that people are largely open to self-assessing and self-managing their sleep issues. This could be realised through the assessment tools we developed in combination with a

self-management approach, which could be developed in the future.

Affected individuals react differently to the same treatments. This suggests that it is best to follow a personalised approach in order to consider the individuality of each person.

6.4.4 Technology Usage

Most affected individuals in our study had a positive opinion towards technology despite having had limited exposure to it in the past. The individuals who had used devices for treatment reported positive effects. Doctors saw the advantage of technology in assessment and treatment, as it provides more information to help diagnose sleep diseases, assess pain levels, and adjust treatments. They also pointed out the importance of improving current technologies, such as sleep stage detection, to make them more reliable.

6.4.5 Differences in Doctors' and Affected Individuals' Perspectives

Overall, doctors and affected people mostly agreed on the discussed topics, but some notable differences were observed.

Doctors often had perspectives of their patients' sleep status that differed from the affected individuals' perspectives. Especially for participants with cancer, doctors did not seem concerned about their sleep issues. They claimed that cancer-affected individuals are only affected from a specific pain level threshold upwards, but affected individuals claimed that they faced issues at lower pain levels. Most of the doctors are generally not very concerned about sleep issues and are not instructed to assess patients' sleep quality. Only some doctors are aware of the current lack of assessment and would like to see a change from the standard of liberal medicating to cause assessment and treatment.

Technology was observed differently. Affected individuals felt positively about technology if they had experience with it; otherwise, they had no opinion. Doctors also felt positively but were often keen to point out issues in current technology and suggest areas for improvement.

Sleep assessment for individuals with a chronic disease is essential, as shown through our interviews. There is clearly a connection between chronic diseases and sleep

issues. In our research, it was commonly reported that chronic diseases change sleep behaviour. There is a significant gap between the awareness of affected individuals regarding their sleep issues and doctors' responses.

6.4.6 Limitations of the Study

This study was performed in just one hospital, potentially creating a bias in terms of race and ethnicity.

Gender was equally distributed for doctors but has a strong bias towards males for affected individuals. We did not investigate the gender difference in the perspective of sleep issues. However, there certainly could be differences, as women and men often perceive symptoms differently. It was shown in [223] that biological factors drive gender differences in sleep behaviour and disorders.

A translator was necessary for the interviews; the interviewer and interviewee did not share a cultural or language background. No audio data were recorded, as this was preferred by interviewees. The language barrier and lack of recordings together constitute limitations and surely influenced our results to some degree. Verbatim analysis from audio recordings constitute a more in-depth investigation, and with translations a certain mode of expression can be lost. This was counteracted, however, by comparing the notes taken by the translator and interviewer after each session, where discrepancies were addressed by consulting the translator. The sample of 18 participants is small, but our main goal was to capture general impressions from a diverse group; for that, our sample was certainly sufficient.

6.5 Summary and Future Directions

The chapter contributes to knowledge by developing an in-depth understanding of relevant stakeholders perspectives on sleep. Doctors' and individuals' opinions strongly matched on the sleep issues we asked about. Some doctors, however, were unaware of their patients' sleep issues or were not equipped to assess them any further. As expected, the reported daily life effects and sleep issues varied based on participants' conditions; this is in line with the literature. The environment was often mentioned as a strong influence factor, likely due to shallow sleep. Incorporating daily technology is essential for experts in the field, but users in our study were hardly familiar with technology, so it was not an aspect we could pursue.

Our findings confirm the results on sleep stages from Chapter 5—health status definitively plays a role. Our findings in Chapter 5 are in line with other studies on changes in sleep architecture stemming from certain chronic diseases [77, 78]. Additionally, we found that many current experts are not convinced regarding the reliability of automated sleep stage detection for PSG, as it requires an experienced scorer. Our fine-grained sleep stage recognition system aims to improve sleep stage recognition in a natural environment to avoid these doubts over reliability. Furthermore, one of the doctors we interviewed emphasised the importance of home-based, IoT-enabled assessment, which supports the design of our framework in Chapter 3. Sleep positions were mentioned as an influence factor for sleep quality, which confirms the importance of our investigation on higher-granularity sleep positions in Chapter 4. The findings encourage investigations into technology for sleep behaviour assessment and remote diagnostics. Sleep can change and affect daily behaviour as a result of chronic diseases. These effects can be used to explore further sleep and wake behaviour in relation to chronic disease detection, which we investigate in Chapter 7. Importantly, the work is a starting point to investigate individual needs and differences in perspective among doctors and affected individuals in more depth. We found that the individuality needs to be considered to provide effective treatment and assessment (which we also stressed to a certain extent in chapters 3 and 5). Additionally, our findings suggest that self-assessment could minimise improper treatment approaches, such as preventive medication.

In the following chapter, we investigate the possibility of automating chronic disease detection from sleep-wake behaviour. Based on our findings, we develop a home-based approach that detects the onset of chronic diseases. This approach can be used in our framework to provide a personalised and user-centred tool.

Chapter 7

Chronic Disease Detection from Sleep-Wake Behaviour

7.1 Introduction

In sections 2.2 and 2.4.5, we reviewed research on medical conditions and sleep that suggested a bidirectional and complex relationship between them. It has been clinically observed that people suffering from chronic diseases, such as diabetes [21, 26], CKD [32], hypertension [79], arthritis, and stroke [26], usually have trouble with sleep, such as difficulty falling or staying asleep and daytime sleepiness [26]. Our findings in Chapter 6 are clearly in line with the literature, as they show (1) relations between sleep behaviour and chronic disease as well as (2) the need to detect changes in circumstance (diseases). The relationship between chronic disease and sleep behaviour is recognised by both doctors and affected individuals. Many chronic diseases show evidence of correlations with sleep-wake behaviour and there is a rising interest in making use of these correlations for early warning systems. The diagnosis and early detection of medical conditions would lead to earlier starts for clinical interventions [70] and, therefore, an increased likelihood of success [41]. It is not possible to monitor sleep information flow continuously through traditional medical channels (doctor or hospital visits). Therefore, to reduce healthcare costs and improve lives, we propose a home-based monitoring system that is optimal for both patients and doctors. Our proposed framework is capable of continuously monitoring sleep and adapting to users' changing circumstances (e.g., developing a chronic disease). To provide this functionality, changes in circumstances must be assessed as soon as possible; the integration of a system capable of detecting changes in circumstances is essential. Already-known relationships with sleep-wake behaviour could be used to predict chronic diseases.

In this chapter, we address objective 8 by developing a method for chronic disease detection that entails mining sleep-wake behaviour with deep learning. Specifically, an LSTM is designed for this investigation, as it is able to deal with temporal data and multi-dimensional features. The differences between diseases are likely manifested in (1) the actigraphy sleep-wake patterns of the individuals, comprising information of the SE and sleep duration [131]; and (2) the clinical feature characteristics, such as BMI. Therefore, with actigraphy, features are extracted from consecutive days of monitoring and then enriched with data on patients' clinical history. The fusion of medical knowledge (existing correlations between clinical history and chronic disease development) with computational technology creates a strong basis to detect early and advanced stages of various chronic diseases. The backbone of our approach is made up of the multi-dimensional clinical data features.

This chapter contributes to the literature in two ways. First, we apply a computational technology that makes decisions from objective measurements. We design LSTMs specific to our investigations of interest. In contrast to previous work [18], our use of LSTMs enables us to incorporate the temporal aspects of the actigraphy data. Second, we fuse clinical history and actigraphy data to introduce a multi-dimensional feature vector. By fusing these data, we make use of existing medical knowledge. Furthermore, we perform tests on multiple chronic diseases. We analyse our results in depth by exploring significant differences in clinical features for correctly- and incorrectly classified individuals. These clinical features indicate a relationship with different disease stages and likely contain additional information. Through this novel approach, we address the challenges facing chronic disease detection and the limitations of previous work. These limitations include the lack of (1) fusion of medical knowledge with computational technology; (2) investigations into early disease detection; and (3) implementation of a reliable method that can handle temporal data. This approach contributes to the emphasis and personalisation layers of our sleep behaviour assessment framework. Furthermore, adjustments to changing circumstances in our framework's modular design become feasible for chronic-disease-affected individuals with this approach.

In Section 7.2, we describe the approach for chronic disease detection, including the initial method configuration. In Section 7.3, we explain the dataset and describe the test settings. Section 7.4 presents the results, which are further discussed in Section 7.5. Finally, Section 7.6 offers a summary of the work and future research directions.

7.2 Chronic Disease Detection Approach

In this section, we present an approach to detect chronic diseases in early and advanced stages, considering sleep-wake behaviour and clinical history data. In other words, medical knowledge and computational technologies are fused to enhance current diagnostic decision-making.

Sleep-wake behaviour can be measured continuously using an actigraph unit to provide time-series data. Time-series data are data that have time dependency; in our case, those are activity counts and white light data. Activity counts indicate the intensity of activeness over time. White light tells us the level of light at a specific time, helping to judge the day-night rhythm. In order to detect anomalies in sleep patterns, longer time periods must be considered, because looking at one-day solely cannot reflect the indicators of regularity. We decided on a deep learning method for time-series data classification—an LSTM (as described in Section 2.4.1). An LSTM can learn temporal dependencies in observations. It can link causes over time, as the method allows for continuous learning by applying gates which are able to memorise information [155, 156]. This is a key aspect that other machine learning algorithms lack; they are unable to handle delayed or sparse indications. Time-series data has already seen success in diagnostics [158], following an architecture described by Graves et al. [224].

We designed a specific architecture for the LSTM method presented in Figure 7.1. First, embedding is performed by applying a word-2-vec method for the time-series data (activity counts and white light). Embedding can extract useful features to reflect similarities and relationships within data. Originally, words are translated into a vector representation [225, 226]. Activity counts and white light data are used as inputs for the word-2-vec approach to improve the results. The technique converts the data into a smaller set of easier-to-use features, which are then used as inputs for the LSTM [225, 226]. Second, the extracted embedding features are fused with another source of information: clinical history data. Clinical history data are used as additional constant features over the entire time-series per person. This means, at each time point, the clinical history information stays the same per individual. Details on the features used can be found in Section 7.3.2. The combination of actigraphy data and clinical history data is used as an input for the LSTM. Third, the LSTM design is followed. The design is key to provide success and avoid overfitting. Various techniques can counteract overfitting. We integrated some of them into our design: (1) regularisation methods such as dropout and

(2) a balance between a powerful and, therefore, larger network (more layers and neurons) and the issue of overfitting, which is more likely when more parameters should be learned from fewer data. Therefore, a complex problem, such as in our case, requires multiple hidden layers. Our LSTM consists of three LSTM layers, each using a dropout strategy for training followed by a dense layer to flatten and fully connect. Dropout is used to ignore neurons during the training phase and prevent overfitting [227]. The rate of dropout can be set between 0.0 to 0.9 (0.0 correlates to no dropout). For example, setting the dropout rate to 20% means that two in ten inputs will be randomly excluded. Dense layers apply a matrix-vector multiplication, resulting in a transformation of the data into a lower-dimensional feature vector. The last step is a softmax activation, which normalises scores to probabilities for classification of two or three classes. Two classes are (a) diseased and (b) healthy, and three classes are (a) pre-disease-affected, (b) diseased, and (c) healthy. The two-class problem sees disease detection distinguishing between affected and healthy individuals, whereas the three-class problem sees early disease detection distinguishing between affected, pre-disease-affected, and healthy.

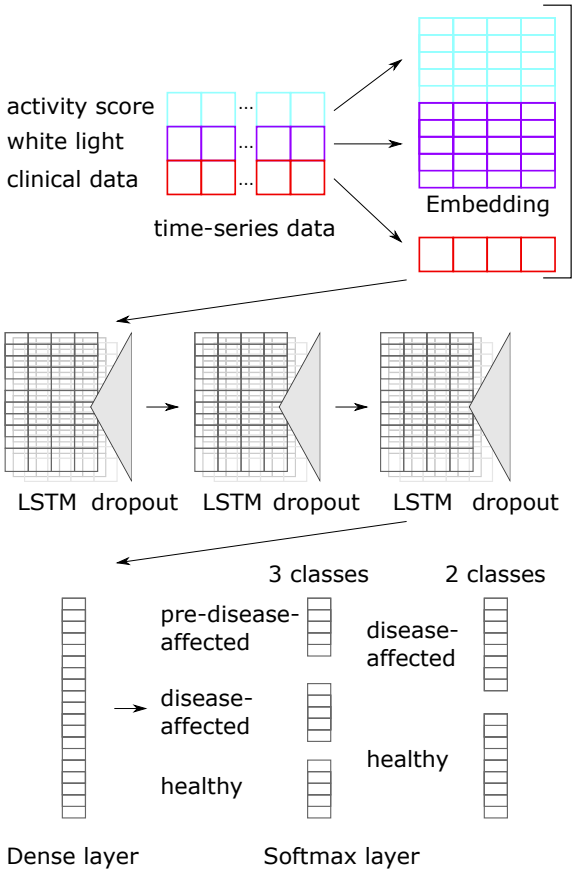


Figure 7.1: The LSTM architecture for disease detection.

7.3 Evaluation of the Proposed Chronic Disease Detection Approach

In this section, we discuss the used data and pre-processing and describe the tests performed to evaluate the approach.

7.3.1 Datasets

For our investigation, we use the HCHS dataset described in Section 3.4.1. The actigraphy data are available in 30-second intervals, which provides 2,880 activity count values per day (30 seconds x 2 per minute x 60 per hour x 12 per day = 2,880). The data for six consecutive days per person are used in our analysis. This is necessary to provide a balanced dataset for training purposes, as not all participants have seven full days of data. Initial results show that using the same amount of data for each individual results in better outcomes, which is in line with previous work [18].

The proof of concept is investigated for four different chronic diseases and conditions: (1) diabetes, (2) hypertension, (3) CKD, and (4) sleep apnoea. These are divided into three different classes each: (a) healthy, (b) pre-disease-affected (early stage), and (c) diseased (see Figure 7.2(a)). Healthy means they do not have the disease investigated for the specific classification task but can be affected by another disease. In the used dataset only two classes are available for hypertension.

