UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA
FACULTAD DE INGENIERÍA
MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA

Adaptive Model to Predict Sleep Quality from the Personalized Selection of Sleep Hygiene Factors. An Approach from the Machine Learning Algorithms.

T H E S I S

for the Degree of
DOCTOR IN SCIENCES

Presents:

ARTURO JESÚS LAFLOR HERNÁNDEZ

DIRECTOR OF THESIS:
DRA. MABEL VÁZQUEZ BRISEÑO.

Ensenada B.C. **julio de 2018**

UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA

FACULTAD DE INGENIERÍA
UNIDAD ENSENADA

Adaptive Model to Predict Sleep Quality from the Personalized Selection of Sleep Hygiene Factors. An Approach from Machine Learning Algorithms.

TESIS

Que para obtener el grado de doctor en ciencias presenta:

Arturo Jesús Laflor Hernández

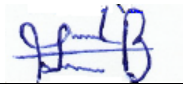
Aprobada por:

Mabel Vázquez Briseño
Directora de tesis

Juan Iván Nieto Hipólito
Miembro del comité

Everardo Gutiérrez López
Miembro del comité

Roberto Conte Galván
Miembro del comité



Armando García Berumen
Miembro del comité

Ensenada Baja California, México. Julio de 2018

Abstract of the Arturo Jess Laflor Hernández thesis, presented in partial fulfillment of the requirements for the degree of Doctor in Sciences. Ensenada, Baja California, México. Julio de 2018.

ADAPTIVE MODEL TO PREDICT SLEEP QUALITY FROM THE PERSONALIZED SELECTION OF SLEEP HYGIENE FACTORS.AN APPROACH FROM THE MACHINE LEARNING ALGORITHMS

Abstract approved by:

Dra. Mabel Vázquez Briseño
Principal advisor

Many social and behavioral phenomena difficult to study years ago, today find explanation when researchers model them through Machine Learning Algorithms. This is possible by the daily interaction between people and devises capturing and storing personal and environmental data. Capacity of memory, process and ability to interact with remote and powerful servers performing the heavy work enable the analysis of large amounts of data with velocity and precision. Phenomena referred above depends on many features and their relations which sometimes they are not linear. We present an adaptive model to predict the Sleep Quality from the critical Sleep Hygiene Factors (SHF) selection. Methodology involved a transactional study to identify the critical SHF in the study population, and a longitudinal study to validate the generalization of the selected SHF to particular cases. The no generalization of the SHF forces us to consider the complete set of the SHF as source data and to design into the predictive model a feature selection algorithm (fsXLR) to discriminate the non-relevant SHF. fsXLR extracts the critical factors finding most variance explanation through the implementation of the Bagging and Best-Search techniques to the results of the three well known algorithms XGBOOST, LASSO and RF.

The assessment we perform shows that the algorithm explains at least the 90% of the variance of the study data and identify critical factors classified in primary, secondary and tertiary. We observe than hypothetically in a real scenario, the minimum time required of monitoring to converge varies from one person to another. For instance, these studied cases required 20, 41 and 46 weeks.

Acronyms:

XGBOOST: *eXtreme Gradient Boosting*, **LASSO:** *Least Absolute Shrinkage and Selection Operator*, **RF:** *Random Forest*, **fsXLR:** *Feature Selection XBOOST, LASSO and RF*,

Dedication

This research is dedicated to...

Jesus my Creator and my Master who told me: "*seeks wisdom as silver, and search for her as for hid treasures? for the merchandise of it is better than the merchandise of silver, and the gain thereof than fine gold*" (Proverb 2:4,3:14 KJV [adapted]);

My dearest wife, who supports me in the practical and emotional, and encourages me to continue fighting for the improvement. Thank you for loving me, for giving me strength, for the sacrifice of living as students after 10 years of marriage and for devoting so much time to the care of our children;

My beloved kids, Bianca and Fer who inspire me to growth day after day. They are my motivation to seek the continuous development of my spiritual, physical and intellectual faculties;

My parents, who laid the foundations to love the truth, honesty, respect ... and taught me to see *The Education* as a worthy and a precious road;

My brother who encourages me to believe that I can achieve what I propose;

All my friends and partners who support me in various areas and circumstances.

LAHA

Acknowledgments

I would like to express my sincere gratitude to...

Universidad Linda Vista for giving me time and financial support.

Universidad Autónoma de Baja California and Consejo Nacional de Ciencia y Tecnología (CONACYT) for economic, technological and academic support.

My advisors for their contributions and guidance during the development of this research.

My teammates in the DAIH group for sharing the pleasure and enthusiasm of immersing ourselves in the fascinating world of the data analysis.

R Acknowledgements

I want to offer a special recognition to all people involved in the design and development of R [1], Rstudio [2] and all developers of R packages for their great contribution to this work. Without their support, this work would be impossible to finish in time.

Table 1: List of R packages used.

dplyr	[3]
e1071	[4]
FNN	[5]
foreach	[6]
FSelector	[7]
ggplot2	[8]
ggpubr	[9]
glmnet	[10]
gplots	[11]
grid	[12]
magrittr	[13]
Matrix	[14]
mice	[15]
neuralnet	[16]
onewaytests	[17]
outliers	[18]
party	[19, 20, 21]
plotly	[22]
pROC	[23]
randomForest	[24]
reshape	[25]
SDMTools	[26]
shiny	[27]
shinythemes	[28]
VIM	[29]
xgboost	[30]

List of Figures

List of Tables

List of Algorithms

Contents

Chapter 1

Problem Statement

1.1 Background and context

Social Networks (SN), Instant Messaging Systems (IMS), Internet of Things (IoT) and service provided by the cloud are concepts that society has naturally incorporated into their lifestyle. This has significantly changed the interaction of human beings with their environment. Digital Mobile Entities (DME) such as smart phones, tablets and smart watches equipped with various sensors, enrich the person environment by expanding it to remote scenarios and people. Human beings and DME maintain a symbiotic interaction promoting the develop of applications and technological platforms that influence personal, corporate and public decision-making [38].

These days we find in DME allies helping us to store moments, remembering appointments, interpreting messages from any language, understanding new concepts, moving efficiently and safely through unknown cities, studying courses in idle time spaces, and the list could be easily extended thinking in personal services. To [39] this is a new era of ubiquity in which human beings become a minority generation and reception of meaningful data. DME consume and interpret most information to offer better alternatives in logistics, automotive and air traffic control, disaster prevention, flora and fauna monitoring, climate monitoring, among other fields of global importance.

In addition to the aforementioned, those related to health are among the most important

and sensitive areas. Organizations such as The Organization for Economic Co-operation and Development (OECD) and The World Health Organization (WHO) have published not encouraging statistics on health and their projections for the well being at the immediate future are worrying [40, 41]. Science advances in the search for more sophisticated treatments to face these challenges, while public and private health institutions promote changes in lifestyle worldwide as a means of reversing the trend. Mobile technology has founded in the lifestyle approach an important growth niche. Of the thousands of dozens of applications designed for the health area, the minority has purposes of clinical scope, most is occupied by those applications which goal is to intervene in the health of users from the lifestyle perspective [42]. For instance, researchers have been proposed projects monitoring and motivating healthy eating habits [43, 44, 45, 46]; applications monitoring and promoting physical activity [47, 48, 49]; and systems monitoring sleep parameters, and promoting SQ [50, 51, 52].

We worked in this study aiming to contribute to the health area from the data analysis perspective using Machine Learning Algorithms (MLA). We explored sleep by hypothesizing that lifestyle influence the SQ which has been associated with the most important dimensions of well being. For example, [53, 54, 55, 56, 57, 58, 59, 60] have associated poor sleep to cardiovascular and degenerative diseases that mainly attack the physical dimension; [61] associates poor sleep to intellectual dimension because it may causes concentration lacking and impaired memory; likewise the repercussion in social and economic field can be found in studies that associate poor sleep quality to changes in mood and depression [62]; automobile and industrial accidents and low work and academic performance are other consequences attributed to poor sleep quality [63, 64, 65, 66].

We began the project motivated and sustained in two aspects mainly: 1) Our worry due to the current statistical data on health and the uninspiring trends that OECD and WHO project for the coming years; 2) Our interest to promote the lifestyle as a means of improving health of people. We support the second motivation on researchers showing that not only genes intervene in the genetics of organisms. Also personal experiences can generate marks in genetic material and these

marks can be transmitted to future generations. This takes more relevance because researchers have found evidence that epigenetic mechanisms intervening in physiological and pathological processes linked with various degenerative diseases [67, 68, 69]. This is slowly leading us to the paradigm of holistic medicine to the detriment of reductionist medicine supported by biological paradigm in a dualistic world view of the human being.

Physicians have studied SQ through qualitative and quantitative constructs among which stand out clinically, The Structured Clinical Interview for the Statistical Diagnosis of Mental Disorders (SCID) [70] and, the Pittsburgh Sleep Quality Index (PSQI) [71], the most used construct by researchers to measure SQ in various clinical and non-clinical studies [72]. In the clinical area the specialists use these constructs along with tools measuring objective parameters of SQ such as latency, efficiency and duration; and physiological parameters such as melatonin and adenosine levels; and heart and breath rates among others indicators.

Currently there are two technological tools approved clinically for the monitoring and diagnosis of sleep disorders. PSG, the oldest and most reliable laboratory test, and recently, the actigraphy that has been approved by organisms such as The American Academy of Sleep Medicine (AASM) and The American Sleep Disorders Association (ASDA) especially for sleep monitoring in populations where PSG is not ideal (for example, children and elderly) or in diagnosis for sleep disorders requiring a long time follow-up [73, 74].

Today, researchers explore new methods to address the issue by exploiting small devices, ubiquitous environments and Wireless Sensor Networks (WSN). Technology designers have proposed mobile applications and systems combining sensors adapted to the environment with software installed on personal devices (computers, tablets, smart phones and wearable electronic devices). The approaches include systems for monitoring sleep and lifestyle parameters; systems to diagnose sleep disorders such as insomnia and Obstructive Sleep Apnea (OSA); and persuasive systems that aim to motivate users to adopt behaviors favoring the SQ [75, 76, 77].

Nathaniel Kleitman in 1939 conceptualized lifestyle habits and environmental conditions associated to sleep in a term he called SH [78]. Later in 1977 Peter Hauri takes up the concept

and proposed specific rules to promote sleep supported in SH [64]. We found in the literature that these behaviors and environmental parameters are around of 20; in this work we called SHF when referred to, both individually and in groups.

Since then, several studies have been carried out to find a association between SQ and the whole set of SHF, SQ and subsets of SHF or SQ and a single SHF.There are divergent positions among researchers on treating sleep problems through SH since the results of various investigations are not conclusive. For example [79] and [80] conclude that SH does not contribute significantly to the treatment of insomnia, while [81, 82, 83] find an association between SH and SQ. We are astonished that some studies have found no association between SQ and SH while other works studying association between a single SHF and SQ have found evidence favoring their hypothesis. (See Section 2.1.2). This divergence allows us to continue doing more research to find evidence for or against that SQ can be understood to some extent by SH.

In this work we hypothesize that it is possible to explain the complex variable of Sleep Quality in health people by observing the Sleep Hygiene Factors. In this sense, it is feasible we develop a system threefold purpose: 1) to monitor the SHF of a Person Interested in Improving their Lifestyle (PIMEV); 2) to estimate their SQ trough an inference model; and, 3) to send recommendation messages aimed to influence in the SH of the PIMEV. Figure 1.1 conceptually represents such a system. PIMEV and system exchange messages through a DME with the purpose of achieving alignment on SH and SQ. The success of the interaction is that the system realizes a SHF-SQ association and the PIMEV recognizes the influence of SQ in their health. Reciprocal feedback between these two entities serves to achieve convergence over the time.

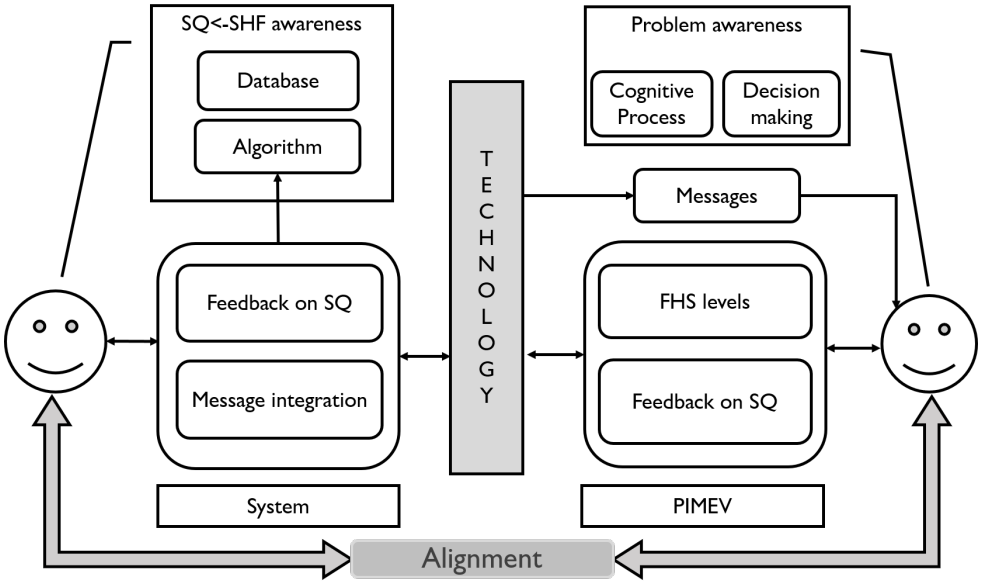


Figure 1.1: Conceptual model of the system. The system is able to process the data coming from the user through the inference model and emits messages supporting the PIMEV in the decision making process. The PIMEV is able to interpret the messages and act accordingly.

1.2 Objectives

Design an adaptive model to predict the quality of sleep of a PIMEV from the personalized selection of a set of his or her most influential sleep hygiene factors.

In order to fulfill the general objective, the following specific objectives were proposed:

1. To define and describe the variables SQ and SH theoretically in the context of this study.
2. To perform a systematic review on technology for sleep and to identify the evaluation methods.
3. To identify the SHF that characterize the SP through a cross-sectional study.
4. To select an MLA by comparing their feasibility of implementation and their efficiency in accuracy and time to predict SQ, using SHF as predictive features.

5. To validate the results of the cross-sectional study in particular cases through a longitudinal study.

6. To design an algorithm to build personal models to estimate SQ by selecting a subset of SHF through the analysis of the personal data of study subjects.

7. To evaluate the prototype through the efficiency in predict SQ.

1.3 Methodology

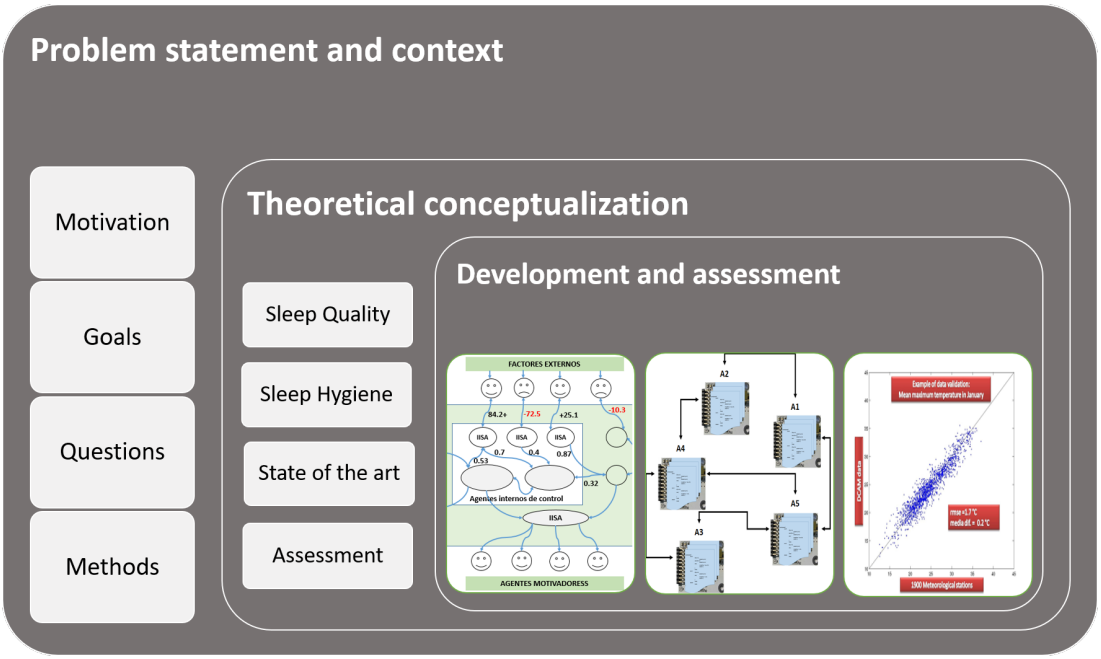


Figure 1.2: Problem statement and context.

