POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE
Laurea Magistrale in Ingegneria Biomedica

Exploration of multi modal datasets towards sleep quality analysis

Author: Xhesjana Mufali

Student ID: 10585961
Advisor: Prof. Elena De Momi
Co-advisors: Erika Lutin, Emilie Pattyn
Academic Year: 2021 -22

ABSTRACT

Sleep is a recurring daily state essential for a person's health and wellbeing. Nowadays, sleep is being researched and monitored due to all the functional and economic implications of sleep deprivation. Other than sleep disorders, sleep can be disturbed by psychosocial issues such as stress. That is why there is a growing interest in finding new ways to monitor sleep, e.g., using wearable sensors that capture physiological signals like electrodermal activity (EDA). EDA is considered a sensitive index for emotional processing since it reflects sympathetic activity which is activated by the stress response of the body. However, despite the major impact of stress on sleep, there is a lack of EDA studies in the context of sleep quality studies. Therefore, the first aim of this master thesis is to research the possible relationship between EDA signal and sleep quality.

EDA is the result of the activation of the eccrine sweat glands which activate due to thermoregulation or 'emotional' sweating. The second objective of this thesis is to explore the relationship between EDA and skin temperature during both day and night.

Physiological data from wearable sensors (such as the electrocardiogram, skin conductance, skin temperature, activity...) have been collected within IMEC to research their use for sleep analysis. Therefore, a sleep detection algorithm was already developed before the master thesis.

In this master thesis, research was executed with as goal to adapt this algorithm so it would accurately classify sleep without sleep diaries as input. Also, the newly adapted algorithm was validated against other open-source sleep detection algorithms to evaluate sleep classification efficiency. The research questions, previously mentioned, were investigated using data exploration and statistical analysis (e.g., Pearson correlation), machine learning (k-nearest neighbours, decision trees, random forest, support vector machine and multi-layer perceptron), principal component analysis (PCA) and multi-level models (linear mixed regression models).

For the first research question, a relationship between EDA and sleep quality was not really observed. 3 out of the 6 tested machine learning models (k-nearest neighbours,

support vector machine, and multi-layer perceptron) did not perform well (i.e., area under ROC < 0.5). However, the best performing machine learning model was a decision tree that had an accuracy of 71% and f1 score of 72% for the classification of sleep quality using EDA. Also, PCA was applied to the feature dataset, but there weren't additional findings. For the multi-level models, a significant relationship (p-value<0.05) between EDA and sleep quality was not found. However, there were some limitations to this work su the as absence of EDA dynamics in the model (since only the mean EDA was considered for the whole night), a small dataset, lack of confounders in the model, such as temperature and mean motion (which influence the EDA signal), and possible incorrectness of the sleep quality parameters since the sleep algorithm used was not validated with polysomnography (PSG). Therefore, further research is needed to conclude these findings.

Regarding the second objective, a significant relationship between EDA and skin temperature was found for using both data from the day and the night. Moreover, as expected, thermoregulation was not the only process causing the EDA signal since the variance explained by skin temperature was less than 1%. Therefore, the effect size of this interaction is very small. It was also confirmed that motion and day/night status significantly affect EDA signal. When motion and day/night variables were included in the model, the explained variance by the fixed effects increased to 8%. Moreover, a different 'interaction' between EDA and skin temperature during day and night was also observed. However, additional data is needed in order to confirm these results. In future work, data related to general arousal should be included in the models in order to effectively say if the EDA signal from the wrist is related to emotional sweating.

SOMMARIO

Il sonno è uno stato che ricorre quotidianamente, essenziale per la salute e il benessere di una persona. Al giorno d'oggi, il sonno è oggetto di ricerca e monitoraggio a causa di tutte le implicazioni funzionali ed economiche derivanti dalla privazione del sonno. Oltre ai disturbi specifici del sonno, questo può essere disturbato anche da problemi psico-sociali come lo stress. Per questo motivo, c'è un crescente interesse nel trovare nuovi metodi per monitorare il sonno, ad esempio, utilizzando sensori indossabili che catturano segnali fisiologici come l'attività elettro-dermica (EDA). L'EDA è considerato un indicatore sensibile dell'elaborazione emotiva poiché riflette l'attività simpatica attivata dalla risposta allo stress. Tuttavia, nonostante il grande impatto dello stress sul sonno, mancano studi EDA nel contesto degli studi sulla qualità del sonno. Pertanto, il primo obiettivo di questo lavoro di tesi è quello di ricercare la possibile relazione tra il segnale EDA e la qualità del sonno.

L'EDA è il risultato dell'attivazione delle ghiandole sudoripare eccrine che si attivano per termoregolazione o sudorazione "emotiva". Il secondo obiettivo di questa tesi è esplorare la relazione tra il segnale EDA e la temperatura cutanea, sia di giorno che di notte.

I dati fisiologici registrati da sensori indossabili (come l'elettrocardiogramma, la conduttanza cutanea, la temperatura cutanea, l'attività...) sono stati raccolti all'interno di IMEC per ricercarne il loro uso nell'analisi del sonno. Pertanto, un algoritmo di rilevamento del sonno è già stato sviluppato prima della suddetta tesi di laurea.

In questo lavoro si è studiata la possibilità di adattare questo algoritmo già esistente in modo da classificare accuratamente il sonno senza diari del sonno come input. Inoltre, l'algoritmo appena adattato è stato validato rispetto ad altri algoritmi di rilevamento del sonno open source per valutarne l'efficienza di classificazione del sonno. Le domande di ricerca, precedentemente menzionate, sono state studiate utilizzando l'esplorazione dei dati e l'analisi statistica (ad esempio l'indice di correlazione di Pearson), il machine

learning (k-nearest neighbours, alberi decisionali, random forest, support vector machine, multi-layer perceptron, ecc.), PCA (principal component analysis) e modelli multilivello (modelli di regressione lineare misti).

Per la prima domanda di ricerca, non è stata realmente osservata una relazione tra EDA e qualità del sonno. 3 dei 6 modelli testati (k-nearest neighbours, support vector machine, multi-layer perceptron) non hanno funzionato bene (area sotto la curva ROC <0,5). Tuttavia, il modello con le migliori prestazioni è stato decision tree con un'accuratezza del 71% e un punteggio f1 del 72% per la classificazione della qualità del sonno mediante EDA. Anche la PCA è stata applicata al set di dati, ma non sono stati rilevati ulteriori risultati. Per i modelli multilivello non è stata trovata una relazione significativa (p-value < 0,05) tra EDA e qualità del sonno. Tuttavia, esistono alcune limitazioni a questo lavoro come un set di dati di piccole dimensioni, la mancanza di fattori confondenti nel modello come la temperatura e il movimento medio (che potrebbero influenzare il segnale EDA) e la possibile scorrettezza dei parametri della qualità del sonno poiché l'algoritmo del sonno utilizzato non è stato convalidato con polisonnografia (PSG). Pertanto, sono necessarie ulteriori ricerche per concludere questi risultati.

Per quanto riguarda il secondo obiettivo, è stata trovata una relazione significativa tra EDA e temperatura cutanea sia durante il giorno che la notte. Inoltre, come previsto, la termoregolazione non era l'unico processo che causava il segnale EDA poiché la varianza spiegata dalla temperatura cutanea era inferiore all'1%. Pertanto, l'effetto dimensionale di questa interazione è molto piccola. È stato riscontrato che anche il movimento e lo stato giorno/notte influiscono significativamente sul segnale EDA. Quando nel modello vengono incluse le variabili movimento e giorno/notte, la varianza spiegata dagli effetti fissi aumenta all'8%. Inoltre, è stata osservata anche una diversa "interazione" durante il giorno e la notte tra l'EDA e la temperatura cutanea. Tuttavia, sono necessari ulteriori dati per confermare questi risultati. In lavori futuri, i dati relativi all'arousal generale dovrebbero essere inclusi nei modelli per dire se effettivamente c'e' una relazione significativa tra l'EDA del polso e la sudorazione emotiva.

Table of Contents

Abstract ..... **i**

Sommario ..... **iii**

1. Introduction and objectives **1**

2. Literature review **3**

2.1 Importance of sleep and effects of sleep deprivation 3

2.2 Sleep monitoring techniques 4

2.2.1 Sleep architecture 4

2.2.2 Polysomnography 4

2.2.3. Actigraphy 5

2.2.4 Sleep detection algorithms based on actigraphy 7

2.2.5 Wearable technology used for sleep monitoring. 9

2.3 Sleep-related physiological signals 9

2.3.1 Electrocardiogram (ECG) 9

2.3.2 Electrodermal activity (EDA) 11

2.3.3. Skin temperature 13

2.2.4. Other sleep-related physiological signals 13

3. Materials and methods 15

3.1 IMEC's sleep detection algorithm analysis and research 15

3.1.1 Dataset explanation 15

3.1.2 IMEC's sleep detection algorithm 16

3.1.3 SPT window algorithm 17

3.1.4 Random Forest sleep detection algorithm 18

3.1.5 Validation of the IMEC's sleep detection algorithm 18

3.2 Exploration of the relationship between GSR signal and sleep efficiency 19

3.2.1 Features extracting for model development 19

3.2.2 Missing data function 19
3.2.3 Dataset preparation 21
3.2.4 Data exploration 22
3.2.5 Machine learning models used in the analysis 23
3.2.6 Linear mixed models 25
3.3 Exploration of the relationship between GSR and skin temperature 25
3.3.1 Data exploration 26
3.3.2 Linear mixed models 26

4. Results 28
4.1 IMEC's sleep detection algorithm analysis and research 28
4.1.1 IMEC's sleep detection algorithm results 28
4.1.2 SPT window results 29
4.1.3 Validation of IMEC's sleep detection algorithm 31

4.2 Exploration of the relationship between GSR signal and sleep efficiency 34
4.2.1 Missing data function 34
4.2.2 Data exploration 35
4.2.3 Machine learning models implementation and results 38
4.2.4 Mixed model's results 42
4.3 Exploration of the relationship between GSR and skin temperature 44
4.3.1 Data exploration 44
4.3.2 Linear Mixed modelling 45
4.3.3 Relationship of GSR and lagged temperature 48
5. Discussion 51
5.1 IMEC's sleep detection algorithm analysis and research 51
5.1.1 SPT window 51
5.1.2 Missing data function 54
5.2 Exploration of the relationship between GSR signal and sleep efficiency 54
5.3 Exploration of the relationship between GSR and skin temperature 56
5.3.1 Data exploration and machine learning models 56
5.3.2 Mixed models 57
5.3.3 GSR and lagged temperature 58
6. Conclusions 59
6.1 IMEC's sleep detection algorithm analysis and research 59

6.2 Exploration of the relationship between GSR signal and sleep efficiency.....60
6.3 Exploration of the relationship between GSR and skin temperature 61
List of abbreviations 66
List of figures 69
List of tables 71
Bibliography 73

CHAPTER 1

INTRODUCTION AND OBJECTIVES

Sleep is a daily recurring state which is essential for a person's health and wellbeing. Sleep deprivation affects an individual's performance, safety, and quality of life. Other than daytime sleepiness, sleep loss comes with several other consequences including reduced cognitive performance (e.g., attention and reaction time), higher stress levels and health problems (e.g., obesity, depression, cardiovascular diseases, and insulin resistance). Besides negative health impact, sleep loss has also an economic cost associated with doctor visits, hospital services, prescriptions, and over-the-counter medications.

Nowadays, sleep is continuously being researched and monitored due to all these functional and economic implications of sleep deprivation. The causes of sleep loss can be attributed not only to lifestyle factors and sleep disorders but also to psychosocial issues such as stress. Electrodermal activity (EDA) measures, which refer to autonomic changes in the electrical properties of the skin, are highly applicable to emotions and stress research [36]. That is why, there is a growing interest in finding new ways to monitor sleep, e.g., using wearable sensors that capture physiological signals like electrodermal activity (EDA). EDA is result of activation of the eccrine sweat glands which in turn are related either to thermoregulation or ''emotional'' sweating [65] [66]. ''Emotional'' sweating occurs mostly in the palmar and plantar sweat glands and relates to either external or internal stimuli that generate stress or, more generally, arousal. This because EDA is mediated solely by the sympathetic branch of the autonomic nervous system which is supporting the stress ("fight or flight") response to threatening events and is a central component of emotional experience and, by extension, cognition and behavior.

EDA has been utilized to investigate numerous areas such as orienting and attention, learning and conditioning, psychopathology, personality disorders, individual differences, brain asymmetry and laterality, cognitive functions, as well as neuropsychology. However, there are still lack of studies of EDA in the context of sleep quality studies. That is why, in this master thesis project the relationship between EDA and sleep quality measure (i.e., sleep efficiency) will be investigated by means of statistical analysis (e.g., Pearson correlation), machine learning (k-nearest neighbours, decision trees, random forest, supported vector machine, multi-layer perceptron, etc.) and multi-level models.

Physiological data from wearable sensors (such as the electrocardiogram, skin conductance, skin temperature, activity...) have been collected within IMEC to research their use for sleep analysis. Therefore, a sleep detection algorithm was already developed before the master thesis.

In this master thesis, it was researched whether it was possible to adapt this algorithm so it would accurately classify sleep without sleep diaries as input. Also, the newly adapted algorithm was validated against other open-source sleep detection algorithms to evaluate sleep classification efficiency.

As depicted before, EDA is the result of the activation of the eccrine sweat glands which activate due to thermoregulation or 'emotional' sweating. The second objective of this master thesis is to explore the relationship between EDA from the wrist and skin temperature during both day and night by means of statistical analysis (e.g., Pearson correlation), and multi-level models (linear mixed models).

This thesis work is organized as follows. Chapter 2 is dedicated to literature review which starts with a brief overview of the importance of sleep studies. Afterwards, a description of the current monitoring techniques together with their advantages and disadvantages is reported, emphasizing the method employed in this work and the state of art of the sleep detection algorithms. Finally, an overview of the most important physiological signals employed in the sleep studies, more specifically in this work is presented.

Chapter 3, which is dedicated to the materials and methods needed to research the raised questions, is splitted in three modules. The first module is dedicated to the analysis and research of IMEC's sleep detection algorithm. It starts with an overview of the available dataset, followed by a description of IMEC's sleep detection algorithm and all methods researched aiming to make the algorithm independent from the sleep diaries and to validate its performance. The second module, which is focused on the exploration of the possible relationship between sleep efficiency and GSR, provides a description of all the data exploration process and the modelling implemented in Python and R. The last module of this chapter provides a detailed description of the methods used for the analysis aiming to explore the possible relationships between GSR and skin temperature.

Chapter 4 describes the results of the analysis made for the three aforementioned modules separately. It is followed by chapter 5 which reports a detailed description and interpretation of the results obtained in chapter 4, combined with eventual limitations for each analysis made.

Finally, chapter 6 brings a summary of the acquired results, conclusion obtained and limitations of this master thesis work . Additionally, suggestions are proposed to further improve and research the analysis performed in this work.

CHAPTER 2

LITERATURE REVIEW

2.1 Importance of sleep and effects of sleep deprivation

Sleep is a daily recurring state which is essential for a person's health and wellbeing. About a third of an individual's life is spent sleeping in which the body recovers, the muscles relax, and memory is stored [1]. Sleep loss, which refers to a sleep duration of less than the average need of 7 to 8 hours per night, affects an individual's performance, safety, and quality of life [1]. Other than daytime sleepiness, sleep loss comes with several other consequences. Different studies [2],[3] have shown that sleep loss can result in various situations: from reduced performance in simple measures of cognition like attention and reaction time to deficits in judgment and complex decision making. In fact, 20% of serious car accidents worldwide are associated with driver sleepiness, independent of alcohol effects [1]. Sleep loss also affects mood, the emotional and psychosocial interpretation of events, and aggravates stress levels [4],[5]. Sleep deprivation can even lead to health problems such as hypertension, obesity, depression, cardiovascular diseases, and insulin resistance [1],[6]. The causes of sleep loss can be attributed to lifestyle (e.g., prolonged working hours, irregular sleep schedules, shift work, jet lag), environmental factors (e.g., light exposure, load noise, temperature, humidity), psychosocial issues (e.g., social isolation, loneliness) and sleep disorders [7],[8],[49]. In the third edition of the International Classification of Sleep Disorders, there were defined seven categories of sleep disorders: insomnia, sleep-related breathing disorders, central disorders of hypersomnolence, circadian rhythm sleep-wake disorders, parasomnias, sleep-related movement disorders, and other sleep disorders [9]. Up to 70 million people in the US and around 45 million people in Europe suffer from a chronic sleep disorder (i.e., sleep deprivation lasting for at least 3 nights a week for a month or longer) [7]. Other than a negative health impact, sleep loss and sleep-related disorders have also an economic cost associated with doctor visits, hospital services, prescriptions, and over-the-counter medications [7].

Nowadays, due to all the functional and economic implications of sleep deprivation to all sleep is being continuosly researched and monitored. The next paragraph contains a review of the state of the art of current sleep monitoring techniques together with a brief overview of the advantages and disadvantages of each of them.

2.2 Sleep monitoring techniques

2.2.1 Sleep architecture

Different sleep monitoring techniques provide different information about sleep like for instance sleep stages . Therefore, a brief overview of sleep architecture is needed. Sleep 'architecture' consist of four main stages: N1, N2, N3 and REM (rapid-eye movement) [12],[13]. The first 3 stages (N1, N2, N3) are characterized by non-rapid eye movement (NREM) and progressively deeper sleep [12]. The first stage of human sleep is generally stage 1 (N1), which is characterized by light sleep, duration of few minutes (1-5 min) [13], slow rolling eye movement, and muscles contractions. Stage 2 (N2) which is characterized by a deeper state of sleep compared to stage 1, usually precedes deep sleep (N3) which is also referred to as slow wave sleep (SWS) [13]. The features of deep sleep (N3) are low muscle tone and oscillations of less than 4Hz in the electroencephalography (EEG) [13]. The fourth stage is characterized by rapid eye movements (REM) and low muscle tone [12],[13]. There is a cyclic alternation between NREM and REM sleep during the night, with a cycle duration of approximately 90 min [13].

2.2.2 Polysomnography

The gold standard for sleep studies is polysomnography (PSG), which is a technique able to assess wakefulness from sleep, sleep stages, sleep quality parameters, and different sleep disorders (e.g., apneas, hypopneas, periodic leg movements) [10],[11]. To do so, PSG records simultaneously multiple physiological signals such as electroencephalogram (EEG), electrocardiogram (ECG), electro-myogram (EMG), electrooculogram (EOG), respiration, oxygen saturation (SpO2) and body movements [14],[15] (see figure 1.1).

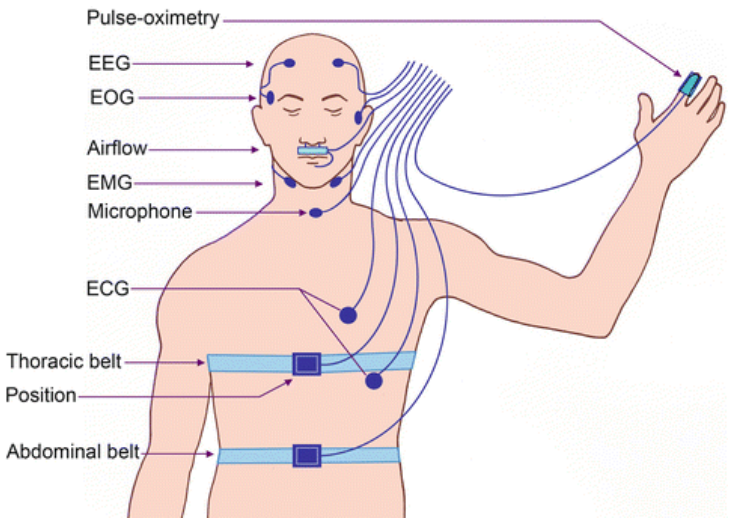

Figure 2. 1 An overview of the traditionally recorded signals during a PSG procedure [Source: Razjouyan et al. [10]]

A human sleep expert, by visually considering EEG, EOG, and EMG, assigns to each 30s time segment a sleep stage according to the rules of the American Academy of Sleep Medicine (AASM) [18] or Rechtschaffen and Kales (RK) [19]. According to the AASM rules [18], 5 stages are identified: Wake, N1, N2, N3, and REM [20]. If scoring instead is performed according to Rechtschaffen and Kales rules [18], NREM sleep is subdivided into stages 1–4 where stages 3–4 correspond to N3 [13].

Other recorded signals such as respiration and oxygen saturation are useful for detecting sleep disorders like for instance sleep-related breathing disorders [17].