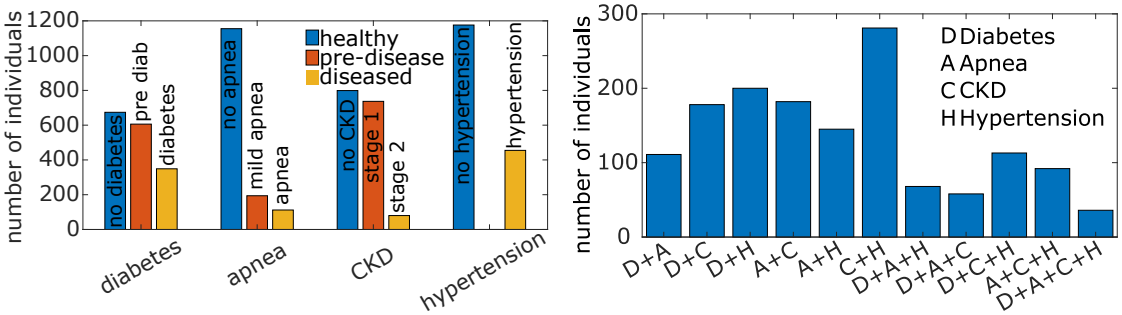


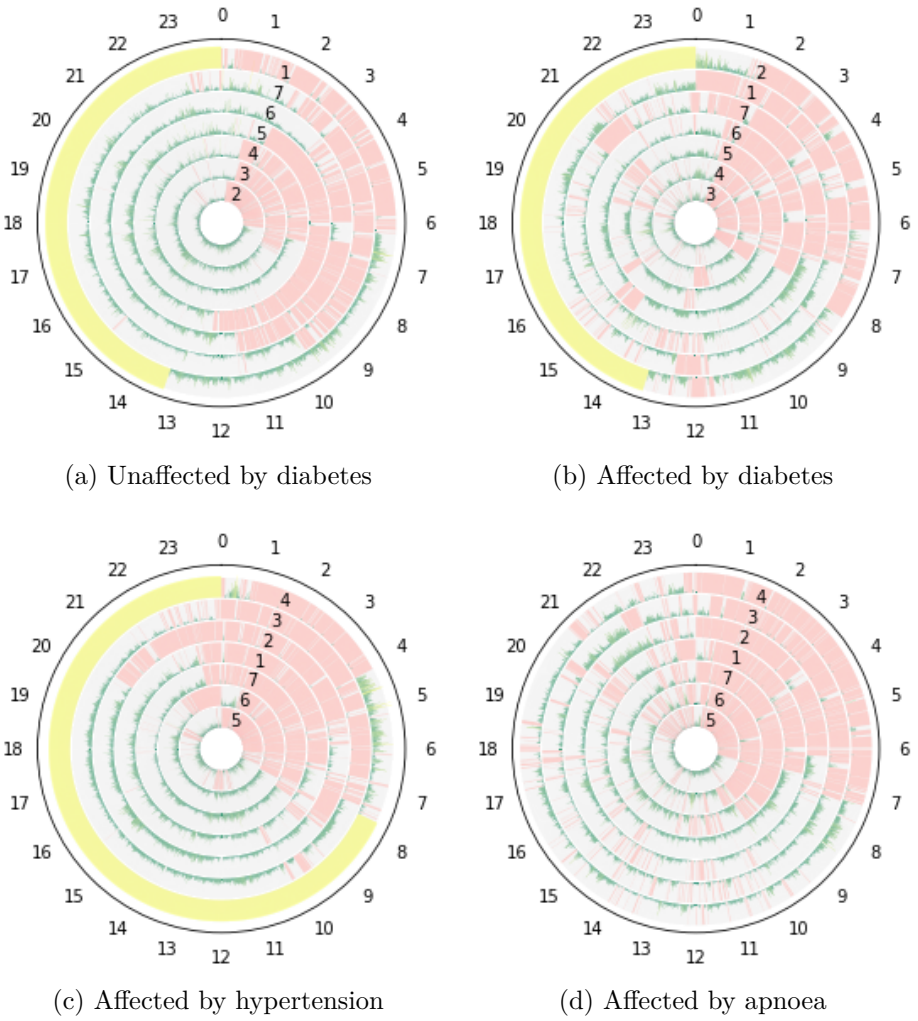
Figure 7.2: (a) The number of people per disease separated by healthy, mildly affected, and affected. (b) Comorbidities in the dataset based on diabetes, apnoea, CKD, and hypertension.

(1) Diabetes is characterised by hyperglycemia [228]. Pre-diabetes is defined as the elevation of plasma glucose above the normal range but below that of clinical di-

abetes [228]. People affected by diabetes are more likely to have trouble sleeping due to the heightened glucose level that influences sleep. Furthermore, lethargy and insomnia can be caused by diabetes [229]. (2) Hypertension is defined by abnormal blood pressure. Hypertension is classified by a systolic pressure of 140 mm Hg or higher or a diastolic pressure of 90 mm Hg or higher [74]. Hypertension is associated with habitual short sleep duration. Furthermore, insomnia is correlated with increased hypertension risk [79]. (3) CKD is a severe health problem defined as kidney damage and is divided into five stages indicated by glomerular filtration rate [76]. In our analysis, only stages one and two are investigated. A wide range of symptoms affects the patients leading to, e.g., fatigue, sleep disturbances, restless legs, and increased urination during the night [32]. (4) OSAS is defined as the upper airways becoming repeatedly blocked during sleep [64]. Classification is based on the AHI, where the events per hour are given. In our use-case, >4% oxyhemoglobin desaturation is used for the AHI rating. In the test case, three groups were used: no, mild, and moderate to severe apnoea (for a full description see Section 2.2.3).

In Figure 7.2(b), the number of people in the dataset with more than one disease and the various combinations of them are provided (e.g., apnoea and diabetes, or diabetes and hypertension). Many comorbidities are present, though some are more prominent than others and, therefore, can influence the classification, as the healthy people can potentially be affected by another disease. However, this is not within the scope—we do not consider comorbidities during our classification.

To give an overview of the data and the individual differences, circle plots are given between affected and healthy individuals in Figure 7.3. The plots show activity count data (green) over the intervals of being awake (grey) and asleep (red) over seven consecutive days. Yellow stands for unavailable data due to differences in the end times. The plots illustrate non-diabetes-affected (Figure 7.3(a)) and diabetes-affected (Figure 7.3(b)) as well as hypertension-affected (Figure 7.3(c)) and apnoea-affected (Figure 7.3(d)). The images depict data from affected individuals (note, not the pre-disease-affected individuals). These images are examples of individuals classified correctly by our approach. Note that there are personal differences between individuals but that general conclusions can still be drawn for specific disease groups, as the results will show in Section 7.4. The use of diseased and healthy individuals is an approach applicable to the real world, because often not all underlying diseases are known. It should be mentioned that this results in a group of individuals who do not have the specific disease but can also not be classified as fully healthy (e.g., they can be unaffected by apnoea but still have diabetes, as comorbidities are common; refer to Figure 7.2(b)).



(a) Unaffected by diabetes

(b) Affected by diabetes

(c) Affected by hypertension

(d) Affected by apnoea

Figure 7.3: Differences in individual sleep patterns affected by different diseases over seven days; activity counts shown in green; sleep shown in red; wake periods shown in grey; absence of available data indicated by yellow.

7.3.2 Processing

Data of interest needs to be pre-processed, including (1) actigraphy data (activity levels and white light) and (2) clinical data, such as BMI, family history of diabetes, and presence of hypertension, diabetes, or asthma. Activity levels and white light data are used for all disease-specific tasks and applied clinical history data are adapted to each disease. For CKD, risk factors include diabetes, high blood pressure, and obesity [230]; as a result, these factors are included as clinical history information (diabetes, hypertension, and BMI). For sleep apnoea, comorbidities are typically obesity, diabetes, hypertension, and sometimes asthma [231]; therefore, the corresponding factors are included as clinical history information (BMI, diabetes,

hypertension, and asthma). For diabetes, obesity and family history of diabetes are important [232] (BMI and family history of diabetes). For hypertension, obesity and diabetes are relevant [233, 234]; that is reflected in the clinical history information (BMI and diabetes).
The data is normalised for training and testing during the classification process.

7.3.3 Chronic Disease Detection Analysis

In this chapter, an LSTM approach is followed. We are mainly interested in comparing training and testing performances in different settings as well as investigating potential features for improvements to the approach. Furthermore, we compared our approach with existing results from Aggarwal et al. [18], which used CNNs. Our approach differs in two ways from that of Aggarwal et al. [18]. First, we follow a method designed for time-series data, and second, we incorporate clinical history data to look at important relationships already established in the literature (refer to Section 7.3.2).

As with any LSTM model, the parameters and the architecture (details in Section 7.2) need to be adjusted empirically, which we did through initial experiments. Therefore, multiple settings are investigated, including (1) number of neurons in the hidden layer, (2) embedding settings, (3) dropout rate, and (4) different optimisers. Based on initial experiments with various neuron sizes, we selected the following layer sizes: 1^{st} layer input size of [400+clinical history factors] and layer output size of 200; 2^{nd} layer input size of 200 and layer output size of 100; 3^{rd} layer layer input size of 100 and layer output size of 50. Furthermore, we chose a dropout rate of 0.2 after experimenting with values between 0.0 and 0.9. To reach a batch size that allows for training per person, the batch size for training and testing is six, as six days are available per individual. Splitting the daily data from 2,880 sequences into 180 (dimension per day, 180x16) or 240 (dimension per day, 240x12) implies batch size training rates of 96 (16*6) and 72 (12*6), respectively (see Figure 7.4). The sequence length of 2,880 (one day) was not used any further after initial tests due to its long training time not resulting in increased accuracy. The model was investigated with an Adam optimisation method [235], as initial experiments ruled out Root Mean Square Propagation for our study. Embedding size was set empirically to 200, where only the activity levels and white light data are embedded. This makes it possible to represent the data sequence in a feature vector [226] with size 200, creating an input vector for the LSTMs of size [400+clinical history factors] consisting of 200 features for activity counts, 200 features for white light and the

various clinical history factors for each disease (see details in Section 7.3.2). The training process was stopped when loss did not decrease within eight intervals (with a maximum of 50 epochs).

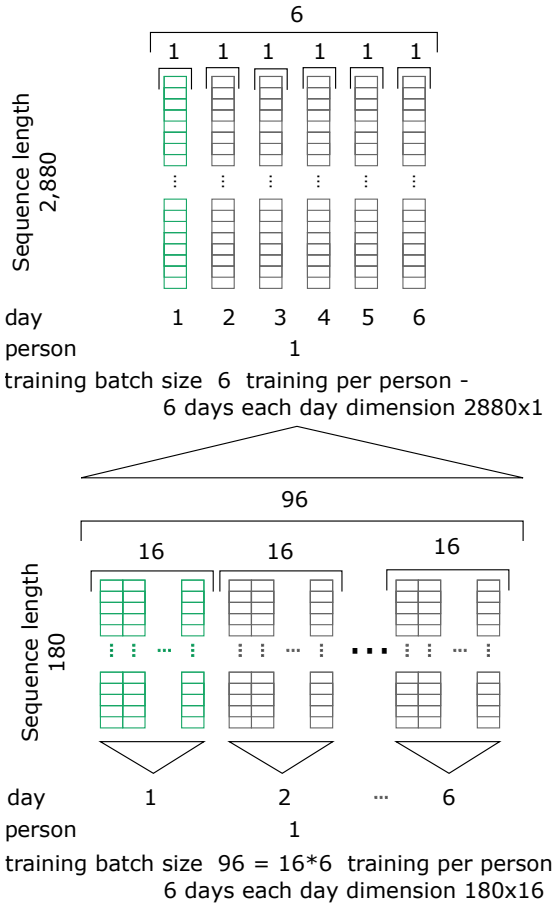


Figure 7.4: Training batch size transformation for a sequence length of 180.

The first test investigates the optimal sequence length. The sequence lengths tested were 180 and 240 for each disease-specific classification for two and three classes.

Repeatedly incorrect classification of certain participants indicates that we should investigate the incorrectly classified participants in more detail to extract differences. These differences can be viewed as potential sources for misclassification that can be used as additional information for improving the classification approach. Participants labelled as 'incorrectly classified' are those that were classified incorrectly more than two times during testing. The significant differences in clinical history factors are further investigated for incorrectly and correctly-classified diseased individuals. The clinical data were investigated using an independent two-samples t-test for different means, variances, and sample size (Welch's t-test). The independent variable is diseased, correctly classified/diseased, incorrectly classified. The dependent variable is one of 38 clinical factors pre-established from the available variables

in the dataset [202, 236]. When using a significance level of 5%, different sets of clinical factors can be extracted that have significant differences between the correctly classified and the incorrectly classified. These are provided with the p-value of the statistical test. These parameters include, among others, BMI, family history of diabetes, and hypertension.

The third test investigates the potential improvement that could be reached by including the knowledge of having six days of data per person. This is done based on (1) counts and (2) average. The data per person for a sequence length of 180 is 180x96. Method (1) counts how often a class is classified between, e.g., intervals of length 96; the one with the highest amount, or the first one in the list if there are two, is reassigned to the whole interval to 96 values. Method (2) takes the average of the interval using a threshold of 0.5.

Furthermore, we compared the outcomes of the optimal scenario, based on the sequence length test, with results from related work. In Aggarwal et al. [18], 'task-specific' models are those that are trained end-to-end for the disorder task at hand; 'multi-task learning' models are trained jointly with all of the disorder prediction tasks (insomnia, apnoea, diabetes, and hypertension).

To guarantee reproducibility, we repeat our experiments. The test results present the averaged outcome for each test, which follows a leave-n-people-out approach. The value n is 10% of the available diseased individuals for each disease-specific classification. This means a 10-fold CV for participants is conducted (i.e., training on 90% of the data and testing on 10%). Furthermore, each test is repeated six times, meaning 60 validation performances are conducted. Per fold, the best 'AUC' and best 'accuracy' of all epochs are chosen to give a 10-fold outcome that was averaged over the six repeats (further referred to as 'averaged best AUC' and 'averaged best accuracy'). The reported outcomes are accuracy, precision, recall, F_1 score, and AUC.

The following investigations are performed: (1) disease detection based on a two-class problem tested with different sequence lengths for apnoea, CKD, hypertension, and diabetes; (2) early disease detection for diabetes and apnoea, detecting early and advanced stages, resulting in a three-class classification problem tested with different sequence lengths; (3) investigation of significant clinical history factors for incorrectly and correctly classified diseased individuals; (4) improvement of accuracy by including knowledge of six days of data per person; and (5) comparison of our overall results to those from a related work.

7.4 Results

In this section, the results of the different tests are presented for disease and early disease detection.

7.4.1 Disease Detection

For disease detection, the LSTM approach is explored for a two-class classification. We provide the performance and findings per disease. In Table 7.1, the performance of each disease investigation is detailed. This includes the batch size as both 180 and 240 averaged over the best accuracy or best AUC; we also include here the precision, recall, F_1 score, AUC, and accuracy results.

Table 7.1: Results for different sequence lengths are given for the tested chronic diseases. Each table shows the averaged CV outcomes for precision (P), recall (R), F_1 score, AUC with SD, and accuracy (acc), for 'averaged best AUC' and 'averaged best accuracy' (refer to Section 7.3).