In order to meet the specific objectives, we planned the project in three stages. Each one raised research questions to guide their development.

First stage: We planned this stage covering the first two objectives in the intention of understand theoretically SH and SQ. We identify the relationship between variables involved and supporting the main hypothesis that it is feasible to find an association between SHF

and SQ. The question we associated with this objective asks: is it feasible to investigate if SH influence to some extent on SQ? Likewise, we studied technology proposals designed to address some aspects related to the quality of sleep and how researchers assess it. The questions we associated are: how has technology evolved to study sleep and its disorders? how should the emerging technology related to SQ be evaluated?

Second stage: In this stage we carried out a cross-sectional study to discover if there are a set of SHF characterizing the SH of the SP and influencing their SQ. Moreover, we use the collected data to evaluate eight MLA and chose the one producing the more efficient models. To guide the process, we posed the following questions: is there a set of factors characterizing the SP so that we can train more efficient models than if we train them using the complete set of SHF? Which MLA is the most appropriate algorithm to model this problem in terms of prediction accuracy and processing time?

Third stage: We performed this stage aiming to test if the results obtained through the cross-sectional study generalize to a data form personalized monitoring. The question we raised to cover this purpose is: to what extend do the SHF obtained from the SP characterizes the SQ of a PIMEV observed individually? The second purposed we posed is to design an adaptive model to personalize the selection of SHF and predict SQ from a personal monitoring. The questions we proposed to cover this purpose are: is it possible to identify a subset of SHF producing personalized models to predict SQ by reaching at least 90% of efficiency? If possible, how much observation time is it necessary?

1.3.1 Strategies to answer the research questions

- Is it feasible to investigate if SH has any influence on SQ? How technology designs to study SQ has evolved? How should the emerging technology related to the quality of sleep be evaluated? We answered these three questions through a theoretical study of SQ, SH and the evolution of technological development associated with this problem.

- Is there a set of factors that characterize SP and allow us to train more efficient models than if we train them using the complete set of SHF? We conducted a cross-sectional study in the SP as described in Section 3.1. We designed and implemented a robust Feature Selection (FS) algorithm to identify the most influential set of SHF in the SP. We validated the results by comparing the prediction efficiency of the models we trained using selected SHF versus those models we trained using the complete set of SHF (See Sections 4.1.1, 4.2 and 4.2.1).

- Which MLA has the best performance in prediction accuracy and processing time to model this problem? We answered this question by comparing accuracy in prediction and execution time for each variant of MLA using the selected SHF in the previous stage as data source of training.

- To what extent do the SHF obtained from the SP characterize the SQ of a PIMEV observed individually? To answer this question we carried out a longitudinal study. Three volunteers measured their SQ for 30 days observing their physiological parameters and sleep patterns through an electronic device. The same number of days the volunteers captured their SH (21 SHF). After, we extrapolated extrapolated the data until we reach enough records to train models for each subject (see 5.3).

- Is it possible to identify the best subset of SHF to predict SQ using a custom model? We design a process to generate customized models that would allow us to estimate SQ individually. The prototype included the design of the algorithm Feature Selection exploiting XGBOOST, LASSO and Random Forest (fsXLR) (see 3) to identify the most relevant SHF in each subject and choose a subset of those factors that would explain a high percentage of the variance of the data (See 5.4).

- How much observation time is necessary? We perform tests using the extrapolated data as giving information. The sections 5.5 and 5.6 show the process in detail.

1.4 Limitations

The longitudinal study was limited to three volunteers due to the available devices to measure SQ and the time required to realize the observation. Also, we use extrapolated data to train models instead of observed data which implies two assumptions: 1) the volunteers have consistent SH habits; 2) The device that measures the quality of sleep will give similar results in the same person for similar patterns in SH habits.

1.5 Contributions

Despite the evidence of the impact of poor sleep quality on health its study has not had the same relevance as other areas in biological medicine. Credits assigned by academics to sleep in the curricula of medicine courses in various universities around the world is no significant [84]. However, this has been changing recently. As we describe earlier, many studies associate SQ with various conditions affecting human well being. Moreover, [85] and [86] published statistics in the E.U. and Mexico showing us that the time devoted to the sleep is below seven hours. This suggest a generalized deprivation of this cellular restoration state, since the consensus of the experts after many studies matches an average of eight hours for adults. (see 2.1).

Furthermore, we found studies showing evidence from various latitudes that insomnia is more common that we think. For example, [87] cited a research on adults from 10 countries founding that 31.6% had insomnia and an additional 17.5% had symptoms of pre-insomnia. Another study reported in the same work suggests that 56% of Americans, 31% of Western Europeans and 29% of Japanese suffer from sleep problems and most do not have a clinical diagnosis to help them manage their situation.

We contribute through this project from the data science exploring the benefits that the lifestyle can generate to the SQ and consequently to the health. We posed our contribution from the various involved areas in this study.

Social area: Depression and anxiety, two illness having a negative impact on the social inter-

actions could have their causes in a poor SQ [88, 89]. Likewise, people having a poor SQ cause various work and traffic accidents [64]. We proposed in this work, the develop of a technology tool to help people enrolling in habits that may improve their SQ.

Scope of public health: Diverse studies have associated with the sleep deprivation and the poor quality of sleep, the deterioration of intellectual abilities and brain-related diseases such as Alzheimer's and Parkinson [54, 55]. Also, there are studies associating the poor SQ with increased risk of Type II diabetes [56, 57], cancer [59, 60] and cardiovascular diseases [53]. We develop a prototype web page for training models to predict SQ using as data input the SHF from a personal monitoring. We show through the SHF selection and the models training by our web page that the variance of the data of the volunteers participating in the study can be explained by a set of SHF influencing their SQ. Then, we showed that it is possible to design a system to send messages to the users favoring the awareness on to improve their SH.

Scientific-Technological Scope: 1) we searched the association between SH and SQ by analyzing data from a Likert survey through MLA while most studies use statistical techniques. 2) we proposed and evaluated an adaptive model to predict SQ from the personalized selection of SHF aiming to send specific and personal messages to the users. Most proposes send messages to give general recommendations to the users while others focus on a specific SHF. 3) we develop a recursive algorithm to optimize the FS process by integrating Bagging(expand/merge) and Best-search(forward/backward) techniques to combine three FS algorithms for continuous variables (eXtreme Gradient Boosting (XGBOOST), Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF)). (see Chapter 5, Section 5.4).

Chapter 2

Sleep (Quality + Hygiene + Technology)

2.1 Sleep

The Oxford English Dictionary defines Sleep as "*a condition of body and mind which typically recurs for several hours every night, in which nervous system is inactive, the eyes closed, the postural muscles relaxed, and consciousness practically suspended.*" Physicians divided the sleep in three stages repeated cyclically in a normal night. The first cycle starts with a short time lapse called sleep latency occurring from the wakeful state until the appearance of the first sleep stage.

The three stages fulfilling important physical and intellectual functions of normal sleep cycle are Light Sleep (LS), Deep Sleep (SWS) and Rapid Eye Movement Sleep (REM). For example, [65, 66] shows evidence of the participation of the three stages in the memory consolidation. Advances in sleep medicine have found that each stage contributes individually in the restoration process for the well being. For instance, studies have shown that most Growth Hormone (GH) production occurs during the SWS stage, being GH the main substance in the restoration process [90]. When SWS stage is repeated normally in the best number of cycles, the individual wake up rested and energized to carry out their daily activities. Contrary, malfunctioning of process

in any of three stages can affect various aspects. For example, REM sleep disorders can lead to other cognitive problems such as lack of concentration, attention difficulties and mood changes.

In people who sleep well, REM and SWS do not connect with each other, LS serves as a transition stage between them. Also, LS is the transition stage between the person asleep and awake in quality sleep cycles. People who do not complete the cycle of Fig. 2.1 may suffer from a sleep disorder and they must be clinically attended.

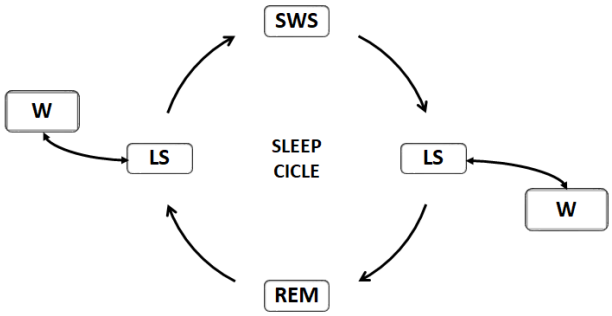


Figure 2.1: Sleep cycle stages and transitions in normal sleep.

In healthy people with good sleep patterns, the number of sleep cycles in a night and the time they take vary depending mainly of the age. In an expert consensus after conducting a systematic review including 321 studies, The National Sleep Foundation (NSF) published in 2015 the Table 2.1 containing the recommendations of sleep duration by age. For example, adults with healthy sleep repeat a sleep cycle of around 90 minutes four to six times. However, the sleep duration is only a dimension of sleep health as we will see later.

Consistent deprivation of sleep time and violation of natural rules in any of its dimensions may cause several health problems in the short and long term. Various investigations have found the connection between sleep and cell recovery which had derived in causal studies finding the influence of sleep on physical health. As the Merriam Webster dictionary says in its sleep definition: "*the power of the body are restored in this natural state of unconsciousness*". [31] proposes a model showing a reciprocal interaction between the different dimensions of sleep and health. The model suggests that what happens in any of the dimensions of sleep has influence in genetic, epigenetic, molecular and cellular processes; which in turn results in modifications

Table 2.1: Sleep duration recommendations in hours by expert panel [33].

Age	**Recommended**	**May be appropriate**	**Not recommended**
Newborns	14 to 17	11 to 13	Less than 11
0-3 mo		18 to 19	More than 19
Infants	12 to 15	10 to 11	Less than 10
4-11 mo		16 to 18	More than 18
Toddlers	11 to 14	9 to 10	Less than 9
1-2 y		15 to 16	More than 16
Preschoolers	10 to 13	8 to 9	Less than 8
3-5 y		14	More than 14
School-aged children	9 to 11	7 to 8	Less than 7
6-13 y		12	More than 12
Teenagers	8 to 10	7	Less than 7
14-17 y		11	More than 11
Young adults	7 to 9	6	Less than 6
18-25 y		10 to 11	More than 11
Adults	7 to 9	6	Less than 6
26-64 y		10	More than 10
Older adults	7 to 8	5 to 6	Less than 5
$\geq$ 65 y		9	More than 9

to the process occurring at the level of organism's systems, which it ends up manifesting itself as a disease, health or alteration in functions of the person. Likewise, as the model in Fig. 2.2 shows, the changes occurring in any of the three blocks on the right can affect any of the sleep dimensions of block one.

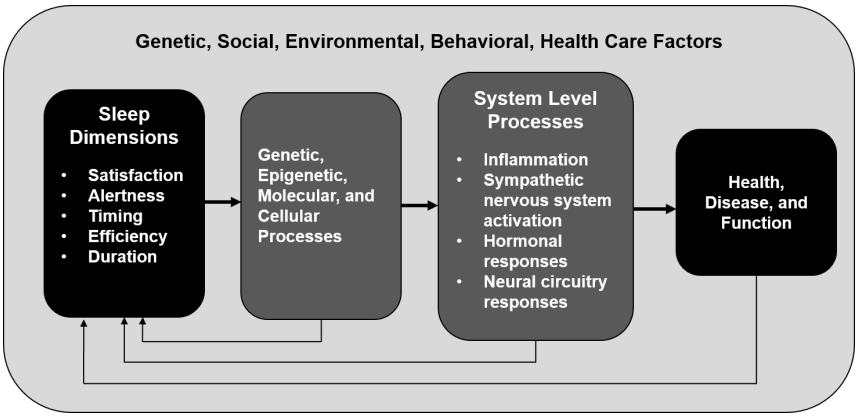


Figure 2.2: Conceptual model of sleep health [31].

We summarize in Table 2.2 some correlational studies cited by [31] associating five indicators of sleep health with diseases affecting physical, emotional, cognitive and social components of human well-being[1].

Table 2.2: Examples of correlational studies on sleep-measures and diseases.

	Quality	Alertness Sleepiness Napping	Timing	Efficiency	Duration
Mortality	✓	✓	✓	✓	✓
Metabolic Syndrome	✓		✓	✓	✓
Obesity					✓
Diabetes/impaired glucose metabolism	✓		✓	✓	✓
Hypertension	✓			✓	✓
Coronary heart disease	✓	✓	✓	✓	✓
Depression	✓			✓	
Impaired neurobehavioral performance		✓			✓
Accidents			✓		

[1]Dimensions of Well-being according to Harvard Framework of Wellbeing: Physical, relational, emotional, financial, spiritual, environmental, intellectual and vocational. (https://wellness.huhs.harvard.edu/)

Other studies show a connection between objective and subjective variables of sleep and components of the people performance in the wakeful state, for example, [91] reports a link between sleep and fatigue, and, sleepiness and concentration in two populations (adults and young students). Objective parameters include (*'total sleep time'*, *'total wake time'* and *'sleep efficiency'*); while subjective features are: Psychological laden sleep variables (*'nocturnal tension'*, *'distress about sleep problems'*); and psychological adjustment (*'Depression'* and *'anxiety'*). Readers can see the complete results in the article, we show in Table 2.3 only positive and significant correlation ($correlation > 0.4, p < 0.05$).

Table 2.3: Results of objective and subjective sleep variables vs fatigue, sleepiness and concentration.

	Older adults			**Students**		
	F	**S**	**C**	**F**	**S**	**C**
Total wake time	.43		.41			
Nocturnal tension						.45
Distress about sleep problem	.56		.51	.41		.47
Depression		.48				
Anxity		.47			.54	.54

F: Fatigue. S: Sleepiness. C: Concentration.

Besides the correlational studies, there are works finding relation of cause-effect between SQ and some diseases, being SQ the independent variable. For example: Poor SQ can cause or contribute to the development of cardiovascular diseases [53], Alzheimer [54], Parkinson [55], Obesity [56, 57] and Cancer [59, 60]. We explain next briefly the rationale supporting those findings:

Alzheimer: The glymphatic system helps drain waste from toxins and other substances that accumulate during the waking period in the brain. One of these substances is the $\beta - amyloid$ protein, founded in abnormal quantities in the brain of Alzheimer's patients.

Cardiovascular diseases: A person suffering OSA has no an adequate breathing during the night, in fact, they stop breathing for periods of time that vary depending on the severity

of the illness. When oxygen stops flowing properly, the heart rhythm is affected and the blood does not flow normally. Stagnant blood can produce clots that in turn have the ability to cause heart attacks and pulmonary embolisms among other problems.

Obesity: Among the functions that the brain performs during sleep, is the regulation of two hormones closely linked to eating habits. Ghrelin the hormone responsible for appetite and leptine, the saciety hormone. Studies have shown that as hours of sleep decrease, ghrelin increases and leptin decreases. This hormonal disorder causes in the people a large appetite and a lower saciety. The combination normally will result in weight gain, which can lead to obesity, being this condition a prelude of type II diabetes.

Buysse after much research makes it clear that sleep is a major component of human well being and he defines Sleep Health as...

> a multidimensional pattern of sleep-wake-fullness, adapted to individual, social, and environmental demands, that promotes physical and mental well-being. Good sleep health is characterized by subjective satisfaction, appropriate timing, adequate duration, high efficiency, and sustained alertness during waking hours [31].

Due to the existing evidence on the health of sleep and its influence on the health, this work takes the concept as a dependent variable of lifestyle and hypothesizes about the possibility that in healthy people "*SH is a predictor of the sleep health*". Because we did not found a validated scale for measuring sleep health, we decided to work with the sleep quality concept. From our position, sleep quality is the concept closest to sleep health because it includes most of the components of its dimensions. We put together the SQ and SH concepts in a single questionnaire to obtain the data for the analysis. We describe the scales we used and the process we performed for the data collecting in Section 3.1.

2.1.1 Sleep Quality

Sleep quality is a complex concept to define and measure because it involves a large number of variables. There are different proposals to measure the quality of sleep. From those that base the measurement on physiological parameters as does PSG, as well as those that propose to measure it by means of questionnaires measuring subjective drowsiness in wakefulness state. Each proposal has advantages and disadvantages, and physicians often use two or three methods together to complement the studies.

For example, PSG efficiently measures physiological parameters to estimate SQ, however, these measurements could not be enough to understand the issue since the specialist performed the test in a single night. Thus, the physicians often complement the results, surveying the people through qualitative constructs asking for their sleep parameters behavior in a longer time. These data together, empower the specialist to give a better diagnosis. On the other hand, by means of the questionnaires the physicians can identify if apart from physical implications, the patients suffer at a psychosomatic level.