As forementioned, PSG can also provide sleep quality measures such as total sleep time (TST), sleep efficiency (SE), and wake after sleep onset (WASO) [11]. The total sleep time (TST) represents the total amount of sleep time during which the patient is in bed with recording equipment activated [16]. In other words, TST represents the sum of N1, N2, N3, and REM sleep and is expressed in minutes [16]. Another sleep parameter provided by PSG is sleep efficiency, which refers to the percentage of total time in bed actually spent asleep. It is computed as in equation 1.1. Sleep efficiency indicates how well the patient slept, but it does not distinguish frequent, brief episodes of wakefulness [16].

$$Sleep\ Efficiency\ = \frac{Total\ Sleep\ Time\ (min)}{Total\ Recording\ Time\ (min)} \qquad (1.1)$$

Sleep latency represents the duration of time (in minutes) between the start of the recording and the first epoch scored as sleep [16]. Finally, wake after sleep onset (WASO) represents the periods of wakefulness occurring after sleep onset (i.e., first 30s epoch scored as sleep) [16].

Although PSG is very useful in sleep studies because it allows a very accurate assessment of sleep architecture, quality, and disorders, unfortunately, it comes with some important drawbacks. An important limitation of PSG is its high costs, due to highly trained personnel and technology, which makes it impractical for long-term sleep monitoring and unfeasible for large-scale population research [6],[21]. PSG is a sophisticated technology that requires that the subject spends the night in a sleep laboratory and makes use of a high number of attached sensors that affect the normal sleep of the subject under study [21]. All these considerations suggest considering alternative monitoring techniques when it comes to sleep studies.

2.2.3. Actigraphy

In the early 1990s, sleep researchers developed a technique called actigraphy to study sleep using wearable devices [6]. Actigraphy is a validated method of objectively measuring sleep parameters using a non-invasive accelerometer [70][71]. The

accelerometer is housed in a small device generally placed on the wrist, ankle, or chest able to record movement for extended periods (days to several weeks) [22],[23] (see figure).

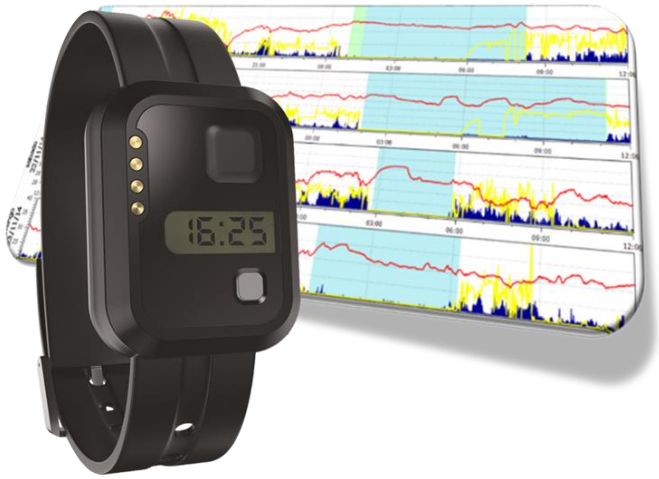

Figure 2. 2: Example of a wrist-watch actigraphy sleep monitoring device and the data it acquires [Source: [72]]

These devices provide a binary classification of wake or sleep (e.g., in 1 min epochs) and rely on the basic premise that the presence of movement indicates wakefulness while the absence of movement indicates sleep [22]. In early studies done by the American Sleep Disorders Association (ASDA), it was found that actigraphy is a valid and reliable sleep assessment method in specific domains of sleep research and sleep medicine [23]. For example, Sadeh et al. [25] concluded that actigraphy could be a useful tool for diagnosing insomnia, circadian rhythm disorders, or excessive sleepiness but not for routine diagnosis, assessment of severity, and management of sleep disorders. In another validation study including healthy participants Sadeh et al. [26] obtained an agreement with PSG greater than 90%. In a more recent study [23], which considered publications regarding the validation and reliability of actigraphy from 2002 to 2010, it was concluded that compared to PSG, actigraphy has reasonable results in assessing sleep-wake patterns in normal individuals with average or good sleep quality. In the same study [23], it was deduced that these results were more questionable in special populations (e.g., elderly people, individuals with major health problems, or individuals with poor sleep quality). The main problem related to the validity of actigraphic sleep/wake classifications is the capability to detect wakefulness during sleep periods for certain devices, algorithms, and populations [23]. On the other hand, actigraphy comes with a lot of benefits compared to PSG. It is non-obtrusive, cost-effective, and easy to set up for long-term monitoring [6],[10]. These advantages make it useful for large population studies where PSG is not feasible [6],[11]. In sleep studies, these devices are commonly worn throughout the

whole day and the participants are usually asked to fill in a sleep diary (e.g., time in bed, sleep onset, and waking up time) [24]. Just like PSG, actigraphy-based devices are also able to provide sleep quality measures such as total sleep time (TST), sleep efficiency (SE), and wake after sleep onset (WASO) [11]. In actigraphy studies the total sleep time represents the total amount of time that the subject is asleep minus the amount of time the person awakens [14]. This includes the time from sleep onset to sleep offset [16]. Sleep onset time is defined as the first minute of 15 continuous minutes of sleep after a self-reported bedtime while the sleep awakening time is the last minute of 15 continuous minutes of sleep that is followed by 30 minutes of movement [14]. Sleep efficiency (see equation 1.2) is computed as the ratio of the total sleep time and the total minutes in bed where total minutes in bed represents the amount of time that the subject spends asleep as well as the amount of time the subject takes to fall asleep, that is, latency. Latency is defined as the period between the self-reported time to bed and the sleep onset [14]. Total sleep time instead is computed by subtracting the wake after sleep onset (WASO) from the duration of the sleep period. WASO is the sum of all moments of wakefulness lasting longer than 5 minutes [14].

$$Sleep\ Efficiency\ = \frac{TotalSleepTime}{TotalMinutesinBed} = \frac{||Sleep\ period||-WASO}{||Sleep\ period||+Latency} \qquad (1.2)$$

2.2.4 Sleep detection algorithms based on actigraphy

Nowadays, actigraphy is being used increasingly in clinical settings since, as depicted before, it has the advantage of providing objective information on sleep habits in the patient's natural sleep environment [73]. There are various validated algorithms for estimating sleep/wakefulness based on movement data collected using actigraphy [74].
The following paragraph provides a brief overview of some of the most famous sleep detection algorithms that make use of actigraphy, starting from the first developed algorithms to the latest trends on this topic.

Webster et al. [46] proposed a method to detect sleep/wake in 1-minute intervals exploiting wrist actigraphy. The actigraphs used by them were set to run in zero-crossing with a data storage epoch of 2 seconds. After several experiments, they obtained equation 1.3 in which T(i) represents the sum of the 2s activity values of the recorder in 1 min at time i. The subject is classified as awake per 1 min interval if D>=1.

$$D = 0.025[0.15 * T(i-4) + 0.15 * T(i-3) + 0.15 * T(i-2) + 0.08 * T(i-1) + 0.21 * T(i) + 0.12 * T(i+1) + 0.13 * T(i+2)] \qquad (1.3)$$

Cole et al. [47] aimed to replicate and expand the work of Webster in order to use the algorithm in normal and sleep disordered subjects. They used actigraphy data and did a

detection of sleep/awake in 1 min intervals, compared their results with the gold standard, and developed a computer program that would find the optimal parameters for the following model (see equation 1.4)

$$D = P[W_{-4} * A_{-4} + W_{-3} * A_{-3} + W_{-2} * A_{-2} + W_{-1} * A_{-1} + W_0 * A_0 + W_{+1} * A_{+1} + W_{+2} * A_{+2}] \quad (1.4)$$

P is a scale factor for the entire equation, W_0, W_{-1} W_{+1} and A_0, A_{-1} A_{+1} represent, respectively, the weights and activity scores for the present minute, for the previous minute, and for the following minute. The optimal parameters which are shown in equation 1.5 obtained an accuracy of 87.05%.

$$D = 0.00001[404 * A_{-4} + 598 * A_{-3} + 326 * A_{-2} + 441 * A_{-1} + 1408 * A_0 + 508 * A_{+1} + 350 * A_{+2}] \quad (1.5)$$

To deal with some misclassifications of wake instead of sleep, some corrections were developed by Webster et al [46] and implemented: 1) if at least 4 min are scored as wake, the next 1 min scored as sleep is also scored as wake, 2) if at least 10 min are scored as wake, the next 3 min scored as sleep are rescored to wake, 3) if at least 15 min are scored as wake, the next 4 min scored as sleep are rescored to wake, 4) 6 min or less scored as sleep surrounded by at least 10 min before and after as wake, are rescored to wake, and 5) 10 min or less scored as sleep surrounded by at least 20 min before and after as wake, are rescored as wake. The best results are obtained by combining rule 1) and 5) or rule 1) with rule 4).

Van Hees et al. [48] proposed a novel method that instead of using the magnitude of the accelerometer as input like previous studies did, uses the derived arm angle (see equation 1.6) to identify sleep. Basically, sleep is considered a period with low frequency changes of the arm angle.

$$angle = \left(tan^{-1} \frac{a_z}{\sqrt{a_x^2 + a_y^2}} \right) * 180/\pi \quad (1.6)$$

In equation 1.6 a_x, a_y and a_z are the median values of the three orthogonally positioned raw acceleration axes in g-units calculated on a rolling time window of five seconds. The estimated angles were averaged considering 5 second windows followed by making the difference between these 5 second values. When the change between successive epochs of 5 seconds is less than 5 degrees for at least 5 min, a period of inactivity or potential sleep period is considered. They used the polysomnography data from 28 persons to validate their algorithm and reached an accuracy of 83%.

In the recent years, new sleep detection algorithms have been developed that employ machine learning and artificial neural networks [22][57]. These methods are being

investigated since they can be accurate, robust and have low cost [22]. Sundararajan et al. [57] proposed a sleep classification which uses a random forest model. Their model was able to detect sleep in unseen data reaching an F1 score of 73.93% compared with PSG.

2.2.5 Wearable technology used for sleep monitoring

Other than PSG and actigraphy which are considered objective measurement of sleep, there are even other wearable sleep trackers that include smartphones, in-bed sensors, and contactless sensors [75]. These devices use similar technology to actigraphy devices and are used for physical activity and sleep tracking purposes [6]. They are relatively low cost and are available without prescription or clinical recommendations [75]. Other benefits of these devices are that data can be collected at any time in an easy way (i.e., users simply wear the device) and without the need of specialized technicians processing the data [75]. Nowadays, the number of these commercially wearable devices has grown rapidly since their first introduction over a decade ago [43]. However, these devices present drawbacks related to their validity, accuracy and reliability in measuring the various sleep parameters. One popular commercial device in this context is Fitbit (Fitbit Inc, San Francisco, CA, USA) [43] for which there have been different studies aiming to assess its validity in sleep studies. In particular, Haghayegh et al [45] conducted a systematic review of publications reporting the performance of wristband Fitbit models in assessing sleep parameters and sleep stages. From this study, it was deduced that in a comparison with polysomnography, Fitbit overestimates total sleep time (range of approximately 7 to 67 mins) and sleep efficiency (range of approximately 2% to 15%), and underestimates wake after sleep onset (range of approximately 6 to 44 mins). It was concluded that Fitbit models show promising results when it comes to differentiating wake from sleep. They can give a gross estimate of sleep parameters, but they show limited specificity and are not a substitute for PSG [45].
Even though sleep wearables may have a key role to better characterize and understand sleep, within the framework of precision medicine [75], further work is needed to investigate the potential use and performance, pros and cons, and limitations of these novel sleep trackers, particularly in sleep disorder populations [75]

2.3 Sleep-related physiological signals

There are different physiological signals useful in sleep studies including EEG, ECG, respiration signal, EDA, skin temperature and acceleration signal. An overview of most of these signals is presented in the following paragraph, focusing on the signals which will be employed in this master thesis project.

2.3.1 Electrocardiogram (ECG)

An electrocardiogram (ECG) represents a non-invasive recording of the electrical activity of the heart from the surface of the body [27]. A normal ECG waveform includes several complexes known as P, Q, R, S, and T waves which represent various stages of the heart cycle [28] (see figure 2.3). From the ECG it is possible to extract the heart rate variability (HRV) which represents the fluctuations of the time intervals between two consecutive heartbeats (R-R) (i.e., the time the heart needs to complete one cycle) [28]. HRV is a useful signal to study since it is related to the outflow of the Autonomic Nervous System (ANS) [28],[29].

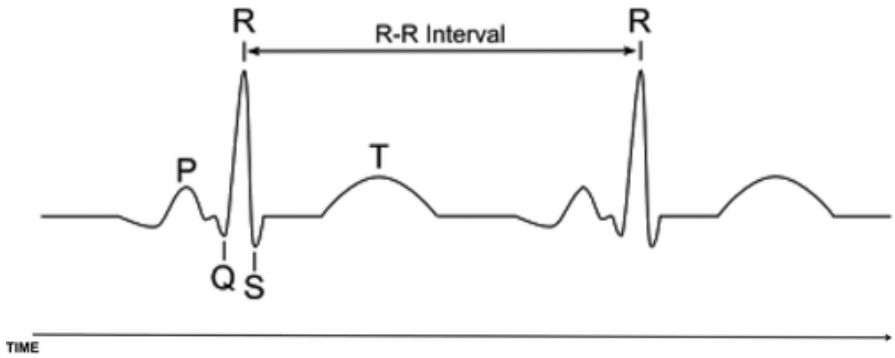

Figure 2. 3 Standard ECG waveform and RR (inter-beat) interval [Source: Sattar et al [27]]

In the physiology of sleep, an important role is played by the autonomic nervous system (ANS), which in this context, regulates cardiovascular functions during sleep onset and transitions to different sleep stages [30]. Cardiac activity is modulated by the combined effects of both branches of the ANS, i.e., sympathetic and parasympathetic autonomic nervous systems [31].

Different studies [32], [33] have concluded that HRV changes during sleep, in particular, significant changes can be seen between REM and non-REM stages. It was revealed that a higher parasympathetic tone occurs during normal non-REM (i.e., healthy subjects) and a shift towards sympathetic predominance during normal REM sleep happens.

The tools used for the analysis of the HRV include classical linear approaches in time and frequency domain, but also non-linear approaches like measures of complexity, entropy derived indices, etc. [30].

In time-domain analysis, other than mean and standard deviation, other parameters like the standard deviation of normal-to-normal intervals (SDNN) can be computed, which represents the overall variability of HRV [34]. There are other measures computed from the differences between consecutive heartbeats like the root mean square successive difference (rMSSD), number of interval differences of successive heartbeats greater than

50 ms (NN50), and proportion of NN50 (pNN50, NN50 divided by total number of heartbeats) [30] which describe the parasympathetic branch [30].
For what concerns frequency domain, the Fourier Transform can be computed, which decomposes the ECG signal into its contained frequencies to build a spectral power spectrum for each frequency [31]. The heart rate variability time series is characterized by three main frequency components: a very low-frequency component (VLF) (<0.04 Hz), a low-frequency component (LF) (0.04–0.15 Hz), and a high-frequency component (HF) (0.15 - 0.4 Hz) [30]. The VLF component is a marker of humoral and hormonal fluctuations; the LF component is associated with both sympathetic and parasympathetic modulation while the HF component, which corresponds to the respiratory peak rate (0.18–0.4 Hz), represents short-term HR variation and is modulated by parasympathetic activity only [30]. Another useful index in the frequency domain is the LF/HF ratio which reflects the sympatho-vagal balance [35] (which reflects the autonomic state resulting from the sympathetic and parasympathetic influences).
HRV is also characterized by non-linear dynamics that characterize autonomic cardiovascular control and which cannot be captured by the traditional linear approaches like power spectral analysis [30]. The complexity of HRV can be measured by different non-linear approaches, but however entropy-derived methods (e.g., approximate entropy, sample entropy, corrected conditional entropy, and Shannon entropy) have been recently applied for the assessment of autonomic cardiovascular complexity during sleep [30].

2.3.2 Electrodermal activity (EDA)

Electrodermal activity (EDA) also known as galvanic skin response (GSR) refers to autonomic changes in the electrical properties of the skin and is usually expressed in conductance units (micro siemens (μS) or micromho (μohm)) [36]. EDA is the result of the activation of the sweat glands, which are distributed on the surface of the body with a higher concentration on the hands and feet, and which are controlled by the sympathetic branch of the autonomic nervous system [37].
The activation of the sweat glands causes hydration which results in a decrease in the resistance of the skin. The primary function of the eccrine sweat glands is thermoregulation (i.,e to enable evaporative heat)[65]. However, Kuno et al. [67] suggest that besides thermoregulatory sweating, there is also the ‘‘emotional’’ sweating which occurs mostly in the palmar and plantar sweat glands [66].
The stimulation of “emotional” sweating is primarily psychological (i.e., there are no changes in the physical environment) and relates to either external or internal stimuli that

generate stress or, more generally, arousal. Boucsein et al. [36] states that "EDA measures are highly applicable to emotions and stress research". This is because EDA is the only autonomic psychophysiological signal not contaminated by parasympathetic activity [36] and is mediated solely by the sympathetic branch of the autonomic nervous system which is supporting the stress ("fight or flight") response to threatening events and is a central component of emotional experience and, by extension, cognition and behaviour [69]. Therefore, EDA is widely used as a sensitive index of emotional processing and sympathetic activity [36]. Due to this, EDA has been utilized to investigate numerous areas such as orienting and attention, learning and conditioning, psychopathology,

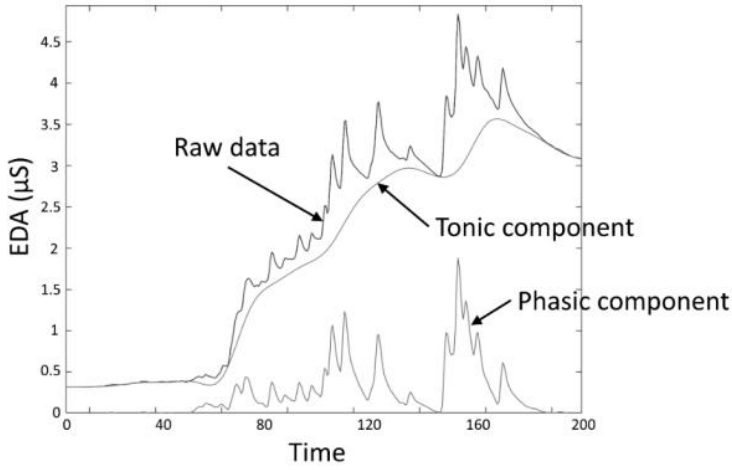

Figure 2. 4: EDA data decomposition into tonic and phasic components [Source: Braithwaite et al [36]]

personality disorders, individual differences, brain asymmetry and laterality, cognitive functions, as well as neuropsychology [36] [67].The recording of the electrodermal activity can be done either by measuring the electrical potential produced or by applying an external current and recording the change in conductivity [38]. The EDA complex, whose frequency band is 0-2Hz, incorporates both tonic and rapid phasic components [36]. The tonic component refers to long-term fluctuations of EDA and is best characterized by changes in skin conductance level (SCL) [39] and is thought to represent general changes in autonomic arousal, like general level of emotional state and level of stress [36]. Since SCL changes slowly, the measurement intervals should be long, i.e. from tens of seconds to minutes [36]. Boucsein et al. [76] suggest that by computing for instance the mean of the tonic component (SCL), an estimate of the general psychophysiological status of the subject can be obtained. According to Boucsein et al. [36] simply averaging the SCL across the signal is not useful because there is going to be an overestimation of SCL due to the presence of SCR (rapid phasic components). He suggests that SCR amplitudes should be subtracted from the SCL in order to have a true estimation of SCL. Another measure of the tonic component is the frequency (1-5 per/min during rest and over 20 per/min during high arousal situations) and amplitude of the nonspecific skin conductance responses (NSSCRs) [36]. Nonspecific skin conductance

represent the number of SCRs which are the result of spontaneous fluctuations in the EDA, non-related to a specific stimulus, even though they have the same characteristics as stimulus-related SCRs and are considered a tonic measure [41]
Phasic measures are related to fast-changing components of the EDA signal in a short time window (orders of seconds) and are reflected by the skin conductance response (SCR) (see Figure 2.4) [41]. SCRs might be related to stimulus-specific responses or non-specific responses [41]. Usually, measures of the elicited amplitude of the SCR are used to evaluate the activity of an event-related stimulus [36]. In these cases, particular attention should be paid because the responses might be related to non-specific events [36].
To separate the phasic component (SCR) from the tonic one, different approaches can be employed, for instance a high pass filter (0.05 Hz) can be applied. After applying the high pass filter SCR can be detected by employing a threshold that could be within a range of 0.01µS - 0.03µS [36].