(a) Diabetes

Size	Avg.	P	R	F_1	AUC(SD)	Acc.
180	acc	0.70	0.77	0.73	0.75(±7e-3)	0.71
180	AUC	0.69	0.74	0.70	0.76(±7e-3)	0.69
240	acc	0.69	0.79	0.73	0.76(±0.02)	0.71
240	AUC	0.67	0.77	0.72	0.77(±0.02)	0.69

(b) Hypertension

Size	Avg.	P	R	F_1	AUC(SD)	Acc.
180	acc	0.60	0.64	0.61	0.60(±0.01)	0.60
180	AUC	0.58	0.60	0.58	0.60(±0.01)	0.58
240	acc	0.58	0.60	0.58	0.56(±2e-3)	0.58
240	AUC	0.55	0.57	0.55	0.57(±2e-3)	0.55

(c) Apnoea

Size	Avg.	P	R	F_1	AUC(SD)	Acc.
180	acc	0.82	0.77	0.78	0.82(±0.05)	0.79
180	AUC	0.79	0.75	0.76	0.85(±0.05)	0.77
240	acc	0.83	0.75	0.78	0.81(±0.07)	0.79
240	AUC	0.81	0.70	0.74	0.84(±0.07)	0.76

(d) CKD

Size	Avg.	P	R	F_1	AUC(SD)	Acc.
180	acc	0.80	0.71	0.73	0.72(±0.15)	0.75
180	AUC	0.75	0.63	0.66	0.78(±0.15)	0.69
240	acc	0.79	0.62	0.67	0.68(±0.07)	0.71
240	AUC	0.74	0.55	0.59	0.73(±0.07)	0.66

avg.- averaged over, acc.- accuracy

Diabetes Our experiments on 'averaged best accuracy' and 'averages best AUC' show that, for diabetes, similar results can be achieved with sequence lengths of 240 and 180; they result in an AUC of 0.76(±0.02) and accuracy of 0.71 for 240. A length of 240 has a slightly higher SD than one of 180 (for 'averaged best accuracy'; see Table 7.1a). Therefore, the best sequence length based on accuracy and SD is classified as the 180 sequence length case.

Table 7.2: Significant clinical factors analysis outcome of the independent two-sample t-test is given for each disease, based on the diseased incorrectly classified and the diseased correctly classified individuals. ↑ indicates a higher level in the incorrectly classified case and ↓ a lower level in the incorrectly classified.

Disease	Clinical Factors(p-value)	
Diabetes	↑ BMI(0.02)	↑ Fam. Hist. Diab.(2.1e-09)
	↑ Pneumonia(0.018)	
Hypertension	↑ BMI(1.8e-17)	↓ Total Drinks per Week(0.012)
	↑ Apnoea/Hypopnoea Index(4.8e-05)	↑ Treatment of Hypert.(0.013)
	↓ Alcohol Use (0.028)	↑ Physical Activity Level(0.036)
Apnoea	↑ BMI(2e-3)	↑ 4-level Hypertension(3e-3)
	↑ Hypertension(1e-3)	↑ 3-level Diabetes(0.04)
	↓ Diabetes Diag.(0.025)	↓ Antidiabetics(0.036)
	↑ Antihypertensives(0.025)	↑ Cigarette Use(0.017)
	↑ Hypertension Treatment(0.015)	
CKD	↑ BMI(0.032)	↑ Antihypertensives(0.025)
	↑ Hypertension(1.4e-3)	↑ Treatment of Hypert.(0.012)
	↓ Airflow Limitation(0.044)	↑ Physical Activity Level(0.043)
	↑ 4-level Hypert.(0.021)	↑ Stroke(0.036)

For diabetes detection, 120 diseased individuals are incorrectly classified as healthy individuals over all tests and 205 are correctly classified as diseased. For the healthy individuals, 107 are incorrectly classified from 325 individuals randomly selected from all healthy individuals in the dataset (see Table 7.3).

Table 7.3: Counts for incorrectly/correctly classified individuals during testing for diseased and healthy with a sequence length of 180; more than two times incorrect constitutes incorrectly classified.

Disease	Counts per Person					
	Healthy			Affected		
	overall	correct	incorrect	overall	correct	incorrect
Diabetes	325	218	107	325	205	120
Hypertension	475	349	126	475	332	143
Apnoea	85	72	13	85	62	23
CKD	60	49	11	60	43	17

In Table 7.2, three variables are significantly different in the diseased case comparing correctly classified and incorrectly classified. These are BMI, family history of diabetes, and pneumonia. The group that is incorrectly classified sees: (1) higher average BMI (at 33.3, than the group that is correctly classified, at 31.6); (2) higher family history of diabetes (0.9, almost double the 0.5 for the correctly classified

group); and (3) higher pneumonia (0.22 over 0.12).

The results are improved when the knowledge of six days per person is incorporated. This results in an accuracy increase of about 1%, to 72.5% and 72.7% for 180 and 240, respectively. This makes the combined score the best optimisation technique for the diabetes case (see Table 7.4).

Table 7.4: Results when the knowledge of six days per person is incorporated. Each entry shows accuracy calculated by using the highest count of labels per person (counts) and the best score per person (average); details can be found in Section 7.3.3.

Disease	Size	Accuracy	Combined Amount (Counts)	Combined Score (Average)
Diabetes	180	0.714	0.725	0.725
	240	0.712	0.726	0.727
Hypertension	180	0.603	0.615	0.620
	240	0.578	0.587	0.589
Apnoea	180	0.794	0.813	0.810
	240	0.791	0.808	0.800
CKD	180	0.750	0.756	0.767
	240	0.709	0.728	0.722

Hypertension The performed comparison with Aggarwal et al. [18] shows that our approach improves the precision, recall, and F_1 score but can reach only similar accuracy for sequence length 180 in the multi-task case (see Table 7.5).

Overall, hypertension performs least well for the two-class problems (see Table 7.1b). The best outcome is achieved using 180; it results in an accuracy and AUC of 0.60 with minimal variation. Our investigation showed that there were 143 participants incorrectly classified more than two times with sequence length 180 and 332 were correctly classified as diseased. For the healthy individuals, 126 were incorrectly classified from a randomly selected 475 from all healthy individuals (refer to Table 7.3). When using a significance level of 5%, six variables are significantly different between the diseased individuals who were correctly classified and those who were incorrectly classified: BMI, alcohol use (yes/no), total drinks per week, AHI, treatment of hypertension, and physical activity level (see Table 7.2). The group that is incorrectly classified sees: (1) higher BMI (34 over 30 in the correctly classified group); (2) lower total drinks per week (1.28 compared to 2.56); (3) higher AHI (7.5 over 3.9); (4) higher treatment of hypertension (0.62 over 0.5); (5) lower alcohol use (2.05 compared to 2.23); and (6) higher physical activity level (2.52 over 2.4).

Table 7.5: Comparison between task-specific and multi-task results from Aggarwal et al. [18] and our results for apnoea, hypertension, and diabetes.

Experiment	Model	Precision	Recall	F_1	Accuracy
Hypertension					
Task-specific	CNN	0.44	0.31	0.37	0.69
Multi-task	CNN	0.48	0.41	0.44	0.61
180	LSTM	0.60	0.64	0.61	0.60
240	LSTMM	0.58	0.60	0.58	0.58
Apnoea					
Task-specific	CNN	0.31	0.63	0.42	0.55
Multi-task	CNN	0.40	0.48	0.43	0.68
180	LSTM	0.82	0.77	0.78	0.79
240	LSTM	0.83	0.75	0.78	0.79
Diabetes					
Task-spec	CNN	0.40	0.51	0.45	0.41
Multi-task	CNN	0.46	0.47	0.46	0.44
180	LSTM	0.44	0.58	0.50	0.47
240	LSTM	0.44	0.57	0.51	0.48

An improvement of 1–2% in accuracy can be achieved when using the combined scores method, resulting in 0.59(240) and 0.62(180) accuracy (see Table 7.4).

Sleep Apnoea The best outcome for sleep apnoea can be reached using 'averaged best accuracy', resulting in similar outcomes for sequence lengths of 180 and 240 (see Table 7.1c). An accuracy of 0.79 and an AUC of 0.82 can be reached. Our investigation showed that there were 23 incorrectly classified as healthy and 62 correctly classified as diseased. For the healthy individuals, 13 were incorrectly classified from 85 randomly selected participants from all healthy individuals. The incorrectly diseased group has significant differences from the correctly diseased group in the following variables: BMI, 4-levels hypertension, hypertension, hypertension treatment, 3-levels diabetes, diabetes diagnosis, antidiabetics, antihypertensives, and cigarette use (see Table 7.2). The incorrectly classified group sees: (1) higher BMI (38 over 33); (2) higher 4-level grouped hypertension (3 over 2.3); (3) higher hypertension (0.8 over 0.4); (4) higher 3-level grouped diabetes (2.3 over 2.0); (5) lower diabetes diagnosis (2.2 compared to 2.6); (6) lower antidiabetics (0.27 compared to 0.07); (7) higher antihypertensives (0.47 over 0.2); (8) higher cigarette use (1.9 over 1.5); and (9) higher hypertension treatment (0.6 over 0.3).

An improvement to accuracy is depicted in Table 7.4 with the combined amount technique reaching an accuracy of 0.813 and 0.808 for 180 and 240, respectively. In comparison, the results of Aggarwal et al. [18] in Table 7.5 can be improved considerably by using our approach, from an accuracy of 0.68 (CNN) to 0.79 (LSTM) and precision from 0.40 (CNN) to 0.82(180) or 0.83(240) (LSTM).

CKD For CKD detection, the best model can be classified as the 180 sequence length model with an accuracy of 0.75 and AUC of 0.72 (see Table 7.1d). ROCs were interpreted and showed a higher SD for the individual true positive rates, which represents the differences in epochs and folds.

An accuracy of 0.77 can be reached with the 180 sequence length model using the combined score technique; an accuracy of 0.73 can be reached with the 240 sequence length model using the combined amount technique (see Table 7.4).

Our investigation showed that there were 17 participants incorrectly classified and 43 participants correctly classified as diseased (for sequence length 180). Of the healthy individuals, 11 were incorrectly classified. Eight variables are significantly different: BMI, antihypertensives, hypertension, treatment of hypertension, airflow limitation, physical activity level, 4-level grouped hypertension, and self-reported prevalent stroke (see Table 7.2). The incorrectly classified group sees: (1) higher BMI (35 over 30); (2) higher 4-level grouped hypertension (2.9 over 2.3); (3) higher hypertension (0.8 over 0.4); (4) lower airflow limitation (0 compared to 0.1); (5) higher physical activity level (2.6 over 2.3); (6) higher prevalent stroke (0.3 over 0.03); (7) higher antihypertensives (0.6 over 0.3); and (8) higher treatment of hypertension (0.7 over 0.3).

Overall, our disease detection approach shows good outcomes for diabetes, apnoea, and CKD. For hypertension, the overall outcome is less promising (see tables 7.1b and 7.5. The results of the two-class tests for hypertension and apnoea were compared to the results from Aggarwal et al. [18]. Our hypertension results showed no improvements in accuracy over Aggarwal et al. [18], though they did show improvements in precision, recall, and F_1 score. For apnoea detection, our approach outperformed the process from Aggarwal et al. [18] using CNN (see Table 7.5).

7.4.2 Early Disease Detection

The early detection tests show the outcomes of the three-class problems: healthy, pre-disease-affected, and diseased. As the SD of CKD was already high for the two-

class problem, we concentrated on diabetes and sleep apnoea. The outcomes show potential for early detection but should be seen as preliminary. Further research could expand upon these findings with additional features or classification models.

Diabetes In Table 7.5, the three-class problem of diabetes is compared to the related work of Aggarwal et al. [18]; our approach reaches an improvement in accuracy of up to 7% from 0.41 (CNN: Task-specific) and 0.44 (CNN: Multi-task) to 0.47 (LSTM: 180) and 0.48 (LSTM: 240). Additionally, recall and F_1 score show improvement overall. Precision has similar outcomes.

For early detection of diabetes, similar outcomes can be achieved for sequence lengths of 180 and 240—averaged accuracy of 0.47 and 0.48, and averaged AUC of 0.63 and 0.61, respectively (see Table 7.6a). Accuracy can be improved to 0.48 (180) and 0.49 (240) using the count method.

Table 7.6: Sequence lengths tests for early disease detection with a multi-class approach, using one class against the others (e.g., C1 describes C1 against all other classes); C1 is healthy, C2 is diseased, C3 is pre-disease-affected; avg. describes the average over all classes and all repeats.

(a) Diabetes

Size	Precision				Recall				F_1				AUC(SD)				Acc.
	C1	C2	C3	avg.	C1	C2	C3	avg.	C1	C2	C3	avg.	C1	C2	C3	avg.	
180	0.48	0.49	0.35	0.44	0.58	0.59	0.50	0.58	0.53	0.56	0.34	0.50	0.66	0.66	0.51	0.62	0.47
240	0.48	0.50	0.35	0.44	0.58	0.60	0.51	0.57	0.53	0.58	0.35	0.51	0.65	0.67	0.53	0.62	0.48

(b) Apnoea

Size	Precision				Recall				F_1				AUC(SD)				Acc.
	C1	C2	C3	avg	C1	C2	C3	avg	C1	C2	C3	avg	C1	C2	C3	avg	
180	0.59	0.55	0.37	0.48	0.57	0.56	0.46	0.57	0.58	0.66	0.34	0.56	0.71	0.68	0.49	0.63	0.53
240	0.56	0.49	0.38	0.46	0.56	0.55	0.47	0.56	0.59	0.50	0.30	0.52	0.69	0.64	0.49	0.61	0.54

Sleep Apnoea For sleep apnoea, the best results can be achieved with a sequence length of 240, which results in an averaged accuracy of 0.54 and an averaged AUC of 0.61 (see Table 7.6b). Accuracy can be improved to 0.55 (180) and 0.56 (240) using the count method. There were 9 participants who were repeatedly incorrectly classified for the pre-disease-affected case, 24 for the diseased case, and 18 for the healthy case. Overall, 85 individuals per group are used for training. The features that show significant differences between the incorrectly and correctly classified groups are as follows: BMI (p-value: 0.012), antihypertensive (0.028), hypertension (0.0018), hypertension treatment (0.029), 3-level diabetes (0.049), diabetes diagnosis (0.019), family history of diabetes (0.038), pneumonia (0.017), and 4-level hypertension (7.6e-4). The incorrectly classified group sees: (1) higher BMI (38 over 34);

(2) higher antihypertensive (0.5 over 0.2); (3) higher hypertension (0.8 over 0.4); (4) higher hypertension treatment (0.58 over 0.3); (5) higher 3-level diabetes (2.4 over 2.0); (6) higher family history of diabetes (0.7 over 0.4); (7) higher pneumonia (0.3 over 0.04); and (8) higher 4-level hypertension (3.0 over 2.3). We can see that similar features are significant in the two-class apnoea problem.

7.5 Discussion

In this section, our results are discussed and limitations are examined.

7.5.1 Disease Detection

Overall, smaller training batch sizes of 180 result in better performance. This makes sense, as there is a trade-off between batch size and training epochs. As we set the epochs to maximum 60, it seems that 180 reaches the optimum within these available epochs. Our approach recognised sleep apnoea best with an accuracy of 81%. This is reasonable, as sleep apnoea is only related to sleep/wake habits. Additionally, diabetes and CKD demonstrate potential to be detected by sleep-wake behaviour. However, hypertension shows low recognition rates; this is potentially related to the high amount of comorbidity in the dataset, especially with hypertension patients (see Figure 7.2(b)). Minimal variations in the detection of hypertension with low performance suggest that specific groups cannot be detected. This encourages further research into the incorrectly classified participants to extract differences in those.