In this work, we measured sleep quality in two ways that were enough to achieve the objectives we proposed. To measure sleep quality individually in the study subjects, we used a device that makes an estimation of the variable by reading physiological parameters (heart rate and respiratory rate), as well as objective data including the duration of sleep, the latency, the number of movements and the times that subjects woke up during the night. To measure the quality of sleep in the SP, we used the PSQI questionnaire which characterizes the quality of sleep in seven dimensions: duration, efficiency, disturbances, latency, diurnal dysfunctions due to drowsiness, use of sleep medications and a self-reported of sleep quality perception (for a detailed description of the questionnaire see Section 3.2.1).

2.1.2 Sleep Hygiene

Sleep Hygiene is a concept introduced by Nathaniel Kleitman in 1939 and taken up in 1977 by Peter Hauri who proposed rules to promote sleep [64]. Various studies have been conducted around

this concept in recent years and experts in sleep have referred it in different ways preserving the essence of the original idea. We show some of them in the following paragraphs:

- SH refers to those behaviors that are believed to promote improved quantity and quality of sleep [92].
- SH covers basic advice on how the patient should behave in order to sleep well (such as reducing caffeine use before bedtime, keeping the bedroom dark and quiet, avoiding alcohol as a sleep aid, advice concerning food intake before bedtime, among others [80].
- SH is a collection of behaviors and environmental conditions that aim to ensure a restorative and good quality sleep and to avoid or to treat certain sleep disorders[79].
- SH is the term sleep experts use to refer to the practices that are believed to promote good sleep quality [85].
- SH is defined as a set of behavioral and environmental recommendations intended to promote healthy sleep and was originally developed for use in the treatment of mild to moderate insomnia [87].
- SH refers to the general rules of behavioral practices and environmental factors that are consistent with good quality sleep [93].
- SH Education is intended to provide information about lifestyle (diet, exercise, substance use) and environmental factors (light, noise, temperature) that may interfere with or promote better sleep. Sleep hygiene also may include general sleep facilitating recommendations, such as allowing enough time to relax before bedtime, and information about the benefits of maintaining a regular sleep schedule *P.J. Hauri cited by* [78].

Based in the definitions above and the literature reviewed, we understand and use for the concept of Sleep Hygiene the following definition: "*Sleep Hygiene is a set of behaviors and environmental conditions that physicians believe people should practice for improving their SQ*

and be less vulnerable to suffer a sleep disorder." Next, we explain the reason to associate the SHF with the SQ.

Nap: Recommendation: Do not nap, if it cannot avoid it, it should not last more than 30 minutes. Rationale: Naps can cause sleep problems because they influence the mechanisms controlling circadian rhythms that determine sleep time, duration and depth of sleep [78]. Studies have found association between poor sleepers and the habit of napping [94].

Sleep time/Wake up time: Recommendation: Have regular times to sleep and wake up as well as getting up immediately after awakening. Rationale: The idea behind having regular hours to sleep and wake up is to maximize the synchronization of circadian rhythms, the physiological impulse of sleep and the very act of sleeping. The studies have shown that sleep delay and irregularities in the schedule to go to bed and wake up, causes lower duration and less continuity of the sleep, and, desynchronization in other endogenous circadian rhythms [87, 64].

Exercise: Recommendation: Do aerobic exercise in the mornings and avoid intense exercise before bedtime. Rationale: The researchers have found that exercising helps people to avoid awaking by disturbing agents such as noise and climate. Additionally, people who exercising achieve a more efficient cellular restoration. [78]. A study published in the American Journal of Physiology in the 2015, searched association between exercising time and physiological parameters of SQ such as body temperature, melatonin levels and heart rate. The researchers observed that a group of subjects exercising in the afternoon-night maintained body temperatures higher compared with other group exercising in the morning. Likewise, the heart rates of the people in the first group remained above normal for several hours after exercising and their melatonin concentration levels were lower than usual. One conclusion of the study is that moderate daily exercise produces differential effects in the circadian rhythm of melatonin, nocturnal body temperature and heart rate, depending on the time that it is done. This may is due to the morning exercise activate the parasympathetic

system while night exercise increases sympathetic activity during sleep hours [95].

Tobacco Recommendation: Avoid nicotine within previous hours till bedtime. Rationale: Although the researchers have no clear how nicotine acts in sleep-wake processes, they have found it does. [96] conducted a study 1994 and reported that nicotine decreased the sleep time by 33 min, sleep efficiency from 89.7 to 83.5%, percent rapid eye movement sleep from 18.8 to 15.1%, and prolonged sleep latency from 6.7 to 18.2 min. Beside to studies showing improved sleep in people who quit smoking, there are studies demonstrating that nicotine ingestion causes sleep fragmentation; likewise physicians knows that people using patches in smoking cessation therapies are prone to suffer from insomnia [92, 97].

Alcohol: Recommendation: Avoid alcohol within four hours before bedtime. Rationale: Often, people fall asleep more quickly after drinking alcohol, nevertheless, researchers have found that people who drink alcohol before to go to the bed are prone to have a more fragmented sleep, waking up more times during the night and waking up earlier in the morning. Furthermore, studies have shown the suppressive effect of alcohol on REM sleep [78, 92].

Caffeine: Recommendation: Avoid caffeine products within the previous six hours of bedtime. Rationale: Adenosine is an endogenous substance capable of exercising regulatory actions on the circuits of the sleep-wake cycle. Adenosine builds up during prolonged wakefulness and decreases during sleep. Caffeine promotes the waking state because it has an antagonistic role in blocking adenosine receptors in the central nervous system, which modifies the amount and intensity of sleep [92, 98, 99].

Stress: Recommendation: Do not go to sleep if you feel stressed, anxious or nervous. Any additional casual events of stress, intellectual activities like making plans or schedules and worrying may cause stress in some people. Rationale: Stress is an adaptive mechanism of the organism to respond appropriately to stimuli from the environment. When the organism activated this mechanism, various chemical substances are secreted to alter and

suppress processes until the person has attended the emergency. One of the responses to these chemical changes produced by stress is the activation of the Hypothalamo-Pituitary-Adrenal (HPA) axis and the adrenomedular sympathetic system. Both with important influence on the regulation of circadian sleep-wake cycle. Moreover, studies have found a connection between the magnitude of repercussions on sleep circadian rhythm and the time at which the stressful event occurs. Over time, if the stress events are prolonged or uncontrollable, mechanisms for recovering of the homeostasis fails, and the stress may causes different pathological states including sleep disorders and mood[90].

Arousal activities before sleep time: Recommendation: Do not do activities stimulating your wakefulness system. This includes: play video games, Internet use, paying bills, study and doing mental works. Rationale: Performing activities like these before going to sleep, can activate the alertness state and cause secretions of substances that physiologically interfere with the ideal state to sleep. Some of these activities are associated with stress states, and others involve interaction with devices emitting light (especially blue), and others imply both. (See Stress and Light exposure to read about the physiological changes that researchers linked with these activities)

Bed use: Recommendation: Use bed only for sleep and sexual activities. Rationale: To have a good night's sleep people should to avoid any stimuli in the immediate environment. They must associate the bedroom and bed only with sleeping to favor internally controlling their sleeping and waking hours [100].

Uncomfortable bed: Recommendation: Sleep in a comfortable bed. Rationale: Poor quality pillows and mattresses, old or unsatisfactory for individual preferences, can cause pain or discomfort, which makes them a sleep disturber [78].

Light exposure: Recommendation: Avoids any light exposure during sleep; if impossible, use a sleeping mask. The hours before sleep are also important; after sunset, the light should be as dim as possible to promote sleep. You should turn off all electronic devices; or at

least, sleep specialist recommended installing filters that block the blue light. Rationale: Light is a stimulant of brain activity that alters the timing of the circadian clock's. Light physiologically activates alertness, which can cause delays in sleeping time and poor sleep quality. Furthermore, light captured by human eye photo-receptors suppresses producing melatonin, thereby inactivating the biochemical signal of darkness in the brain [101, 102, 103].

Noisy: Recommendation: Sleep in an environment as quiet as you can, if eliminating environmental noise is out of your hands, use earplugs. [78]. Rationale: Irregular noises 40 decibels onward can be disruptive to sleep according with the National Sleep Foundation [76].

Environmental temperature: Recommendation: Sleep in a cool environment or sleep as cold as you feel comfortable. Rationale: The homeostatic system lowers the body temperature during the sleep; a high environment temperature can interfere with the internal processes happening during sleep [95]. The National Sleep Foundation suggests sleeping in a temperature range of $12 - 24°C$ [76].

Dinner: Recommendation: Eat a light snack before bedtime, avoiding heavy foods. Rationale: The sleep is the space that the organism uses for the cellular recovery, the main objective is restoring and cleaning the body systems. Heavy dinners will make that digestion (one of the most demanding energy tasks) preventing the best cellular restoration. On the other hand, light Snacks based in carbohydrate raises the glycemic index and elevate insulin causing a state of drowsiness that favors the sleep if they are ingested about four hours before going to bed. Nuts and other seeds ingested at evening also favor sleep due to their melatonin content [104, 78].

There is insufficient evidence in the literature to support that sleep hygiene alone is an effective treatment for insomnia [64, 87]. Some researchers have found evidence to argue against. For instance [79] have found no association between sleep quality and awareness of SH and [80] hypothesized with positive results that a self-help book for insomnia is better treatment than

SH advise. However, it is interesting that many studies as we shown above, founded association between physiological parameters of sleep and individual SHF. On the other hand, several studies show association between SQ and SH [81, 82, 83]. For example, in a three month longitudinal study comparing changes in SQ in workers with insomnia, the researchers founded that people in the group combining SH practices with education on sleep improved significantly their SQ more than people in the group instructed through education alone [105]. Likewise [64] shows that a program combining sleep hygiene education with cognitive behavioral exercises such as relaxation training, sleep restriction, and stimulus control is an effective method of non-pharmacological intervention for patients with insomnia.

The efficiency of SH in the treatment of insomnia and other sleep disorders is debatable. The subject gives us the opportunity of continue investigating and building on the influence that lifestyles have on health and the challenge of finding ways to measure it.

The following section describes technological proposals that researchers have designed with the aim of contributing to improve the SQ. The proposals consider people suffering a sleep disorder, and people who think they sleep well but are interested in improving their sleep health habits.

2.2 Sleep and Technology

Systems for improving sleep are going up due to increasing knowledge of how sleep influence the quality of life [75]. Technology has contributed to this clinical area in diverse ways, addressing sleep issues from different approaches. From old PSG, through actigraphy, to mobile applications involving environmental sensors. As the years pass, more innovative technology emerge to monitor sleep and diagnose sleep disorders. Furthermore, persuasive systems have also been designed to induce people to change habits to improve their SQ. Emerging technology includes mobile applications, wearable monitoring devices, and, sensor networks working together to mobile applications and desktop systems, among others.

2.2.1 Methodology

We systematically analyze technology designed to address sleep domain by following the methodology we show in Figure 2.3. We consulted four databases (SpringerLink, ScienceDirect, Institute of Electrical and Electronics Engineers (IEEE), and Association for Computing Machinery (ACM)) with the query: ('*technology*' OR '*software*' OR '*computing*' OR '*app*' OR '*ubiquitous*') AND ('*sleep problem*' OR '*sleep disturbances*' OR '*sleep behavior*' OR '*sleep deprivation*' OR '*sleep hygiene*' OR '*sleep disorders*'). We added a filter in ScienceDirect and SpringerLink to limit papers to the Neuroscience and Computer Science domains. We included in the study English journals papers from 2010 to 2015. We excluded publications in workshops, doctoral consortium, and book chapters.

We obtained in the first search 1185 papers, 151 from ACM, 69 from IEEE, 125 from SpringerLink, and 840 from ScienceDirect. We read the titles and abstract to retain only those papers in the scope of our study. We show in Table 2.4 the distribution of the 60 papers we selected after the analysis.

Table 2.4: Distribution of studies per databases.

Databases	**Found**	**Selected**	**Rate**
ACM	151	17	11%
ScienceDirect	840	13	2%
IEEE	69	14	20%
SpringerLink	125	16	13%
Total	**1185**	**60**	**5%**

In the next step, we perform a full text's analysis. We excluded articles whose main goal is to test an electronic device and those having more medical than technological perspective. At the end we selected 18 papers to perform the study. We chose 17 through the criterion above and we added one more due to the number of citations we found of it in reading the full text. We summarized the source distribution in Table 2.5.

We analyzed these works to identify the relevant concepts and their relationship in the sleep

Table 2.5: Studies selected after last filter.

Databases	**Obtained**	**Downloaded**	**Selected**
ACM	151	17	6
ScienceDirect	840	13	1
IEEE	69	14	5
SpringerLink	125	16	5
Total	**1185**	**60**	**17**

and technology scope. We synthesized in the Figure 2.4 the information we obtained from the analysis.

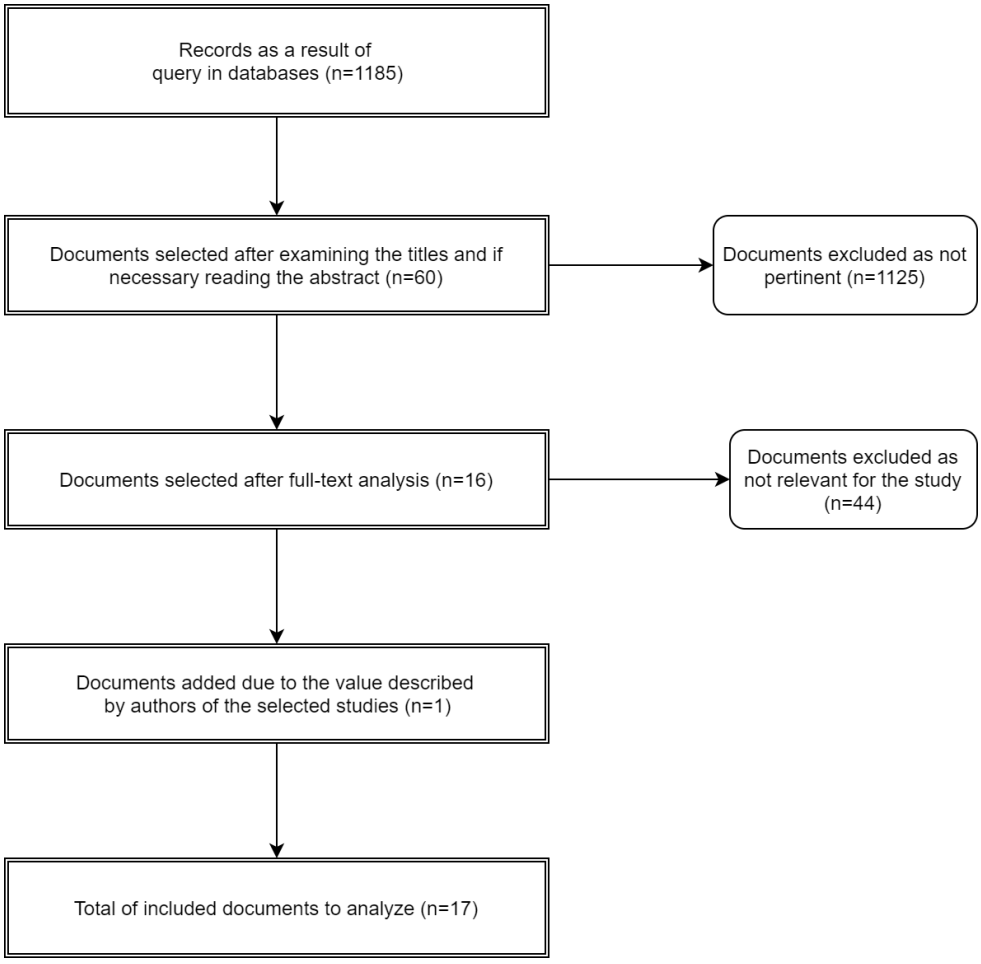


Figure 2.3: Flowchart of process to select reviewed studies.

2.2.2 Mental Map of Relevant Concepts

Fig. 2.4 shows a mental map representing the relationship between the most relevant concepts of technological systems designed to address sleep problems. We found that the systems are mainly aimed at addressing three aspects related to sleep from three different perspectives: a) Monitoring Sleep Patterns and/or Monitoring Sleep Hygiene b) Diagnose Sleep Disorders, and c) Persuading users to make changes in SH to improve SQ.