2.3.3 Skin temperature

Skin temperature represents the temperature of the outermost surface of the body which in normal human skin temperature (on the trunk of the body) varies between 33.5 and 36.9 °C. The skin temperature results to be lower over protruding parts, like the nose, and higher over muscles and active organs [77]. Surface skin temperature in humans varies alongside ambient temperature, internal temperature and conditions affecting both the skin and underlying structures [77]. Skin temperature often plays a significant role in affecting thermoregulatory processes (i.e., the mechanism by which mammals maintain body temperature with tightly controlled self-regulation independent of external temperatures [78]). For instance, when ambient temperature is high, cutaneous blood flow is increased (vasodilation), facilitating the transfer of internal body heat to the skin. The processes of evaporation and convection of sweat cause a loss of body heat from the skin surface to the environment, providing an effective means for lowering body temperature [79].

For what concerns sleep, skin temperature play an important role in sleep regulation [80]. It is known that sleep onset is usually preceded by a decline in core body temperature [81] which is promoted by the enhancement of heat loss which results in an increase in the distal skin temperature related with distal skin vasodilation [81].

2.3.4 Other sleep related physiological signals

Another signal that is useful in sleep studies, especially as a sleep quality indicator is respiration signal, which together with heart rate (HR), are known to vary greatly during

sleep, and have a close relationship with each sleep stage [82]. In particular, respiration signal is useful in detecting sleep disorders such as snoring and sleep apnea [83].

Chapter 3

MATERIALS AND METHODS

The following chapter describes the materials and methods used to tackle the raised research questions, splitting the thesis work in three modules. The first module is dedicated to the analysis , research of possible independence from sleep diaries and validation of IMEC's sleep detection algorithm. It starts with an overview of the available dataset, followed by a description of the sleep detection algorithm provided by IMEC and all methods used to make the algorithm independent from the sleep diaries. Afterward, the project is focused on the exploration of the relationship between sleep efficiency and GSR, describing all the data exploration process and the modelling implemented in Python and R. The last part of M&M will provide a detailed description of the methods used for the analysis aiming to explore the possible relationships between GSR and skin temperature.

3.1 IMEC's sleep detection algorithm analysis and research

3.1.1 Dataset explanation

The available dataset was previously collected and contains data from a wristband and a chest patch. The data from the wristband, also called minifornax data, includes the 3 axis accelerometer data (x,y,z), the GSR data, and the skin temperature data (see figure 3.1). Afterwards, a quality indicator is computed for the GSR, which is considered 1 if GSR data is of good quality and 0 otherwise. The data from the chest patch, also called stingray data, includes the 3-axis ACC data and the ECG data. The total dataset includes 56 Users and for each user there are available several days (5 to 9) days (where for one day is intended a 24
hours interval starting from midnight). The participants under study are asked to provide their sleep diary data which consists of recording sleep parameters such as their time in bed, sleep onset, and waking up time

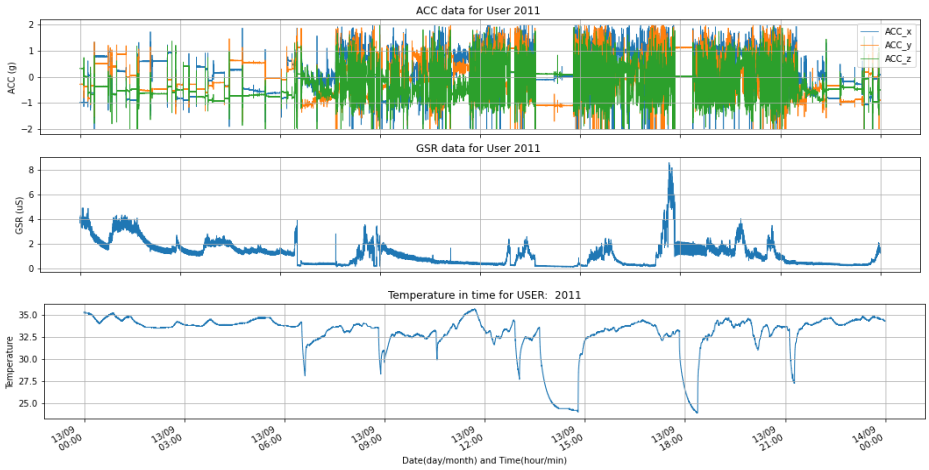

Figure 3. 1: Data of day 4 for user 2011 during the whole day (i.e., 24h). The first graph represents the 3 axis ACC data in (g), the second graph represents the GSR data in (uS), and the third graph represents the skin temperature in (°C).

The original data from the devices which were saved in hdf5 format were extracted and converted into a pandas data frame by means of a self-written function: hdf5_to_pd_dataframe.

3.1.2 IMEC's sleep detection algorithm

IMEC's sleep detection algorithm exploits the 3-axis accelerometer data (minifornax or stingray) and the sleep diaries provided by the subjects to make a binary classification: wake or sleep in 5s time intervals (see figure 3.2). In figure 3.2 it can be seen the output of the sleep detection algorithm for user 2019 for all the available days , in which the blue line is 1 whenever the subject under study is sleeping, zero otherwise. The violet graph represents the accelerometer data. The algorithm provides an output between the time asleep and time awake provided by the subject (green lines in the figure 3.2). The algorithm, which is partially based on the research of Webster et al. [46] and the modifications introduced by Cole et al. [47], relays on the premise that the presence of movement indicates wakefulness while the absence of movement indicates sleep.

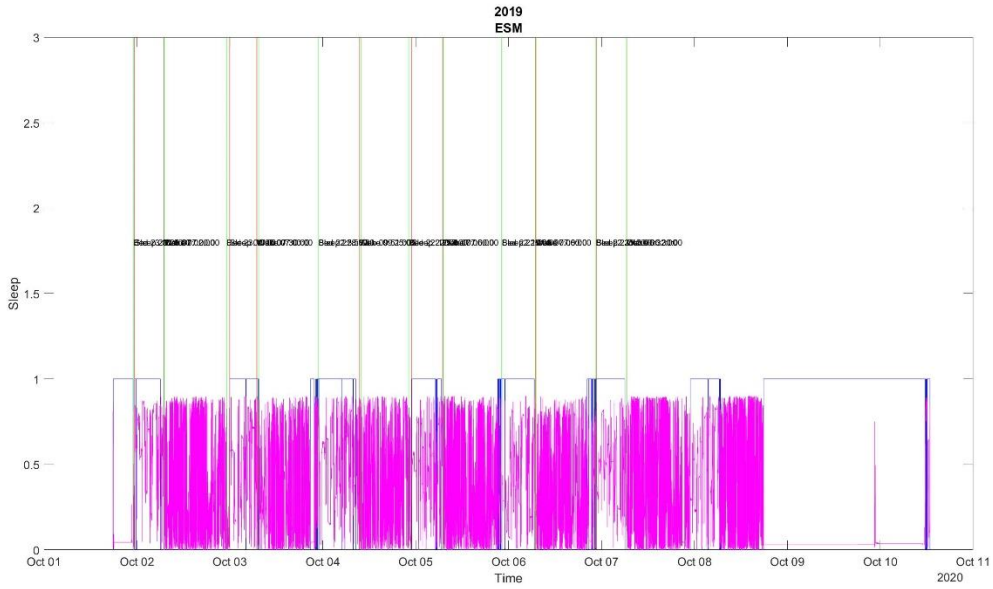

Figure 3. 2: Results of the sleep detection algorithm. The blue line represents the output of the sleep detection algorithm. The violet graph represents the accelerometer data. The green/yellow lines represent the ESM data.

The input of the algorithm is the 3 axis accelerometer data (fs = 32 Hz): x, y, z which are used to compute the ACC magnitude (equation 3.1). Afterward, the ACC magnitude is converted in 5-second intervals by taking the median value.

$$ACC = \sqrt{x^2 + y^2 + z^2} \quad \text{(Equation 3.1)}$$

As depicted before, this sleep wake classification model is based on some empirical formulas developed by Webster et al. [46] combined with rules developed by Cole et al. [47], For more details regarding the formulas, see [46][47].

3.1.3 SPT window algorithm

The sleep period time (SPT) window represents the time window that starts when the subject goes to sleep until waking up. To detect the SPT window an algorithm was proposed by Van Hees et al. [24] whose schematic is reported in figure 3.3.

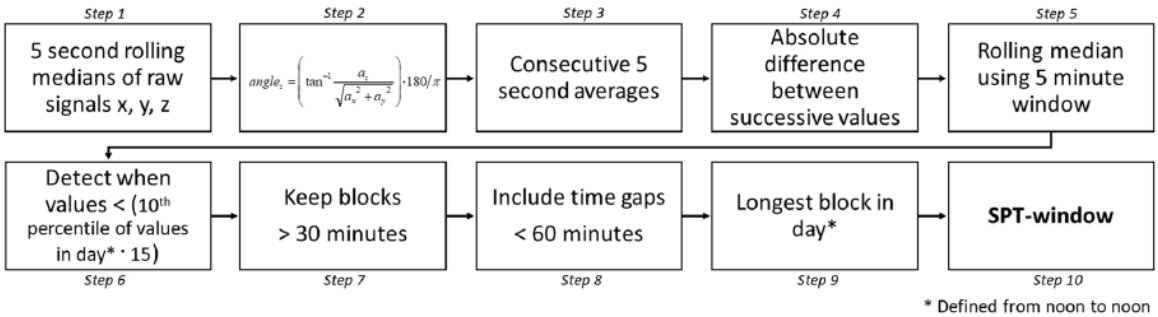

Figure 3. 3: Steps of the SPT- window detection algorithm

The algorithm takes as input the raw accelerometer data coming from axis x,y, and z. As a first step, it computes the 5 second rolling median of the raw accelerometer and afterward computes the angle z as shown in equation (3.3).

$$angle = \left(tan^{-1} \frac{a_z}{\sqrt{a_x^2 + a_y^2}} \right) * 180/\pi \qquad (3.3)$$

Steps 3 to 5, which make the algorithm invariant to the potentially unstandardized orientation of the accelerometer relative to the wrist, consist of calculating the 5-minute rolling median of the absolute differences between successive 5-second averages of the z-angle. Step 6 consists of defining a personal threshold using the physical activity of 24 hours (noon to noon). This threshold is computed by taking the 10^{th} percentile of the output of step 5 and serves to discriminate between sleep from wake. More specifically, if the output of step 5 is lower than the threshold sleep is predicted, otherwise wake is predicted. The next step is to detect continuous 30 min blocks of sleep using this threshold. Afterward, the final SPT window is created by unifying these 30 min sleep blocks if there are separated by a maximally 60 min of wake. The SPT window algorithm implemented in this master thesis work performs only between 8pm to 10am (time interval found empirically) in order to avoid day sleepiness.

3.1.4 Random forest sleep detection algorithm

The gold standard for sleep studies is PSG [10], so ideally the validation of a sleep detection algorithm would consist of a quantitative comparison with PSG. In absence of PSG data for the validation of IMEC's sleep detection, a validation was done by using the sleep detection algorithm developed by Sundararajan et al. [57]. This sleep detection algorithm was trained by implementing a random forest (RF) model on data from 134 adult participants who wore a wristband containing an accelerometer during a one-night polysomnography. The algorithm was able to distinguish sleep-wake states with an F1 score of 73.93% on a test set of 24 participants. Compared to the Van Hees et al. [48] algorithm (that use zero-crossing techniques), this algorithm reports a more accurate sleep prediction. However it is less interpretable.

3.1.5 Validation of the IMEC's sleep detection algorithm

The output of the RF algorithm, developed by Sundararajan et al [57], was compared quantitatively with the results of IMEC's sleep detection algorithm. The prediction window of the RF algorithm was modified to match the 5 seconds of IMEC's sleep detection algorithm. The two algorithms were compared within the SPT using the accuracy and Cohen's kappa with the results of IMEC's sleep detection algorithm as

ground truth. A second comparison was done after applying median smoothing in 30s window on the results of the random forest sleep detection algorithm.

3.2 Exploration of the relationship between GSR signal and sleep efficiency

The following sections will present the whole process towards model development which aims to explore the possible relationship between the GSR signal and the sleep efficiency. It starts with a brief introduction of feature extraction from the dataset, necessary for the modelling process. It continues with the function that deals with missing data, dataset preparation, the performed data exploration and finally it presents the actual models implemented, which consist of machine learning and mixed models.

3.2.1 Features extracting for model development

To proceed with model development, all the necessary features were extracted from the dataset by the thesis supervisor. The feature dataset consists of several features, calculated in rolling 5 minute windows with a step size of 1 min. The features extracted from the GSR signal were: general features including maximum of GSR (GSR_max) , minimum of GSR (GSR_min), mean of GSR(GSR_mean),median of GSR(GSR_median), standard deviation of GSR (GSR_std), squared difference of GSR (gsr_scdiff2) and features related to the SCR (i.e., skin conductance responses) including number of responses (gsr_scr), area (gsr_scr_area), duration (gsr_scr_dur), magnitude (gsr_scr_mag), response rate (gsr_scr_rr), standard deviation of GSR SCR (gsr_std) and the phasic component of GSR (gsr_scph).
. The mean and std deviation of the GSR quality indicator was also extracted. For skin temperature, the following features were computed: maximum of temperature (temp_max) , minimun of temperature (temp _min), mean of temperature (temp _mean), median of temperature (temp _median), standard deviation of temperature (temp_std), and the slope (temp_slope). The features were calculated without considering the presence of missing data, thus missing data was removed after feature calculation which will be discussed in the next paragraph.

3.2.2 Missing data function

The presence of missing data is related to time periods in which the device is on, but it is not worn by the subject under study.
In the available dataset, the missing data is present in two forms: data missing for the whole day/night (i.e., data missing for more than 16 hours) (see figure 3.4) or data missing for a few hours (0 to 4 hours) during the day (see figure 3.5)

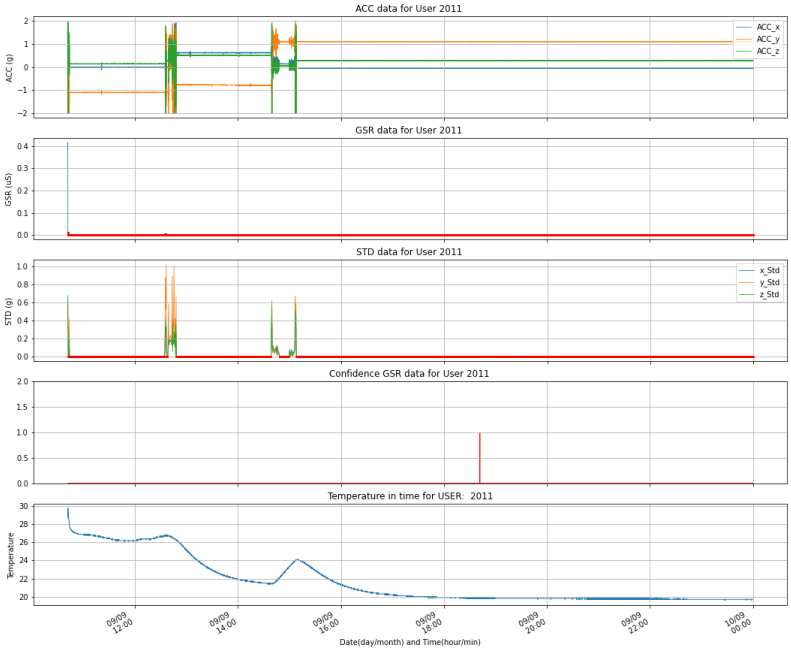

Figure 3. 4: Missing data (ACC,GSR, STD of ACC, GSR quality indicator and skin temperature)for more than 16h

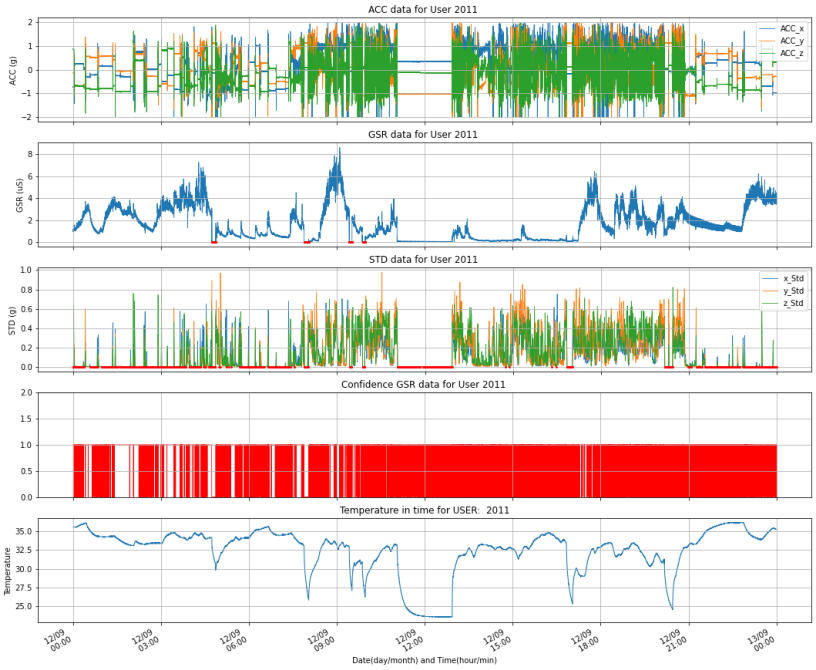

Figure 3. 5*: Missing data* (ACC,GSR, STD of ACC, GSR quality indicator and skin temperature)for a few hours during the day

A self-written function was developed in Python to detect and deal with the missing data in the dataset. The first part of the function had to deal with missing data for the whole day (i.e., data missing for more than 16 hours). To do so, the GSR quality indicator and skin temperature were employed. The function considers the data missing for that day, if both the sum of the quality indicator during the whole day was less than 1500000 (GSR quality indicator is made by 0 and 1 and this sum corresponds to data missing for more than 16h) and if the mean skin temperature was less than 28 degrees. The second part of the function dealt with missing data for some hours (0 to 4) but not for the whole day (see figure 3.4). For this purpose, the standard deviation (STD) of ACC data (x, y, and z axis), the quality indicator of the GSR data, and skin temperature were employed. As a first step, the STD of the ACC data was taken in 30 seconds and it was checked how many axes (of x, y, and z) had a STD lower than 0.08. If in a 10 min window, at least 2 of the 3-axis ACC data had a STD of ACC less than 0.08 for the whole, the window was considered missing, if one of the following conditions was satisfied: the mean skin temperature in the window was less than 30 degrees or the STD of the skin temperature in the window was less than 0.6. Once the blocks with the missing data were detected, the features corresponding to these blocks were eliminated. The thresholds described within this function were empirically optimized.

3.2.3 Dataset preparation

Once missing data was removed from the feature dataset, the dataset was prepared for analysis, by dividing it into data during sleep and data during wake using the sleep detection algorithm outcomes. Afterwards, the mean value of each feature for the whole night (or day) was computed in order to couple it to the sleep efficiency of that night.
Since the range of the sleep efficiency is 0 to 1, one could research the relationship between the features and the target using regression methods. However, the problem was converted into a classification problem because it is hypothesized to give better results [62] using a threshold of 0.85 [6]. So, if the calculated sleep efficiency of a night was lower than 0.85, the night was considered as bad sleep quality corresponding to class 0. Otherwise, the night was considered as good sleep quality and got class 1. By applying the previous actions, a dataset consisting out 72 observations remained.
To remove collinearity between features, features with a correlation higher than 0.9 were removed in further analysis where the feature that had the highest correlation to the target (i.e., sleep efficiency) was maintained. Afterwards the dataset was normalized using the standard scaler which standardizes the features by removing the mean and scaling to unit variance.
Since logistic regression requires the data to be normal, a correction for normality was performed. Different transformations were used to normalize the data including

log(1+data), square root and Yeo-Johnson. Yeo-Johnson transformation allows for zero and negative values of y. λ can be any real number, where $\lambda = 1$ produces the identity transformation (see equation 3.4).

$$\psi(\lambda, y) = \begin{cases} \big((y+1)^{\lambda} - 1\big)/\lambda & \text{if } \lambda \neq 0, y \geq 0 \\ \log(y+1) & \text{if } \lambda = 0, y \geq 0 \\ -\big[(-y+1)^{2-\lambda} - 1\big]/(2-\lambda) & \text{if } \lambda \neq 2, y < 0 \\ -\log(-y+1) & \text{if } \lambda = 2, y < 0 \end{cases} \quad (3.4)$$

3.2.4 Data exploration

Once the dataset was prepared, data exploration was performed to investigate eventual relationships between physiological signals, more specifically the GSR signal, and the calculated sleep efficiency. First, visual data exploration was performed by making different graphs and evaluating the relationships between the features and the target. Also, the Pearson correlations between the features and the sleep efficiency were calculated. Some attention was given to cleaning further the dataset (of bad quality GSR data) and also to select the right features of the GSR among all the features considered until now. To do so, the considered features of the GSR signal were plotted for each user (see figure 3.6) and they were observed in order to detect features that behave in a similar way (i.e that are correlated) or bad quality data.