In general, diseased individuals are more likely to be incorrectly classified than healthy individuals. This can be caused by existing comorbidities that are more strongly related to sleep behaviour changes. More information is needed, such as from the significantly different clinical history factors between the groups. Some of these significant different parameters might be able to improve the recognition rate in the future; especially, as some of them have shown relations with specific diseases as found in literature. For example, CKD patients have a higher risk for strokes [237]; this matches with the extracted clinical factors for significant differences between the individuals that are incorrectly and correctly classified. Because of the smaller data sample in the CKD case, a higher number of clinical factors are found to be significantly different, as smaller variations have a higher impact.

Integrating the knowledge from the data into the process reaches around 2% improvement in accuracy, which is a reasonable outcome. We consider segments with a length of 180 timepoints and classify them as being either diseased or healthy. We then incorporate the knowledge that 96 of these segments belong to the same person, consequently, only one overall classification is necessary per 96 segments. The individuals segment-based classifications could later be extended to provide a certainty about how likely the person was classified as diseased.

Compared to Aggarwal et al. [18], our approach results in better precision, recall, and F_1 score but lower accuracy for hypertension. These results could stem from the fact that the embedding used was not designed for activities, but for words. This could be improved by, e.g., emulating the activity embedding proposed by Aggarwal et al. [18].

7.5.2 Early Disease Detection

Early disease detection has proven to be a challenge. We can improve current accuracy by up to 7% but there is still significant room for improvement and for further research in this area.

Overall, we conclude that the pre-disease condition is harder to classify. Early detection is complex and in need of further insight from more features or data sources. Potential sources could be other sensor devices (e.g., pulse oximeter) or the presented significantly different clinical factors.

Interestingly, the significant different clinical history factors for sleep apnoea are similar to the ones in the two-class classification, the list is extended by the family history of diabetes and pneumonia. This finding suggests that sleep apnoea classification from two and three classes have similar identifiers, which strengthens the integration of further features within the assessment.

7.5.3 Limitations

Two limitations of this experiment are the usage of a pre-defined embedding function and the lack of an evaluation of different loss functions. Initial experiments on pre-defined embedding showed good results and were used for further investigations. The word-2-vec embedding method could be improved by using a designed act-2-vec method, as presented by Aggarwal et al. [18] for the activity count and white

light data. Based on the literature, the categorical cross-entropy approach has been used for classification, but others have not been tested.

The decision was made to use the same amount of data for each individual—six days each. This decision resulted in some limitations, as (1) some individuals have a full weekend in their dataset while others have just one day and (2) not all participants share the same consecutive days sequence, which can lead to difficulties when training the classification algorithm. This could likely be overcome by including weekdays as a feature; however, this was outside the scope of the study.

A detailed interpretation of why the approach of Aggarwal et al. [18] is improved is not possible through our assessment. As we only compared our overall approach, we introduced a limitation. The main scope was to investigate the performance of our overall approach and the clinical features that influence that performance. We were not interested in whether the inclusion of clinical features or the time-series approach were behind the improvement.

The use of diseased and healthy (unaffected by investigated disease) is a real-world approach. It needs to be mentioned, as this can result in a group that is unaffected by a specific disease but is also unable to be classified as fully healthy.

The significant feature analysis for early disease detection could offer more insight by grouping the individuals who were incorrectly classified into two groups, (1) healthy and (2) pre-disease-affected; this, however, was not within the scope of this research.

7.6 Summary and Future Directions

This chapter contributes to knowledge with a chronic disease detection approach. We have investigated the hypothesis of chronic disease detection based on the correlations between sleep-wake behaviour and chronic diseases. An LSTM algorithm fusing actigraphy data with clinical features was applied, leading to a multi-dimensional feature vector for effective disease detection. We have tested and evaluated various use case scenarios, including two-class disease detection problems (healthy versus diseased) and three-class early disease detection problems (healthy, pre-disease-affected, and diseased) for four common chronic diseases: hypertension, diabetes, sleep apnoea, and CKD. Our approach shows promising results, varying by use case due to differences in chronic diseases characteristics. As assumed, there is no one-size-fits-all approach for chronic diseases; however, using specific clinical features for each disease can address the varying characteristics to a certain extent. Nevertheless,

our approach significantly contributes to the literature; it is an improvement over existing research in terms of performance and clearly demonstrates the feasibility of predicting chronic diseases from sleep-wake behaviour. Further investigations will be needed to enable early detection, however.

The availability of a mechanism, such as ours, that properly incorporates relations between sleep and chronic diseases, as well as clinical history factors, is important to diagnosis diseases. This mechanism enables our framework to automatically adapt to changes in health status resulting in a personalised sleep behaviour assessment. Our findings indicate that it is possible to detect diseases by fusing knowledge sources. This can encourage researchers in the future to (1) conduct in-depth evaluations to determine any other chronic diseases that can be detected using sleep-wake behaviour analysis and (2) investigate the reasons for why some of them are less accurate than others. We provided salient features for various chronic diseases and determined which clinical data play significant roles in disease detection; these provisions should be expanded upon in the future. Early disease detection is a developing field that is highly important to the medical field. Our work lays a foundation for future research approaches involving time-series analysis for early disease detection. To our knowledge, our approach was one of the first to propose and investigate a method to classify various early diseases by fusing medical knowledge from clinical history data with actigraphy measurements. Further work is still necessary to incorporate comorbidities and the presented significantly different clinical data parameters.

The integration of a chronic disease detection mechanism makes our adaptable framework possible.

Chapter 8

Conclusions and Future Work

Sleep helps us to relax and recharge. Consequently, if the quality of our rest is inconsistent or inadequate, our health is adversely affected (e.g., by a weakened immune system and a higher risk of chronic diseases). These adverse effects often manifest in sleep issues, such as sleep deprivation and apnoea, which are common in the population and can have severe impacts. These issues can be classified by abnormalities if sleep behaviour is properly monitored and assessed.

Technological advancements have enabled people to continuously monitor and investigate their sleep behaviour in their natural sleeping environments. This ability has provided many new insights. Academics and commercial actors investigate technical solutions to provide meaningful information and potentially prompt early interventions. Modular, personalised, and adaptable approaches are vital for reliable and effective home-based sleep assessment. These technical solutions facilitate predictive and personalised healthcare.

We discuss the conclusion from our conducted research in Section 8.1, and the made contributions and the relevance of our work in Section 8.2. Furthermore, we identify open issues and opportunities for future research in Section 8.3 and provide concluding remarks in Section 8.4.

8.1 Summary of Work

In Chapter 1, we established the research background on sleep behaviour assessment, outlined the motivations behind our investigations, and addressed the challenges our work intends to overcome. Moreover, we described the overall aims and objectives of this thesis.

To meet our aims and objectives, we reviewed the literature for current open is-

sues and challenges in the field. We presented the key challenges we identified in Chapter 2; these included the lack of a sleep behaviour assessment framework, an accurate method for sleep position detection, a sleep stage analysis process, a sleep regularity algorithm, visualisation standards, and a chronic disease assessment based on correlations with sleep issues. The subsequent chapters described the work done to address these open issues.

The overarching aim of this thesis was the development of a home-based sleep behaviour assessment framework. We developed an overall framework that can support future researchers through an integration of the most crucial sleep behaviour aspects with the following properties: being adaptable to technology developments, applicable in existing home environments (e.g., smart homes), and designed to aim at avoiding user misinterpretation. Chapter 3 presented this framework design and detailed its layered structure, components, and characteristics. The main benefit of this design is its adaptability to users' needs, doctors' recommendations, and available data sources. In addition, we proposed (1) a sleep regularity algorithm to investigate the important aspects of regularity (objective 2) and (2) a visualisation concept aimed at avoiding misinterpretation (objective 3). The framework aims to represent an overall picture of sleep by assessing it and setting it in context with related factors such as chronic diseases, environmental changes, and routines. Furthermore, we discuss the necessity and challenged of seamlessly integrating our approaches with existing environments. The framework serves as the backbone for chapters 4–7. Each one targeted individual components of the framework and addressed the identified challenges from the literature review. Future work must apply this framework in a real-world setting and test it under various conditions, e.g., in a smart home.

Our fourth objective was to establish an approach that could detect higher-granularity sleep positions, which are common in the population. The integration of leg positioning with a reliable and computing-efficient algorithm facilitated a real-time assessment of more complex sleep positions. In our approach, features are extracted from stable state data and a GMLVQ model is applied. In sleep position recognition, it is crucial to find a balance between the number of integrated sensors (which potentially influence sleep behaviour) and the potential for detecting more complex positions (which provide more information about sleep behaviour). This approach was evaluated using wearable devices on six individuals with both personal and general models. The comparative results showed that the approach is applicable in both simulated and real-world settings. One limitation comes from the exclusive use of orientation data; this could be overcome in the future by incorporating abso-

lute or relative positions.

Our fifth objective was to investigate sleep-wake behaviour by addressing the key issues with standard approaches: one-model-fits-all approaches, exclusion of medical knowledge, and investigations are often limited to one data source. To close these gaps, we adopted an adaptive learning approach. We developed a fine-grained sleep stage detection approach to enable more reliable detection of sleep-wake stages. The personalised sleep patterns were learned using data from multiple sources, including personal and physiological sources. The personal datasets are a fusion of clinical history and genetic information, such as race, gender, and health status. Furthermore, to assess sleep behaviour, sleep parameters must be reliably measured from extracted sleep-wake patterns. Consequently, we effectively developed a personalised sleep parameter extraction process (objective 6). To incorporate the differences between groups of individuals, we used dynamic instead of static thresholds that could better represent these differences. The strength of this approach lies in the fusion of computational technology with medical knowledge about factors that influence sleep architecture. The approach was extensively investigated using a large and diverse dataset to improve the performance to be more in line with PSG assessment. This investigation demonstrated the importance of personalised approaches in sleep behaviour analysis and the positive effects that come from including influence factors. The fusion of physiological data sources brings together the advantages of actigraphy (which performs well on healthy participants) and HRV (which performs well on diseased individuals). The approach is adaptable to available data but is dependent on access to personal data. Future work should consider comorbidities and investigate the fusion of multiple physiological data sources even further.

Our objective to detect chronic diseases from sleep-wake analysis was achieved twofold. First, a thematic analysis was performed on sleep issues and their relationships with chronic diseases, in addition to treatment satisfaction and technology acceptance. In Chapter 6, we presented the outcomes from interviews differentiating between the perspectives of doctors and individuals suffering from a medical condition. Our findings confirm that there are relations between sleep and chronic diseases and, therefore, that it is possible to use sleep-wake behaviour to detect chronic diseases. Furthermore, results supported our framework design through individuals' responses regarding (1) the importance of home-based IoT-enabled assessment (addressed in Chapter 3); (2) the importance of in-depth sleep position detection (addressed in Chapter 4); (3) changes in sleep architecture (see Chapter 7); and (4) current limitations of sleep-stage approaches (addressed in Chapter 5). Second, the main finding in the relations between sleep and chronic diseases was further pursued. We pro-

posed a method using sleep-wake information to detect chronic diseases that aimed to distinguish between early and advanced stages. The integration of multiple data sources bringing relevant information together made this research fairly advanced. The benefit of our approach comes from the usage of (1) the temporal aspect of our data in the design of the approach and (2) the multi-dimensional feature vector. Another advantage is the in-depth analysis of features that influence performance. The approach was evaluated using the HCHS dataset and showed promising results. In future work, the salient features should be reassessed by being incorporated in the approach; additionally, more medical conditions and potential influences from comorbidities should be included.

To summarise, we designed a framework for sleep behaviour assessment. Furthermore, we investigate improvements to individual parts of this framework. This framework has been designed to work with a variable number of data sources depending on availability, and considers users' needs, as discussed in Chapter 3. The integrated technologies are personalised; individual aspects are incorporated to provide a solid estimate of sleep status. This was achieved by combining medical knowledge with computational technologies and incorporating multiple data sources and influence factors for sleep position detection, sleep stage recognition, and chronic disease detection. The incorporation of a regularity assessment and a visualisation component were aimed at providing an efficient and understandable framework structure.

8.2 Revisit of Contributions

Our research significantly advances computational sleep behaviour analysis. Specifically, we have made eight significant contributions to the literature: (1) a modular layered sleep behaviour assessment framework that incorporates developed sleep analysis approaches; (2) a sleep regularity algorithm for individual sleep parameters; (3) a user-dependent visualisation concept with varying levels of abstractions; (4) a higher-granularity sleep position detection approach that considers leg movement; (5) a fine-grained sleep stage detection approach using multiple data sources; (6) a personalised sleep parameter extraction process with dynamic thresholds; (7) in-depth understanding on sleep and chronic disease relations with findings confirming the worth in using sleep-wake behaviour for chronic disease detection; and (8) a sleep-wake behaviour-based chronic disease detection method including multi-dimensional features.

The first contribution is a *modular layered sleep behaviour assessment framework* that addresses the lack of a guideline for accurate and understandable sleep assessment. The layered design provides a flexible structure tailored to individual needs and facilitates the integration of new technological advances. The modular design is essential, as individuals are influenced differently based on their clinical history and daily routines. The framework monitors the main elements of sleep and influence factors to enable self-management as well as personalised and predictive healthcare. This opens up new potential areas of investigation, e.g., the investigation of correlations between different elements in the framework. The personalised approach has the potential to help predict anomalies faster, provide alerts, and offer tailored recommendations. The framework is ideal for sleep behaviour assessment because it is scalable, robust, and efficient. The framework can deal with data from multiple channels and incorporates clinical history and preferences, thus allowing for personalised medical and behavioural assessment.

The second contribution is a *sleep regularity algorithm* that addressed the limitations of the standard two-day regularity assessment. Regularity describes the state of having a regular sleep behaviour, and when measured over longer period of time it is of great value in sleep assessment. Sleep regularity can provide insights into sleep disorders. To guarantee a reliable assessment, we use long-term assessments for individual sleep parameters. This constitutes a new way of assessing sleep regularity, which has the advantage of comparing behaviour over multiple days. This offers new opportunities to determine differences and anomalies in sleep rhythm by incorporating various factors, such as day of the week, season, or moon phase.

The third contribution is a *personalised visualisation concept*, which is adaptable to the needs of users and doctors. Currently, information about sleep is presented without context, and often lead to incorrect conclusions. Appropriate visualisation benefits users by allowing for self-management and doctors by enabling them to provide accurate recommendations, diagnosis, and treatments. To avoid misinterpretation and provide details where required, we present a new way of visualisation that uses a modular and layered structure. Sleep parameters are set in context by incorporating health status, phenotypical characteristics, users' preferences, and doctors' recommendations.