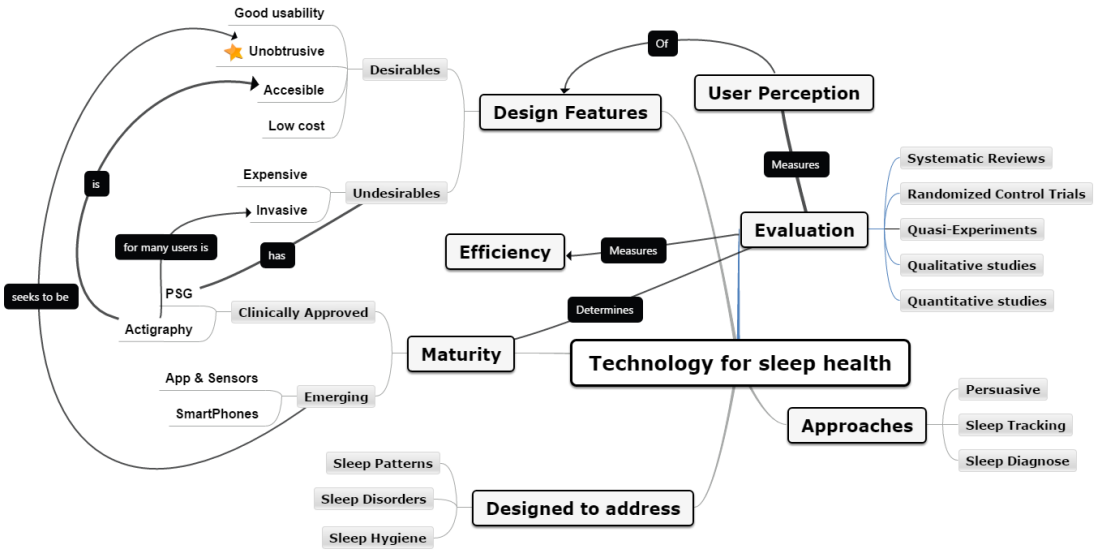


Figure 2.4: Mental map of relevant concepts involved in technology for sleep health.

Monitoring systems are aimed to observe sleep patterns, such as positions or complex movements [106]. Typically they record the data that physicians can see in real time or on demand. Some monitoring systems perceive patterns on sleep time such as breath and heart rate, another capture data related with behaviors of PIMEV in diurnal activities that may affect their sleep. Others are able to send information to physicians periodically or when irregularities in patterns are detected.

Diagnostic systems provide reports that physicians can use when they are attending a sleep disorder. Reports often offer desirable benefits, especially the possibility of reducing hospital-related costs [107]. For instance, Sleep Apnea Monitor helps to diagnose OSA with high probabilities of success before physicians order advanced and expensive sleep tests [108].

Persuasive systems supported on proven psychological theories are intended to take the role of coaches. For instance [100] proposes to use Cognitive Behavioral Therapy (CBT) in persuasive applications to help people when the causes of the illness came from the emotional dimension. The rationale behind is that many of the problems that are related to beliefs, attitudes and emotions, can be treated by therapies, avoiding the necessity of pharmacological treatments. Insomnia is one of the sleep disorders that on repeated occasions has its origin in emotional factors or lifestyle habits that patients can be address and correct through an appropriate therapy. Persuasive systems use strategies such as serious games, social activities with group challenges, and timely reminders motivating people to improve sleep habits [75]. The purpose of changing habits in adults is questionable [109], however, researchers such as [110] and [111] have founded that persuasive systems can engaged the users in healthier sleep habits by motivating them with information about their sleep behavior.

Another relevant subject we analyzed in this literature review is the user perception about comfort, accessibility, and usability respect to the systems. In the users' opinion there are 'desirable' and 'not desirable' features in these systems. Most sought-after emerging technology design are non-intrusive as we show in Figure 2.4, followed by ease of use, accessibility and low cost. Characteristics such as easy accessibility and low cost are related to the purchasing power of the users and both, emerging and mature technology manage to reach it for their target population.

We make the difference between mature and emerging technology by the level of the evaluations to which the researchers submitted their systems. PSG and Actigraphy have reached maturity and have undergone evaluation protocols that have given them the status of approval by international agencies to use them in clinical protocols. In contrast, we call emerging technology

those proposals not evaluated by clinical protocols. The evaluations to which the researchers submitted their systems aims to measuring the efficiency in their proposed goals and discovering weaknesses and strengths in their design. Researchers can identify aspects to refine their proposals for future versions through these evaluations. Also, they identify features that that they should eliminate due to their little or no popularity among users. We found that the researchers evaluated their system to proof efficiency in reach their proposed goals and capture the user perception. The assessments reached at most the level of 'quasi-experiments'; no system was evaluated through an Randomized Controlled Trials (RCT).

In the next section, we describe the technology we classified as mature. Later in Section 2.4, we make a broader analysis of the way in which technological systems aimed at sleep problems should be evaluated and we propose a guide rubric to evaluate emerging developments.

2.3 Systems classified by its maturity

In this section, we describe 'approved' and 'not approved' technology designed to address sleep quality issues (see Fig. 2.5). We describe PSG and actigraphy in the first subsection while in the second we speak about two emerging technology: the smart-phone applications and the EBS which combine sensors with software in mobile, desktop and web applications.

2.3.1 Clinically approved technology

Polysomnography

AASM considers PSG to be the gold standard test for diagnosing sleep disorders in clinical studies [51]. PSG provides an accurate estimation of SQ through various sensors (see Fig. 2.6) including electroencephalography to monitor brain activity, electromyography to perceive muscle contraction, electrocardiography to detect heart rate, oximetry to determine blood oxygen levels, microphone for snoring detection and a camera to monitor body movement [112]. PSG is a laboratory test used alongside subjective surveys and face-to-face interviews between patients

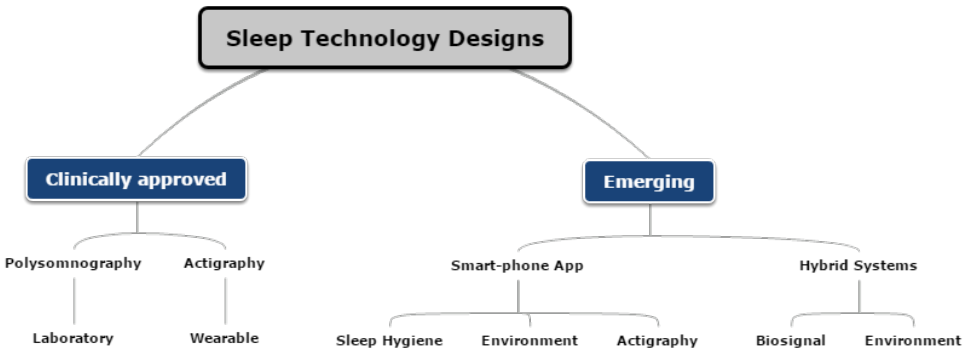


Figure 2.5: Classification of sleep technology designs by its maturity.

and specialists.

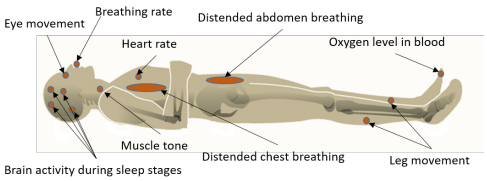


Figure 2.6: Principal data obtained through PSG

To diagnose sleep disorders using PSG, the patient must sleep in an artificial environment or, at the best, at home with sensors placed on the body [113]. Despite its high cost and obtrusiveness [112], PSG is frequently used by physicians because it provides valuable data through an in-depth study necessary for diagnosis various sleep disorders [107, 114, 115]. PSG can determine with precision the quality of each sleep stage in a night of monitoring; however, it cannot be used for long-term monitoring, such as is required for diagnosing insomnia [116].

Actigraphy

Actigraphy is a technique less expensive and intrusive than PSG [116]. It is used to track sleep-/wake behavior patterns including frequency, quality and duration [110]. It is a useful technique for people who need to collect evidence about their sleep patterns for future medical appointments. If feasible physicians prefer actigraphy instead of PSG to monitor sleep [74]. ASDA has accepted actigraphy for clinical use in specific domains, including the identification of periodic limb movements in sleep, sleep-disordered breathing in elderly people [74], and primary insomnia [116]. AASM also has approved actigraphy for use in delineating sleep patterns in healthy children, given the cumbersome nature of PSG for long-term monitoring in this population [73]. On the other hand, actigraphy cannot be used for identifying OSA [117].

Actigraphy assumes that it is possible to detect whether a person is sleeping through analyzing their body stillness or movement. However, its main drawback is in the threshold to detect movements. Often, it is incapable of differentiating whether a person is asleep or in complete rest, leading to lower accuracy [73]. Through experimental studies, researchers reported an average accuracy of 88% to 96% to infer the sleep time [51]. [74] reported a reasonable sensitivity of up to 60% compared to PSG in small time intervals; in long-term evaluation, the error was lower and reliability was higher. Unlike PSG, actigraphy has the advantage of allowing a person to sleep in their natural environment during monitoring [106, 117, 118].

We describe in the next subsection the emerging technology, aiming to monitor, diagnosis and rehabilitation of sleep disorders.

2.3.2 Emerging Designs

High costs, an unnatural operating environment [113], and infeasibility in several situations like in long-term monitoring [116, 112], are some disadvantage of PSG motivating researchers to propose other technology to monitor sleep. On the other hand, the increase of portable devices with high storage and processing capacity, as well as the proliferation of micro hardware with embedded software, makes it feasible to think the design of new technology to attend the sleep

health area. The audience that can be benefited and the diversity of the problems that can be addressed by these technologies grows. Likewise, some of the difficulties of traditional technology have been paved away.

Smartphone applications

Currently, dozens of smart phone applications exist to address sleep problems; many have the intention of influence in improve the SQ of people interested in this issue. In this work we describe a few to show the diversity of the designs, and the potential of the technological development in this area of study. Integration of sensors in smart phones and the symbiosis between people and mobile devices [108, 119] have led researchers to design new systems to monitor and track sleep, to measure sleep duration, to estimate sleep quality, and to influence in modifying SH behaviors. Some applications rely only on the Graphic User Interface (GUI) to interact with users interchanging information. Others use built-in phone sensors to obtain data from the users' context and physical environment. Some other mobile applications such as Sleep eDiary [120] incorporate physicians' talk to provide feedback or remote coaching.

Applications such as Best Effort Sleep (BES) [121] and SleepMiner [122] infer the sleep patterns of users in an unobtrusive manner. The main goal of these applications is to infer SQ from a persons' lifestyle patterns. BES monitors lifestyle patterns by tracking the duration of time in which the phone exists in a state during a normal day (phone-locked, phone-switched-off, phone-charging, for example). Users are unaware of the monitoring process because the devices carried out it in the background without user noticing while they interact with the smart phone normally. On the other hand, SleepMiner is a short-term monitoring application that infers sleep quality through a data process of collected information, using Factor Graphs Model and the Sum-Product-Algorithm. The application captures information related to daily activities, living environment, and social activities from users to make an inference of sleep quality.

Using a different approach, there are applications that monitor users when they are asleep and use some metric to determine the SQ. For instance, iSleep inferring the user SQ by detecting

sleep-related events through identify acoustic signals with the smart phones' microphone while the user is asleep [51]. SleepfulApp is another smart phone application designed to measure sleep efficiency supported on the well-regarded work of Riley et al. [123], where the system and user interact through speaker-microphone. The users start the action by telling to the system the time of going to sleep, stop the application after waking up and interact with the application each time they exit or return to bed. After users say to the system that they are lying down to sleep, the system occasionally emits low audio tones and waits for a response. The last response from the user determines the time that the user fell asleep and the system can calculates the 'sleep latency'. The audio tone continue periodically during the night to determine the time that the user remained asleep. Other applications such as Somnometer [110], use the interaction between the user and the smartphone in a less intrusive manner to SleepfulApp. Somnometer combines manual and automatic monitoring; the user captures its perception of sleep rating while the system determines duration of sleep by sensing trough a mobile application, the time when user has gone to bed and waking up.

Additionally to monitor sleep duration and time to go to bed and waking up, [110] proposes to intervene in improve sleep patterns of users by increasing their sleep habits awareness. The hypothesis is that by sharing sleep patterns to others users in a social network, awareness about healthy sleep habits will increase among friends. These systems include the persuasive concept; the researchers have the intention of intervene in treatments through the system to improve the sleep quality.

The underlying concept is that users can use mobile applications alone to persuade them to sleep better. In this line of thought a qualitative study [124], suggested using rewards to motivate users and keep them engaged in improving sleep habits as BuddyClock does [125]. Another applications such as ShutEye [85] and Bewell [111], persuade users with advice to improve daily SH to have a better nocturnal rest.

Other persuasive systems include a high level of interaction between physicians and patients through technology. Some of these applications are supported by a specific theory of persuasion

such as [126] and [127] propose. [127] implements CBT for insomnia through a mobile application tracking SH. This strategy includes data analysis to transform raw data into high-level information, which is more objective and precise than information obtained through interviews and it has less user burden than writing in a diary or notebook. [126] adds the participation of caregivers being able to enter appropriate activities for a patient using a web service application. The patients capture their activities in the morning and in the evening, while the application maintains adherence through the messages generated by the social and machine persuasion modules.

Beun [100] highlights important issues to consider in the design of new couching technology for health improvements. He recognizes that communication between humans offers a superior information channel than the system-human interaction; however, he argues that smart phone applications remain with the user longer, an advantage that we should exploit. In this sense, it is important that we differentiate activities that the smart phones applications can monitor and intervene from activities that require physician intervention. Beun suggested a mixed-track design involving communication among users, smart phone applications and physicians. Design of these environments imply that applications are able to identify when users require a physician intervention to send notifications. In the same way, physicians must be able to send messages and query users for information through systems. Sleep eDiary [120] implements the above concepts with the goal of assessing and diagnosing insomnia. The application has an interface, a database of relevant messages-questions, and, a data transfer protocol in a secure and intuitive software. Data transfer refers to an open communication channel between physicians and patients with insomnia. This approach allows a subjective and remote evaluation of sleep, reducing the number of face-to-face visits and, thus, costs.

Environment and body systems

Systems integrating sensors, applications and small devices have been proliferate recently being useful in various areas. Among the most exploited areas is smart medical homes for the elderly [128, 112]. Likewise, in sleep research there exists various technological developments where

designers combine applications and light sensors, microphones, accelerometers, pressure and pulsation sensors [128, 124, 118]. Sensors are placed in the phones, wearable technology, skin devices, mattresses, pillows, and sleep-related home objects as shows Fig. 2.7 [108, 106]. One of the goals in EBS is to reduce the obtrusiveness by acquiring parameters from the users' body and the user environment while he/she is asleep.

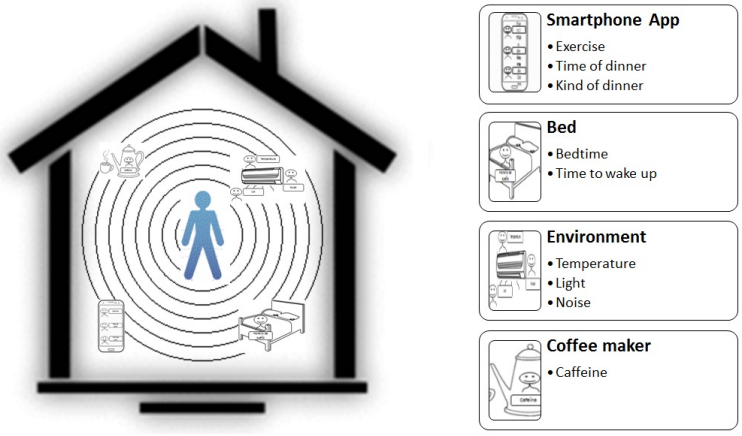


Figure 2.7: EBS generating an ubiquitous environment to capture sleep hygiene factors.

Due to EBS integrate two or more technologies, robustness is a mandatory feature. It means that the failure of one sensor or malfunctioning of a routine embedded in a small device does not affect the whole system [128]. Examples of EBS are [106] and [115]. The first one, uses a suit of Radio Frequency Identification (RFID) sensors for limb monitoring, while the second work uses a suit of 14 sensors in a face-band to observe users' bio-signals emulating a home PSG. We consider obtrusive these two approaches since they obtain data from users by having direct physical contact with the skin of users.

Others systems such as [76] and [129] use a camera as an unobtrusive sensor to identify the movement of users during sleep to infer the sleep patterns. In order to be less intrusive, [130] designed a system composed of a set of sensors placed in a band that users should attach to the mattress. In this way, sensors can monitor sleep and identify patterns without human intervention. In the same context, [112] proposes an alternative, unobtrusive and innovative

way to estimate the quality of sleep. The system measures breath and heart rate using radio frequency signals from a wireless network. This creative design avoids costs of extra sleep-monitoring devices since most people have a wireless network at home.

Using systems to obtain contextual information to infer sleep patterns, diagnose sleep disorders, and intervene with the intention of influence to improve sleep quality is a new research field with many opportunities for innovation [75].