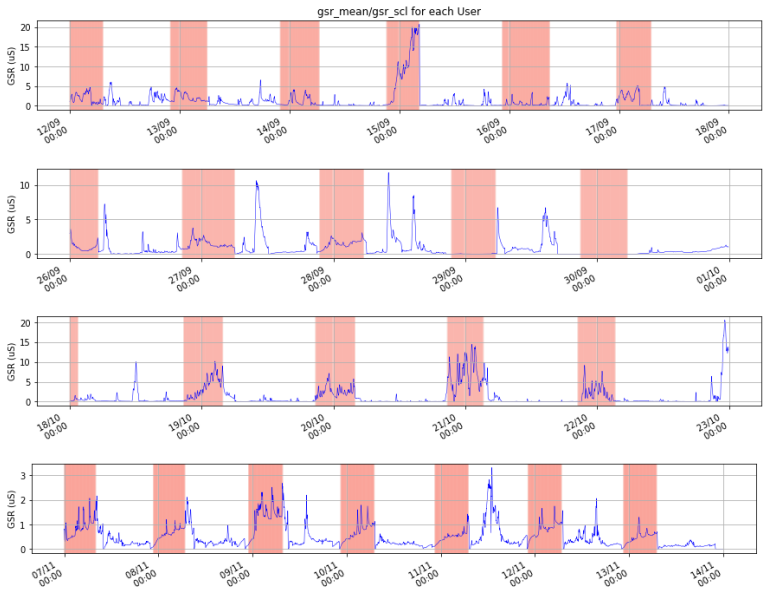

Figure 3. 6: GSR mean data plotted for all available days for user 2011,2017,2021 and 2023

3.2.5 Machine learning models used in the analysis

To explore the relationship between the physiological features and sleep efficiency, different machine learning models were implemented including naive bayes, logistic regression, KNN, decision tree, random forest and multilayer perceptron (MLP). Also, principal component analysis (PCA) was applied. A brief description of the functioning of each of these models is described in the following paragraph.

Naive Bayes classifier

This machine learning algorithm is based on Bayes theorem which describes how the probability of an event is evaluated based on prior knowledge of conditions that might be related to the event (see Equation 3.5)

$$\mathrm{P}(Y \mid X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (3.5)$$

In Equation 3.5, P(Y|X) is the probability of an event Y, given that event X has already occurred, P(X) is the probability of event X, P(Y) is the probability of event Y and P(X|Y) is the likelihood of event X given a fixed value of Y [50].
For a classification problem, the Bayes Classifier will predict class 1 if P(Y=1|X=x0) is higher than 0.5, and class two otherwise. This model is simple and fast, but it assumes independence among feature variables which may not always be the case [50].

Logistic regression

Logistic regression is similar to Bayes' classifier as it also predicts the probability that Y is associated with an input variable X. It uses the logistic function and fits the parameters B0 and B1 using the maximum likelihood technique (see Equation 3.6). After evaluating the two parameters (B0 and B1), the logistic function can be used to predict the target variable probability p(xi) for a given input xi.

$$P(X) = \frac{e^{\beta_0+\beta_1 X}}{1+e^{\beta_0+\beta_1 X}} \quad (3.6)$$

Decision tree

This classification algorithm consists of dividing the dataset into segments based on certain feature variables. The threshold values for these divisions are usually the mean or mode of the respective feature variables (if they are numerical). The criteria that are used to split the data are based on two measures that test the purity of the split (a segment of

the dataset is pure if it has data points of only one class). The first measure is the Gini index that measures total variance across the N classes and the other measure is cross-entropy. The dataset is spitted into segments by minimizing one of the previous measures.

Random forests

Random forests are a learning method which can be used for classification and regression problems. It works by constructing a multitude of decision trees. For classification tasks, the output of the random forest is the class selected by most of the trees while for regression tasks, the mean prediction of the individual trees is returned. Random forests aim to prevent overfitting to the training set which is often a disadvantage of decision trees [63].

K-nearest neighbour

This algorithm works by using a proximity measure to make predictions or classifications about the grouping of an individual data point. When used in a classification problem, the observation is classified by a plurality vote of its neighbours. For instance, if k = 1, then the observation is simply assigned to the class of that single nearest neighbour. In k-NN regression, value is the average of the values of k nearest neighbours.

Multilayer perceptron (MLP)

Multilayer perceptron is a class of feedforward artificial neural network (ANN). They refer to networks composed of multiple layers of perceptrons (with threshold activation). An MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Each node of MLP is a neuron that uses a nonlinear activation function, except the input node. To perform training, MLP utilizes a supervised learning technique called backpropagation. The presence of multiple layers and non-linear activation distinguish MLP from a linear perceptron. Their main characteristic is that they can distinguish data that is not linearly separable [63].

Principal Component Analysis (PCA)

The last step of this analysis in Python by means of machine learning was applying PCA to all the GSR features (GSR signal features + GSR quality indicator features).
Principal component analysis is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning. It consists of an orthogonal transformation that converts a set of possibly correlated features into a new set of linearly uncorrelated features which are called principal components. Dimensionality reduction can be

obtained by projecting each data point onto the first few principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible [59]. Practically, PCA is used to compress information to store and transmit data more efficiently.

All these machine learning models were implemented in Python before and after applying the PCA. For every model a grid search was performed to optimize the parameters to generate the best model in terms of accuracy. Also, 5-fold cross- validation was used to avoid overfitting. Accuracy, f1 score, and recall were used to evaluate and compare the results of different models for both the train and test set. Also, the receiver operating curve (ROC), area under the ROC curve (AUC) and confusion matrix were considered.

3.2.6 Linear mixed models

Linear mixed models were implemented in R in order to explore the eventual relationship between GSR signal and sleep efficiency.
Linear mixed models are an extension of simple linear models that contain both fixed and random effects and are particularly used when there is presence of correlated data in statistical analyses. This may be due to grouping of subjects, e.g., students within classrooms, or to repeated measurements on each subject over time or space, or to multiple related outcome measures at one point in time [50].
Fundamentally, mixed models incorporate both fixed and random effects. A Fixed effect models assume that the explanatory variable has a fixed or constant relationship with the response variable across all observations. While on the other hand, a random-effects model assumes that explanatory variables have fixed relationships with the response variable across all observations, but that these fixed effects may vary from one observation to another [51].

The matrix notation of a mixed model is reported in Equation (3.7)

$$y = X\beta + Zu + \varepsilon \qquad (3.7)$$

In this equation y is a known vector of observations, B is the unknown vector of fixed effects, u is the unknown vector of random effects and e is the unknown vector of random errors.
Like any other model, linear mixed model makes assumptions which have to be satisfied in order for the linear mixed model to be meaningful. These assumptions consist in absence of collinearity, absence of influential data points, homoscedasticity (or absence of heteroskedasticity) and normality of residuals.

3.3 Exploration of the relationship between GSR and skin temperature

3.3.1 Data exploration

To tackle this research question some basic data exploration was performed in R. First, in order to see if there is effectively a correlation between temperature and GSR the Pearson's correlation is computed for every day and night for all the 69 observations.
The Pearson correlation is a measure of the strength of the linear relationship between two variables. It can range between -1 to 1, with a value of -1 meaning a total negative linear correlation, 0 being no correlation, and + 1 meaning a total positive correlation [53]. Continuing with data exploration, the boxplots of the temperature and GSR were computed for both day and night for all the subjects.

3.3.2 Linear mixed models

As a first step, before jumping into the mixed modelling process, it was inquired whether including a random effect in the model is effectively justified. Random effects, as stated before, are added to solve the problem of non-independence for each subject. They characterize idiosyncratic variations due to individual differences. To be able to see if they are effectively needed, the comparison between the Akaike Information Criterion (AIC) of the base-line model without random intercepts and of the model with random intercepts was performed.
Afterwards, the linear mixed models were implemented in R for both day and night. To be able to compare different models or to see if a fixed effect is actually relevant for the model, Anova test are performed, since p-value for mixed models is not as straightforward as linear models. Anova tests are basically likelihood ratio test, whose logic is to compare the likelihood of two models with each other. To see if a fixed effect is actually relevant for that model the baseline (null) model (without the fixed effect/predictors we are interested in) and the full model with the fixed effect are compared by means of anova test. It is concluded that the fixed effect is significant for the model if the difference between the likelihood of these two models is significant.
To be able to interpret the results of the models other indexes like the marginal and conditional R-squared (R2) are used.
The marginal R-squared indicates how much of the model's variance is explained by the fixed effects part only, while the conditional R-squared takes both the fixed and random effects into account and indicates how much of the model's variance is explained by the "complete" model.

Chapter 4

RESULTS

4.1 IMEC's sleep detection algorithm analysis and research

4.1.1 IMEC's sleep detection algorithm results

The results of IMEC's sleep detection algorithm for 10 days are reported in Table 4.1. IMEC's sleep detection algorithm gives in output this table in which it reports the results of the algorithm and also the ESM data filled by the participants. It also provides some sleep parameters such as sleep efficiency, WASO etc.
The results of IMEC's sleep detection algorithm will be considered in the following paragraphs in the process of validation.

Sleep time ESM	Sleep time Algorithm	Awake time ESM	Awake time Algorithm	Time In bed (in min)	Sleep Efficiency	WASO (in min)
20.08.2020 00:30:00	20/08/2020 00:34	20.08.2020 06:36:00	20/08/2020 06:32	448	0.81	24.8
20.08.2020 22:43:00	20/08/2020 22:44	21.08.2020 06:05:00	21/08/2020 06:05	523	0.84	12.2
22.08.2020 00:30:00	22/08/2020 01:08	22.08.2020 06:45:00	22/08/2020 06:45	495	0.78	19.9
23.08.2020 00:30:00	23/08/2020 00:30	23.08.2020 09:05:00	23/08/2020 08:57	600	0.92	14.5
24.08.2020 00:00:00	24/08/2020 00:00	24.08.2020 07:30:00	24/08/2020 07:02	535	0.78	33.0
24.08.2020 22:33:00	24/08/2020 23:38	25.08.2020 07:29:00	25/08/2020 07:20	520	0.78	23.4
25.08.2020 22:30:00	26/08/2020 00:28	26.08.2020 08:00:00	26/08/2020 08:00	570	0.69	32.0
11.09.2020 22:45:00	11/09/2020 23:19	12.09.2020 08:15:00	12/09/2020 04:38	610	0.52	14.0
12.09.2020 23:00:00	12/09/2020 23:00	13.09.2020 07:00:00	13/09/2020 07:00	515	0.98	0
13.09.2020 23:00:00	13/09/2020 23:18	14.09.2020 07:30:00	14/09/2020 05:05	588	0.56	33.7

Table 4.1: Results of IMEC's sleep detection algorithm for 10 days. The first and the third column represent the ESM data. The second and the forth column represent the outcomes of the IMEC's sleep detection algorithm while three last columns represent the sleep parameters computed by the algorithm.

4.1.2 SPT window results

The implemented algorithm in Python calculates the SPT window (i.e., time instant representing the start and end of sleep) and compares the results with the sleep diaries. The algorithm seeks for the SPT window between 8pm and 10 am (time interval which is found empirically). This comparison consists of a visual (see Figure 4.1) and quantitative representation. The quantitative comparison consists of computing the difference in minutes between the prediction of the SPT window algorithm and the available sleep diary.

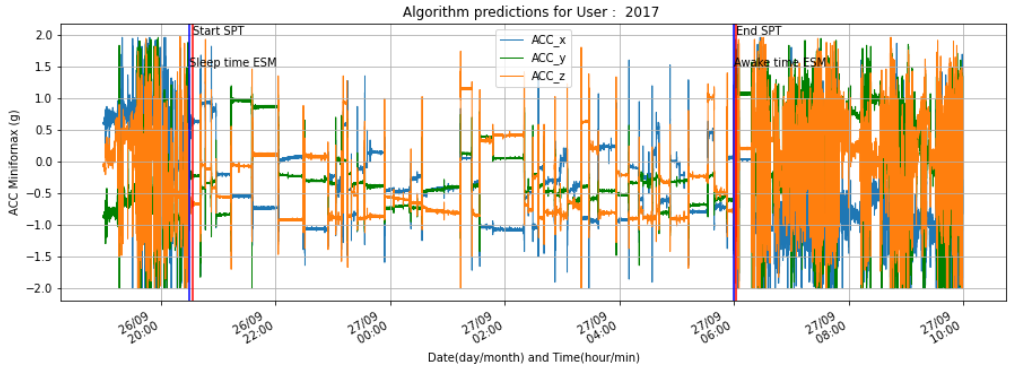

Figure 4. 1: Comparison of SPT window algorithm results and sleep diaries; the blue vertical lines represent the sleep diaries (sleep and wake time) while the red lines represent the SPT window algorithm results

The performance of the algorithm was tested on 6 different Users and 19 nights (not all the available nights were tested due to time constraints). The quantitative results are reported in Table 4.2, where error_sleep/error_awake represents the absolute difference in minutes between the start of sleep/end of sleep predicted by the SPT window algorithm and the sleep time/awake time reported by the subjects under study.

User	Day	Error_sleep (in minutes)	Error_awake (in minutes)
2032	6	0.6	11.5
2032	7	5.4	32.5
2034	2	11.5	47.3
2034	3	29.8	40.9
2034	4	13.8	3.4
2034	5	32.2	10.1
2011	5	205.5	20.9
2011	6	37.6	21.4
2017	4	7.4	3.7
2017	5	6.6	1.0
2017	6	17.6	27.0
2035	5	12.9	22.6
2035	6	6.0	1.8
2035	7	74.0	10.0
2037	2	77.5	3.3
2037	3	43.0	46.3
2037	4	71.0	4.0
2037	5	14.0	114.9
2037	6	6.6	79.3

Table 4.2: Results of comparison between SPT window and sleep diaries. Error_sleep/Error_awake represents the absolute difference in minutes between the start of sleep/end of sleep predicted by the SPT window algorithm and the sleep time/awake time reported by the subjects under study.

From the Table 4.2 it can be seen that for 6 out of 19 nights (31.6%) the difference between the sleep time ESM and the start of SPT window (error_sleep) is less than 10 min (like in Figure 4.1). For another 31.6% the error_sleep is between 10 min to 30 min, while for the remaining 36.8% the error is more than 30 min (like in Figure 4.2). In the same way for 6 out of 19 nights (31.6%) the difference between the awake time ESM and the end of SPT window (error_awake) is less than 10 min, for 36.8% the error_awake is between 10 min and 30 min and for the 31.6 % it is more than 30 min. The results on average were: 35.4 ± 47.9 min for the difference between the sleep time ESM and the start of SPT window and 50.2± 29.7 min for difference between the awake time ESM and end of SPT window.

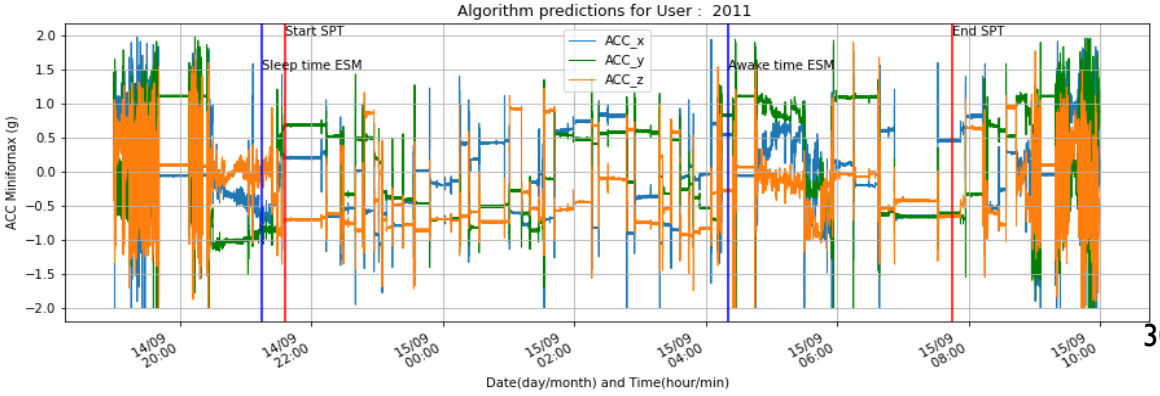

Figure 4. 2: Bad results of the SPT window algorithm; the blue vertical lines represent the sleep diaries (sleep and wake time) while the red lines represent the SPT window algorithm results

4.1.3 Validation of IMEC's sleep detection algorithm results

The validation algorithm, implemented in Python, aims to compare the results of the sleep detection algorithm provided by IMEC with the results of another sleep detection (Sundararajan et al. [57]) algorithm found in literature as explained in M&M [see section 3.1.4].
The validation process consisted of a quantitative analysis (by means of accuracy and Cohen's kappa) but also of a graphical representations (see Figure 4.3). As reported in M&M [see section 3.1.4], the validation process was performed for both cases: with and without the 30s smoothing on the results of the random forest sleep detection algorithm.

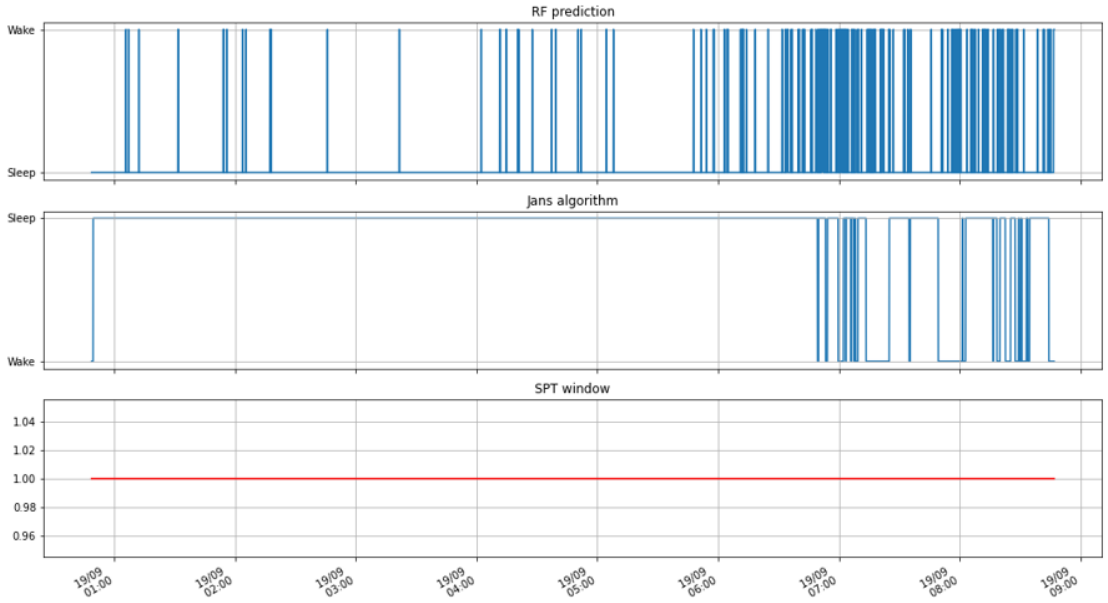

Figure 4. 3: a) Results of the Random Forest sleep detection algorithm [57]; b)Results of IMEC's sleep detection algorithm; c) SPT window within which the validation process in performed

The first part of the comparison (i.e., without performing any smoothing on the results of the Random Forest sleep detection algorithm) was performed considering 10 different Users and 50 nights (see Table 4.3). The mean of accuracy and Cohen's kappa obtained in this case were respectively: 0.96 (std= 0.025) and 0.25 (std= 0.16). Regarding instead the second part of the validation process (with the 30s smoothing the results of the Random Forest sleep detection algorithm) it was performed using 7 Users and 32 nights (see Table

4.4). The mean of accuracy and Cohen's kappa obtained in this case were respectively: 0.97 (std= 0.024) and 0.25 (std=0.2).