The fourth contribution is a *higher-granularity sleep position analysis approach.* Research on wearable devices has mainly considered just four basic sleep positions and excluded the limbs. This is problematic, as leg positions are essential to describe some of the most common positions. Our approach, which is applicable in the

real-world, incorporates these leg positions and can accurately extract characteristic patterns for individual positions. GMLVQ is applied on actigraphy data, providing a mobile assessment at low computational costs. We provided a comparative study to see the differences in results between real-world and simulated settings. A general model is promising for real-world application, as it captures the variability of the positions well. More complex positions give additional information about sleep, which is a starting point for behavioural interventions, e.g., treatment of sleep apnoea. Standard medical assessments of sleep positions are based on subjective questionnaires. Our approach provides objective assessment, opening the possibility of investigations on the correspondence between subjectively reported positions and objectively recognised positions. Furthermore, this approach will enable future researchers to explore the reasons behind waking up and falling asleep in particular positions.

The fifth contribution is a *fine-grained sleep stage detection approach*. Standard sleep stage approaches ignore known factors that influence sleep architecture, overlook the advantage of personalisation, and are limited to information from just one physiological data source. Influence factors can be represented through a fusion of data sources, including sensor data, clinical history, and genetic information. Personal sleep-wake patterns are captured in models trained on groups of individuals. This approach supersedes current one-model-fits-all approaches and is adaptable to individuals and available data sources. Notably, the fusion of multiple data sources has shown considerable advantages. Findings include the improvement of sleep-wake recognition at home through the consideration of gender, health status, as well as through the inclusion of multiple physiological data sources. The results demonstrate the importance of personalised approaches in sleep-wake detection. Future researchers could extend this approach and explore more complex sleep stage detection.

The sixth contribution is a *personalised sleep parameter extraction process*. We addressed the current limitation of static thresholds for parameter extraction from sleep-wake stages in actigraphy scoring. We did this by introducing dynamic thresholds and incorporating race, gender, and health status information to achieve personalised sleep parameter extraction. Results indicate that our approach provides a better match for SOL to the gold standard, PSG, than the method presented in [201, 202]. Our new method of extracting sleep parameters should encourage future researchers to deliver meaningful results that match the population of interest.

The seventh contribution is made up of the *in-depth understanding on sleep and*

chronic disease relations. Generally, questionnaires are used to assess correlations between sleep and chronic disease. Of course, questionnaires cannot match the breadth of information available through interviews. Semi-structured interviews were conducted on the currently under-represented Chinese population. We sought to provide a comprehensive picture from individuals affected by a disease as well as doctors, who are usually not consulted. This allows for examining the differences between their opinions. Our thematic analysis on sleep issues, causes, impacts, assessment, treatment, and technology usage, indicated relationships between sleep behaviour and chronic diseases. Doctors and affected individuals generally agreed on most topics and confirm these relationships. However, doctors were often unaware of their patients' sleep issues; this stems from the lack of sleep assessment and behavioural treatment. These findings should encourage future researchers to objectively investigate chronic diseases and sleep-wake behaviour.

The eighth contribution is a *sleep behaviour-based chronic disease detection approach*. There have been few serious approaches to use sleep-wake behaviour for chronic disease detection. However, medical knowledge and the temporal aspect of the data is usually not integrated; hence, we construct multi-dimensional features from clinical history and sensor data. An LSTM capable of handling temporal data is adapted to consider sleep-wake behaviour information. We investigated early and advanced stages of different diseases. The method showed promising results, though they varied by use case due to differences in chronic disease characteristics. We provide significantly different clinical features which might improve the performance. This should encourage future researchers to investigate these features in more detail. Our work shows the potential and importance of early-stage disease detection from sleep-wake behaviour. This approach emphasises the need for time-series analysis. Furthermore, it demonstrates the value of fusing medical knowledge with computational approaches in chronic disease detection.

8.3 Open Issues and Future Work

Throughout this thesis, several opportunities and open issues have been identified and highlighted. The proposed approaches to solve existing challenges in computational sleep behaviour analysis must be further investigated. In this section, we discuss some of the areas that need improvement and potential directions for future research.

First, we presented personalised approaches that achieve higher accuracy. However,

to achieve a proper balance between personalised assessment and generalisability, we provide models that can be used for groups of individuals. Personalisation can limit direct comparisons between individuals from different groups. Nevertheless, this is in line with the current assessment from doctors who differentiate between healthy from diseased individuals with respect to sleep behaviour. Still, though, more enrichment is required; approaches must consider other influence factors, including comorbidities, age, and day of the week. This could be achieved by developing (1) a health status feature that summarises individuals' information and (2) a weekday feature; these could be integrated as supplemental features. Most importantly, we provide strong evidence that personalised models are ideal in many instances, and this will influence future research.

Second, the combination of sensors was explored, and found to be successful, e.g., for sleep stage detection. We explained how to fuse different data sources by merging information into equal interval steps. Further work on our data fusion approach is necessary and is an interesting area to pursue due to its observed performance effects. However, the best possible combination of multiple sensors needs to be determined to provide sufficient information but avoid discomfort and influence on natural sleep habits. This is difficult but the insight that can be gained through it is valuable, e.g., the supplementary information from more complex positions in combination with doctor's advice can help user's in sleep-management. Future research could assess other sensors or develop entirely new ways of monitoring. For sleep stage detection, sensors which capture new features could be beneficial, e.g., which monitors breathing changes to detect when people go into a deeper sleep. Trends indicate that smart clothing, single-channel EEG devices, and reliable non-wearables will be available on the market in the next few years—researchers could make use of them.

Third, our approaches can be easily adapted to suit other investigations in sleep research. Bed exits, e.g., can be detected using our sleep positions approach with some additional labels. Our sleep-wake detection approach can be use to train models for other sleep stages by following multiple two-class problems. Our chronic disease approach can be used to investigate other diseases, such as Parkinson's, or Alzheimer's disease. It has shown promising results using time-series analysis in combination with medical knowledge. Our work on disease detection in early and advanced stages could encourage research to develop new techniques and features for chronic disease diagnosis from sleep-wake behaviour. Furthermore, developed approaches in this thesis could help to objectively measure certain parameters that are currently only measured subjectively, such as sleep position.

Fourth, improving home-based technologies is essential not only to address the aforementioned issues but also to seamlessly integrate methods into existing smart homes and devices. Therefore, the challenges facing data extraction, pre-processing, and occupancy (discussed in Chapter 3) must be overcome by new solutions (e.g., the presence of multiple individuals affecting measurements). For example, extracting HRV from PPG instead of ECG (see Chapter 5) would allow it to be measured with a smartwatch. Access to smart home devices is valuable to embed sleep behaviour assessment in activity recognition approaches and to detect potential connections. The correlations between sleep and various factors of daily life, are not yet extensively incorporated. If embedded into existing environments, however, this could soon become feasible. Nevertheless, the incorporation into existing environments will help to make sleep assessment easier to interpret, when potential causes can be targeted.

Fifth, in our framework, we integrate various elements of sleep, some of which require further attention. For example, a method that can reliably provide in-depth information about the impact of the environment still needs to be developed. The inclusion of IoT sensors and long-term monitoring will involve changes in healthcare, such as in sleep disorder diagnosis, and alternative self-management approaches (instead of medication). Hence, in future work, this framework must be tested in a long-term study in a real-world setting. This will help to further acquire the best set of to-be-determined parameters for the used LSTM and thresholds for the sleep parameter extraction. Essentially, providing the information should not be the last step—instead, the last step should be extracting adequate feedback, suggestions, and potential interventions that can easily be followed by users in the future.

Sixth, standard approaches are limited by following existing guidelines. This involves, e.g., assessment in 30-second intervals when a smaller range could potentially detect shorter underlying structures, such as K-complexes and sleep spindles. The detection of these could reveal behavioural changes and will be the focus of future research. Methods are validated with data scored by one technician; multiple scores have an interscorer agreement of 80% [160]. These could be replaced by (1) standards introduced from pattern recognition approaches or (2) models representing different scorers' styles and fusing the outcomes, if not a more accurate, at least a more consistent labelling could be reached.

Seventh, data analysis is, of course, limited by the available data. Our sleep position detection is limited by the data only containing orientation, which could be overcome, in the future, by more complex methods able to monitor the absolute

or relative positions. In several studies throughout this thesis (chapters 3, 5, and 7), we assumed the availability of the necessary data. In reality, however, they are not necessarily shared even though they can be easily obtained. Generally, smaller datasets are restricted in that they result in the outcomes being less generalisable (see Chapter 4). We found that the more diverse a dataset, the more representative the performance. A potential solution is a crowdsourcing approach. Development in consumer wearables will adapt to allow users to share sleep behaviour data with researchers. Diverse datasets also make possible the investigation of influences from comorbidities.

We expect 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 assessment will soon vanish, opening up opportunities that will potentially lead to transformations in future healthcare. This change will, however, require close collaboration and knowledge-sharing among healthcare professionals, researchers, and users.

8.4 Concluding Remarks

This thesis has contributed significantly to knowledge in the areas of home-based sleep behaviour analysis. This has been achieved through the development of enabling technologies within a comprehensive framework. Specifically, this thesis advances approaches such as more complex sleep position detection, fine-grained sleep stage recognition, parameter-specific sleep regularity assessment, personalised sleep parameter extraction, personalised data visualisation, and chronic disease detection from sleep-wake behaviour, and a framework is provided through which researcher can address various challenges. The literature was reviewed in-depth for the mentioned areas from which open issues were highlighted. These open issues were addressed by our five conducted studies, which culminated in eight contributions to the literature (see Section 8.2). Importantly, several opportunities for future work were indicated and discussed in Section 8.3 to encourage further research into this important subject. Using computational methods to assess and interpret available sleep behaviour data is key for the development of predictive and personalised healthcare methods. Sleep-related issues add high costs to health care and affect individuals differently. As more individuals than ever are affected by sleep problems, demand is rapidly increasing for home-based sleep assessment that delivers adequate, reliable, and personalised sleep evaluation. We are confident that future research will benefit

from the approaches, algorithms, processes, and methods developed and evaluated in this thesis.

Bibliography

[1] M. Häggström and F. Michel, "Human_body_silhouette.svg, Upper_body_front.png, Background made transparent." `https://commons.wikimedia.org/wiki/File:Polysomnografie_Body_en_106.svg`, 2013. Accessed: 2018-11-30.

[2] P. Robinson, "Steel bed.svg, colour adapted and included into overall picture." `https://commons.wikimedia.org/wiki/File:Steel_bed.svg`, 2007. Accessed: 2018-12-03.

[3] American academy of sleep medicine task force, "Sleep-related breathing disorders in adults : recommendations for syndrome definition and measurement techniques in clinical research," *Sleep*, vol. 22, pp. 667–689, 1999.

[4] A. Rechtschaffen and A. Kales, "A Manual of Standardized Terminology, Techniques and Scoring System for Sleep Stages of Human Subjects," *US Government Printing Office, US Public Health Service, Washington DC*, vol. 204, pp. 976–977, 1968.

[5] 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," *Journal of Ambient Intelligence and Humanized Computing*, vol. 9, no. 4, pp. 897–917, 2018.

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

[7] J. R. Lunsford-Avery, M. M. Engelhard, A. Marie Navar, and S. H. Kollins, "Validation of the Sleep Regularity Index in Older Adults and Associations with Cardiometabolic Risk," *Scientific Reports*, vol. 8, 2018.

[8] K. S. Park and S. H. Choi, "Smart technologies toward sleep monitoring at home," *Biomedical Engineering Letters*, vol. 9, pp. 73–85, Feb. 2019.

[9] A. Dias, L. Gorzelniak, J. Rudnik, D. Stojanovic, and A. Horsch, "Detecting periodic limb movements with off-the-shelf accelerometers: A feasibil-

ity study," *Studies in Health Technology and Informatics*, vol. 192, no. 1-2, pp. 152–156, 2013.

[10] 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 and uHealth*, vol. 4, no. 4, p. e125, 2016.

[11] D. Wang, K. Wong, G. C Dungan, P. Buchanan, B. Yee, and R. Grunstein, "The Validity of Wrist Actimetry Assessment of Sleep With and Without Sleep Apnea," *Journal of Clinical Sleep Medicine*, vol. 4, pp. 450–5, 2008.

[12] J. Accardo, *Sleep in Children with Neurodevelopmental Disabilities: An Evidence-Based Guide*. Springer International Publishing, 2019.

[13] 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 *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (New York, NY), pp. 4654–4657, IEEE, July 2013.

[14] S. Matsui, T. Terada, and M. Tsukamoto, "Smart Eye Mask: Sleep Sensing System Using Infrared Sensors," in *Proceedings of the International Symposium on Wearable Computers*, (New York, NY), pp. 58–61, Acm, 2017.

[15] 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," *Physiological Measurement*, vol. 39, no. 12, p. 124005, 2018.

[16] 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 *Proceedings of Long Island Systems, Applications and Technology*, (New York, NY), pp. 1–6, IEEE, 2015.

[17] X. Sun, L. Qiu, Y. Wu, Y. Tang, and G. Cao, "SleepMonitor: Monitoring Respiratory Rate and Body Position During Sleep Using Smartwatch," *Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 3, pp. 1–22, 2017.

[18] K. Aggarwal, S. R. Joty, L. Fernández-Luque, and J. Srivastava, "Co-Morbidity Exploration on Wearables Activity Data Using Unsupervised Pre-training and Multi-Task Learning," *Computing Research Repository*, 2017.

[19] 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," *Physiological Measurement*, vol. 35, no. 1, pp. R1–r57, 2013.

[20] M. J. Deen, "Information and communications technologies for elderly ubiquitous healthcare in a smart home," *Personal and Ubiquitous Computing*, vol. 19, no. 3, pp. 573–599, 2015.

[21] K. L. Knutson, "Role of Sleep Duration and Quality in the Risk and Severity of Type 2 Diabetes Mellitus," *Archives of Internal Medicine*, vol. 166, no. 16, pp. 1768–74, 2006.

[22] S. C. Maung, A. El Sara, C. Chapman, D. Cohen, and D. Cukor, "Sleep disorders and chronic kidney disease," *World Journal of Nephrology*, vol. 5, no. 3, pp. 224–32, 2016.

[23] E. Kasasbeh, D. S. Chi, and G. Krishnaswamy, "Inflammatory aspects of sleep apnea and their cardiovascular consequences," *The Southern medical journal*, vol. 99, no. 1, pp. 58–67; quiz 68–69, 81, 2006.

[24] M. Zimmerman, J. B. McGlinchey, D. Young, and I. Chelminski, "Diagnosing Major Depressive Disorder I: A Psychometric Evaluation of the DSM-IV Symptom Criteria," *The Journal of nervous and mental disease*, vol. 194, pp. 158–163, Mar. 2006.

[25] D. J. Schwartz, W. C. Kohler, and G. Karatinos, "Symptoms of Depression in Individuals With Obstructive Sleep Apnea May Be Amenable to Treatment With Continuous Positive Airway Pressure," *Chest*, vol. 128, no. 3, pp. 1304–1309, 2005.

[26] D. Foley, S. Ancoli-Israel, P. Britz, and J. Walsh, "Sleep disturbances and chronic disease in older adults: Results of the 2003 National Sleep Foundation Sleep in America Survey," *Journal of Psychosomatic Research*, vol. 56, no. 5, pp. 497–502, 2004.