2.4 How to evaluate the Sleep Technology

The Journal of Physiological Measurement published in June 2013 a review on application designed to attend sleep issues in which Behar et. al. conclude that: "*With the exception of simple questionnaires, no existing sleep-related application available for smartphones is based on scientific evidence*"[77]. This assertion matches partially with our conclusions exposed in Section 2.3 that beyond of PSG and actigraphy, emerging technology such as Smart Phone Applications and EBS, have not credentials to contribute as clinical tools neither as treatments nor for diagnosis. We say 'partially' because it is true that no one has clinical credentials, nevertheless that doesn't means they don't have assessments through scientific methodologies. Next paragraph discuss about it in a conciliatory perspective.

To begin the analysis, we integrate the Table 2.6 after studying three frameworks of methodologies to demonstrate scientific evidence in health area [34, 36, 35].

Systematic Reviews has the highest level of evidence, while RCT is ranked in the second place in both categories of evaluation while Observational Studies is ranked beside to RCT in the efficiency category. Methodologies ranked in levels III to V are categorized as fair or poor for health. Even when methods in levels three and four follow logical and systematic process to evaluate efficiency, their results are not conclusive mainly because they cannot be generalized. Generalization of results to large populations is the main goal in clinical studies and Systematic Reviews, RCT and Observational Studies to reach it successfully. This area aims its procedures and treatments to be applicable to all human beings, which is debatable, but not the subject of

Table 2.6: Evidence level of methodologies of evaluation in the health field based on [34, 35, 36].

Efficiency		**Perception of the user**	
Level	**Methodology**	**Level**	**Methodology**
I	Systematic Reviews (SR)	I	Systematic Reviews (SR)
II	Randomized Controlled Trials (RCT) Observational studies (OS)	II	Randomized Controlled Trials
III	Non-Randomized Control Trials (NRCT) Before and after studies (BAS)	III	Cross sectional surveys Focus groups (FG) Phenomenological study
IV	Case of study Correlational study (CS) Single qualitative study (SQS)	IV	Expert Opinions (EO) Other qualitative designs
V	Expert Opinion (EO)		

this work.

Nevertheless the relevance of Systematic reviews they are limited to studies broadly addressed. In addition, their interpretative nature introduces some of subjectivity. This leads to RCT being considered the most effective methods to produce scientific evidence. RCT have the highest level of internal validity due mainly to two essential components: 1) The comparison of results among groups, at least two, the control group and the experimental one; 2) The random choice of the study subjects [35]. Additionally, RCT include rigorous procedures to determine the intervention time and the strategy to calculate the sample size to guarantee the confidence of the results. The calculation of the sample imply three important parameters the significance level of test (α), the power of test ($1 - \beta$) and a pertinent effect-size to guarantee the practical convenience of implementing a new treatment (Δ). Another relevant feature of the RCT is that they include procedures to avoids '*confounding variables*' which ensures that all variables included in experiment contribute to explain the variability of data.

Under these assertions, emerging systems designed to handle sleep problems have two options

to test scientific evidence of clinical efficiency: to undergo RCT or observational studies. The affirmation of Behar builds in this sense, in the conclusion of his article encourage designers of applications to venture into the clinical field to provide a better benefit to users. Undoubtedly, that health-oriented applications become clinically approved and recommended by physicians is the ultimate goal of many developers, however all processes must mature to consolidate. Mobile applications and EBS are emerging technology in the process of consolidation.

Klansja et al. [109] from the Human Computer Interface (HCI) area, argues that is not feasible to evaluate emerging technology through RCT studies in the early stages of design. RCT would have a high cost in resources while the outcomes would be not comparable in quality. Instead of preparing a rigorous study, researchers in the early stages would prepare evaluations that provide information in two directions: 1) The efficiency that developments have regarding the goals proposed in the current stage of design; 2) The perception that users have of the efficiency of use, ease of use, intrusiveness and comfort. In exploratory studies, researchers can obtain this level of evidence through quasi-experiments, correlational studies, case studies and qualitative evaluations, among others. The outcomes through these evaluations do not provide evidence of causality, but establish relations between variables and helps the designers to identify patterns [35].

In concordance with the arguments above, Evans et.al. [34] introduces in addition to efficiency, two important aspects for evaluating in an intervention: 'appropriateness' and 'feasibility'. Its framework proposes the evaluation of these concepts through correlational studies, focus groups, before and after studies, phenomenological studies, expert opinion, among others. They propose to use RCT in final stages of design.

It is relevant to note that the questions proposed by Evans et al. to evaluate appropriateness and feasibility are qualitative. In the case of appropriateness, the questions are: "*What is the experience of the consumer?*", "*What health issues are important to the consumer?*", "*Does the consumer feel the outcomes as beneficial?*". The above questions are similar to those used in HCI to evaluate applications implemented in various domains. They can be answered with responses

based on the perceptions of the users. It is possible to merge frameworks of health and HCI areas to build an appropriate methodology to evaluate technology applications in the sleep research domain.

Another point in favor of use alternative methods to evaluate emerging systems is that RCT are not able to respond all type of questions. There is a lot of valuable information that researchers can obtain through alternative scientific methods of evaluation. Such is the case of quasi-experiments, quantitative and qualitative studies helping to respond important questions maintaining the research quality [35]. The key is to identify the type of question that is posed, since there exist appropriate methods to diverse questions [36]. By selecting adequate methodology to the type of question, the evaluations will produce valuable outcomes providing reliable evidence for proposed goals.

Many outcomes in these evaluations will be used by researchers to understand the impact that the systems produce in people closely related with the issue [34]. Outcomes in these stages do not provide clinical evidence, however, contribute by providing guidelines to the next design stages where more rigorous evaluations will be required. Collaterally, the scientific community interested in the same subject, will design from a more robust platform [109].

Researchers in the area of technology for sleep health, must be careful when describing goals they want to reach in their current stage of work. The main goal and the questions associated to it define the appropriate evaluation for research. It avoids the thought of emerging technology ignoring scientific evidence, but demonstrates scientific evidence depending on its design stage.

After analyzing arguments exposed above, we propose the Table 2.7 to evaluate systems for sleep in their early stages of development.

Table 2.7: Proposed framework to evaluate emerging systems designed to attend sleep health.

	Methodology	
Level	**Efficiency**	**Perception of the User**
High	Non-Randomized Control Trials (NRCT) Before and after studies (BAS)	Cross sectional surveys Focus groups (FG) Phenomenological Study
Medium	Case study Correlational Study (CS) Single Qualitative Study (SQS)	Expert Opinions (EO) Other qualitative designs
Low	Expert Opinion (EO)	
Poor	Poor Methodology/Does not report efficiency evaluation	Does not report a user perception evaluation

2.5 Emerging Systems Evaluation

Guided by the concepts exposed in [131, 132, 36, 35, 34, 109], we evaluated on Table 2.8 the eighteen emerging systems described in Section 2.5. The first five columns are based in the following criteria:

1. It is possible to make some tests in short term and with small sample sizes to obtain results that contribute to knowing whether technology is achieving aims that researchers are supposed to be doing.

2. It is feasible to compare techniques, algorithms and implementations to validate their efficacy and test intervention strategies.

3. The most important issues in short evaluation tests, are the conclusions obtained from users regarding their experiences when using the systems.

The last two columns classify systems in a level of evidence based on Table 2.7

Table 2.8: Evaluation of methodology used to assess systems designed to attend sleep health.

System	Met its Proposed Goals	Sample	Weeks of Intervention	LEEP	Proof	Efficiency	Users' Perception
[85]	*yes*	12	4	ED	DI	Medium	High
[122]	*yes*	15	4	EP	DS	Medium	Low
[126]	*yes*	5	6	ED	DS	Medium	High
[121]	*yes*	8	1	ED	DS	Medium	High
[127]	*yes*	18	—	ED	tt	Low	High
[51]	*yes*	7	$\frac{1}{2}$ - 2	ED	DS	High	Low
[76]	*yes*	4	2	ED	DS, DI	Medium	High
[125]	*yes*	8	6	ED	tt	High	High
[111]	*yes*	27	3	ED	tt	High	Medium
[123]	*yes*	26	1	ED	PC,tt	High	High
[115]	*yes*	–	–	EB	–	Medium	Low
[114]	*yes*	1	$\frac{1}{2}$	EP	ML,BS	Medium	Poor
[112]	*yes*	6	12	ED	DS	High	Low
[129]	*yes*	7	—	EP	ML	High	Low
[130]	*yes*	10	1	ED	ML,tt	High	Low
[106]	*yes*	–	–	EB	–	Medium	Low
[110]	*yes*	8	6	ED	AV,PC	Medium	High
[120]	*yes*	–	–	EB	–	Low	Low

LEEP: Level of description of the evaluation procedure; **ED:** Explained in Detail; **EP:** Explained Partially; **EB:** Explained Briefly; **ML:** Machine Learning; **tt:** Student Test; **AV:** ANOVA Test; **PC:** Pearson Correlation; **DS:** Descriptive Statistics; **DI:** Deductive/Inductive

Systems are far from meeting the RCT requirements or equivalent. Only [112] has an intervention time comparable to the shortest interventions of RCT [132]. On the other hand, none evaluations in this list have a sample size close to the minimum size required by RCT. However, all the authors of the evaluated systems argued they reached their proposed goals. Supported in Table 2.7, 7/18 systems were evaluated at '*High Level*', 9/18 at '*Medium Level*' and 2/18 at

'*Low Level*'.

Even though the systems were evaluated through a scientific methodology that does not meet the requirements for clinical evidence, the designers can use the results to achieve more satisfactorily goals in the next version of the system. Furthermore, information that developers obtained from the users perception help to them to improve the new designs. This information provides competitive advantages that third parties should analyze to make decisions especially in the services and business fields. For example, thanks to qualitative studies of the systems, we identified that technology has high level of acceptance when it is not intrusive. This is a valuable finding that helps researchers in sleep technology to design systems with emphasis on the particular concept of '*unobtrusiveness*'.

On the other hand, some weaknesses were found and it is important to highlight them. From eighteen evaluated systems, a 47% do not include a qualitative study. From all systems with a qualitative study, only three ([126, 76, 85]) reported a methodology plan supported theoretically. Furthermore, no article explains the reason of the sample size, neither why or how the duration time of intervention was chosen.

Chapter 3

Collecting and Preparing Data

In the previous chapter, we defined the concepts of SH and SQ, likewise we found in the literature that researchers differ on the influence of SH on the SQ. Further, we search questionnaires to measure these two variables and technological tools clinically approved to support the diagnosis and treatment of sleep disorders. Likewise, we discussed on emerging technology available to people for supporting in improve sleeping.

We describe in this chapter the cross-sectional study and the data preparation to fulfill partially the strategy to identify the most influential SHF in the SQ of the SP. In Section 3.1 we describe the methodology in order to obtain data on SHF and SQ from the SP. We obtained the data in a sample composed of high school and university professors from six academic institutions in México. In Section 3.2 we explain the process of cleaning the collected data in order to prepare them and use them as data input of the MLA to produce predictive models. These processes are fundamental to obtain success in modeling through MLA and consumes from 50% to 70% of the total project time [133, 134].

Fig. 3.1 shows the strategy we planned in three stages to the data acquisition and preparing.

Data Acquisition and Data Coding: In this stage we used a questionnaire to survey the SP in order to generate the first dataset for the study. After survey the SP we obtained a raw dataset, that we validated and coded in numeric format for further analysis (see Sections

3.1 and 3.2.1).

Data Quality Analysis: Using the methodology of [135], we separated features into continuous and categorical in order to generate a data quality report for each group. We measured the quality of continuous features through: Quantity, Missing Values, Cardinality, Minimum Value, First Quartile, Median, Third Quartile, Maximum Value, Mean and Standard Deviation. Likewise, we measured the quality of categorical features through nine measures: Quantity, Missing Values, Cardinality, Mode, Frequency of the Mode, Percentage of the Mode, Second Mode, Frequency of the Second Mode, Percentage of the Second Mode. Additionally, we generated histograms and box-plots for continuous features in the aims of identify outliers (See Fig. 3.3).

Data Quality Plan (DQP) In this stage, we performed each item of the DQP. We concluded with the data quality report showing the cleanliness of the new dataset when is compared with the same report before implementing the data quality plan (See Section 3.2.4).

3.1 Data collection

3.1.1 Population

We surveyed volunteer teachers from six universities a long of México. The sample included two southern provinces (Chiapas and Yucatán), two in the centre (San Luis Potosí and Puebla) and two in the north (Baja California and Nuevo León) of the country. We calculated the sample assuming an infinite population (N), a significance level of $90\% - 95\%(1.645 \geq Z \leq 1.96)$, an error level of $5\%(e = 0.05)$ and a proportion of 50% (p=q=50%) due to lack of previous studies showing the behavior of the SQ in the SP. After surveying the sample, we obtained a dataset of 338 records, enough to reach a significance level of 93.45%. Finally, we reached a dataset of 304 instances, decreasing our significance level to $\approxeq 92\%$.

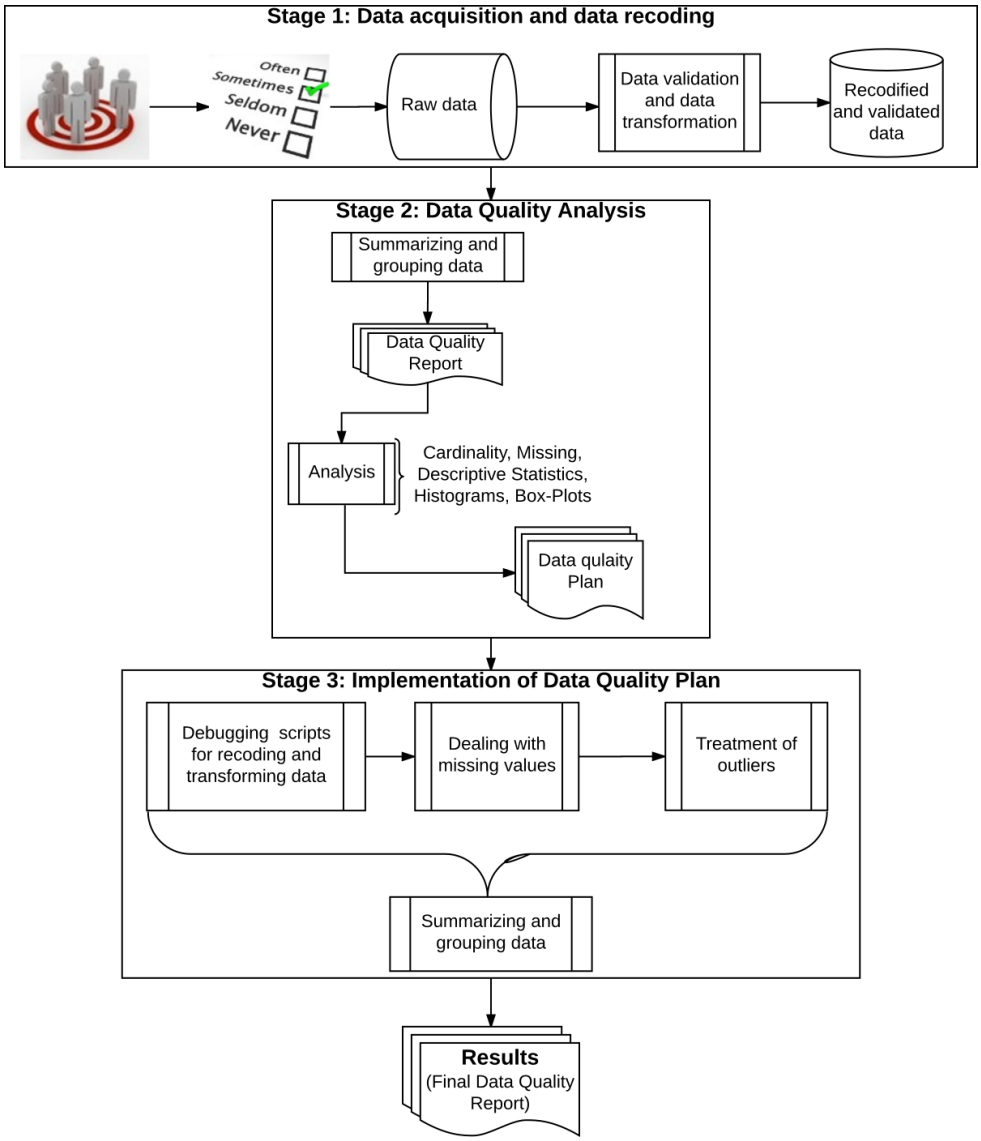


Figure 3.1: Acquisition data preprocessing flow diagram.

3.1.2 Questionnaire

In order to obtain data to explore the relationship between SH and SQ, we use a three section questionnaire. The first section includes demographic (n=6), emotional (n=1) and health (n=1) questions. The second section is the PSQI questionnaire (n=18) and the third section is an edited version of the Sleep Hygine Index (SHI) questionnaire (n=21).