User	Day	Accuracy	Cohen's kappa
2013	1	0.89	0.23
2013	2	0.94	0.24
2013	3	0.93	0.11
2013	4	0.92	0.2
2013	5	0.90	0.23
2013	6	0.95	0.11
2017	4	0.98	0.09
2017	5	0.99	0.19
2017	6	0.99	0.14
2021	5	0.98	0.45
2021	6	0.97	0.55
2021	7	0.97	0.35
2031	1	0.98	0.03
2031	2	0.92	0.30
2031	3	0.96	0.07
2031	4	0.98	0.08
2031	5	0.91	0.29
2031	6	0.93	0.37
2031	7	0.98	0.35
2055	1	0.96	0.13
2055	2	0.93	0.36
2055	3	0.98	0.15
2055	4	0.97	0.09
2055	5	0.96	0.17
2055	6	0.96	0.11
2055	7	0.98	0.63
2023	4	0.95	0.49
2023	5	0.92	0.50
2023	6	0.94	0.44
2023	7	0.95	0.30
2034	2	0.96	0.29
2034	3	0.98	0.24
2034	4	0.97	0.43
2037	1	0.98	0.03
2037	2	0.97	0.12
2037	3	0.94	0.21
2037	4	0.98	0.16
2037	5	0.95	0.37
2037	6	0.94	0.18
2037	7	0.97	0.32
2035	5	0.96	0.50

2035	6	0.97	0.32
2035	7	0.95	0.42
2047	1	0.97	0.04
2047	2	0.98	0.17
2047	3	0.97	0.19
2047	4	0.97	0.11
2047	5	0.99	0.03
2047	6	0.95	0.57
2037	4	0.98	0.16
2037	4	0.98	0.16

Table 4.3 : Results (accuracy and cohen's kappa) of the validation algorithm without performing any smoothing

User	**Day**	**Accuracy**	**Cohen's kappa**
2023	4	0.96	0.54
2023	5	0.93	0.54
2023	6	0.95	0.48
2023	7	0.96	0.30
2034	2	0.97	0.28
2034	3	1.00	0.34
2034	4	0.99	0.52
2037	1	0.99	-0.01
2037	2	0.98	0.14
2037	3	0.96	0.20
2037	4	0.99	0.18
2037	5	0.96	0.31
2037	6	0.95	0.16
2037	7	0.98	0.33
2013	1	0.90	0.25
2013	2	0.95	0.27
2013	3	0.95	0.04
2013	4	0.93	0.16
2013	5	0.92	0.25
2013	6	0.97	0.11
2013	7	1.00	0.08
2035	1	1.00	-0.01
2035	2	1.00	0.00
2035	5	0.97	0.61
2035	6	0.99	0.45
2035	7	0.97	0.55
2047	1	0.99	-0.01
2047	2	1.00	0.18
2047	3	0.98	0.14

2047	4	0.99	0.07
2047	5	1.00	-0.01
2047	6	0.97	0.70
2037	4	0.99	0.18

Table 4.4 : Results (accuracy and cohen's kappa) of the validation algorithm with smoothing procedure

4.2 Exploration of the relationship between GSR signal and sleep efficiency

4.2.1 Missing data function

The signals employed for the missing data function were the standard deviation of ACC data, quality indicator of the GSR, and the skin temperature. Different thresholds were used to detect missing values. It was empirically determined that the data would be classified as missing for that day whenever the sum of the GSR quality indicator would be less than 1500000(which corresponds to data missing for 16h) and the skin temperature would be less than 28 degrees for the whole day. In the same way it was determined that the threshold for the STD of the ACC data was 0.08 in order to split missing and non-missing data. The results were checked visually since the ground truth for the missing data was unknown. The current thresholds were found to successfully detect the missing data without overfitting too much to this dataset. Figure 4.4 shows how based on these conditions the 10 min missing intervals were detected

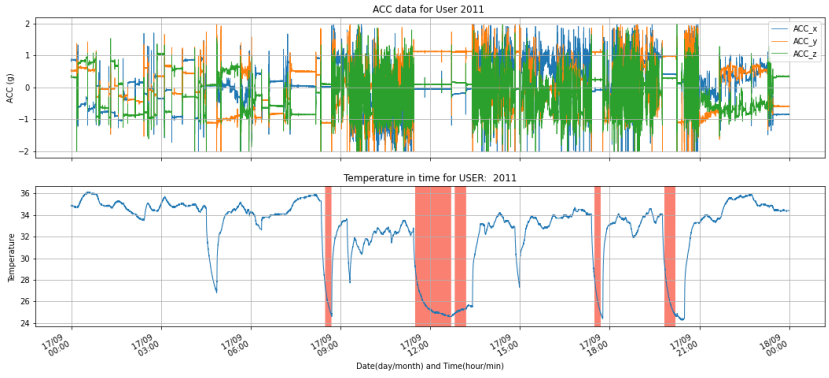

Figure 4. 4: Results of the missing data function. The red sections in the second graph represent the data that is considered missing by the missing data function

4.2.2. Data exploration

The data exploration process, as elaborately explained in M&M [see section 3.2.4], consisted of computing the linear correlations between GSR features. It was found that most of the GSR features (e.g., GSR mean, GSR max, GSR min, GSR std) had a high correlation between them, (i.e., correlations higher than 0.8). The features with a correlation higher than 0.8 were eliminated in the further analysis. The eliminated features were: gsr_scph, gsr_scr_dur, gsr_scr_area, gsr_scr_rr, gsr_median, gsr_max, gsr_min.

Also, correlations between the features and the sleep efficiency were computed (see Table 4.5). From Table , there can be seen that most features have a correlation lower than 0.1. The mean correlation with the target variable is -0.036 (std: 0.05). Correlations lower than 0.1 were not significant (alpha=0.05) in this case. These result indicate a very weak linear relationship between GSR features and sleep efficiency. After elimination of the correlated features, the dataset was reduced to 73 observations and 5 features, namely: gsr_scdiff2, gsr_mean/gsr_scl, gsr_scr, gsr_scr_mag, gsr_std.

Feature	**Correlation**
Sleep efficiency	1.000
gsr_scph	0.043
gsr_scdiff2	-0.003
gsr_mean/gsr_scl	-0.010
gsr_max	-0.023
gsr_scr_mag	-0.069
gsr_std'	-0.095
gsr_scr'	-0.096

Table 4.5: Pearson correlations between sleep efficiency and GSR features

Afterwards the dataset was normalized using the standard scaler. In Figure 4.5, boxplots of the data before and after standardization are shown. After the standardization of the dataset it can be observed that the individual features look more like normally distributed data (e.g. Gaussian with 0 mean and unit variance) compared to the data before standardization, which is in fact the main goal of standardization.

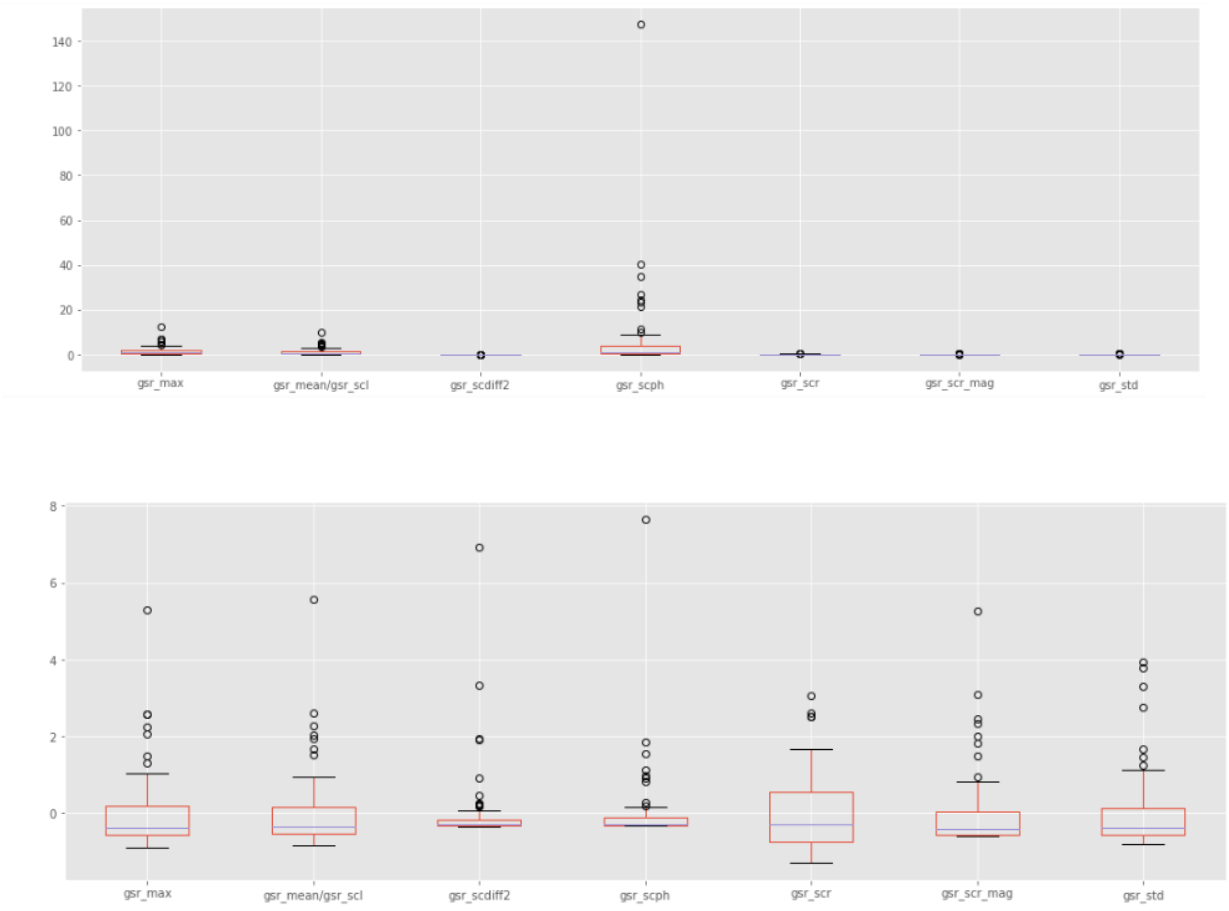

Figure 4. 5: Boxplots of the GSR features before and after standardization by means of standard scaler

Afterwards visual data explorations were made between each feature and the target variable (see Figure 4.6). It can be observed from the histograms of the features for each class in Figure 4.6 that there isn't a clear separation between the two classes. Also from the figure 4.7 it can be observed the same thing.

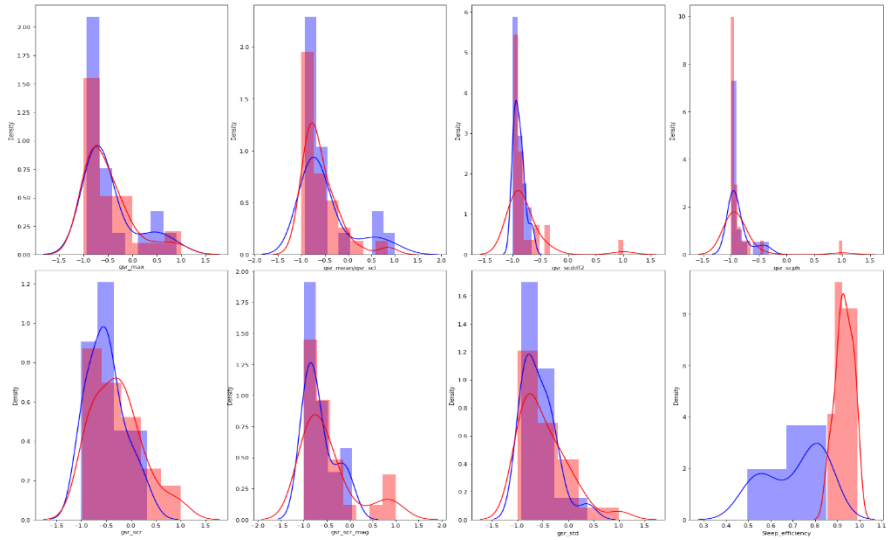

Figure 4. 6: Histograms of each class for every GSR feature.

Figure 4. 7: Bivariate plot that represent the relationship between every couple of GSR features highlighting the class

Therefore, from this simple data exploration no clear separation between the classes 0 (low sleep efficiency) and 1 (high sleep efficiency) was observed.

4.2.3 Machine learning models implementation and results

To explore the possible relationship between sleep efficiency and GSR signal different machine learning models were implemented in Python. Before doing so, it was checked if the dataset was unbalanced, and it resulted that there were 29 observations classified as 0 and 44 classified as 1. The dataset was not very balanced, but since the available dataset is relatively small, it was decided to continue with the dataset as it is, without trying to balance it in order not lose data.
Practically, the models implemented for this analysis included: KNN, logistic regression, decision tree, random forest ,and multi-layer perceptron (see Table 4.6).

	TEST			**TRAIN**			
	Accuracy	**Recall**	**F1 score**	**Accuracy**	**Recall**	**F1 score**	**AUC**
Logistic regression	0.6	1.0	0.74	0.6	1.0	0.76	0.56
KNN	0.54	0.7	0.64	0.7	0.9	0.8	0.45
Desicion tree	0.50	0.7	0.62	0.73	0.93	0.80	0.40
SVM	0.6	1.0	0.74	0.6	1.0	0.76	0.56
MLP	0.5	0.70	0.62	0.90	1.0	0.92	0.37
Random Forest	0.54	0.61	0.61	0.84	0.90	0.87	0.51

Table 4.6 : Results (accuracy,recall, f1 score) of machine learning models on both test and train set

From the table 4.6, it can be observed that AUC metric for most of the machine learning models is less than 0.5 which means that the models are performing worse than a random classifier. Another problem of these models is that they perform relatively good for the train set but not that good for the test set, which indicates presence of overfitting. It seems that the model performing better among all the considered models is logistic regression. Since logistic regression requires the data to be normal, an additional correction for normality was performed since this did not happen before this analysis. Different transformations were used to normalize the data including log(1+data), square root and Yeo-Johnson. The performance of each method was analysed by visually comparing the distributions before and after the transformation. The best performing one was found to be the Yeo-Johnson transformation (see figure 4.8)

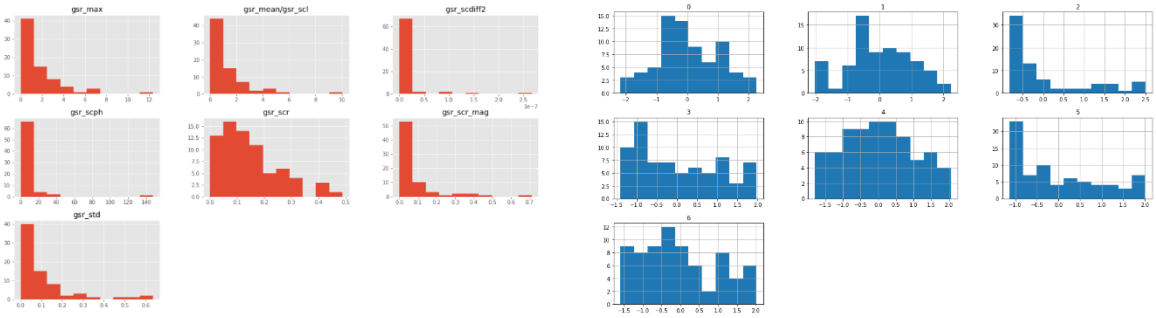

Figure 4. 8: GSR feature distribution before and after Yeo-Johnson transformation.

However, the results of the model did not improve even after normalizing the data. Logistic regression in this case has also the drawback of only predicting the sleep efficiency in test set as 1 as it can be seen from the confusion matrix (see Figure 4.9). This is an additional proof that the model does not perform very well.

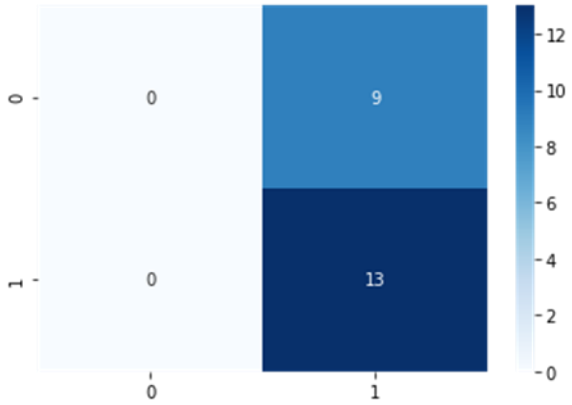

Figure 4. 9: Confusion matrix of logistic regression model

To improve the results of the models, some attention was given to further cleaning the dataset (of bad quality GSR data) and also to perform a second GSR feature selection. To do so, the GSR features considered until now were plotted for each user (see Figure 4.10) and they were observed in order to detect features that behave in a similar way (i.e that are correlated) or bad quality data.

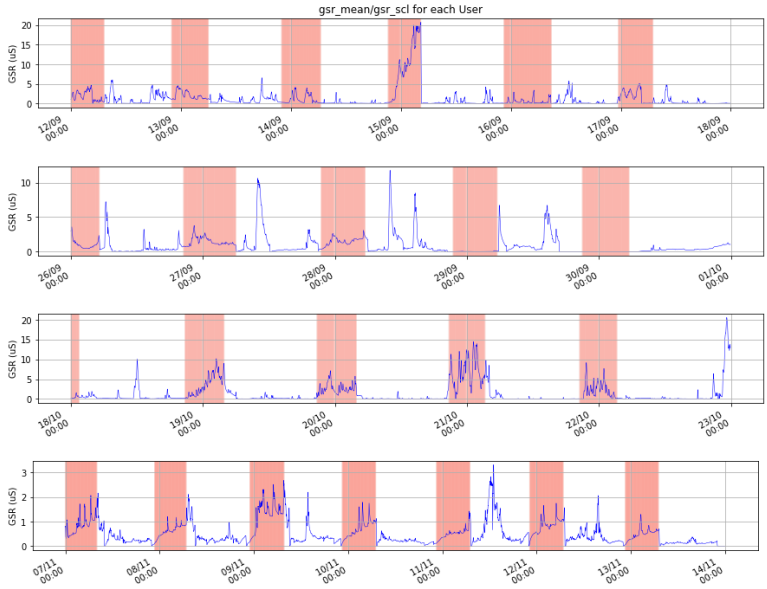

Figure 4. 10: GSR mean data plotted for all available days for user 2011,2017,2021 and 2023(each row represents the GSR mean data for that user)

It was observed that there were some days in the data that were of bad quality ,i.e., amplitude very close to zero (see Figure 4.10, second graph) . This was also confirmed by the quality indicator of GSR, which is lower for those days compared to the other ones (see Figure 4.11).

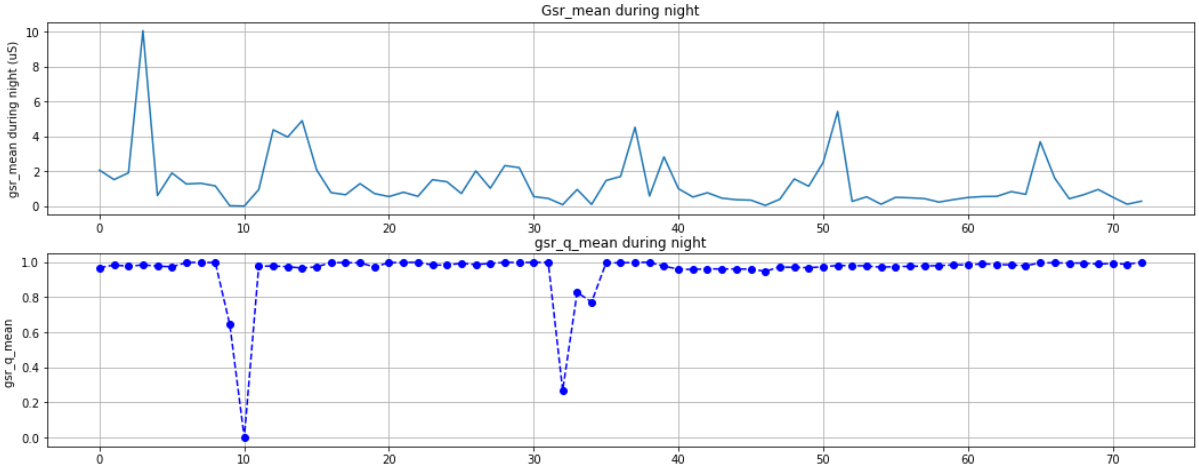

Figure 4. 11: GSR mean feature and quality indicator feature for each night

The days with a quality indicator lower than 0.8 were eliminated, in this way the dataset was reduced from 73 observations to 69. What was also observed from this visual analysis is that some of the features behave very similarly, which means that the models that were implemented before might suffer of collinearity (see Figure 4.12). For instance, as it can be seen in the Figure 4.12, GSR mean and GSR std behave in similar way, therefore contain almost the same information. To confirm this result quantitively, the correlations between the features were again calculated, but this time, the features correlated with more than 0.5 were eliminated. The remaining features were: GSR mean , GSR scr and GSR_diff2.