[27] Arthritis Foundation, "Arthritis Facts." Available at `http://www.arthritis.org/about-arthritis/understanding-arthritis/arthritis-statistics-facts.php`. Accessed: 29-10-2017.

[28] Arthritis Foundation, "Sleep and Pain." Available at `https://www.arthritis.org/health-wellness/healthy-living/managing-pain/fatigue-sleep/sleep-and-pain`. Accessed: 17-02-2020.

[29] V. Jha, G. Garcia-Garcia, K. Iseki, Z. Li, S. Naicker, B. Plattner, R. Saran, A. Y. M. Wang, and C. W. Yang, "Chronic kidney disease: Global dimension and perspectives," *The Lancet*, vol. 382, no. 9888, pp. 260–272, 2013.

[30] K. Havas, A. Bonner, and C. Douglas, "Self-management support for people with chronic kidney disease: Patient perspectives," *Journal of Renal Care*, vol. 42, no. 1, pp. 7–14, 2016.

[31] G. Nigam, M. Camacho, E. Chang, and M. Riaz, "Exploring sleep disorders in patients with chronic kidney disease," *Nature and Science of Sleep*, vol. 10, 2018.

[32] C. Thomas-Hawkins and D. Zazworsky, "Self-Management of Chronic Kidney Disease," *American Journal of Nursing*, vol. 105, no. 10, pp. 40–48, 2005.

[33] J. Savard and C. M. Morin, "Insomnia in the Context of Cancer: A Review of a Neglected Problem," *Journal of Clinical Oncology*, vol. 19, no. 3, pp. 895–908, 2001. Pmid: 11157043.

[34] M. Hafner, M. Štěpánek, J. M. Taylor, W. M. Troxel, and C. van Stolk, "Why Sleep Matters-The Economic Costs of Insufficient Sleep: A Cross-Country Comparative Analysis," *Rand health quarterly*, vol. 6, no. 4, p. 11, 2016.

[35] D. Hillman, A. Scott Murphy, and L. Pezzullo, "Economic cost of sleep disorders," *Sleep*, vol. 29, pp. 299–305, 2006.

[36] S. Watson and K. Cherney, "The Effects of Sleep Deprivation on Your Body." `https://www.healthline.com/health/sleep-deprivation/effects-on-body#1`, 2019. Accessed: 2019-08-20.

[37] 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 Research*, vol. 200, no. 2, pp. 795–801, 2012.

[38] M. de Zambotti, A. Goldstone, S. Claudatos, I. M. Colrain, and F. C. Baker, "A validation study of Fitbit Charge 2^{TM} compared with polysomnography in adults," *Chronobiology International*, vol. 35, no. 4, pp. 465–476, 2018.

[39] 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 *2013 6th International Conference on Biomedical Engineering and Informatics*, (New York, NY), pp. 121–126, IEEE, Dec. 2013.

[40] K. R. Lorig and H. R. Holman, "Self-management education: History, definition, outcomes, and mechanisms," *Annals of Behavioral Medicine*, vol. 26, no. 1, pp. 1–7, 2003.

[41] M. Novak, L. Costantini, S. Schneider, and H. Beanlands, "Approaches to Self-Management in Chronic Illness," *Seminars in Dialysis*, vol. 26, no. 2, pp. 188–194, 2013.

[42] R. Ravhichandran, 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," *Personal Informatics & Self-Tracking*, pp. 6864–6875, 2017.

[43] U. Reimer, S. Emmenegger, E. Maier, and Z. Zhang, "Recognizing Sleep Stages with Wearable Sensors in Everyday Settings," *3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health*, pp. 172–179, 2017.

[44] D. Waltisberg, B. Arnrich, and G. Tröster, *Sleep Quality Monitoring with the Smart Bed*, pp. 211–227. London: Springer London, 2014.

[45] 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 *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 6864–6875, Acm, 2017.

[46] 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 Computing and Applications*, vol. 29, no. 8, pp. 1–16, 2018.

[47] 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. 1012, 2017.

[48] 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 Transactions on Biomedical Engineering*, vol. PP, no. 99, p. 1, 2016.

[49] 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 *Proceedings of the EMBEC & NBC*, (Singapore), pp. 1049–1052, Springer Singapore, 2017.

[50] M. Rosenberger, M. Buman, W. Haskell, M. McConnell, and L. Carstensen, "Twenty-four Hours of Sleep, Sedentary Behavior, and Physical Activity with Nine Wearable Devices," *Medicine and Science in Sports and Exercise*, vol. 48, pp. 457–465, 3 2016.

[51] V. Ibàñez, J. Silva, and O. Cauli, "A survey on sleep questionnaires and diaries," *Sleep Medicine*, vol. 42, pp. 90–96, 2018.

[52] J. D. Lorrain, D.; Koninck, "Sleep Position and Sleep Stages: Evidence of Their Independence," *Clinical Medicine - Sleep*, vol. 21, pp. 335–340, June 1998.

[53] R. Yousefi, S. Ostadabbas, M. Faezipour, M. Farshbaf, M. Nourani, L. Tamil, and M. Pompeo, "Bed posture classification for pressure ulcer prevention," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 7175–7178, 2011.

[54] 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," *European Journal of Heart Failure*, vol. 17, no. 12, pp. 1302–1309, 2015.

[55] M. T. Cibulka and J. Threlkeld-Watkins, "Patellofemoral Pain and Asymmetrical Hip Rotation," *Physical Therapy*, vol. 85, no. 11, pp. 1201–1207, 2005.

[56] 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.

[57] P. Barsocchi, "Position recognition to support bedsores prevention," *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 1, pp. 53–59, 2013.

[58] J. Lee, M. Hong, and S. Ryu, "Sleep Monitoring System Using Kinect Sensor," *International Journal of Distributed Sensor Networks*, vol. 11, no. 10, pp. 1–9, 2015.

[59] P. Barsocchi, M. Bianchini, A. Crivello, D. L. Rosa, F. Palumbo, and F. Scarselli, "An unobtrusive sleep monitoring system for the human sleep behaviour understanding," in *7th IEEE International Conference on Cogni-*

tive Infocommunications (CogInfoCom), (New York, NY), pp. 91–96, IEEE, Oct. 2016.

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

[61] 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 *HealthCom 2008 - 10th International Conference on e-health Networking, Applications and Services*, (New York, NY), pp. 165–168, IEEE, July 2008.

[62] A. H. Khandoker, M. Palaniswami, and C. K. Karmakar, "Support vector machines for automated recognition of obstructive sleep apnea syndrome from ECG recordings," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 37–48, 2009.

[63] 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 Journal of Biomedical and Health Informatics*, 2018.

[64] National Heart, Lung and Blood Insitute, U.S. Department of Health and Human Services, "Apnea: Diagnosis-Treatment." `https://www.nhlbi.nih.gov/health-topics/sleep-apnea`. Accessed: 2018-04-19.

[65] T. Grimm, M. Martinez, A. Benz, and R. Stiefelhagen, "Sleep position classification from a depth camera using Bed Aligned Maps," in *Proceedings of the 23rd International Conference on Pattern Recognition*, (New York, NY), pp. 319–324, 2016.

[66] A. Doryab, J. K. Min, J. Wiese, J. Zimmerman, and J. I. Hong, "Detection of behavior change in people with depression," *AAAI Workshops Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pp. 12–16, 2014.

[67] A. G. Harvey, "Insomnia: Symptom Or Diagnosis?," *Clinical Psychology Review*, vol. 21, no. 7, pp. 1037–1059, 2001.

[68] A. G. Harvey and N. Tang, "(Mis)perception of Sleep in Insomnia: A Puzzle and a Resolution," *Psychological bulletin*, vol. 138, pp. 77–101, 10 2011.

[69] T. F. de Araújo, D. C. Jarrin, Y. Leanza, A. Vallières, and C. M. Morin, "Qualitative studies of insomnia: Current state of knowledge in the field," *Sleep medicine reviews*, vol. 31, pp. 58–69, 2017.

[70] H. Tunstall-Pedoe, "Preventing Chronic Diseases. A Vital Investment: WHO Global Report. Geneva: World Health Organization," *International Journal of Epidemiology*, vol. 35, no. 4, 2006.

[71] M. E. Falagas, K. Z. Vardakas, and P. I. Vergidis, "Under-diagnosis of common chronic diseases: prevalence and impact on human health," *International Journal of Clinical Practice*, vol. 61, no. 9, pp. 1569–1579, 2007.

[72] T. P. Ryan, J. A. Sloand, P. C. Winters, J. P. Corsetti, and S. G. Fisher, "Chronic kidney disease prevalence and rate of diagnosis," *The American Journal of Medicine*, vol. 120, no. 11, pp. 981–6, 2007.

[73] T. J. Lyons and A. Basu, "Biomarkers in diabetes: hemoglobin A1c, vascular and tissue markers," *Translational Research*, vol. 159, no. 4, pp. 303–312, 2012.

[74] Mayo Clinic, "High blood pressure (hypertension): Diagnosis." `https://www.mayoclinic.org/diseases-conditions/high-blood-pressure/diagnosis-treatment/drc-20373417`, 2018. Accessed: 2018-04-19.

[75] C. Laske, H. R. Sohrabi, S. M. Frost, K. López-de Ipiña, P. Garrard, M. Buscema, J. Dauwels, S. R. Soekadar, S. Mueller, C. Linnemann, S. A. Bridenbaugh, Y. Kanagasingam, R. N. Martins, and S. E. O'Bryant, "Innovative diagnostic tools for early detection of Alzheimer's disease," *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, vol. 11, no. 5, pp. 561–578, 2015.

[76] A. S. Levey, J. Coresh, E. Balk, A. T. Kausz, A. Levin, M. W. Steffes, R. J. Hogg, R. D. Perrone, J. Lau, and G. Eknoyan, "National Kidney Foundation Practice Guidelines for Chronic Kidney Disease: Evaluation, Classification, and Stratification," *Annals of Internal Medicine*, vol. 139, pp. 137–147, 2003.

[77] Y.-E. S. Ju, J. S. McLeland, C. D. Toedebusch, C. Xiong, A. M. F. Fagan, S. P. Duntley, J. C. Morris, and D. M. Holtzman, "Sleep quality and preclinical alzheimer disease," *JAMA Neurology*, vol. 70, no. 5, pp. 587–593, 2013.

[78] 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 Neurology*, vol. 14, no. 1, p. 76, 2014.

[79] D. A. Calhoun and S. M. Harding, "Sleep and Hypertension," *Chest - American College of Chest Physicians*, vol. 138, no. 2, pp. 434–443, 2010.

[80] M. Gullvåg, K. H. Gjeilo, N. Fålun, T. M. Norekvål, R. Mo, and A. Broström, "Sleepless nights and sleepy days: a qualitative study exploring the experiences of patients with chronic heart failure and newly verified sleep-disordered breathing," *Scandinavian Journal of Caring Sciences*, vol. 0, no. 0, 2019.

[81] J. Li, Y.-S. Yao, Q. Dong, Y.-H. Dong, J.-J. Liu, L.-S. Yang, and F. Huang, "Characterization and factors associated with sleep quality among rural elderly in China," *Archives of gerontology and geriatrics*, vol. 56, no. 1, pp. 237–243, 2013.

[82] D. L. Bliwise, D. J. Foley, M. V. Vitiello, F. P. Ansari, S. Ancoli-Israel, and J. K. Walsh, "Nocturia and disturbed sleep in the elderly," *Sleep Medicine*, vol. 10, no. 5, pp. 540–548, 2009.

[83] A. S. Walters and D. B. Rye, "Review of the Relationship of Restless Legs Syndrome and Periodic Limb Movements in Sleep to Hypertension, Heart Disease, and Stroke," *Sleep*, vol. 32, pp. 589–597, 5 2009.

[84] Y. Hayashino, S. Yamazaki, M. Takegami, T. Nakayama, S. Sokejima, and S. Fukuhara, "Association between number of comorbid conditions, depression, and sleep quality using the Pittsburgh Sleep Quality Index: Results from a population-based survey," *Sleep Medicine*, vol. 11, no. 4, pp. 366–371, 2010.

[85] A. von Ruesten, C. Weikert, I. Fietze, and H. Boeing, "Association of Sleep Duration with Chronic Diseases in the European Prospective Investigation into Cancer and Nutrition (EPIC)-Potsdam Study," *Plos One*, vol. 7, pp. 1–9, 1 2012.

[86] S. Wang, B. Li, Y. Wu, G. Ungvari, C. Ng, Y. Fu, C. Kou, Y. Yu, H. Sun, and Y.-T. Xiang, "Relationship of sleep duration with sociodemographic characteristics, lifestyle, mental health, and chronic diseases in a large Chinese adult population," *Journal of Clincal Sleep Medicine*, vol. 13, no. 3, pp. 377–384, 2017.

[87] B. S. Shackell, R. Jones, G. Harding, S. Pearse, and J. Campbell, "'Am I going to see the next morning?' A qualitative study of patients' perspectives of sleep in COPD," *Primary care respiratory journal : journal of the General Practice Airways Group*, vol. 16, pp. 378–83, 1 2008.

[88] J. J. Herdegen, *Chronic Disease and Sleep Architecture*, pp. 163–169. Boston, MA: Springer US, 2006.

[89] M. Hepburn, P. C. Bollu, B. French, and P. Sahota, "Sleep Medicine: Stroke and Sleep," *Journal of the Missouri State Medical Association*, vol. 116, no. 6, pp. 527–532, 2018.

[90] B. V. Fortner, E. J. Stepanski, S. C. Wang, S. Kasprowicz, and H. Durrence, "Sleep and Quality of Life in Breast Cancer Patients," *Journal of Pain and Symptom Management*, vol. 24, no. 5, pp. 471–480, 2002.

[91] G. H. Louie, M. G. Tektonidou, A. J. Caban-Martinez, and M. M. Ward, "Sleep disturbances in adults with arthritis: Prevalence, mediators, and subgroups at greatest risk. Data from the 2007 National Health interview survey," *Arthritis Care & Research*, vol. 63, no. 2, pp. 247–260, 2011.

[92] 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," *Computers in Biology and Medicine*, vol. 103, pp. 71–81, 2018.

[93] A. Procházka, J. Kuchyňka, O. Vyšata, P. Cejnar, M. Vališ, and V. Mařík, "Multi-Class Sleep Stage Analysis and Adaptive Pattern Recognition," *Applied Sciences*, vol. 8, no. 5, p. 697, 2018.

[94] 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 Transactions on Biomedical Circuits and Systems*, vol. 11, no. 1, pp. 189–202, 2017.

[95] M. Peker, "An efficient sleep scoring system based on EEG signal using complex-valued machine learning algorithms," *Neurocomputing*, vol. 207, no. May 2016, pp. 165–177, 2016.

[96] M. Shahin, B. Ahmed, S. T. Hamida, F. L. Mulaffer, M. Glos, and T. Penzel, "Deep learning and insomnia: Assisting clinicians with their diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 6, pp. 1546–1553, 2017.

[97] 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," *Biomedical Signal Processing and Control*, vol. 42, pp. 107–114, 2018.