Demographic Data

In this section, we ask for '*age*', '*gender*', '*occupation*', '*kind of work*', '*religion*' and '*civil status*'. We add these items with exploratory purposes; may researchers use our results in future studies addressing the same subject. The item 'kind of work' additionally to provide demographic information, it provides relevant information about sleep, since physical activity impacts the SQ [97]. Options for asking this question are: *'intellectual'*, *'physical'*, *'more intellectual than physical'* and, *'more physical than intellectual'*. May we do not use this item because our SP, however, the item could be useful in other studies.

There are two items we added to the demographic section contributing with relevant information to our study. The first asks the respondent if they are living in crisis. The crisis could be financial, or related to emotional problems, such as the loss of a family member or a divorce. Since such crises often affect the quality of sleep, we added this question with filter purposes. People suffering a crisis should be removed from the dataset for bias reasons before generating the prediction models. The second is a health condition question we used to ask the people if they suffer from a chronic degenerative disease. Diseases such as diabetes, hypertension or depression may alter the sleep quality, producing bias in the analysis. For instance, a disease could physiologically affect the quality of sleep if it alters the heart rate; on the other hand, diabetes could alters sleep if the person gets up several times to the bathroom at night. Similar to 'emotional question', records of people suffering any of these diseases should be removed before training the model.

Sleep Quality

For the second section we used 18/19 items from the Spanish version of the PSQI, a questionnaire widely used to estimate the quality of sleep in the clinical and non-clinical population [71]. The PSQI is considered the gold standard scale to evaluate subjective sleep quality [122]; Physicians has used it in several populations showing internal consistency, test-retest reliability, construct validity and empirical validity [58]. The PSQI evaluates the quality of sleep using eighteen items grouped into seven components: subjective sleep quality, sleep duration, sleep disturbances, use of sleep medication, day time dysfunction and sleep latency. Each component is computed according with a formula provided by the author. The results are in a scale of 0 to 21. According with the interpretation guide from 0 to 5 means "*good quality of sleep*" and from 6 to 21 means "*poor quality of sleep*". We removed the last item of the original version asking about some specific sleep disorders (e.g. sleep apnea) for two reasons: 1) A respondent roommate must answer the item; 2) The item does not contribute numerically in the original scale according with the PSQI instructions.

Sleep Hygiene

SHI is a scale designed to measure the SH in a non-clinical population. Its theoretical basis is founded in the criteria of The International Classification of Sleep Disorders (ICSD) to diagnose an inadequate sleep hygiene. The original scale has reported an internal validity of $\alpha = 0.71$, and high reliability in test-retest assessments [136]. We edited and translated (see Section 3.1.2) this questionnaire from its original version of 13 items in English to a version of 21 items in Spanish. We added items to obtain more precised answers without altering the SHI target as we explain in Section 3.1.2.

SHI Edition

- Following the structure and meaning of the item number four *'I exercise to the point of sweating within 1 h of going to bed'*, we added two items to the questionnaire: *'I exercise*

to the point of sweating during the morning' and *'I exercise to the point of sweating during the afternoon'*. Our main purpose is to know whether the exercise at morning/afternoon are correlated with the SQ.

- The item six in the original questionnaire says: *'I use alcohol, tobacco, or caffeine within 4 hours of going to bed or after going to bed'*. Instead, we wrote one item for each substance (alcohol, tobacco and caffeine).

- The item 11 in the original scale asks about an uncomfortable bedroom due to four environment factors (light, noise, heat and cold). Instead, we included four items asking about of the bedroom environment.

- Due to reports we founded in the literature about the impact of dinner on sleep quality [97, 137, 87, 104]; we added one item asking about this practice.

The edited SHI consists of 21 items grouped into four components: 'Stress Factors' (five items), 'Disruptor Factors' (five items), 'Circadian and Homeostatic Factors' (eight items), 'Drug Factors' (three items).

SHI Translation

We translated the SHI based on the guidelines of [138]. We performed a triple blind as Fig. 3.2 illustrate. A bilingual researcher in the education area (A) translated the questionnaire from English to Spanish, another bilingual professional with experience as an interpreter between Mexicans and Americans (B) returned the Spanish translation to the English language, and a team conformed by two professionals in the health field (C) compared the questionnaire obtained by the translation of the B translator against the original questionnaire. The team C wrote comments regarding those items that did not match in meaning. We attended all suggestions and corrections by iterating the process until we reached satisfactory results from the perspective of the reviewers.

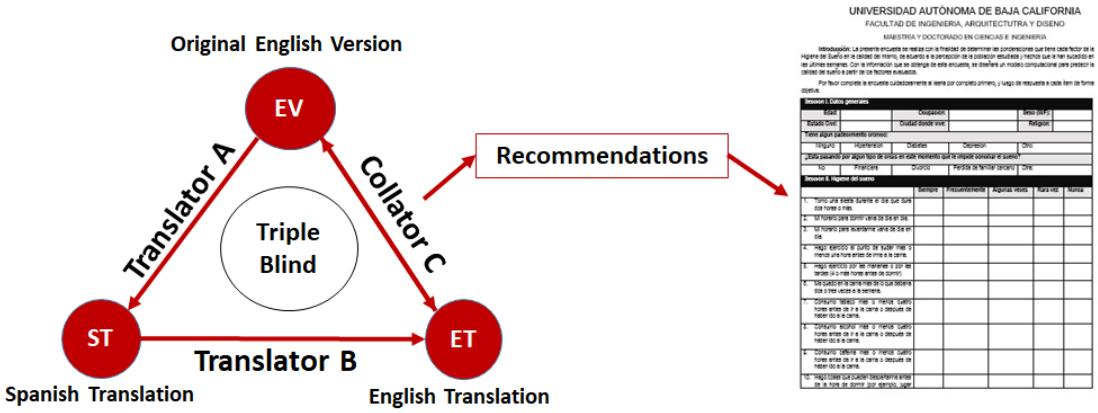


Figure 3.2: Translation process of SHI scale.

Validity and Reliability

Two professionals in the health and three with experience in validity of Spanish questionnaires validated the translated survey. They qualified the items as clear ($\mu = 4.5$) and pertinent ($\mu = 4.7$) by using a scale of 1 to 5. We carried out a pilot test on a randomly sample of 30 subjects to calculate the internal validity of the instrument. The Cronbach's alpha ($\alpha = 0.69$) of the pilot test is acceptable and consistent with that reported by [136] for the original SHI scale. After, we proceeded to survey the SP and store the data in a raw dataset.

3.2 Data Preparation

3.2.1 The Dataset

We cleaned and formated the raw dataset trough the following tasks:

- Renaming of the columns.
- Validating the columns containing information of time and age.
- Coding the PSQI and SHI questionnaires from text to numeric format for further analysis. We convert the textual words: *'Never'*, *'Rarely'*, *'Some Times'*, *'Often'*, *'Always'*, to 0,1,2,3,4 respectively.

This process yielded a 48 columns dataset distributed as the Table 3.1 shows.

Table 3.1: Dataset distribution by type of features

Type	**Quantity**	**Columns**
CharID	1	1
Categorical	41	[3-9] and [14-48]
Continuous	6	2 and [10-13]

After this process, we proceeded to calculate the new columns that summarize the PSQI and SHI scales according with the authors instructions. After this process, we obtain a dataset composed of one ID, 50 categorical and 12 continuous features[1] as Table 3.2 shows.

Table 3.2: Dataset composition by type of features after computing columns of PSQI and SHI scales

Group	**Categorical**	**Continuous**	**Total**
ID	0	1	1
Demographic data	7	1	8
PSQI	14	4	18
SHI	21	0	21
Scale PSQI	8	1	9
Scale SHI	0	5	5
Total	50	12	62

We show in the following tables the structure of the dataset for the data analysis. Table 3.3 describes the demographic features, Table 3.4 contains the description of the PSQI features, while Table 3.5 describes the seven components and the two target features (**SQTT** and **SQCL**) of the PSQI scale. Table 3.6 describes the SHI scale and Table 3.7 shows the four components of this scale and its target feature called **SHTT**.

[1] Sum of SHI groups, age and ID are not strictly continuous features, however for this work we are differentiating them from the categorical features resulting from responses to the questions in Likert scale

Table 3.3: Demographic features description.

Section	Feature	Type	Value	Description
ID	email	Text	[A-Z0-9a-z._%+-]+[A-Za-z0-9.-]+.[A-Za-z]{2,64}	Identifier
Demographic	DD1	Continuous	[1-100]	Demographic data for statistical purposes only
	DD2	Categorical	Female, Male	
	DD3		Student, Employer, Teacher, Independent professional, Other	
	DD4		Intellectual, Physical, More intellectual than physical, More physical than intellectual	
	DD5		SDA, Catholic, Jehovah's witness, Evangelic, Other	
	DD6		Married, Single, Divorced, Free Union, Other	
	DD7		No, Hypertension, Diabetes, Depression, Other	Demographic information with filtering purposes
	DD8		No, Financial, Divorce process, Loss of Family, Other	

Table 3.4: PSQI Questionnaire.

Section	Feature	Dimension	Value Range	Description
PSQI	Continuous features			Each feature provides information to PSQI Scale.
	SQ1	SQSE	Real$[1.0 \leq SQ1 \leq 12.99]$	
	SQ2	SQLAT	Integer$[0 \leq SQ2 \leq 60]$	
	SQ3	SQSE	Real$[1.0 \leq SQ3 \leq 12.99]$	
	SQ4	SQDUR	Real$[1.0 \leq SQ3 \leq 12.99]$	
	Categorical features			
	SQ5a	SQLAT		
	SQ5b			
	SQ5c			
	SQ5d			
	SQ5e			
	SQ5f	SQDIS	Level$[0 \leq SQ5[b-j] \leq 3]$. 0 meaning best possible of SQ and 3 meaning worst possible of SQ.	
	SQ5g			
	SQ5h			
	SQ5i			
	SQ5j			
	SQ6	SQRP		
	SQ7	SQMS		
	SQ8	SQDD		
	SQ9			

Table 3.5: PSQI Scale features description.

Section	Feature	Dimension	Range	Scale
PSQI	SQDUR	Sleep Duration		SQCL:
	SQDIS	Sleep Disturbance		$0 < SQTT \leq 5$ = Good
	SQLAT	Sleep Latency		$6 < SQTT \leq 21$ = Poor
	SQDD	Day Dysfunction		
	SQSE	Sleep Efficiency		
	SQRP	SQ Respondent Perception	$[0 < SQ* \leq 3]$	
	SQMS	Consume Medicine for Sleep		
	SQTT	Total of PSQI	$[\ 0 < SQTT \leq 21\ ]$	
	SQCL	Qualitative SQ	[Good/Poor]	

Table 3.6: SHI features description.

Section	Feature	Classification	Range	Purpose
SHI	SH1	Circadians	$[0 \leq SH* \leq 4]$. 0 meaning best possible SH and 4 meaning worst possible SH	Predictive Features. Each feature is a potential provider of information for the model.
	SH2			
	SH3			
	SH4			
	SH5			
	SH6			
	SH7			
	SH21			
	SH8	Drugs		
	SH9			
	SH10			
	SH11	Stress		
	SH12			
	SH13			
	SH19			
	SH20			
	SH14	Environment Disruptors		
	SH15			
	SH16			
	SH17			
	SH18			

Table 3.7: SHI scale feature description.

Section	Name	Range	Scale
	SHCH	$0 < SHCH \leq 32.$	$0 < SHTT \leq 16$ = Very Good
	SHDG	$0 < SHDG \leq 12.$	$17 < SHTT \leq 33$ = Good
Scale SHI	SHSTR	$0 < SHSTR \leq 20.$	$34 < SHTT \leq 50$ = Regular
	SHDIS	$0 < SHDIS \leq 20.$	$51 < SHTT \leq 65$ = Poor
	SHTT	$0 < \sum(SHI) \leq 84.$	$66 < SHTT \leq 84$ = Very Poor

Table 3.8: Equivalence of features, items, abbreviations and keys for SHF.

Factor	Item	Abbreviation	Key
SH1	I take daytime naps lasting two hours or more.	NAP	NA
SH2	I go to bet at different times from day to day.	SLEEP-TIME	ST
SH3	I get out of bet at different times from day to day.	WUP-TIME	WT
SH4	I exercise to the point of sweating at night.	EV-EXERCISE	EE
SH5	I exercise to the point of sweating for the mornings.	MO-EXERCISE	ME
SH6	I exercise to the point of sweating for the afternoons.	AF-EXERCISE	AE
SH7	I stay in bed longer than I should two or three times a week.	STAY-IN-BED	SIB
SH8	I use tobacco within the four hours before bedtime.	TOBACCO	TB
SH9	I use alcohol within the four hours before bedtime.	ALCOHOL	AC
SH10	I use caffeine within the four hours before bedtime.	CAFFEINE	CF
SH11	I do something that may wake me up before bedtime (for example: play video games, use the internet, or clean).	INT-ACT-BS	IABS
SH12	I go to bed feeling stressed, angry, upset, or nervous.	STRESS-BS	SBS
SH13	I use my bed for things other than sleeping or sex (for example, watch television, read, eat, or study).	N-SLEEP-BU	NSBU
SH14	I sleep on an uncomfortable bed (for example: poor mattress or pillow, too much or not enough blankets).	UNC-BED	UB
SH15	I sleep in a bedroom with some degree of lighting.	LUMINOSITY	LU
SH16	I sleep in a noisy environment (for example: busy street, aircraft noise, working machinery).	NOISY	NO
SH17	I sleep in an uncomfortable bedroom due to the cold.	COLD	CO
SH18	I sleep in an uncomfortable bedroom due to the heat.	HEAT	HE
SH19	I do important activities before of going to bed (payment of invoices, study, try to solving problems).	INT-WK-BS	IWKBS
SH20	I think, plan, or worry when I am in bed.	WORRIES-BS	WBS
SH21	I take the dinner close to the hour of bedtime (a dish that may include meat or cheese).	DINNER	DI

3.2.2 Data Quality Analysis

The DQA begins with a data quality report. We separated the features in two groups: continuous and categorical features. The group of continuous features is composed of one demographic feature (**DD1** = AGE) and four features to measure the sleep duration (**SQ1** = *'Time to go to bed'*, **SQ2** = *'sleep latency'*, **SQ3** = *'Time to wake up'*, **SQ4** = *'Time that remains asleep'*).

Table 3.9: Data quality report of continuous features.

Feature	Count	Miss	Card.	Min	Q1	Median	Q3	Max	μ	σ
DD1	338	6	49	16.00	27.00	35.00	44.00	66.00	35.92	11.41
SQ1	338	0	29	0.00	10.00	11.00	11.08	12.50	9.99	2.71
SQ2	338	1	15	1.00	5.00	15.00	20.00	60.00	14.99	11.91
SQ3	338	2	40	0.70	5.08	6.00	7.00	11.00	6.18	1.31
SQ4	338	3	107	-0.05	6.17	6.88	7.75	10.75	6.91	1.35

Table 3.9 shows some irregularities in continuous features. We can see that the minimum value in **SQ4** is negative, since this feature captures the *'time between going to bed and waking up'*, negative values are not valid. Another issue we observed is that the minimum value in **SQ1** (the time to go to the bed) is 0.0. We agreed to capture 12.00 for this feature instead 0.0 when subjects would want to refer to midnight. On the other hand, the standard deviation ($\sigma = 11.91$) for the **SQ2** (*'the time that person takes to fall asleep in minutes'*) is too large, since its mean is ($\mu \simeq 15$).

None of these features has too many missing values, **DD1** is the variable with most NA; however, the missing values represent the 0.018% of the data. If we assume that each missing value is in a different record, we would have 12 records with missing values, which represents the 0.036% of the total records in the dataset. As this percentage is small, and we have no previous works describing this population through these features, records with missing values could be deleted.

We see good cardinality in the continuous variables, even though when the ratio between cardinality and the number of records is far from 1.0 ($DD1 = 0.14, SQ1 = 0.09, SQ2 = 0.04, SQ3 = 0.12, SQ4 = 0.32$). The nature of data justifies this fact, because the expected responses for these questions are in a small range of values. For instance, **SQ1** (time when a person go to

bed) takes a small range of values because people commonly answer this kind of question with rounded values. For example, people generally respond that they went to bed at 9:00, 10:30, 10:45 or 11:00, even when they actually went to bed at 9:03, or 10:32. **SQ3** and **SQ4** features have similar nature.