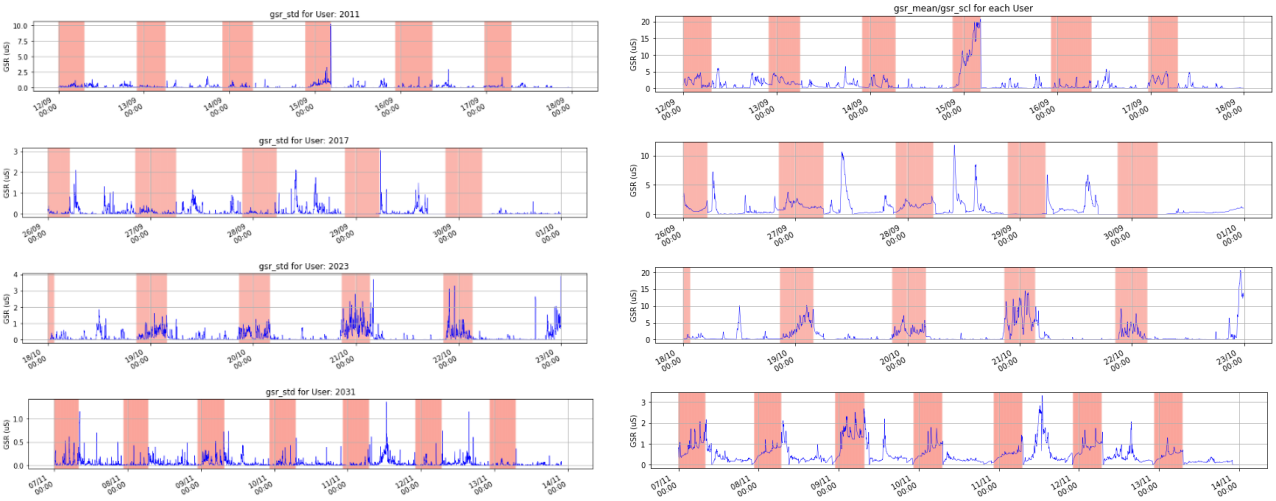

Figure 4. 12: GSR std and GSR mean plotted for user 2011,2017,2023,2031. Every row represents the GSR std and GSR mean for that user.

The previous models were implemented by using only these three features (see Table 4.7)

	TEST			**TRAIN**			
	Accuracy	**Recall**	**F1 score**	**Accuracy**	**Recall**	**F1 score**	**AUC**
Logistic regression	0.57	1.0	0.72	0.6	1.0	0.73	0.40
KNN	0.47	0.50	0.52	1.00	1.00	1.00	0.47
Desicion tree	0.71	0.67	0.72	0.83	0.75	0.84	0.70
SVM	0.6	1.0	0.72	0.6	1.0	0.73	0.61
MLP	0.5	1.00	0.72	0.60	1.00	0.73	0.31
Random Forest	0.38	0.66	0.55	0.70	0.92	0.78	0.41

Table 4.7: Results (accuracy,recall, f1 score) of machine learning models on both test and train set after dataset cleaning

From this second analysis it can be observed (see Table 4.7) that the model giving the best results was the decision tree. This model gives a good AUC, relatively good metrics (compared to the results of the other models) and small difference between the train and test set.

PCA analysis result

In order to prevent collinearity, there was opted to apply PCA. PCA is applied to all the features (GSR signal features + GSR quality indicator features) and only the first 3 principal components were included in the models, which explain more than 90% of the variance. The model that gave the best results in this case was the decision tree. The performance of the model decreases if the quality indicator's features are not included in the model (see Table 4.8). However, the AUC curve increases in this case passing from 0.74 to 0.77.

	TEST			**TRAIN**			
	Accuracy	**Recall**	**F1 score**	**Accuracy**	**Recall**	**F1 score**	**AUC**
PCA with GSR Quality indicator Features	0.76	0.91	0.81	0.75	0.82	0.79	0.74
PCA without GSR Quality indicator features	0.66	0.58	0.66	0.85	0.78	0.86	0.77

Table 4.8: Decision tree results (accuracy,recall, f1 score) of machine learning models on both test and train set before and after applying PCA

4.2.4 Mixed model's results

To evaluate and interpret the relationship and effect size between GSR and sleep efficiency while taking into account the variability that there is for each user, there was decided to tackle the same research question using statistics, more specifically, using linear mixed effect models. Before the modeling, some analyses in R were executed in

order to evaluate the variability between users (see Figure 4.13). It was observed that both mean value of sleep efficiency of each user and the variability within each user is different between different users. The same trends are seen for the mean GSR (see figure 4.13).

First, the possible relationship between GSR mean and sleep efficiency was modeled by means of a linear mixed model. An anova test was performed in order to compare the baseline model (random model) and the

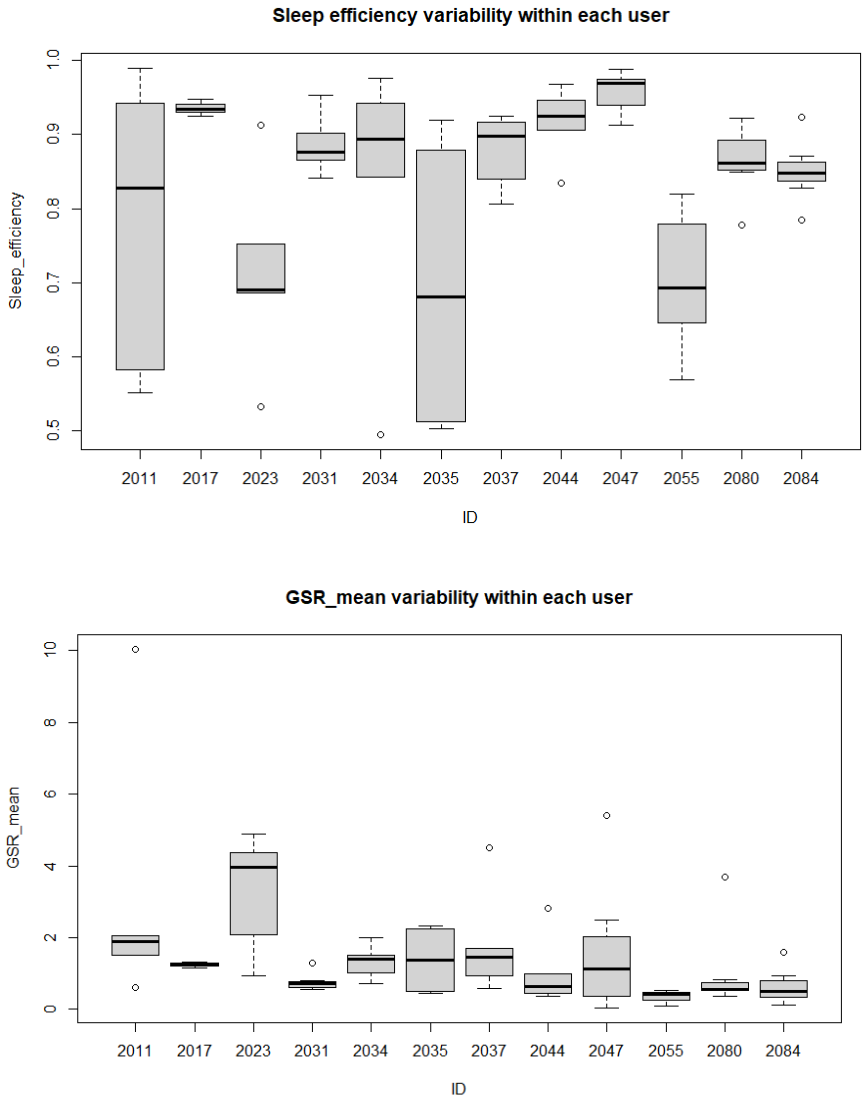

Figure 4. 13: Sleep efficiency and GSR mean boxplots for every user

full model (sleep efficiency ~ GSR mean). The p-value of the anova test was not significant (p-value = 0.38). Also, the relationship between GSR_scr and sleep efficiency was explored, but even in this case there weren't any significant results (p-value= 0.62). To optimally combine the multiple GSR features, which have been showed to be highly correlated among each other (see section 4.2.3), PCA was applied. The first three

components of PCA were included in the model but the model remained insignificant (p-value=0.7).

4.3 Exploration of the relationship between GSR and skin temperature

4.3.1 Data exploration

To tackle this research question, some basic data exploration was first performed in R. Pearson's correlations were computed for every day and night between temperature mean and GSR mean for all the 69 observations. It was observed that the correlations during night were significant (p-value<0.05) for 52 nights out of 69 of which 19 nights (out of 52) had a negative correlation. For the correlation between temperature and GSR during day, it was observed that for 60 days out of 69 the correlation was significant, and that 12 days out of these 60 days had a negative correlation.
Afterwards, the boxplots of the temperature for both day and night were computed (see figure). It was observed (see figure 4.14) that skin temperature was generally higher during night than during day for almost all the users which was confirmed by the boxplots of temperature during day and night for all the users together .

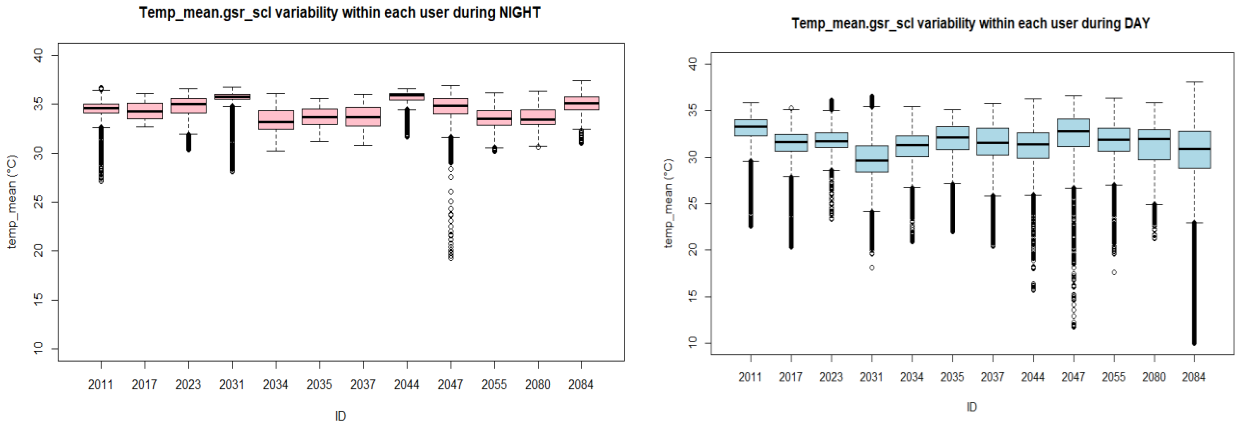

Figure 4. 14: Boxplots of mean temperature during day and night for each subject

The same analysis can be made for the GSR signal (see figure 4.15). From these graphs it can be observed that also the GSR signal appears to be slightly higher during night than during day and also the variability of GSR is higher during night compared to day.

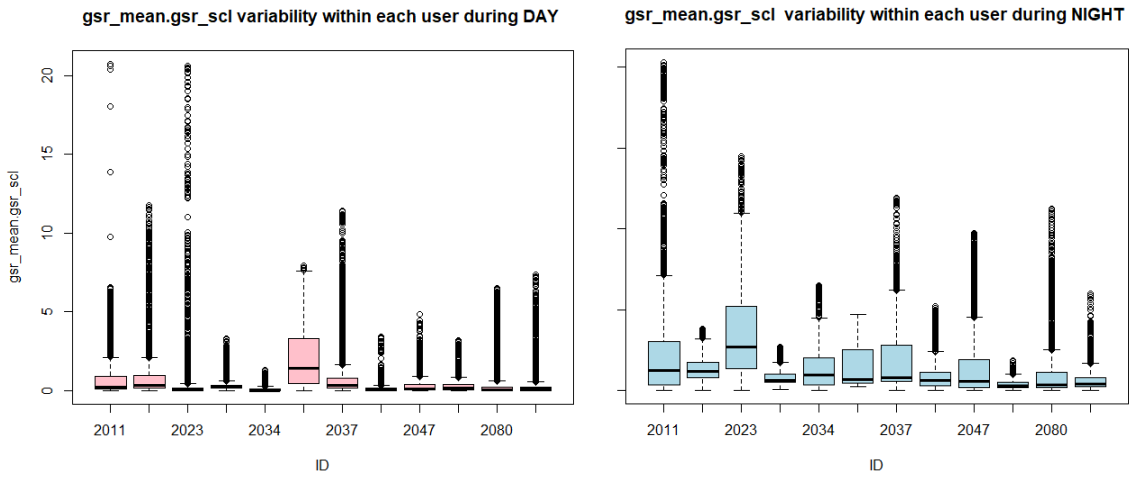

Figure 4. 15: Boxplots of GSR during day and night for each subject

Finally a last step involved testing if the inclusion of a random effect structure with random intercepts was justified by comparing the AIC of the baseline model without random intercepts and of the model with random intercepts (respectively 352014.8 and 360740). Since the AIC of the model with random intercepts was lower, the usage of random intercepts is justified.

4.3.2 Linear Mixed modelling

The significant correlations between GSR signal and temperature, reported in the former section [see section 4.3.1], were further researched by means of the linear mixed models. The analysis was done separately for day and night. This model was also controlled for motion by adding motion as an interaction effect with mean temperature, since motion has been shown to also increase mean GSR [56]. Then, the anova tests were performed to see if the interaction between temperature and motion is a significant fixed effect for both the day and the night.

The linear mixed effect model of the day data showed that the mean temperature significantly affected mean GSR ($p < 2.2 e-16$) increasing it by about 0.029 µS (std=0.0014 µS) regardless from motion, while for the night data mean temperature also significantly affected mean GSR ($p < 2.2 e-16$) increasing it by about 0.097 µS (std= 0.0092 µS) regardless from motion (see Table 4.9). Also, the conditional and marginal R^2 were

computed, for the night model they were 17.9% and 0.4% respectively while for the day model they were 19.3% and 0.6%.

Afterwards, this model was also controlled for motion by adding the mean of motion as a fixed effect to the model with only mean temperature as a fixed effect. The anova tests showed that the mean motion in the additive model significantly affected mean GSR in both day and night data (p <2.2 e-16 and p <4.6 e-7 respectively). Also R2 conditional and R2 marginal were computed after adding motion (see Table 4.10). It can be observed that R2 marginal remained almost unaltered for the night while it increased from 0.006 to 0.011 for the day (see Table 4.10)

Then, anova tests were performed to see if the interaction between temperature and motion is a significant for both the day and the night and it resulted significant for the day (p-value <2.2 e-16) and non-significant for the night (p-value= 0.0518).

	Model on the day data	**Model on the night data**
Estimated fixed effects (p-value)		
Intercept	-0.3246 µS	-1.8761 µS
Mean temperature	0.029 µS (<2.2 e-16)	0.096µS (<2.2 e-16)
Variance random effects		
ID	0.253	0.729
Residual	1.095	3.432
Explained variance by the model		
Marginal R^2	0.6	0.4
Conditional R^2	19.3	17.9

Table 4.9: Mixed model results for the model with mean temperature as a fixed effect

	Model on the day data	**Model on the night data**
Model only with mean temperature		
Marginal R^2	0.6%	0.4%
Conditional R^2	19.3%	17.9%
Model with mean temperature		

and with mean motion		
Marginal R^2	1.1%	0.5%
Conditional R^2	18.8%	17.6%

Table 4.10: Comparison of R2(marginal and conditional) of the mixed models without and with presence of motion as a fixed effect

To compare the estimates of the relationship between mean GSR and mean temperature between day and night, a binary day-night variable was introduced into a merged model as a fixed effect. The merged model considered all the explored variables until this point (i.e., mean temperature, mean motion, day night variable) and also the possible interactions between them (i.e., Mean temperature*mean motion, Mean temperature*day night, Mean motion*day_night, Mean temperature*mean motion*day night) (see Table 4.11).

It resulted from the anova test that the binary day-night variable significantly affected GSR mean (p <2.2 e-16) decreasing it by about -20.99 µS (std = 4.81 uS) when passing from night to day. Also the triple interaction (i.e., mean temperature*mean motion*day night) significantly affected mean GSR (p_value= 0.000479) increasing it by about -0.471270 µS (std= 0.134944µS). It can be observed by the merged model, that also the interaction mean motion and day night variable seem to significantly affect GSR (p-value< 0.05) changing it by about -21 µS . When it comes to the explained variance by the fixed effects, the full model was able to explain it by almost 8%. This model results to be the one which reaches the highest explained variance by the fixed effects among all the models considered until now.

	Final model
Intercept	4.918604
Mean temperature	-0.005875
Mean motion	-6.464583
Day night variable	-20.991145
Mean temperature*mean motion	0.096618
Mean temperature*day night	0.436138
Mean motion*day_night	21.173356
Mean temperature*mean motion*day night	-0.471270
ID	0.2077
Residual	1.9420
Marginal R^2	7.9%

Conditional R^2	16.8%

Table 4.11 : Full mixed models results

4.3.3 Relationship of GSR and lagged temperature

Generally, it is hypothesised that GSR increase during emotional sweating or activity would be associated with a decrease in temperature (because of the sweating) but there is no hypothesis in literature about the order of these events. So, the relationship between GSR and lagged temperature was explored by implementing different linear mixed models that

have a fixed effect of mean temperature with different lags (2, 4, 6, and 8 min) for both day and night data. The results are reported respectively for day and night in Table 4.12 and Table 4.13 (see Tables 4.12 ,4.13). It can be observed from the tables for both day and night that if the lag increases, the Intercept and the estimate of the temperature stays nearly the same. Also the std of the Intercept and the estimate of the temperature stays nearly the same. Also the std of the intercept and the slope remain almost constant.

	Intercept	**Std Error**	**Temp mean**	**Std Error**
GSR/Temp 2 min lag	-6.9	0.5	0.244	0.014
GSR/Temp 4 min lag	-6.7	0.5	0.238	0.013
GSR/Temp 6 min lag	-6.6	0.5	0.234	0.014
GSR/Temp 8 min lag	-6.5	0.5	0.231	0.014

Table 4.12: Day results of the linear mixed models between GSR and lagged temperature

	Intercept	Std Error	Temp mean	Std Error
GSR/Temp 2 min lag	-3.05	0.34	0.114	7.8e-3
GSR/Temp 4 min lag	-2.98	0.34	0.119	7.9e-3
GSR/Temp 6 min lag	-2.86	0.35	0.108	8.1 e-3
GSR/Temp 8 min lag	-2.74	0.36	0.104	0.008

Table 4.13: Night results of the linear mixed models between GSR and lagged temperature

Chapter 5

DISCUSSION

In this chapter, the results of all the three modules of this master thesis project will be discussed separately.

5.1 IMEC's sleep detection algorithm analysis and research

5.1.1 SPT window

The sleep detection algorithm provided by IMEC has a dependency on sleep diaries, which are not always available. In this master thesis, a first goal was to research whether it was possible to adapt the current algorithm so it would accurately classify sleep without sleep diaries as an input. Therefore, the SPT window algorithm developed by Van Hees et al. [24], was implemented. This algorithm only uses accelerometer data to detect the sleep period time. It was observed from the individual results (i.e., every night) that the algorithm performs well in 32% of the cases (i.e., the difference with ESM data less than 10 min), it performs moderately in 32% of the cases (i.e., the difference with ESM data between 10-30 min) while in 36% of the cases it performs badly (i.e., the difference with ESM data more than 30 min).

In this work, the algorithm seeks for the SPT window only between 8pm to 10 am, which is the most probable time when the individuals sleep. However, this is a limitation of algorithm, which does not consider naps during the day or day sleepers. Another limitation of the algorithm is that it makes basic assumptions about sleep interruptions (i.e., if the subject is wake for less than 1 hour between 30 min sleep intervals, it is considered as sleeping for that one hour). This is not a mild assumption especially in case of subjects with different sleep alterations.

However, the obtained results should be interpreted not by only considering the performance of the algorithm but also the problems related to the available dataset. It happens that the protocols for the sleep diary compilations (which are not always straightforward) are not fully understood by the participants, which leads to possible not correct sleep diaries. This could be the case of this user 2011 in this particular day (see

Figure 5.1).What probably happened in this case is that the person was sleeping (it can be assumed by the low activity of the accelerometer data), probably woke up for few min during the night, and filled the sleep diary considering that time as awake time, even though he/she probably when back to sleep. For this particular night, the algorithm would give a higher error, but not because of the algorithm but due to the probably not correct compilation of the sleep diaries. Another problem with the available dataset is related to the presence of missing data. This becomes a problem for the algorithm for two reasons: first, due to the low amplitude of the missing data, the algorithm misclassifies it as sleep, and secondly, as explained in M&M [see section 3.1.3] the SPT window algorithm makes use of an individual threshold, which due to the presence of missing data will get lower and this would lead to misclassifications. As a matter of fact, the presence of the missing data is very often unavoidable, therefore the algorithm should be improved in order to be more robust to missing values. In this work, it was attempted to impute the missing values (i.e., to substitute 10 min missing intervals with 10 min reference accelerometer data which was considered to be good data), but however, the results did not improve, probably due to the very simple imputation technique implemented. Another possible improvement of the algorithm can be the adaptation of the threshold used by the algorithm for each day /user (taking into account even the presence of missing data), but however, attention should be paid since this might lead to overfitting. Nonetheless these limitations, the SPT window algorithm was implemented in the IMEC's sleep detection algorithm to make the algorithm independent from sleep diaries.