[98] 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," *Chronobiology International*, vol. 31, no. 3, pp. 356–362, 2014.

[99] Z. Zhang and G. Yang, "Monitoring cardio-respiratory and posture movements during sleep: What can be achieved by a single motion sensor," in *12th IEEE International Conference on Wearable and Implantable Body Sensor Networks, BSN 2015, Cambridge, MA, USA*, (New York, NY), pp. 1–6, IEEE, June 2015.

[100] M. Yeo, Y. S. Koo, and C. Park, "Automatic Detection of Sleep Stages based on Accelerometer Signals from a Wristband," *IEIE Transactions on Smart Processing and Computing*, vol. 6, no. 1, pp. 21–26, 2017.

[101] 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," *Physiology & behavior*, vol. 90, pp. 257–266, Feb. 2007.

[102] 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 *Proceedings of the EMBEC & NBC*, (Singapore), pp. 615–618, Springer Singapore, 2017.

[103] 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.

[104] 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," *Journal of Affective Disorders*, vol. 217, no. April, pp. 299–305, 2017.

[105] J. Mantua, N. Gravel, and R. Spencer, "Reliability of Sleep Measures from Four Personal Health Monitoring Devices Compared to Research-Based Actigraphy and Polysomnography," *Sensors*, vol. 16, p. 646, 2016.

[106] A. Tal, Z. Shinar, D. Shaki, S. Codish, and A. Goldbart, "Validation of Contact-Free Sleep Monitoring Device with Comparison to Polysomnography," *Journal of Clinical Sleep Medicine*, vol. 13, no. 3, pp. 517–522, 2016.

[107] 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 *Proceedings of the EMBEC & NBC*, (Singapore), pp. 77–80, Springer, 2017.

[108] B. Vandenberghe and D. Geerts, *Oh, What a Night! Effortless Sleep Monitoring at Home*, pp. 417–424. Cham: Springer International Publishing, 2015.

[109] 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," *Journal of Sleep Research*, vol. 28, no. 4, p. e12789, 2019.

[110] J. Cook, M. Prairie, and D. 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," *Journal of Clinical Sleep Medicin*, vol. 14, no. 5, pp. 841–848, 2018.

[111] Z. Liang and M. A. Chapa Martell, "Validity of Consumer Activity Wristbands and Wearable EEG for Measuring Overall Sleep Parameters and Sleep Structure in Free-Living Conditions," *Journal of Healthcare Informatics Research*, vol. 2, no. 1, pp. 152–178, 2018.

[112] S.-G. 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," *Journal of Psychosomatic Research*, vol. 97, pp. 38–44, 2017.

[113] M. de Zambotti, N. Cellini, A. Goldstone, I. Colrain, and F. Baker, "Wearable Sleep Technology in Clinical and Research Settings," *Medicine and Science in Sports and Exercise*, vol. 51, no. 7, pp. 1538–1557, 2019.

[114] J. Behar, A. Roebuck, J. S Domingos, E. Gederi, and G. D Clifford, "A review of current sleep screening applications for smartphons," *Physiological Measurement*, vol. 34, pp. R29–46, 6 2013.

[115] T. Rahman, A. T. Adams, R. V. Ravichandran, M. Zhang, S. N. Patel, J. A. Kientz, and T. Choudhury, "DoppleSleep: A Contactless Unobtrusive Sleep Sensing System Using Short-range Doppler Radar," in *Proceedings of the International Joint Conference on Pervasive and Ubiquitous Computing*, (New York, NY), pp. 39–50, ACN, 2015.

[116] 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 (Switzerland)*, vol. 17, no. 7, pp. 1–17, 2017.

[117] E. Dafna, A. Tarasiuk, and Y. Zigel, "Sleep staging using nocturnal sound analysis," *Scientific Reports*, vol. 8, 2018.

[118] J. J. Liu, W. Xu, M. C. Huang, N. Alshurafa, M. Sarrafzadeh, N. Raut, and B. Yadegar, "Sleep posture analysis using a dense pressure sensitive bedsheet," *Pervasive and Mobile Computing*, vol. 10, pp. 34–50, 2014.

[119] M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning Sleep Stages from Radio Signals: A Conditional Adversarial Architecture," in *Proceedings of the 34th International Conference on Machine Learning*, vol. 70 of *Machine Learning Research*, pp. 4100–4109, PMLR, August 2017.

[120] 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 *Proceedings of the 2nd International Conference on Computer and Communications*, (New York, NY), pp. 414–417, IEEE, 2016.

[121] Y. Choi, G. Demiris, S. Lin, S. Iribarren, C. Landis, H. Thompson, S. McCurry, M. Heitkemper, and T. Ward, "Smartphone Applications to Support Sleep Self-Management: Review and Evaluation," *Journal of Clinical Sleep Medicine*, vol. 14, no. 10, pp. 1783–1790, 2018.

[122] E. Fino and M. Mazzetti, "Monitoring healthy and disturbed sleep through smartphone applications: a review of experimental evidence," *Sleep and Breathing*, 2018.

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

[124] L. J. Meltzer and A. M. Westin, "A comparison of actigraphy scoring rules used in pediatric research," *Sleep Medicine*, vol. 12, no. 8, pp. 793–796, 2011.

[125] L. J. Meltzer, C. M. Walsh, and A. A. Peightal, "Comparison of actigraphy immobility rules with polysomnographic sleep onset latency in children and adolescents," *Sleep and Breathing*, vol. 19, pp. 1415–1423, 2015.

[126] G. Valenti, A. G. Bonomi, and K. R. Westerterp, "Quality Sleep Is Associated With Overnight Metabolic Rate in Healthy Older Adults," *The journals of*

gerontology. Series A, Biological sciences and medical sciences, vol. 72, no. 4, pp. 567–571, 2016.

[127] Y. Nam, Y. Kim, and J. Lee, "Sleep monitoring based on a tri-axial accelerometer and a pressure sensor," *Sensors (Switzerland)*, vol. 16, no. 5, pp. 1–14, 2016.

[128] A. Khademi, Y. El-Manzalawy, O. M. Buxton, and V. Honavar, "Toward personalized sleep-wake prediction from actigraphy," in *2018 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, pp. 414–417, IEEE-ebms, Mar. 2018.

[129] M. Iwasaki, S. Iwata, A. Iemura, N. Yamashita, Y. Tomino, T. Anme, Z. Yamagata, O. Iwata, and T. Matsuishi, "Utility of Subjective Sleep Assessment Tools for Healthy Preschool Children: A Comparative Study Between Sleep Logs, Questionnaires, and actigraphy," *Journal of Epidemiology*, vol. 20, no. 2, pp. 143–149, 2010.

[130] A. J. K. Phillips, W. M. Clerx, C. S. O'Brien, A. Sano, L. K. Barger, R. W. Picard, S. W. Lockley, E. B. Klerman, and C. A. Czeisler, "Irregular sleep/wake patterns are associated with poorer academic performance and delayed circadian and sleep/wake timing," *Scientific reports*, vol. 7, no. 1, p. 3216, 2017.

[131] L. Matuzaki, R. Santos-Silva, E. C. Marqueze, C. R. de Castro Moreno, S. Tufik, and L. Bittencourt, "Temporal sleep patterns in adults using actigraph," *Sleep Science*, vol. 7, no. 3, pp. 152–157, 2014.

[132] Q. Huang, D. Cohen, S. Komarzynski, X. M. Li, P. Innominato, F. Lévi, and B. Finkenstädt, "Hidden Markov models for monitoring circadian rhythmicity in telemetric activity data," *Journal of the Royal Society, Interface*, vol. 15, no. 139, 2018.

[133] C. Blume, N. Santhi, and M. Schabus, "'nparACT' package for R: A free software tool for the non-parametric analysis of actigraphy data," *MethodsX*, vol. 3, 05 2016.

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

[135] H.-C. Hung, Y.-C. Yang, H.-Y. Ou, J.-S. Wu, F.-H. Lu, and C.-J. Chang, "The association between self-reported sleep quality and overweight in a Chinese population," *Obesity*, vol. 21, no. 3, pp. 486–492, 2013.

[136] J. Shim and S. W. Kang, "Behavioral Factors Related to Sleep Quality and Duration in Adults," *Journal of lifestyle medicine*, vol. 7, no. 1, pp. 18–26, 2017.

[137] D. Gu, J. Sautter, R. Pipkin, and Y. Zeng, "Sociodemographic and Health Correlates of Sleep Quality and Duration among Very Old Chinese," *Sleep*, vol. 33, no. 5, pp. 601–610, 2010.

[138] H. Middelkoop, D. A Smilde-van den Doel, A. Neven, H. A. C. Kamphuisen, and C. P Springer, "Subjective Sleep Characteristics of 1,485 Males and Females Aged 50-93: Effects of Sex and Age, and Factors Related to Self-Evaluated Quality of sleep," *The Journals of Gerontology. Series A, Biological sciences and medical sciences*, vol. 51, pp. 108–15, 6 1996.

[139] M. Kim, C. Chun, and J. Han, "A Study on Bedroom Environment and Sleep Quality in Korea," *Indoor and Built Environment*, vol. 19, no. 1, pp. 123–128, 2010.

[140] P. Batog and M. Badura, "Dynamic of changes in carbon dioxide concentration in bedrooms," *Procedia Engineering*, vol. 57, pp. 175–182, 2013.

[141] A. Malafeev, D. Laptev, S. Bauer, X. Omlin, A. Wierzbicka, A. Wichniak, W. Jernajczyk, R. Riener, J. Buhmann, and P. Achermann, "Automatic Human Sleep Stage Scoring Using Deep Neural Networks," *Frontiers in Neuroscience*, vol. 12, no. November, pp. 1–15, 2018.

[142] D. C. Lim, K. Sutherland, P. A. Cistulli, and A. I. Pack, "P4 medicine approach to obstructive sleep apnoea," *Respirology*, vol. 22, no. 5, pp. 849–860, 2017.

[143] K. Morgan, "Daytime activity and risk factors for late-life insomnia," *Journal of sleep research*, vol. 12, no. 3, pp. 231–8, 2003.

[144] V. K. Kapur, "Personalizing sleep apnea care," *Sleep*, vol. 41, 3 2018.

[145] C. Chen, A. Ugon, X. Zhang, A. Amara, P. Garda, J. Ganascia, C. Philippe, and A. Pinna, "Personalized sleep staging system using evolutionary algorithm and symbolic fusion," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2266–2269, Aug. 2016.

[146] H. Yang, L. Rothrock, S. Tirupatikumara, and R. Singareddy, "VisOSA: Visualizing Obstructive Sleep Apnea symptoms and comorbidities," in *IIE Annual*

Conference and Expo 2013, (United States), pp. 1582–1591, Institute of Industrial Engineers, 2013.

[147] M. Ohayon, E. M. Wickwire, M. Hirshkowitz, S. M. Albert, A. Avidan, F. J. Daly, Y. Dauvilliers, R. Ferri, C. Fung, D. Gozal, N. Hazen, A. Krystal, K. Lichstein, M. Mallampalli, G. Plazzi, R. Rawding, F. A. Scheer, V. Somers, and M. V. Vitiello, "National Sleep Foundation's sleep quality recommendations: first report," *Sleep Health*, vol. 3, no. 1, pp. 6–19, 2017.

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

[149] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[150] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, "SVMs Modeling for Highly Imbalanced Classification," *Trans. Sys. Man Cyber. Part B*, vol. 39, no. 1, pp. 281–288, 2009.

[151] S. Abe, *Support Vector Machines for Pattern Classification*. Springer Publishing Company, Inc., 2nd ed., 2010.

[152] N. Bhatia and Vandana, "Survey of Nearest Neighbor Techniques," *ArXiv*, 2010.

[153] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[154] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers Inc., 3rd ed., 2011.

[155] S. Hochreiter and J. Schmidhuber, "Long Short-term Memory," *Neural Computations*, vol. 9, no. 8, pp. 1735–1780, 1997.

[156] F. A. Gers, J. A. Schmidhuber, and F. A. Cummins, "Learning to Forget: Continual Prediction with LSTM," *Neural Computions*, vol. 12, no. 10, pp. 2451–2471, 2000.

[157] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 3, pp. 189–194, 2000.

[158] Z. C. Lipton, D. C. Kale, C. Elkan, and R. C. Wetzel, "Learning to Diagnose with LSTM Recurrent Neural Networks," *Computing Research Repository*, 2015.

[159] P. Jiang and R. Zhu, "Dual tri-axis accelerometers for monitoring physiological parameters of human body in sleep," *Proceedings of IEEE Sensors*, vol. 6050, pp. 2–4, 2017.

[160] R. S. Rosenberg and S. Van Hout, "The American Academy of Sleep Medicine Inter-scorer Reliability Program: Sleep Stage Scoring," *Journal of Clinical Sleep Medicine*, vol. 9, pp. 81–7, 1 2013.

[161] S. Kampakis, "The Data Scientist-Everything about data science, blockchain and AI." `http://thedatascientist.com/performance-measures-cohens-kappa-statistic/`, 2016. Accessed: 2018-01-07.

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

[163] N. Schaltenbrand, R. Lengellé, 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.

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

[165] K. Fujimoto, Y. Ding, and E. Takahashi, "Sleep stage detection using a wristwatch-type physiological sensing device," *Sleep and Biological Rhythms*, vol. 16, 2018.

[166] K. Aggarwal, S. Khadanga, S. R. Joty, L. Kazaglis, and J. Srivastava, "A Structured Learning Approach with Neural Conditional Random Fields for Sleep Staging," *ArXiv*, 2018.

[167] W.-H. Liao and C.-M. Yang, "Video-based activity and movement pattern analysis in overnight sleep studies," in *Proceedings of the 19th International Conference on Pattern Recognition*, (New York, NY), pp. 1–4, IEEE, 2008.

[168] M. Borazio, E. Berlin, N. KÃijcÃijkyildiz, P. Scholl, and K. Van Laerhoven, "Towards Benchmarked Sleep Detection with Inertial Wrist-worn Sensing Units," in *Proceedings - 2014 IEEE International Conference on Healthcare Informatics, ICHI 2014*, 09 2014.

[169] A. R. Hassan and M. I. H. Bhuiyan, "Automatic sleep scoring using statistical features in the EMD domain and ensemble methods," *Biocybernetics and Biomedical Engineering*, vol. 36, no. 1, pp. 248–255, 2016.

[170] R. Sharma, R. B. Pachori, and A. Upadhyay, "Automatic sleep stages classification based on iterative filtering of electroencephalogram signals," *Neural Computing and Applications*, vol. 28, no. 10, pp. 2959–2978, 2017.

[171] 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, 2013.

[172] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A Model for Automatic Sleep Stage Scoring Based on Raw Single-Channel EEG," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 11, pp. 1998–2008, 2017.

[173] A. Koushik, J. Amores, and P. Maes, "Real-Time Sleep Staging using Deep Learning on a Smartphone for a Wearable EEG," *ArXiv*, 2018.

[174] B. Koley and D. Dey, "An ensemble system for automatic sleep stage classification using single channel EEG signal," *Computers in Biology and Medicine*, vol. 42, no. 12, pp. 1186–1195, 2012.