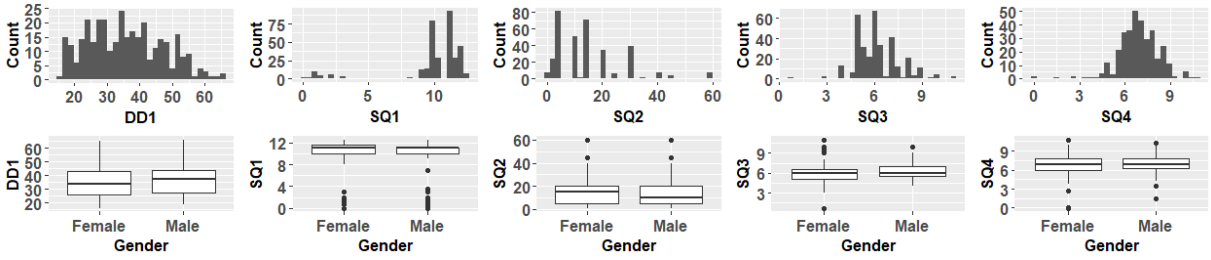


Figure 3.3: Histograms and Box-plots of continuous features.

In Fig. 3.3 we through the box-plots the features containing outliers. We analyzed analyzed them to decide their inclusion/exclusion in the final dataset. These features do not intervene directly in training the model; however, the outliers could indicate some sleep disorder in the respondent, then we performed a process to exclude these records.

We grouped categorical features in SH and Demographic features. We analyzed first the group of SH in the information of Table 3.10.

We found in the report of the quality of SHF (Table 3.10) that 19 of the 21 features have excellent cardinality, all possible values for each item ($[0-5]$) are represented in the data. Feature **SH9** has good cardinality since the answers are distributed among four of five possible values. Nevertheless, we observed two variables with high mode, the 81.66% of respondents, answered *'never (0)'* for **SH9** *'I use alcohol within 4 hours of going to bed or after going to bed.'*, while the 94.38% responded *'never (0)'* for **SH8** *'I use tobacco within 4 hours of going to bed or after going to bed'*. These high modes allow us to dispense with these features for the analysis due to their low variability. The other 19 categorical features are in a good range, **SH10** has the highest mode, 65.68% of the respondents answered *'never (0)'* to the question *'I use caffeine within 4 hours of going to bed or after going to bed'*. The variability on these features makes them eligible to train a model.

This set of data has a small number of missing values, however, we decide to perform an

Table 3.10: Quality Report of Sleep Hygiene Factors.

Feature	Count	Miss	Card.	Mode	MF	M%	M2	MF2	M2%
SH1	338	0	5	1	114	33.73%	0	111	32.84%
SH2	338	0	5	1	116	34.32%	2	116	34.32%
SH3	338	0	5	1	160	47.34%	2	79	23.37%
SH4	338	1	5	0	204	60.36%	1	67	19.82%
SH5	338	0	5	0	176	52.07%	2	64	18.93%
SH6	338	1	5	0	152	44.97%	1	75	22.19%
SH7	338	2	5	0	142	42.01%	1	107	31.66%
SH8	338	0	5	0	319	94.38%	4	7	2.07%
SH9	338	0	4	0	276	81.66%	1	34	10.06%
SH10	338	0	5	0	223	65.98%	1	57	16.86%
SH11	338	1	5	0	104	30.77%	2	78	23.08%
SH12	338	0	5	1	133	39.35%	2	114	33.73%
SH13	338	0	5	2	92	27.22%	1	76	22.49%
SH14	338	0	5	0	210	62.13%	1	61	18.05%
SH15	338	1	5	0	101	29.88%	1	91	26.92%
SH16	338	0	5	0	163	48.22%	1	84	24.85%
SH17	338	1	5	0	222	65.68%	1	82	24.26%
SH18	338	0	5	0	173	51.18%	1	82	24.26%
SH19	338	0	5	0	125	36.98%	1	81	23.96%
SH20	338	0	5	2	117	34.62%	1	90	26.63%
SH21	338	0	5	1	130	38.46%	2	96	28.4%

imputation process since these group of features are the predictive ones of the model. The sum of all features represents the sleep hygiene behavior of people. We decide by theoretical analysis among three algorithms (K-Nearest Neighbors (KNN), RF or Multiple Imputation by Chained Equation (MICE)) to performed the data-imputation.

We found in Table 3.11 that **DD6** has the highest possible mode (100%). We considered this mode indicates an error in some task of the data pre-process that must be analyzed. If the data contains true responses, all respondents are living in *'free union'*, which is very doubtful considering the results of 2015 in the intercensal survey in Mexico [139], where only 15.4% of the adult population responded to live in *'free union'*. Against 84.3% who remain in another civil status and 0.3% who did not responded. Apart of **DD6**, no other feature has irregularity in their cardinality. For **DD7** and **DD8**, the mode captures a high percentage of the data; however, it is desirable for this research. DD7 asks people about suffering some chronic disease and **DD8** asks

Table 3.11: Quality report of categorical-demographic features.

Feature	Count	Miss	Card.	Mode	MF	M%	M2	MF2	M2%
DD2	338	0	2	Female	188	55.62%	Male	150	44.38%
DD3	338	1	5	Teacher	143	42.31%	Employee	70	20.71%
DD4	338	0	4	More intellectual than physical	156	46.15%	intellectual	143	42.31%
DD5	338	2	5	SDA	150	44.38%	Catholic	129	38.17%
DD6	338	0	1	Free Union	338	100%	NA	NA	NA%
DD7	338	0	5	No	284	84.02%	Other	26	7.69%
DD8	338	1	5	No	272	80.47%	Other	34	10.06%

about some crisis disrupting sleep. Our interest is in analyzing healthy people, thus, *'any'* and *'no'* are the best answers for **DD7** and **DD8**. Missing values represent a very low percentage of the data, so it is feasible we eliminate the records with missing values.

3.2.3 Data Quality Plan

We summarizes in Tables 3.12 and 3.13 the issues we founded in data quality analysis and the strategies to attend them.

3.2.4 Data Quality Plan Implementation

Analysis of Scripts for Restructuring and Validating the Raw-dataset

The first step we performed in order to resolve the issues, was the analysis of the code for restructuring and validating the raw dataset. This analysis avoided the suspicion that the wrong data has been generated by code bugs and the persistent errors were adjudicated to the wrong human capture. After this analysis we fixed these three bugs:

1. The code validating civil status had a logical error.

2. The script computing the time of sleep based on *'Time to go bed'*, *'Time to wake up'* and *'Time to fall asleep'*, does not include the case of a person going to bed and got up twelve

Table 3.12: Data Quality Plan for continuous features.

Feature	Data quality issue	Strategy
SQ1	The feature contains wrong values (e.g. 0.0), if it means midnight, the value should be 12.00.	Inspect the validation and transformation scripts of raw data for possible bugs, if any, correct them.
SQ4	The feature contains negative values, which is wrong since this feature captures the wake-up time.	Inspect the validation scripts for bugs, if any, correct them, otherwise, eliminate the records with this issue.
SQ2	Its standard deviation is too large	Execute a process to exclude records containing outliers with the aims of improve the accuracy of the model.
All	Missing values	Imputation is not feasible because we have not previous studies to know the trend of the data. As percentage of missing values is very low, we will eliminate the records with missing values.

hours apart.

3. The script converting the "*data time features*" to "*double*", avoided two critical cases: values in the range of $(-\infty, 0)$ must be NA, and values in the range of $[0, 1)$ must be transformed by adding 12 units.

Tables 3.14 and 3.15 show the results of processes again the raw dataset after correcting the scripts. The differences in involved features are notorious.

The minimum value in **SQ1** is 1.00 instead of 0.0 as before. The minimum value in **SQ4** is not negative and corresponds to the minimum value of **SQ4** in the raw dataset. However wrong data due to a human capture remains in this feature: "*a respondent who said they go to bed at 12:30 and wakes up at 12:42 every day*". On the other hand, the civil status (**DD6**) has congruent values for the SP, the 53.55% of respondents said to be married and 35.21% to be single, in contrast to the previous report where the 100% of the people seemed to have reported living in free union.

Table 3.13: Data Quality Plan for Sleep Hygiene Factors.

Feature	Data quality issue	Strategy
SH8	The mode is very high (>94%)	Analyze the relevance of including this feature in the analysis due its low variability.
SH9	The mode is high (>81%)	Analyze the relevance of including this feature in the analysis due its low variability.
All	Missing values	Use of imputation technique to replace the missing values with the most probable values.

Table 3.14: Data quality report of continuous features after correcting the scripts.

Feature	Count	Miss	Card.	Min	Q1	Median	Q3	Max	μ	σ
SQ1	338	0	27	1	10	11	11.5	12.75	10.17	2.46
SQ4	338	3	107	0.12	6.17	6.89	7.75	11.95	6.94	1.33

Missing Values

We deleted records containing missing values in the continuous and the categorical demographic features as we proposed in the data quality plan (see Table 3.12). After we deleted these records, the dataset was reduced from 338 to 324 rows, which is a reduction of 4.1% from the original.

Deleting Outliers in SQ2 and SQ4

Atypical values in relevant features can affect the accuracy of a model. In this case, outliers in **SQ2**: *'time between a person go to bed and he/she fall asleep'* and **SQ4**: *'time the person spent asleep'* could indicate a sleep disorder. Thus, we exclude these records using the Zscores method implemented in the outliers R package [140, 18]. We use the Z distribution and a range of $\cong$ three standard deviations ($p = 0.99$) to discriminate the outliers.

After we deleted the records containing outliers, the dataset reduced its dimensionality from 324 to 304 records, which represents a reduction of 10.1% from the original. Fig. 3.4 shows histograms and box plots after we deleted the records with outliers in **SQ2** and **SQ4**.

Table 3.15: Data quality report of demographic features after correcting the scripts.

Feature	Count	Miss	Card.	Mode	MF	M%	M2	MF2	M2%
DD6	338	2	5	Married	181	53.55%	Single	119	35.21%

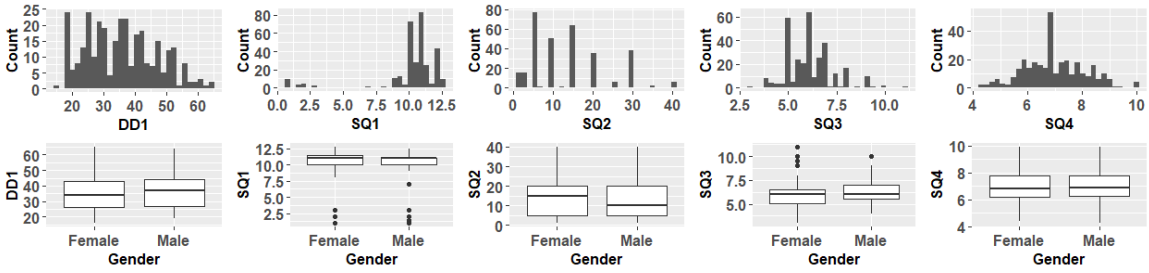


Figure 3.4: Histograms and Box-plots after deleting outliers in DD1,SQ2 and SQ4.

Imputation of Missing Values in SHF

Figure 3.5 shows columns with missing values. The histogram depicts incidence of missing values in each feature and the right panel illustrates its distribution in the dataset. The complete records represent 97.06% of data as shown in the lower part of the panel, while the remaining percentage containing missing values is shown in the rest of the panel indicating the percentages by feature.

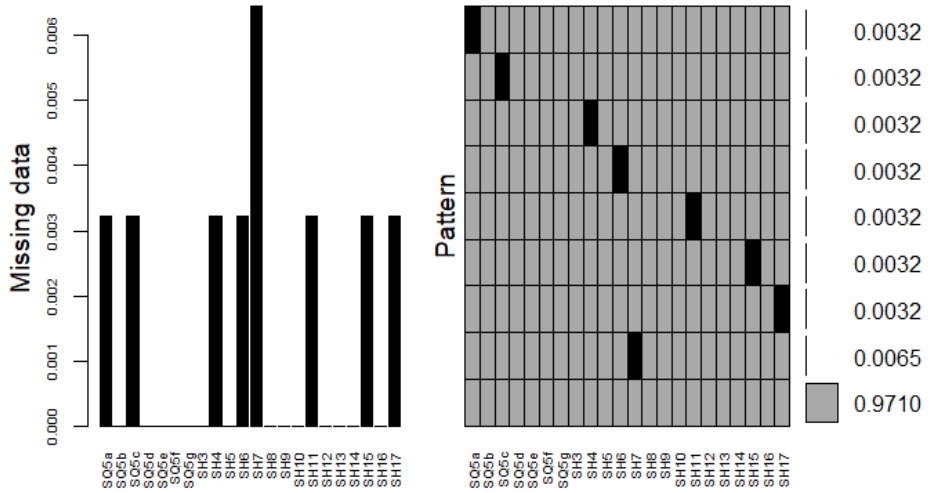


Figure 3.5: Graphical representation of missing values.

Table 3.16 is a report of missing values per rows and columns. In this document we show partially the table for space reasons, however, the columns replaced by "..." are completed.

Table 3.16: Report of missing values in Sleep Hygiene Factors.

	SQ5b	**...**	**SH21**	**SQ5a**	**SQ5c**	**SH4**	**SH6**	**SH11**	**SH15**	**SH17**	**SH7**	
297	1	...	1	1	1	1	1	1	1	1	1	0
1	1	...	1	0	1	1	1	1	1	1	1	1
1	1	...	1	1	0	1	1	1	1	1	1	1
1	1	...	1	1	1	0	1	1	1	1	1	1
1	1	...	1	1	1	1	0	1	1	1	1	1
2	1	...	1	1	1	1	1	1	1	1	0	1
1	1	...	1	1	1	1	1	0	1	1	1	1
1	1	...	1	1	1	1	1	1	0	1	1	1
1	1	...	1	1	1	1	1	1	1	0	1	1
	0	...	0	1	1	1	1	1	1	1	2	9

Table 3.17 shows a summary of relevant columns in Table 3.16.

Table 3.17: Summary of missing values in Sleep Hygiene Factors.

Feature	**SQ5a**	**SQ5b**	**SH4**	**SH6**	**SH7**	**SH11**	**SH15**	**SH17**	**Total**
Missing values	1	1	1	1	2	1	1	1	9

We performed data imputation on 25 features, 21 are the data source of the model, and the other four features reveal relevant information on SQ; nevertheless, they will not be involved in the generation of the model. We use the Polytomous Regression Imputation (PRI) implemented in the mice method [15] for categorical data with more than two levels. We executed the method in a temporal dataset composed of features with missing data. We configured the function to performs 50 iterations generating five datasets to choose the best option of imputation.

We chosen from Table 3.18 the dataset matching more values to the other proposed datasets. To find it, we perform bestDSforImputationbySimilarity function, which finds similarities through euclidean distances between each row of one dataset with respect to the others in the pool. The function returns a two elements list. The first element is a vector with the count of similarities for each dataset, the second element is the index of dataset with highest similarity. We show in Table 3.19 the result of first element in the list.

Table 3.18: Suggested datasets for imputation.

Number of Row	DS1	DS2	DS3	DS4	DS5
ROW 225	1	2	2	2	2
ROW 72	0	2	0	0	3
ROW 12	0	0	0	1	0
ROW 233	1	2	0	1	0
ROW 144	0	0	1	1	3
ROW 183	1	1	4	1	0
ROW 201	0	1	0	0	0
ROW 17	0	3	0	1	2
ROW 146	0	0	0	1	0

Table 3.19: Matching values per dataset.

DS1	DS2	DS3	DS4	DS5
16.00	12.00	17.00	12.00	13.00

After we observed the outcomes of Table 3.19, we chosen the 'DS3' to impute missing data because it has the highest score of similarity. We impute missing values using the 'complete' function of the mice R package.

3.2.5 Results of Data Quality Plan Implementation

Table 3.20 shows the data quality report after implementing the Data Quality Plan to guarantee the quality of the dataset.

Except for high percentages in the mode of **SH8** and **SH9**, all other features have quality in the dataset. We attended the issues of **SH8** and **SH9** through the feature selection process in the next stage described in Chapter 4. At the end of this process, we had a dataset ($m = 304, n = 62$) composed of 12 continuous and 50 categorical features.

Table 3.20: Data Quality Report of Sleep Hygiene Factors after implementing the Data Quality Plan.