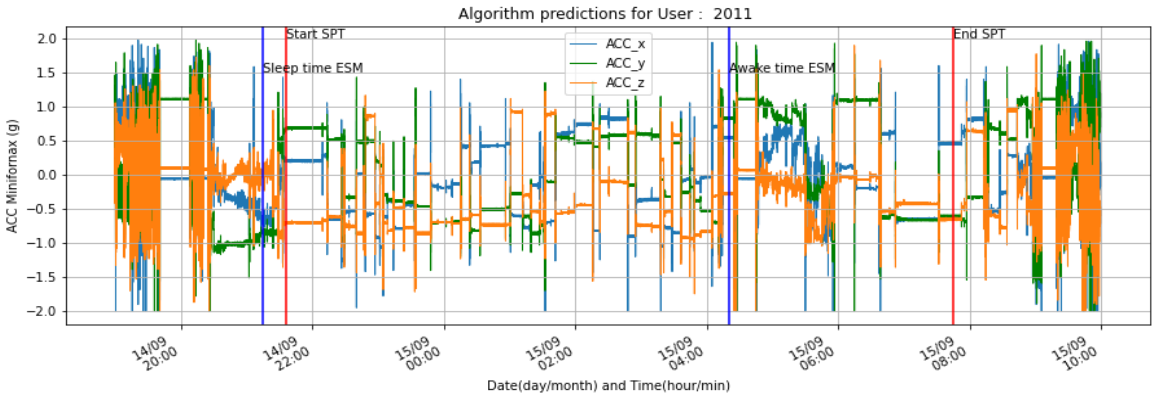

Figure 5. 1: Bad results of the SPT window algorithm; the blue vertical lines represent the sleep diaries (sleep and wake time) while the red lines represent the SPT window algorithm results

5.1.1 Validation algorithm

The validation algorithm, which aims to compare the results of the sleep detection algorithm provided by IMEC with the results of Random Forest sleep detection [57], was performed for both with and without the 30s smoothing on the results of the random forest sleep detection algorithm. To compare the two algorithms more accurately in terms of the Cohen's kappa, the smoothing procedure was performed since it was observed that the Random Forest algorithm was bouncing much more that the IMEC's sleep detection algorithm (see Figure 5.2). The number of nights used to perform the validation algorithm with and without the smoothing procedure were 50 and 32 nights, respectively. The validation was not performed for the whole dataset due to time constrains. However, the results of the comparison were generally good, having high accuracies (i.e. 0.96 (std= 0.025) and 0.97 (std= 0.024) for without and with smoothing respectively). Although these two sleep detection algorithms have different working principles, the results of the validation are promising.

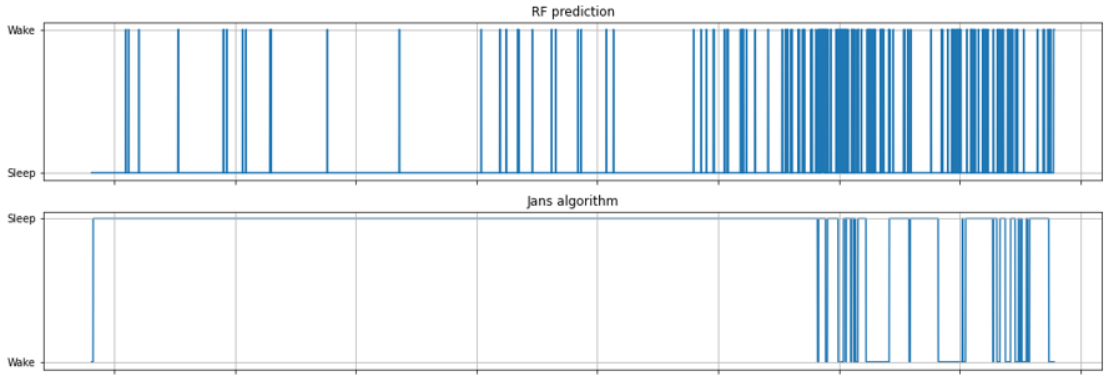

***Figure 5. 2**: a) Results of the Random Forest sleep detection algorithm [57]; b) Results of IMEC'sleep detection algorithm; c) SPT window within which the validation process is performed.*

Limitations to the analysis were the fact that there was no direct PSG data to use for validation, which is the gold standard for sleep measurements. Therefore, an indirect validation with the PSG data was performed by means of Random Forest sleep detection algorithm, made by Sundararajan et al. [57] which was actually tested against PSG for 134 nights and obtained an F1 score 73.93%. A second limitation to the analysis is the fact that even with the 30s smoothing, the visual comparison between IMEC's algorithm

remained bouncier than hoped. To fully validate the sleep detection algorithm, for future work, validation with outcomes of PSG is required.

5.1.2 Missing data function

To detect and remove the missing data, a missing data function was developed that is based on the standard deviation of ACC data, quality indicator of the GSR, and skin temperature. The choices of the inputs were based on literature and logical thinking which will be shortly defended here. The STD of ACC was chosen because it is widely used in actigraphy studies to detect missing values [84]. When the wristband is not worn by the user, but remains on, the skin conductance signal will suddenly drop to zero because the circuit becomes interrupted. The GSR quality indicator will then indicate low quality because of two reasons: 1) the sudden drop in the GSR is not something that is normally seen in GSR and will cross the threshold: maximal decrease of 10%/sec, that is present within the GSR quality indicator algorithm, and 2) because the GSR quality algorithm has a threshold on minimal GSR that is needed, which is on 0.005 μS. Finally, when the wristband is removed by the user, one can also expect a drop in the skin temperature since room temperature is expected to be lower than skin temperature.
Because none of the aforementioned conditions on its own would be sufficient to detect missing data with certainty, e.g. room temperature could be possibly as high as skin temperature when it is very hot outside, the three conditions needed to be fulfilled together in order to classify as missing data. The exact methodology of the algorithm was explained in section [4.1.2].
The thresholds of the mean and std of skin temperature, the std of ACC and the sum of quality indicator were empirically optimized so they adapted to the current dataset without overfitting it. The results of the missing data function were visually found to accurately detect missing data in more than 80% of the nights. However, the limitation of this analysis is the fact that the actual missing moments (the ground truth) was unknown and thus could not be used to optimize the thresholds. In future work, this is something which can be easily done, by setting up a small experiment where the wristband is taken off during predefined moments in time.

5.2 Exploration of the relationship between GSR signal and sleep efficiency

An investigation was performed to study if GSR during night is related with sleep efficiency. From the analysis in Python, it was explored whether GSR signal was able to predict sleep efficiency. The first machine learning models (i.e., without performing PCA) did not perform well since the AUC curve for 3 out of 6 models was less than 0.5, which indicates models predicting worse than a random classifier. The logistic regression and

SVM model had an AUC higher than 0.5, however their accuracy, f1 score, and recall were not good. Particularly their recall, which was 1 in both train and test set, was found unsatisfactory, since this means that the models were predicting only one of the two possible classes. The decision tree model gave the best results (see Table 4.7) because it had a reasonable AUC (70%), relatively good performance metrics (see Table 4.7), and small difference between the train and test set which indicates that there isn't a lot of overfitting.

Afterwards, PCA which compresses, stores and transmits data more efficiently and it solves the problem of collinearity, was applied to the feature dataset with and without the GSR quality indicator features. The results including the GSR quality indicator were relatively good (see table 6.1), which marks the GSR quality indicator as a good predictor of sleep efficiency, which makes sense since the sleep efficiency was computed based on the same data.

	TEST			**TRAIN**			
	Accuracy	**Recall**	**F1 score**	**Accuracy**	**Recall**	**F1 score**	**AUC**
PCA with GSR Quality indicator Features	0.76	0.91	0.81	0.75	0.82	0.79	0.74
PCA without GSR Quality indicator features	0.66	0.58	0.66	0.85	0.78	0.86	0.77

Table 5.1: Decision tree results (accuracy, recall. F1 score) in test and train set before and after applying PCA

The PCA results, when applied to the feature dataset without the GSR quality indicator features, decrease. In this case, it can be observed a higher difference between the metrics in the train and test set which indicates a higher presence of overfitting. From the results of the first part of the analysis, it can be said that there is roughly a relationship between the GSR signal and sleep efficiency

The second part of analysis performed in R, which investigated the relationship between GSR and sleep efficiency by means of linear mixed models, did not find a significant relationship between GSR and sleep efficiency (p-value >0.05), not even after applying PCA.

From the results of the analysis made in both python and R, there wasn't found a clear relationship between sleep efficiency and GSR signal. However, the following limitations should be considered.

Firstly, sleep efficiency was calculated per night, so the GSR features were summarized into one value for whole the night. Thus, in this way a lot of information is lost since it is impossible to capture the complexity of GSR changes during the night by one value. In future work, the dynamics of GSR during night should be considered in this analysis.
Secondly, it might be possible that the available dataset (69 nights) is too small to accurately model this relationship.
Thirdly, in the machine learning phase, only classification models were considered, although the sleep efficiency was a continuous variable. The threshold, of 85%, for making this continuous variable a binary one was taken from literature [6] but was not validated in any way. In future work, the machine learning models should ideally be remade on a larger dataset (previous argument) while both considering sleep efficiency as continuous and binary variable. Ideally, the threshold for binary transformation should be validated by executing a large literature review or by experimental validation.
Fourthly, only models between sleep efficiency and GSR were considered. However, the wristband also acquires data that might confound or influence this relationship, like temperature and motion [36]. In future work, one should consider these confounders in the model in order to more accurately model this relationship and to account for these confounders explained variance of the model.
Lastly, there might be a small mistake in the used sleep efficiencies, since the algorithm was not validated with the golden standard, PSG.

5.3 Exploration of the relationship between GSR and skin temperature during both day and night

5.3.1 Data exploration and machine learning models

A recap of the findings from the first part of data exploration analysis were that: 87% of the days and 75% of the nights, had significant correlations between the mean of GSR and the mean of temperature. Also, negative correlations were observed, more specifically for 19 nights and 12 days. These results, indicate an eventual linear relationship between GSR and temperature which would be further verified in the analysis.
However, if the reason for this interaction is thermoregulation, what should be expected is a positive correlation between skin temperature and GSR signal, because a higher temperature would cause sweating, which in turn would increase the skin conductance, therefore the GSR signal. Nevertheless, there is the presence of negative correlations between these two signals, which might suggest that the thermoregulation is not the only phenomena causing the variability of the GSR signal. Another finding that needs to be discussed is the fact that the number of significant correlated days is higher than the number of significant correlated nights while the number of negative correlations is higher for the night than for the day. What can be hypothesized from these results is that there might be a different “interaction” between temperature and GSR during day and

night. Naively, it can be said that this relationship, which was assumed to be due to thermoregulation, seems to be more stable during day then during night since the correlated days are higher than the correlated nights. It is supported even by the fact that the negatively correlated days are less than the negatively correlated nights (since negative correlation seems to contradict presence of thermoregulation). These hypotheses must be further investigated in this analysis.
From the data exploration, it was observed that the temperature is higher during night than during day which is coherent with other studies like Sano et al. [54]. However, some measures like ambient temperature or of whether the person's wrist was under a blanket are lacking, and this last one is likely to make the skin warmer. Also, the GSR signal seems to be in average higher in amplitude and with a higher variability during night than during day, but further information about the subjects under study is needed to comment this result.

5.3.2 Mixed models

From the mixed models results, there seems to be a significant relationship between temperature and GSR signal for both day and night (i.e., p-value<0.05 . However, the effect size of this relationship is very small since only 0.4% and 0.6% of the variance of the model (for day and night respectively) was explained by the temperature. Also, the explained variance by both the fixed and random effects for both day and night were only less than 20%, which means that there are probably other factors causing the variance of GSR. In fact, according to literature, the GSR signal is a result of the sweat glands whose activation is caused by emotional stimuli or thermoregulation [65] [66]. Therefore, indirectly, it can be assumed that since thermoregulation explains only small part of the variance of the GSR signal, probably emotional stimulus has an influence on the GSR signal too. It should be also said that most of the explained variance was due to the variability that there is between subjects (i.e., inter-variability), which is a coherent result with literature [62].
Motion was considered when exploring the relationship between GSR and temperature since active moments can distort the GSR signal [56]. As expected, the results for both day and night were significant (p-value< 0.05) . However, the interaction between motion and temperature was only found significant in the day model. This result seems logical since physical activity is generally limited during night.
Another point of interest in this study is the eventual different relationship between temperature and GSR during day and night. To compare the estimates of the relationship between mean GSR and mean temperature between day and night, a binary day-night variable was introduced into a merged model as a fixed effect. The merged model considered all the explored variables until this point (i.e., mean temperature, mean motion, day night variable) and also the possible interactions between them (i.e., Mean temperature*mean motion, Mean temperature*day night, Mean motion*day_night,

Mean temperature*mean motion*day night). It resulted from the merged model , that there is a significant difference between the relationship GSR and temperature during day and night (p-value<0.05). To explain this result, i.e., this slightly clearer interaction between GSR and temp during day rather than night it should be considered that according to studies the thermoregulation process is suppressed during REM sleep and persist during NREM [55] and since 20-25% of sleep is REM sleep this might be the reason why the relationship is more consistent during day, here the thermoregulation process is probably persistent during the whole day.

From the merged model result, as reported in result section, it can be observed a significant relationship of GSR mean with mean temperature, mean motion and the day night variable. Also the interactions seem significant (p-value<0.05), but however it seems that the ones that most change the GSR signal are the motion and day night variable. These results once again emphasize the effect of motion in this process, and also of the different interaction that there is between temp and GSR signal during day and during night.

5.3.3 GSR and lagged temperature

Since it is hypothesised that GSR increase during emotional sweating or activity would be associated with a decrease in temperature (because of the sweating) but there is no hypothesis in literature about the order of these events, the relationship between GSR and lagged temperature (with different lags (2, 4, 6, and 8 min)) for both day and night data was explored. What is expected if the premise is correct is that for a lagged temperature, the slope between GSR mean and temperature mean becomes negative. From the results of this analysis, the hypothesis was not confirmed, since the slope between the GSR and temperature for both day and night remain positive for every lag.

However, it needs to be said that only one direction of the relationship was considered (i.e., GSR and lagged temperature) and also a linear relationship was hypothesised. As a future work it can be exploring their relationship with a non-linear model and also considering both directions of the interaction.

Chapter 6

Conclusions

6.1 IMEC's sleep detection algorithm analysis and research

In this master thesis, a first goal was to research whether it was possible to adapt the current sleep detection algorithm so it would accurately classify sleep without sleep diaries as an input. The algorithm implemented was able to have a good to moderate performance for about 75% of the cases, however the algorithm presents several limitations related to robustness in presence of missing data, and the fact that it does not consider day sleepers and naps during day. Also, the algorithm makes very basic assumptions about sleep interruptions which are not mild assumptions when in case of people with different sleep alterations. Therefore as a future work, there is a need to generalize the algorithm for all these particular cases.

Another goal of the first module of this master thesis was to perform the validation of IMEC's sleep detection algorithm by comparing it by an open-source of Random Forest sleep detection [57]. The analysis was performed for both with and without the 30s smoothing. The results of the comparison were generally good, having high accuracies (i.e. 0.96 and 0.97 for both without and with smoothing respectively). Although these two sleep detection algorithms have different working principles, the results of the validation are promising.

Limitations to the analysis were the fact that there was no direct PSG data to use for validation, which is the gold standard for sleep measurements. Therefore, an indirect validation with the PSG data was performed by means of Random Forest sleep detection algorithm, made by Sundararajan et al. [57] which was actually tested against PSG for 134 nights and obtained an F1 score 73.93%. A second limitation to the analysis is the fact that even with the 30s smoothing, the visual comparison between IMEC's algorithm remained bouncier than hoped. To fully validate the sleep detection algorithm, for future work, validation with outcomes of PSG is required.

6.2 Exploration of the relationship between GSR signal and sleep efficiency

The main objective of the second module of this master thesis work was the investigation of the relationship between GSR during night and sleep efficiency in Python and R.

From the results of the first part of the analysis, it can be said that there is roughly a relationship between the GSR signal and sleep efficiency. The second part of analysis (i.e., linear mixed models) did not find a clear relationship between these two since the relationship between GSR and sleep efficiency was non-significant (p-value >0.05).

Therefore, from the analysis made in both Python and R there wasn't found a clear relationship between sleep efficiency and GSR signal. However, some limitations present in this work should be considered.

Firstly, sleep efficiency was calculated per night, so the GSR features were summarized into one value for whole the night . Thus, in this way a lot of information is lost since it is impossible to capture the complexity of GSR changes during the night by one value. In future work, the dynamics of GSR during night so be considered in this analysis.

Secondly, it might be possible that available dataset (69 nights) is too small to accurately model this relationship.

Thirdly, in the machine learning phase, only classification models were considered, although the sleep efficiency was a continuous variable. The threshold, of 85%, for making this continuous variable a binary one was taken from literature [6] but was not validated in any way. In future work, the machine learning models should ideally be remade on a larger dataset (previous argument) while both considering sleep efficiency as continuous and binary variable. Ideally the threshold for binary transformation should be validated by executing a large literature review or by experimental validation.

Fourthly, only models between sleep efficiency and GSR were considered. However, the wristband also acquires data that might confound or influence this relationship, like temperature and motion [36].In future work, one should consider these confounders in the model in order to more accurately model this relationship and to account for these confounders explained variance of the model.
Lastly, there might be a small mistake in the used sleep efficiencies, since the algorithm was not validated with the golden standard, PSG.

6.3 Exploration of the relationship between GSR and skin temperature during both day and night

From the different analysis made in this context, it can be concluded that there is a significant relationship between skin temperature and GSR. However, the size effect of this relationship is very small, which would indicate that thermoregulation is not the only process causing GSR. It can be indirectly concluded that emotional sweating probably influences GSR but additional data is needed to confirm this hypothesis. As a future work, arousal data form the participants could be collected and used together with the skin temperature in order to investigate the effective size effect that these two have in the GSR signal. Also it should be considered that to investigate this relationship, the thermoregulation process was represented strictly by the skin temperature, which is a strong approximation, since a lot of physiological systems participate in the thermoregulation process (e.g., circadian rhythms). Therefore, other physiological signals could be included in the analysis to better represent the thermoregulation process (e.g., HR, body temperature, sleep stages).
Even though the size effect of the relationship between skin temperature and GSR signal was small, it could be observed a significant different interaction between this two during day and during night (i.,e wakefulness and sleep). In particular, a slightly stronger relationship could be observed during day with respect to night. A possible explanation of this, is the near complete inhibition of thermoregulatory process during REM sleep which composes 20-25% of the sleep, however, there is not enough evidence to conclude this due to the fact that we don't know the exact amount of this different interaction and also we don't have the sleep stages.
Also the relationship between GSR and a lagged temperature was explored but the slope between GSR mean and temperature mean did not become negative.
However, it needs to be said that only one direction of the relationship was considered (i.e., GSR and lagged temperature) and also a linear relationship was hypothesised. As a future work it can be exploring their relationship with a non-linear model and also considering both directions of the interaction.