[175] K. A. I. Aboalayon, H. T. Ocbagabir, and M. Faezipour, "Efficient sleep stage classification based on EEG signals," in *Proceedings of Long Island Systems, Applications and Technology Conference*, (New York, NY), pp. 1–6, IEEE, 2014.

[176] Z. Shinar, A. Baharav, and S. Akselrod, "Detection of different recumbent body positions from the electrocardiogram," *Medical and Biological Engineering and Computing*, vol. 41, no. 2, pp. 206–210, 2003.

[177] G.-Z. Liu, K. Li, L. Zheng, W. 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.

[178] P. Jeng and L.-C. Wang, "Stream data analysis of body sensors for sleep posture monitoring: An automatic labelling approach," in *Proceedings of the 26th Wireless and Optical Communication Conference*, (New York, NY), pp. 1–4, IEEE, 2017.

[179] 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," *European Journal of Medical Research*, vol. 23, no. 1, p. 28, 2018.

[180] 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 *Proceedings of the 37th Annual International Conference of the Engineering in Medicine and Biology Society*, (New York, NY), pp. 4942–4945, IEEE, 2015.

[181] M. B. Pouyan, J. Birjandtalab, M. Heydarzadeh, M. Nourani, and S. Ostadabbas, "A pressure map dataset for posture and subject analytics," in *Proceedings of the International Conference on Biomedical Health Informatics*, (New York, NY), pp. 65–68, IEEE-EBMS, 2017.

[182] W. Viriyavit, V. Sornlertlamvanich, W. Kongprawechnon, P. Pongpaibool, and T. Isshiki, "Neural network based bed posture classification enhanced by Bayesian approach," in *Proceedings of the 8th International Conference of Information and Communication Technology for Embedded Systems*, (New York, NY), pp. 1–5, IEEE, 2017.

[183] 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 *Proceedings of the 18th International Conference on e-Health Networking, Applications and Services*, (New Yor, NY), pp. 1–6, IEEE, 2016.

[184] M. Wang, P. Liu, and W. Gao, "Automatic Sleeping Posture Detection in Ballistocardiography," in *Communications, Signal Processing, and Systems*, (Singapore), pp. 1785–1792, Springer Singapore, 2019.

[185] M. Enayati, M. Skubic, J. M. Keller, M. Popescu, and N. Z. Farahani, "Sleep Posture Classification Using Bed Sensor Data and Neural Networks," in *Proceedings of the 40th Annual International Conference of the Engineering in Medicine and Biology Society*, (New York, NY), pp. 461–465, IEEE, 2018.

[186] 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 *Proceedings of the International Conference on Digital Health*, (New York, NY), pp. 75–79, Acm, 2018.

[187] 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 *Proceedings of 14th International Conference on Machine Learning and Applications*, (New York, NY), pp. 837–842, IEEE, 2015.

[188] 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," *Annals of Epidemiology*, vol. 20, no. 8, pp. 629–641, 2010.

[189] G. C. Gutierrez-Tobal, D. Alvarez, A. Crespo, F. Del Campo, and R. Hornero, "Evaluation of Machine-Learning Approaches to Estimate Sleep Apnea Severity from at-Home Oximetry Recordings," *IEEE Journal of Biomedical and Health Informatics*, vol. 2194, 2018.

[190] F. Chung, P. Liao, H. Elsaid, S. Islam, C. Shapiro, and Y. Sun, "Oxygen Desaturation Index from Nocturnal Oximetry: A Sensitive and Specific Tool to Detect Sleep-Disordered Breathing in Surgical Patients," *Anesthesia and analgesia*, vol. 114, no. 5, pp. 993–1000, 2012.

[191] 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 *Proceedings of the 28th International Workshop on Machine Learning for Signal Processing*, (New York, NY), pp. 1–6, IEEE, 2018.

[192] K. Lacourse, J. Delfrate, J. Beaudry, P. Peppard, and S. C. Warby, "A sleep spindle detection algorithm that emulates human expert spindle scoring," *Journal of Neuroscience Methods*, 2018.

[193] Z. Ahmed, S. Garfield, Y. Jani, S. Jheeta, and B. Franklin, "Impact of electronic prescribing on patient safety in hospitals: Implications for the UK," *Clinical Pharmacist*, vol. 8, 5 2016.

[194] M. Hirshkowitz, K. Whiton, S. M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E. S. Katz, L. Kheirandish-Gozal, D. N. Neubauer, A. E. O'Donnell, M. Ohayon, J. Peever, R. Rawding, R. C. Sachdeva, B. Setters,

M. V. Vitiello, J. C. Ware, and P. J. A. Hillard, "National Sleep Foundation's sleep time duration recommendations: methodology and results summary," *Sleep Health*, vol. 1, no. 1, pp. 40–43, 2015.

[195] T. Reichherzer, S. Satterfield, J. Belitsos, J. Chudzynski, and L. Watson, *An Agent-Based Architecture for Sensor Data Collection and Reasoning in Smart Home Environments for Independent Living*, pp. 15–20. Cham: Springer International Publishing, 2016.

[196] J. Doyle, A. Kealy, J. Loane, L. Walsh, B. O'Mullane, C. Flynn, A. Macfarlane, B. Bortz, R. B. Knapp, and R. Bond, "An integrated home-based self-management system to support the wellbeing of older adults," *Journal of ambient intelligence and smart environments*, vol. 6, no. 4, pp. 359–383, 2014.

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

[198] A. van Straten, T. van der Zweerde, A. Kleiboer, P. Cuijpers, C. M. Morin, and J. Lancee, "Cognitive and behavioral therapies in the treatment of insomnia: A meta-analysis," *Sleep Medicine Reviews*, 2017.

[199] A. N. Seiger, K. A. Dudley, M. Rueschman, J. Weng, R. Wang, S. R. Patel, M. Ramirez, J. S. Loredo, Y. Mossavar-Rahmani, A. R. Ramos, P. C. Zee, K. Reid, and D. Sotres-Alvarez, "Reproducibility of a Standardized Actigraphy Scoring Algorithm for Sleep in a US Hispanic/Latino Population," *Sleep*, vol. 38, pp. 1497–1503, 9 2015.

[200] S. Tao, L. Cui, G.-Q. Zhang, D. Mobley, M. Kim, M. Rueschman, R. Mueller, S. Mariani, and S. Redline, "The National Sleep Research Resource: towards a sleep data commons," *Journal of the American Medical Informatics Association*, vol. 25, pp. 1351–1358, 5 2018.

[201] D. E. Bild, D. A. Bluemke, G. L. Burke, R. Detrano, A. V. Diez Roux, A. R. Folsom, P. Greenland, D. R. JacobsJr., R. Kronmal, K. Liu, J. C. Nelson, D. O'Leary, M. F. Saad, S. Shea, M. Szklo, and R. P. Tracy, "Multi-Ethnic Study of Atherosclerosis: Objectives and Design," *American Journal of Epidemiology*, vol. 156, pp. 871–881, 11 2002.

[202] National Sleep Research Resource, "Free research data and tools." `https://sleepdata.org`. Accessed: 2018-05-10.

[203] National Sleep Research Resource, "Multi-Ethnic Study of Atherosclerosis." `https://sleepdata.org/datasets/mesa`. Accessed: 2019-03-14.

[204] D. L. J. De Koninck and P. Gagnon, "Sleep Positions and Position Shifts in Five Age Groups: An Ontogenetic Picture," *American Sleep Disorders Association and Sleep Research Society*, vol. 15, no. 2, 1992.

[205] I. Skog, P. Handel, J. O. Nilsson, and J. Rantakokko, "Zero-Velocity Detection; An Algorithm Evaluation," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 2657–2666, Nov. 2010.

[206] P. Schneider, M. Biehl, and B. Hammer, "Adaptive Relevance Matrices in Learning Vector Quantization," *Neural computation*, vol. 21, no. 12, pp. 3532–3561, 2009.

[207] M. Biehl, P. Schneider, D. Smith, H. Stiekema, A. Taylor, B. Hughes, C. Shackleton, P. Stewart, and W. Arlt, "Matrix relevance LVQ in steroid metabolomics based classification of adrenal tumors," in *20th European Symposium on Artificial Neural Networks (ESANN 2012)* (M. Verleysen, ed.), pp. 423–428, d-side publishing, 2012.

[208] L. Benoist, M. de Ruiter, J. de Lange, and N. de Vries, "A randomized, controlled trial of positional therapy versus oral appliance therapy for position-dependent sleep apnea," *Sleep Medicine*, vol. 34, pp. 109–117, 2017.

[209] G. E. de Vries, A. Hoekema, M. H. Doff, H. A. Kerstjens, P. M. Meijer, J. H. Hoeven, and P. J. Wijkstra, "Usage of Positional Therapy in Adults with Obstructive Sleep Apnea," *Journal of Clinical Sleep Medicine*, vol. 11, pp. 131–137, 2 2015.

[210] Z. Wang, Z. Yang, and T. Dong, "A Review of Wearable Technologies for Elderly Care that Can Accurately Track Indoor Position, Recognize Physical Activities and Monitor Vital signs in real time," *Sensors*, vol. 17, no. 2, p. 341, 2017.

[211] Enflux, "Motion Capture Clothing For Creators and Developers." Online, 2017. Available at `https://www.getenflux.com/`. Accessed: 30-06-2017.

[212] D. Roetenberg, H. Luinge, and P. Slycke, "Xsens MVN: Full 6DOF Human Motion Tracking Using Miniature Inertial Sensors," tech. rep., xsens, 2009.

[213] S. Xu, Y. Zhang, L. Jia, K. Mathewson, K. Jang, J. Kim, H. Fu, X. Huang, P. Chava, R. Wang, S. Bhole, L. Wang, Y. Na, Y. Guan, M. Flavin, Z. Han, Y. Huang, and J. Rogers, "Soft microfluidic assemblies of sensors, circuits, and radios for the skin," *Science*, vol. 344, pp. 70–74, 2014.

[214] T. Rusterholz, L. Tarokh, H. P. A. Van Dongen, and P. Achermann, "Interindividual differences in the dynamics of the homeostatic process are trait-like and distinct for sleep versus wakefulness," *Journal of Sleep Research*, vol. 26, no. 2, pp. 171–178, 2017.

[215] M. Figueiro, L. Sahin, C. Roohan, M. Kalsher, B. Plitnick, and M. Rea, "Effects of red light on sleep inertia," *Nature and Science of Sleep*, vol. 11, pp. 45–57, 5 2019.

[216] Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology, "Heart Rate Variability," *Circulation*, vol. 93, no. 5, pp. 1043–1065, 1996.

[217] National Sleep Research Resource, "Multi-Ethnic Study of Atherosclerosis, HRV Analysis Overview." `https://sleepdata.org/datasets/mesa/pages/hrv-analysis.md`. Accessed: 2018-03-11.

[218] National Sleep Research Resource, "MESA Exam 5 Sleep Data Documentation Guide." `https://sleepdata.org/datasets/mesa/files/documentation/MESA_Sleep_Data_Documentation_Guide.pdf`. Accessed: 2020-01-01.

[219] V. Parro and L. Valdo, "Sleep-wake detection using recurrence quantification analysis," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 8, p. 85706, 2018.

[220] N. Pinheiro, R. Couceiro, J. Henriques, J. Muehlsteff, I. Quintal, L. Gonçalves, and P. Carvalho, "Can PPG be used for HRV analysis?," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 2945–2949, Aug. 2016.

[221] V. Braun and V. Clarke, "Using Thematic Analysis in Psychology," *Qualitative research in psychology*, vol. 3, pp. 77–101, 1 2006.

[222] V. Braun, V. Clarke, N. Hayfield, and G. Terry, *Thematic Analysis*. Singapore: Springer Singapore, 2019.

[223] M. P. Mallampalli and C. L. Carter, "Exploring Sex and Gender Differences in Sleep Health: A Society for Women's Health Research Report," *Journal of Women's Health*, vol. 23, no. 7, pp. 553–562, 2014. Pmid: 24956068.

[224] A. Graves, "Generating Sequences With Recurrent Neural Networks," *Computing Research Repository*, 2013.

[225] Google Code, "Word2vec." https://code.google.com/archive/p/word2vec/. Accessed: 2018-08-23.

[226] T. Mikolov, Q. V. Le, and I. Sutskever, "Exploiting Similarities among Languages for Machine Translation," *CoRR*, 2013.

[227] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, Jan. 2014.

[228] S. M. Grundy, "Pre-Diabetes, Metabolic Syndrome, and Cardiovascular Risk," *Journal of the American College of Cardiology*, vol. 59, no. 7, pp. 635–643, 2012.

[229] Diabetes.co.uk - the global diabetes community, "Diabetes and Sleep." https://www.diabetes.co.uk/diabetes-and-sleep.html. Accessed: 2018-05-02.

[230] S. B. Ghaderian and S. S. Beladi-Mousavi, "The role of diabetes mellitus and hypertension in chronic kidney disease," *Journal of Renal Injury Prevention*, vol. 3, no. 4, pp. 109–110, 2014.

[231] J. A. Pinto, D. K. Ribeiro, A. F. d. S. Cavallini, and G. S. Duarte, Caue and-Freitas, "Comorbidities Associated with Obstructive Sleep Apnea: a Retrospective Study," *International Archives of Otorhinolaryngology*, vol. 20, no. 2, pp. 145–150, 2016.

[232] M. Gatineau, C. Hancock, N. Holman, H. Outhwaite, L. Oldridge, A. Christie, and L. Ells, "Public Health England Adult obesity and type 2 diabetes." https://www.gov.uk/government/publications/adult-obesity-and-type-2-diabetes, 2014. Accessed: 2018-05-01.

[233] M.-J. Kim, N.-K. Lim, S.-J. Choi, and H.-Y. Park, "Hypertension is an independent risk factor for type 2 diabetes: the Korean genome and epidemiology study," *Hypertension Research*, vol. 38, no. 11, pp. 783–789, 2015.

[234] M. Deo, P. Pawar, S. Kanetkar, and S. Kakade, "Prevalence and risk factors of hypertension and diabetes in the Katkari tribe of coastal Maharashtra," *Journal of Postgraduate Medicine*, vol. 63, no. 2, pp. 106–113, 2017.

[235] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *CoRR*, 2014.

[236] L. C. Gallo, F. J. Penedo, M. Carnethon, C. Isasi, D. Sotres-Alvarez, V. L. Malcarne, S. C. Roesch, M. E. Youngblood, M. L. Daviglus, P. Gonzalez, and

G. P. Talavera, "The Hispanic Community Health Study/Study of Latinos Sociocultural Ancillary Study: Sample, Design, and Procedures," *Ethnicity & disease*, vol. 24, no. 1, pp. 77–83, 2014.

[237] T. Dad and D. E. Weiner, "Stroke and Chronic Kidney Disease: Epidemiology, Pathogenesis, and Management Across Kidney Disease Stages," *Seminars in Nephrology*, vol. 35, pp. 311–322, 7 2015.