List of abbreviations

EDA	Electrodermal activity
PCA	Principal component analysis
GSR	Galvanic skin response
PSG	Polysomnography
EEG	Electroencephalography
EMG	Electromyography
ECG	Electrocardiogram
EOG	Electrooculography
TST	Total sleep time
SE	Sleep efficiency
WASO	Wake time after sleep onset
ANS	Autonomic nervous system
HRV	Heart rate variability
VLF	Very low frequencies
LF	Low frequencies
HF	High frequencies
SCL	Skin conductance level
SCR	Skin conductance responses
ESM	Experience sampling method
SPT	Sleep period time
ACC	Accerelerometer
KNN	K-nearest neighbour
MLP	Multi-layer perceptron
STD	Standard deviation

List of figures

Figure 2. 1 An overview of the traditionally recorded signals during a PSG procedure [Source: Razjouyan et al. [10]] 4
Figure 2. 2: Example of a wrist-watch actigraphy sleep monitoring device and the data it acquires [Source: [72]] 6
Figure 2. 3 Standard ECG waveform and RR (inter-beat) interval [Source: Sattar et al [27]] 10
Figure 2. 4: EDA data decomposition into tonic and phasic components [Source: Boucsein et al [36]] 12

Figure 3. 1: Data of day 4 for user 2011 during the whole day (i.e., 24h). The first graph represents the 3 axis ACC data in (g), the second graph represents the GSR data in (uS), and the third graph represents the skin temperature in (°C). 166
Figure 3. 2: Results of the sleep detection algorithm. The blue line represents the output of the sleep detection algorithm. The violet graph represents the accelerometer data. The green/yellow lines represent the ESM data. 177
Figure 3. 3: Steps of the SPT- window detection algorithm 177
Figure 3. 7: Missing data (ACC,GSR and skin temperature) for more than 16h 200
Figure 3. 5: Missing data for a few hours during the day 200
Figure 3. 6: GSR mean data plotted for all available days for user 2011,2017,2021 and 2023 222

Figure 4. 1: Comparison of SPT window algorithm results and sleep diaries; the blue vertical lines represent the sleep diaries (sleep and wake time) while the red lines represent the SPT window algorithm results 29
Figure 4. 2: Bad results of the SPT window algorithm; the blue vertical lines represent the sleep diaries (sleep and wake time) while the red lines represent the SPT window algorithm results 31
Figure 4. 3: a) Results of the Random Forest sleep detection algorithm [57]; b)Results of IMEC's sleep detection algorithm; c) SPT window within which the validation process in performed 311
Figure 4. 4: Results of the missing data function. The red sections in the second graph represent the data that is considered missing by the missing data function 344
Figure 4. 5: Boxplots of the GSR features before and after standardization by means of standard scaler 366
Figure 4. 6: Histograms of each class for every GSR feature. 377
Figure 4. 7: Bivariate plot that represent the relationship between every couple of GSR features highlighting the class 377

Figure 4. 8: GSR feature distribution before and after Yeo-Johnson transformation. 39
Figure 4. 9: Confusion matrix of logistic regression model 39
Figure 4. 10: GSR mean data plotted for all available days for user 2011,2017,2021 and 2023(each row represents the GSR mean data for that user) 400
Figure 4. 11: GSR mean feature and quality indicator feature for each night 400
Figure 4. 12: GSR std and GSR mean plotted for user 2011,2017,2023,2031. Every row represents the GSR std and GSR mean for that user. 411
Figure 4. 13: Sleep efficiency and GSR mean boxplots for every user 433
Figure 4. 14: Boxplots of mean temperature during day and night for each subject..... 444
Figure 4. 15: Boxplots of GSR during day and night for each subject 455
Figure 5. 1: Bad results of the SPT window algorithm; the blue vertical lines represent the sleep diaries (sleep and wake time) while the red lines represent the SPT window algorithm results 522
Figure 5. 2: a) Results of the Random Forest sleep detection algorithm [57]; b) Results of IMEC'sleep detection algorithm; c) SPT window within which the validation process is performed. 533

List of tables

4.1 Results of IMEC's sleep detection algorithm for 10 days. The first and the third column represent the ESM data. The second and the forth column represent the outcomes of the IMEC's sleep detection algorithm while three last columns represent the sleep parameters comuputed by the algorithm 28

4.2 Results of the validation algorithm without performing any smoothing 30

4.3 Results of the validation algorithm with smoothing procedure 33

4.4 Pearson correlations between sleep efficiency and GSR features Results of the validation algorithm without performing any smoothing 34

4.5 Results (accuracy,recall, f1 score) of machine learning models on both test and train set 35

4.6 Decision tree results (accuracy,recall, f1 score) on both test and train set 38

4.7 Mixed model results for the model with mean temperature as a fixed effect 41

4.8 Results (accuracy,recall, f1 score) of machine learning models on both test and train set 42

4.9 Decision tree results (accuracy,recall, f1 score) on both test and train set 46

4.10 Mixed model results for the model with mean temperature as a fixed effect 47

Bibliography

[1] Institute of Medicine (US) Committee on Sleep Medicine and Research; Colten HR, Altevogt BM, editors. Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem. Washington (DC): National Academies Press (US); 2006. 2, Sleep Physiology

[2] Alhola P, Polo-Kantola P. Sleep deprivation: Impact on cognitive performance. Neuropsychiatr Dis Treat. 2007;3(5):553-567

[3] Dinges D, Rogers N, Baynard MD. Chronic sleep deprivation. In: Kryger MH, Roth T, Dement WC, editors. Principles and Practice of Sleep Medicine. 4th ed. Philadelphia: Elsevier/ Saunders; 2005. pp. 67–76

[4] Worley SL. The Extraordinary Importance of Sleep: The Detrimental Effects of Inadequate Sleep on Health and Public Safety Drive an Explosion of Sleep Research. P T. 2018;43(12):758-763.

[5] Minkel JD, Banks S, Htaik O, et al. Sleep deprivation and stressors: evidence for elevated negative affect in response to mild stressors when sleep deprived [published online February 6, 2012] Emotion. 12(5):1015–1020.

[6] Sathyanarayana A, Joty S, Fernandez-Luque L, et al. Sleep Quality Prediction From Wearable Data Using Deep Learning [published correction appears in http://mhealth.jmir.org/2016/4/e130/] [published correction appears in JMIR Mhealth Uhealth. 2016 Nov 25;4(4):e130]. JMIR Mhealth Uhealth. 2016;4(4):e125. Published 2016 Nov 4. doi:10.2196/mhealth.6562

[7] Medic G, Wille M, Hemels ME. Short- and long-term health consequences of sleep disruption. Nat Sci Sleep. 2017;9:151-161. Published 2017 May 19. doi:10.2147/NSS.S134864

[8] Johnson DA, Billings ME, Hale L. Environmental Determinants of Insufficient Sleep and Sleep Disorders: Implications for Population Health. Curr Epidemiol Rep. 2018;5(2):61-69. doi:10.1007/s40471-018-0139-y

[9] Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest. 2014 Nov;146(5):1387-1394. doi: 10.1378/chest.14-0970. PMID: 25367475

[10] Razjouyan J, Lee H, Parthasarathy S, Mohler J, Sharafkhaneh A, Najafi B. Improving sleep quality assessment using wearable sensors by including information from postural/sleep position changes and body acceleration: a comparison of chest-worn sensors, wrist actigraphy, and polysomnography. J Clin Sleep Med. 2017;13(11):1301–1310

[11] Blackwell T, Redline S, Ancoli-Israel S, et al. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. Sleep. 2008;31(2):283-291. doi:10.1093/sleep/31.2.283

[12] The role of sleep and sleep disorders in the development, diagnosis, and management of neurocognitive disorders Michelle A. Miller* Warwick Medical School, University of Warwick, Coventry, UK

[13] Automatic Human Sleep Stage Scoring Using Deep Neural Networks
Alexander Malafeev Dmitry Laptev4, Stefan Bauer4,5, Ximena Omlin2,6, Aleksandra Wierzbicka7, Adam Wichniak8, Wojciech Jernajczyk7, obert Riener2,3,6,9, Joachim Buhmann4 and Peter Achermann1,2,3*

[14] Sathyanarayana A, Joty S, Fernandez-Luque L, et al. Sleep Quality Prediction From Wearable Data Using Deep Learning [published correction appears in http://mhealth.jmir.org/2016/4/e130/] [published correction appears in JMIR Mhealth Uhealth. 2016 Nov 25;4(4):e130]. JMIR Mhealth Uhealth. 2016;4(4):e125. Published 2016 Nov 4. doi:10.2196/mhealth.6562

[15] Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography

[16] Miguel Marino, PhD, Yi Li, PhD, Michael N. Rueschman, MPH, J. W. Winkelman, MD, PhD, J. M. Ellenbogen, MMSc, MD, J. M. Solet, PhD, Hilary Dulin, BS, Lisa F. Berkman, PhD, Orfeu M. Buxton, PhD

[17] Shrivastava D, Jung S, Saadat M, Sirohi R, Crewson K. How to interpret the results of a sleep study. J Community Hosp Intern Med Perspect. 2014;4(5):24983. Published 2014 Nov 25. doi:10.3402/jchimp.v4.24983

[18] Polysomnographic Recording Technique, Sudhansu Chokroverty Sushanth Bhat C. Iber, S. Ancoli-Israel, A. Chesson, and S. F. Quan, "The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specification," 2007

[19] J. Allan Hobson, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," Electroencephalography and Clinical Neurophysiology, vol. 26, p. 644, June 1969.

[20] Chambon, Stanislas, et al. "A Deep Learning Architecture for Temporal Sleep Stage Classification Using Multivariate and Multimodal Time Series." IEEE Transactions On Neural Systems and Rehabilitation Engineering : a Publication of the IEEE Engineering in Medicine and Biology Society, vol. 26, no. 4, 2018, pp. 758-769.

[21] Sleep Disorders ,Rick D. Kellerman MD, in Conn's Current Therapy 2021, 2021

[22] Tilmanne J, Urbain J, Kothare MV, Wouwer AV, Kothare SV. Algorithms for sleep-wake identification using actigraphy: a comparative study and new results. J Sleep Res. 2009 Mar;18(1):85-98. doi: 10.1111/j.1365-2869.2008.00706.x. PMID: 19250177

.

[23] The role and validity of actigraphy in sleep medicine: An update Avi Sadeh* The Adler Center for Research in Child Development and Psychopathology, Department of Psychology, Tel Aviv University, Tel Aviv 69978, Israel

[24] Van Hees, V.T., Sabia, S., Jones, S.E. et al. Estimating sleep parameters using an accelerometer without sleep diary. Sci Rep 8, 12975 (2018). https://doi.org/10.1038/s41598-018-31266-z

[25] Sadeh A, Hauri PJ, Kripke DF, Lavie P. The Role of Actigraphy in the Evaluation of Sleep Disorders. Sleep 1995; 18(4):288-302

[26] Sadeh, A., Alster, J., Urbach, D., & Lavie, P. (1989). Actigraphy Based Automatic Bedtime Sleep-Wake Scoring: Validity and Clinical Applications. Journal of Ambulatory Monitoring, 2, 209-216.

[27] Sattar Y, Chhabra L. Electrocardiogram. 2021 Jul 31. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 31747210..

[28] Sebastiano, Massaro., Leandro, Pecchia. (2019). Heart Rate Variability (HRV) Analysis A Methodology for Organizational Neuroscience. Organizational Research Methods

[29] Shaffer F, Ginsberg JP. An Overview of Heart Rate Variability Metrics and Norms. Front Public Health. 2017 Sep 28;5:258.

[30] Tobaldini E, Nobili L, Strada S, Casali KR, Braghiroli A, Montano N. Heart rate variability in normal and pathological sleep. Front Physiol. 2013 Oct 16;4:294. doi: 10.3389/fphys.2013.00294. PMID: 24137133; PMCID: PMC3797399.

[31] Chouchou F, Desseilles M. Heart rate variability: a tool to explore the sleeping brain? Front Neurosci. 2014 Dec 11;8:402. doi: 10.3389/fnins.2014.00402. PMID: 25565936; PMCID: PMC4263095.

[32] Mendez M, Bianchi AM, Villantieri O, Cerutti S. Time-varying analysis of the heart rate variability during REM and non REM sleep stages. Conf Proc IEEE Eng Med Biol Soc. 2006;2006:3576-9. doi: 10.1109/IEMBS.2006.260067. PMID: 17946573.

[33] Takahashi, K., Kayama, Y., Lin, J. S., and Sakai, K. (2010). Locus coeruleus neuronal activity during the sleep-waking cycle in mice. Neuroscience 169, 1115–1126. doi: 10.1016/j.neuroscience.2010.06.003

[34] Rajendra Acharya U, Paul Joseph K, Kannathal N, Lim CM, Suri JS. Heart rate variability: a review. Med Biol Eng Comput. 2006 Dec;44(12):1031-51. doi: 10.1007/s11517-006-0119-0. Epub 2006 Nov 17. PMID: 17111118

[35] Busek P, Vanková J, Opavský J, Salinger J, Nevsímalová S. Spectral analysis of the heart rate variability in sleep. Physiol Res. 2005;54(4):369-76. Epub 2004 Dec 9. PMID: 15588154.

[36] Braithwaite, J.J., Watson, D.P., Jones, R.O., & Rowe, M.A. (2013). Guide for Analysing Electrodermal Activity & Skin Conductance Responses for Psychological Experiments. *CTIT technical reports series*.

[37] Critchley HD. Electrodermal responses: what happens in the brain. Neuroscientist. 2002 Apr;8(2):132-42. doi: 10.1177/107385840200800209. PMID: 11954558.

[38] Jaime Vila, in Encyclopedia of Applied Psychology, 2004

[39] G. Turpin, T. Grandfield, in Encyclopedia of Stress (Second Edition), 2007

[40] Arousal and Valence Recognition of Affective Sounds based on Electrodermal Activity Alberto Greco, Gaetano Valenza, Luca Citi, and Enzo Pasquale Scilingo

[41] Innovations in Electrodermal Activity Data Collection and Signal Processing: A Systematic Review

[42] Sarchiapone, M., Gramaglia, C., Iosue, M. et al. The association between electrodermal activity (EDA), depression and suicidal behaviour: A systematic review and narrative synthesis. BMC Psychiatry 18, 22 (2018). https://doi.org/10.1186/s12888-017-1551-

[43] Feehan LM, Geldman J, Sayre EC, et al. Accuracy of Fitbit Devices: Systematic Review and Narrative Syntheses of Quantitative Data. JMIR Mhealth Uhealth. 2018;6(8):e10527. Published 2018 Aug 9. doi:10.2196/10527

[44] Montgomery-Downs HE, Insana SP, Bond JA. Movement toward a novel activity monitoring device. Sleep Breath. 2012 Sep;16(3):913–7. doi: 10.1007/s11325-011-0585-y. [PubMed] [CrossRef] [Google Scholar] [Ref list]

[45] Haghayegh S, Khoshnevis S, Smolensky MH, Diller KR, Castriotta RJ. Accuracy of Wristband Fitbit Models in Assessing Sleep: Systematic Review and Meta-Analysis. J Med Internet Res. 2019 Nov 28;21(11):e16273. doi: 10.2196/16273. PMID: 31778122; PMCID: PMC6908975.

[46] Webster JB, Kripke DF, Messin S, Mullaney DJ, Wyborney G. An activity-based sleep monitor system for ambulatory use. Sleep. 1982;5(4):389-99. doi: 10.1093/sleep/5.4.389. PMID: 7163726.

[47] Cole RJ, Kripke DF, Gruen W, Mullaney DJ, Gillin JC. Automatic sleep/wake identification from wrist activity. Sleep. 1992 Oct;15(5):461-9. doi: 10.1093/sleep/15.5.461. PMID: 1455130.

[48] Van Hees VT, Sabia S, Anderson KN, Denton SJ, Oliver J, Catt M, Abell JG, Kivimäki M, Trenell MI, Singh-Manoux A. A Novel, Open Access Method to Assess Sleep Duration Using a Wrist-Worn Accelerometer. PLoS One. 2015 Nov 16;10(11):e0142533. doi: 10.1371/journal.pone.0142533. PMID: 26569414; PMCID: PMC4646630.

[49] Smagula SF, Stone KL, Fabio A, Cauley JA. Risk factors for sleep disturbances in older adults: evidence from prospective studies. Sleep Med Rev 2016;25:21–30. Crossref, Medline, Google Scholar

[50] Galecki, A. & Burzykowski, T. (2013). Linear Mixed-Effects Models Using R. Springer

[51] Gelman, A. & Hill, J. (2006). Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

[52] Rutter N . The dermis. Semin Neonatol 2000;5:297–302.

[53] Williams B, Cremaschi S. Novel Tool for Selecting Surrogate Modeling Techniques for Surface Approximation

[54] Sano, A., Picard, R. W., & Stickgold, R. (2014). Quantitative analysis of wrist electrodermal activity during sleep. International journal of psychophysiology : official journal of the International Organization of Psychophysiology, 94(3), 382–389. https://doi.org/10.1016/j.ijpsycho.2014.09.011

[55] Adam, K., Tomeny, M., Oswald, I., 1986. Physiological and psychological differences between good and poor sleepers. J. Psychiatr. Res. 20, 301–316

[56] T. Westeyn, P. Presti and T. Starner, "ActionGSR: A Combination Galvanic Skin Response-Accelerometer for Physiological Measurements in Active Environments," 2006 10th IEEE International Symposium on Wearable Computers, 2006, pp. 129-130, doi: 10.1109/ISWC.2006.286360.

[57] Sundararajan, K., Georgievska, S., te Lindert, B.H.W. et al. Sleep classification from wrist-worn accelerometer data using random forests. Sci Rep 11, 24 (2021). https://doi.org/10.1038/s41598-020-79217

[58] Liu JC, Verhulst S, Massar SA, Chee MW. Sleep deprived and sweating it out: the effects of total sleep deprivation on skin conductance reactivity to psychosocial stress. SLEEP 2015;38(1):155–159.

[59] Chahboun, S.; Maaroufi, M. Principal Component Analysis and Machine Learning Approaches for Photovoltaic Power Prediction: A Comparative Study. Appl. Sci. 2021, 11, 7943. https://doi.org/ 10.3390/app11177943

[60] [60] Dawson, M. E., Schell, A. M., Filion, D. L. (2007). The electrodermal system. In Cacioppo, J. T., Tassinary, L. G., Berntson, G. G. (Eds.), Handbook of psychophysiology (pp. 200–223). Cambridge, UK: Cambridge University Press

[61] Grice, E. A., Segre, J. A. (2011). The skin microbiome. Nature Reviews Microbiology, 9(4), 244–253. doi:10.1038/nrmicro2537

[62] https://www.media.mit.edu/galvactivator/faq.html

[63] Hastie, Trevor. Tibshirani, Robert. Friedman, Jerome. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York, NY, 2009.

[64] https://www.bcm.edu/news/how-stress-can-affect-your-sleep#:~:text=%E2%80%9CHigh%20levels%20of%20stress%20impair,disrupts%20sleep%2C%E2%80%9D%20Wilson%20explained.

[65] Valkenburg, A., Niehof, S., van Dijk, M. et al. Skin conductance peaks could result from changes in vital parameters unrelated to pain. Pediatr Res 71, 375–379 (2012). https://doi.org/10.1038/pr.2011.72

[66] Rutter N . The dermis. Semin Neonatol 2000;5:297–302.

[67] Kuno, Y. (1956). Human perspiration. Springfield, IL: Charles C Thomas

[68] Hugdahl, K. (1995). Psychophysiology: The mind body perspective. Cambridge, MA: Harvard University Press.

[69] The Body and the Brain: Measuring Skin Conductance Responses to Understand the Emotional Experience

[70] [Morgenthaler T, Alessi C, Friedman L, Owens J, Kapur V, Boehlecke B, Brown T, Chesson A Jr, Coleman J, Lee-Chiong T, Pancer J, Swick TJ; Standards of Practice Committee; American Academy of Sleep Medicine. Practice parameters for the use of actigraphy in the assessment of sleep and sleep disorders: an update for 2007. Sleep. 2007 Apr;30(4):519-29. doi: 10.1093/sleep/30.4.519. PMID: 17520797..

[71] Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, Dulin H, Berkman LF, Buxton OM. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep. 2013 Nov 1;36(11):1747-55. doi: 10.5665/sleep.3142. PMID: 24179309; PMCID: PMC3792393.

[72] https://sleeponthebay.ca/actigraphy-sleep-monitoring/

[73] [73]Martin JL, Hakim AD. Wrist actigraphy. Chest. 2011;139(6):1514-1527. doi:10.1378/chest.10-1872

[74] Desta Fekedulegn, Michael E Andrew, Mingming Shi, John M Violanti, Sarah Knox, Kim E Innes, Actigraphy-Based Assessment of Sleep Parameters, *Annals of Work Exposures and Health*, Volume 64, Issue 4, May 2020, Pages 350–367, https://doi.org/10.1093/annweh/wxaa007

[75] de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable Sleep Technology in Clinical and Research Settings. Med Sci Sports Exerc. 2019;51(7):1538-1557. doi:10.1249/MSS.0000000000001947

[76] W. Boucsein, Electrodermal activity, 2nd ed. Springer Science & Business Media, 2012.

[77] Bierman, William (1936-04-04). "The Temperature of the Skin Surface". Journal of the American Medical Association. 106 (14): 1158.

[78] Osilla EV, Marsidi JL, Sharma S. Physiology, Temperature Regulation. [Updated 2022 May 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK507838/

[79] Zhang, H; Arens, Edward A (2006). "The skin's role in human thermoregulation and comfort". EScholarship. Woodhead Publishing Ltd: 560–602.

[80] Ichiba, T., Suzuki, M., Aritake-Okada, S. et al. Periocular skin warming promotes body heat loss and sleep onset: a randomized placebo-controlled study. Sci Rep 10, 20325 (2020). https://doi.org/10.1038/s41598-020-77192-x

[81] Kräuchi, K. The thermophysiological cascade leading to sleep initiation in relation to phase of entrainment. Sleep Med. Rev. 11, 439–451 (2007)

[82] Adami A., Hayes T., Pavel M. Unobtrusive monitoring of sleep patterns; Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Cancun, Mexico. 17–21 September 2003; pp. 1360–1363. [Google Scholar] [Ref list]

[83] Nam Y, Kim Y, Lee J. Sleep Monitoring Based on a Tri-Axial Accelerometer and a Pressure Sensor. Sensors (Basel). 2016;16(5):750. Published 2016 May 23. doi:10.3390/s16050750