Feature	**Count**	**Miss**	**Card.**	**Mode**	**MF**	**M%**	**M2**	**MF2**	**M2%**
SH1	304	0	5	1	108	35.53%	0	96	31.58%
SH2	304	0	5	1	109	35.86%	2	105	34.54%
SH3	304	0	5	1	146	48.03%	2	70	23.03%
SH4	304	0	5	0	184	60.53%	1	63	20.72%
SH5	304	0	5	0	157	51.64%	2	56	18.42%
SH6	304	0	5	0	135	44.41%	1	69	22.7%
SH7	304	0	5	0	128	42.11%	1	98	32.24%
SH8	304	0	5	0	290	95.39%	3	5	1.64%
SH9	304	0	4	0	253	83.22%	1	29	9.54%
SH10	304	0	5	0	200	65.79%	1	52	17.11%
SH11	304	0	5	0	93	30.59%	2	69	22.7%
SH12	304	0	5	1	122	40.13%	2	98	32.24%
SH13	304	0	5	2	85	27.96%	1	69	22.7%
SH14	304	0	5	0	187	61.51%	1	54	17.76%
SH15	304	0	5	0	93	30.59%	1	79	25.99%
SH16	304	0	5	0	149	49.01%	1	76	25%
SH17	304	0	5	0	198	65.13%	1	78	25.66%
SH18	304	0	5	0	156	51.32%	1	73	24.01%
SH19	304	0	5	0	111	36.51%	1	77	25.33%
SH20	304	0	5	2	105	34.54%	1	85	27.96%
SH21	304	0	5	1	122	40.13%	2	87	28.62%

Chapter 4

Feature Selection

Chapter 4 explains the process and techniques we used to find the relevant SHF in the SP. We deal with this challenge though dimensionality reduction techniques by using different algorithms. There exists two approaches in dimensionality reduction, Feature Extraction (FE) and FS. FE consists in generating a new and small feature space producing new features from the original. The new datasets are not understandable in terms of original ones, rather, new ones are abstractions of originals and their visualization have no practical meaning. On the other hand, FS based on an evaluation criterion extracts a small subset of the relevant features from the original dataset as shown in Fig. 4.1. This subset usually leads to better learning performance, lower computational costs, and better model interpretability by discarding irrelevant, noisy and redundant features [37, 141].

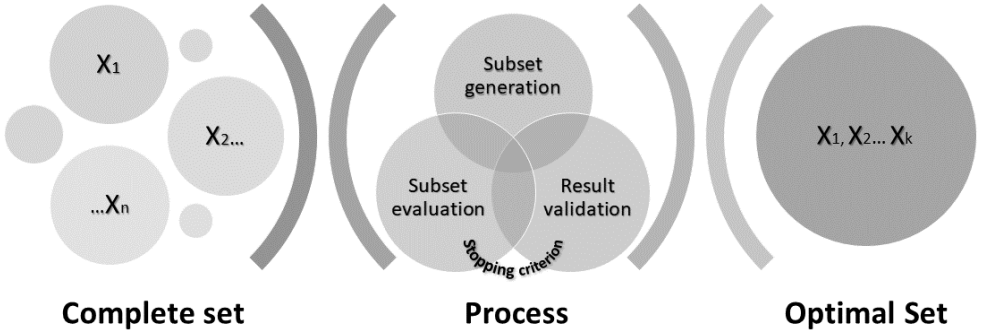


Figure 4.1: Feature selection process.

What approach is better? It depends on the nature and restrictions on the problem. For purposes of this study, the feature selection approach is most appropriate. We have the purpose of decreasing the number of predictive variables due to the high cost of design and infrastructure that means capturing 21 different signals through sensors. If it is possible to characterize a high percentage of the phenomenon through a reduced number of SHF, the design and implementation of a system will be more feasible and less expensive.

4.1 Feature Selection Models

In 1996, Liu et. al. [32] propose two models to achieve reduction of features that researchers have been used to develop diverse algorithms still in force. The Filter Models (FM) (see Figure 4.2) proposes to use attributes to the data domain as criteria of feature selection. They proposed analyzing and deciding over irrelevance/relevance of features based on one of four measures: '*Information Gain (IG)*', '*dependence*', '*distance*' and '*consistency*'.

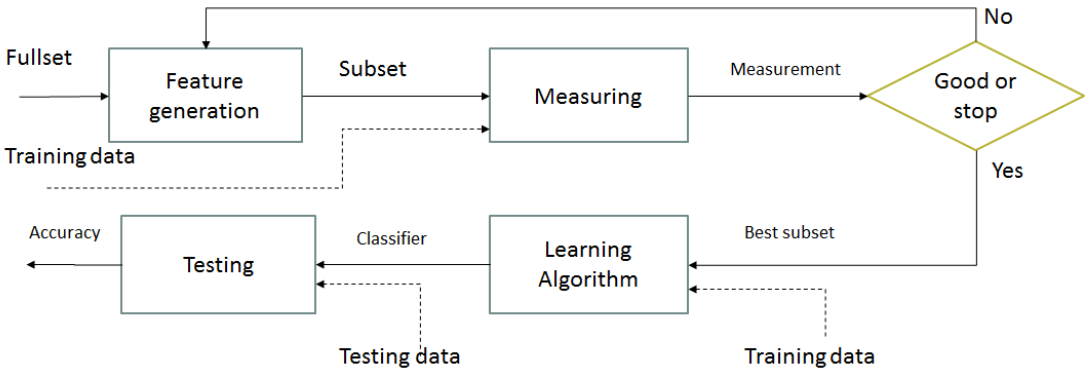


Figure 4.2: Filter Model Liu et. al. [32].

We show in Fig. 4.3 the Wrapper Models (WM) using prediction accuracy as selection criterion. The methods designed under this paradigm commit to the data analyst in this preliminary stage of the learning process, to use a particular classifier to produce the models in the future.

We explore the literature on performance of the algorithms using both paradigms, and we describe the main advantage and disadvantages. Techniques based on FM perform better than others based on WM; however, filter approach ignores prediction accuracy during the feature

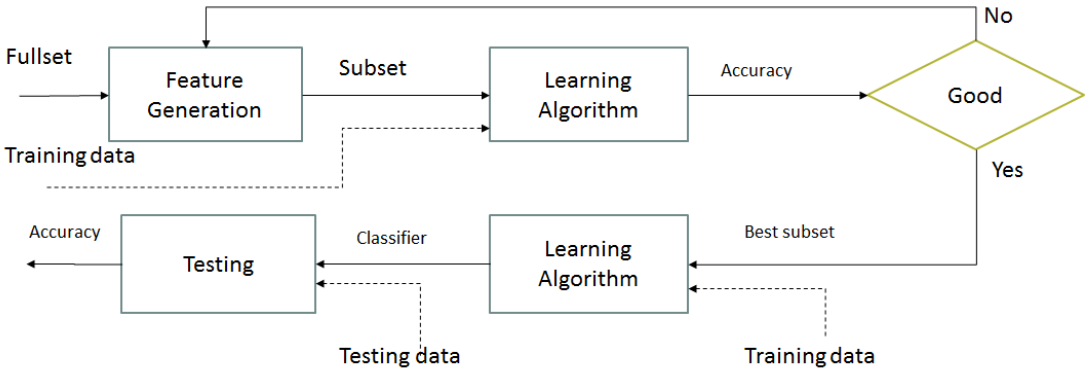


Figure 4.3: Wrapper Model Liu et. al. [32].

selection process [37]. Some analysts prefer not to take this risk, because if prediction accuracy is not achieved at the proposed level, the first steep can be regarded as a waste of time. On the other hand, some researchers argued that selecting features through a classifier, limits the analyst to use another classifier for prediction. In this sense, the analyst should choose at the beginning the classifier to generate the final model. [135] argues to reasons to use FM instead of glswm: FM are less computationally expensive than WM and they have shown generate models achieving good accuracy.

[32] highlights three main dimensions to chose the most appropriate technique of FS and proposes the schema of Fig. 4.4.

Search: It refers to the way in which features are chosen: Deterministic, heuristic or complete.

Scheme: It determines if the search will be forward, backward or randomly.

Measure: It Establishes the way to set the threshold to stop the search of relevant features. (precision, consistency or the classic criteria: distance, IG and dependency).

Description of relevant measures:

Distance: The main goal to use distance, is to find similarities among instances in a dataset. The Equation proposed by Minkowski (see Eq. 4.3) is a generalization of the distances used in MLA. The most common are the particular cases $p = 1$ called Manhattan Distance (see

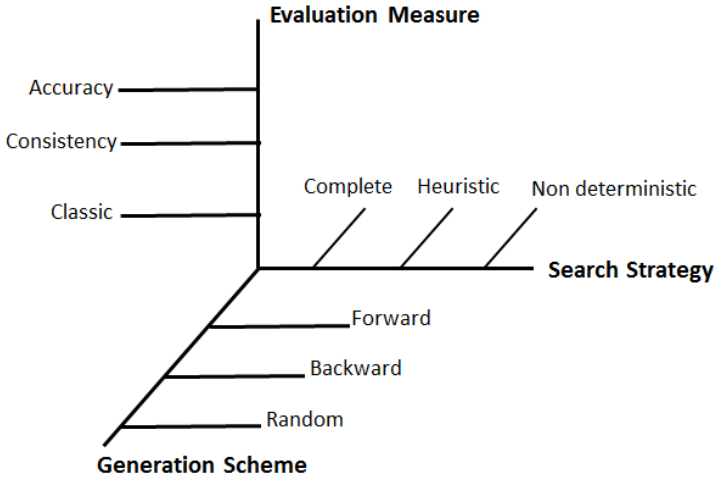


Figure 4.4: Main dimensions in Feature Selection, Liu et. al.

Eq. 4.2) and $p = 2$, the well known Euclidean Distance (see Eq. 4.1)[1]. The implication using different values of p will be noticed in the difference of distances between two values of any feature. Since the distance are directly proportional to the value of p, the large differences between two features in an instance, impact stronger in the result when p grows.

$$Euclidean(A, B) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_1)^2 + \cdots + (a_n - b_n)^2} \tag{4.1}$$

$$Manhattan(a, b) = \sum_{i=1}^{m} abs(a[i] - b[i]) \tag{4.2}$$

$$Minkowski(a, b) = \left(\sum_{i=1}^{m} abs(a[i] - b[i])^p \right)^{\frac{1}{p}} \tag{4.3}$$

Accuracy: Accuracy refers to the successes of a model to predict target of instances in a dataset. It opposes to misclassification error as we can see in 4.4 and 4.5 equations. These two equations normally are represented in a confusion matrix together to a Receiver Operating Characteristics (ROC) curve. They provide understanding and visualization of the

[1] We took the three equations from [135]

specificity and sensibility, two relevant metrics for evaluations of models in health contexts.

$$misclassification - rate = \frac{(FP + FN)}{(TP + TN + FP + FN)} \tag{4.4}$$

$$accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{4.5}$$

Inconsistency: An inconsistency occurs when two instances have the same value in all descriptive features, but they belong to a different class. We can compute two values to measure the inconsistency for a subset of features in a dataset. The first value called the inconsistency count (IC) can be defined as $IC = nM - LCI$, where nM is the number of instances that coincide in all descriptive features, and LCI is the largest class of the classes in this group of instances. The second value is the inconsistency rate defined as $IR = \frac{\sum_{i=0}^{m} IC_i}{N}$, where m is the total of the groups of matching instances in the dataset.

Information Gain: To explain IG, we refer to the Information Entropy concept described in the math theory of communication, developed by Shannon in 1948 [142]. Shannon formulated the Equation 4.6 to quantify the uncertainty level when choosing elements of a vector randomly. In a dataset, the entropy describes the heterogeneity/homogeneity of the target feature. If we can predict the choice of an element of the target feature with a high likelihood of success, entropy is low and vice versa.

The process to compute IG is summarized in the following algorithm:

1. Calculate the total entropy for the target feature through Eq. 4.6.
   For each predictive feature:
   (a) Split the target feature according to the classes of predictive feature.
   (b) Calculate entropy (Ec. 4.6) in each subset of the target feature and times the result by $\frac{sc_t}{n}$ where sc_t is the count of records in the subset of t class and n is the total count of records in the dataset.
   (c) Sum entropies of all subsets.

2. Calculate IG per each predictive feature through subtracting the result of step two, from the total entropy calculated in step one.

$$H(t) = -\sum_{i=1}^{l}(P(t=i)\log_s(P(t=i))) \tag{4.6}$$

In Equation 4.6, $P(t=i)$ is the chance to choose an element of the i class randomly. l is the number of classes in the dataset and $\lg_2(P(c=i))$ is the penalty for the computation depending of the i probability. We use $\log_2$ because we work with information, however, we can chose arbitrary the base for the log. We can calculate the entropy of a dataset with n number of records in a vector with l classes if we compute the product $P(i)\lg_2(P(i))$ per class and finally we sum all the products. The sign at the beginning of the equation is the corresponding sign to the negative term of the product $(-\lg_2(x))$. It has been subtracted from the equation to simplify it. [135].

A third model for feature selection is Embedded Models (EM). Models in this paradigm allow selecting relevant features during the training. EM have the advantage of filter models in terms of low computational cost and the advantage of wrapper models, since the analysts can compute the prediction accuracy during the selection process. [37] describes three types of embedded models we show in Table 4.1.

Table 4.1: Embedded Models (quoted verbatim from [37]).

Model	**Description**	**Reference**
Pruning	Utilizing all features to train a model and then attemp to eliminate some features by setting the corresponding coefficients to 0, while maintaining model performance such as recursive feature elimination using Support Vector Machine (SVM)	[143]
Build-in	Mechanism for feature selection as ID3 and C4.5	[144, 145]
Regularization	Utilizes objective functions that minimize fitting errors and in the mean time force the coefficients to be small or zero.	[146]

These models are representative of theoretical basis where a lot of algorithms for FS in the last twenty years have been fueled.

4.1.1 Feature Selection Process

We proposed a merge-rank method for selecting the best set of features (See Fig. 4.5). We selected six methods for feature selection of categorical features according with the nature of entries in the dataset. Each method selected relevant features based on their own criterion, then we performed a merge with the results of the consistent methods to chose the first 'k' matching features at the beginning of each list. Since each method differs to others in strategy and math, those features prevailing as selected in all consistent methods, are considered the most relevant features. We explain briefly below, the rationale behind of each FS method used in the merge-rank algorithm.

Random Forest: Breiman designed the RF in 2001, a robust algorithm for classification [147]. It uses the Bagging technique[2] to combine a large collection of no correlated decision trees generated through Classification and Regression Tree (CART) algorithm. Each tree is generated from a subset of instances that the algorithm chose randomly with replacement and a random subset of features to avoid correlation. For this work, we use RF for selecting the best subset of features by generating a forest of 1000 trees and measuring the mean of classification accuracy after permuting a set of instances over all trees.

CV.LASSO: The Least Absolute Shrinkage and Selection Operator was formulated in 1996 by Robert Tibshinari [148] based on the Nonnegative Garrote algorithm introduced by Breiman one year early. The LASSO performs a regression with l_1 norm regularization, which ensures that all coefficients of no relevant features will be zero in the process of minimizing the function. All coefficients greater than zero will be part of the best subset of features.

Relief: Relief is a feature weight based algorithm proposed by Kira and Randell in 1992 [149]. This algorithm evaluates the consistence of each feature to describe the target. Relief selects

[2]Combination of learning models increases the classification accuracy.

instances at random, then for each instance it searches the positive and negative nearest neighbor using the Euclidean Distance. In a dataset where the target feature consists of two classes, negative neighbors does not match in the target feature with the selected instances and positive neighbors do. In the next step Relief compares the value of each feature in the instance with each feature in the neighbors. For each feature Relief assigns weights based in the following rule: If the value of selected instance match with the value of positive neighbor the algorithm adds one, otherwise, adds 0. If the value of selected instance matches with the value of a negative neighbor the algorithm adds 0, otherwise, adds 1. Finally, Relief computed the mean per feature and compares it with a given threshold. Relief selects the features above of the threshold.

Best First Search (BFS): This algorithm is based on the Greedy algorithm and was proposed firstly as a feature selection algorithm by Kohavi and John in 1997 [150]. It is a wrapper algorithm in a forward scheme and heuristic strategy. At the beginning of the process BFS predicts the target feature based on each predictive feature. The high accuracy is kept as 'best-accuracy' and the feature that produced the best result is added to the 'optimum subset'. After that, the algorithm performs a recursive process to combine the first selected feature with each of features in the remaining set, and it computes the accuracy of prediction for each combination. If the high accuracy of new results exceed 'best-accuracy' in at least an ϵ (normally 1%), the new feature is added to the 'optimum subset', otherwise, the algorithm stops its search and returns the 'optimum subset'. The algorithm may end at any time and leave out combinations with high accuracy; however, it is a trade-off between accuracy and computational cost. It is a risk worth taking as the algorithm has proven to be robust in various datasets.

Chi Square (χ^2): Chi-square is a widely used statistical test to measure independence between to events described in categorical variables. High values of χ^2 indicate that the hypothesis of independence has no support, which means that tested variables have correlation. χ^2 is used in feature selection by testing each predictor feature against the target feature. After this process, results are ranked descending and features with high scores are selected as the relevant ones.

Information Gain: The rationale behind of the algorithm of selecting features through IG based in the Information Entropy theory consists in chose as relevant, those features that organize the target the most homogeneity (low entropy). We explain Entropy and the the algorithm to calculate IG in Section 4.1.

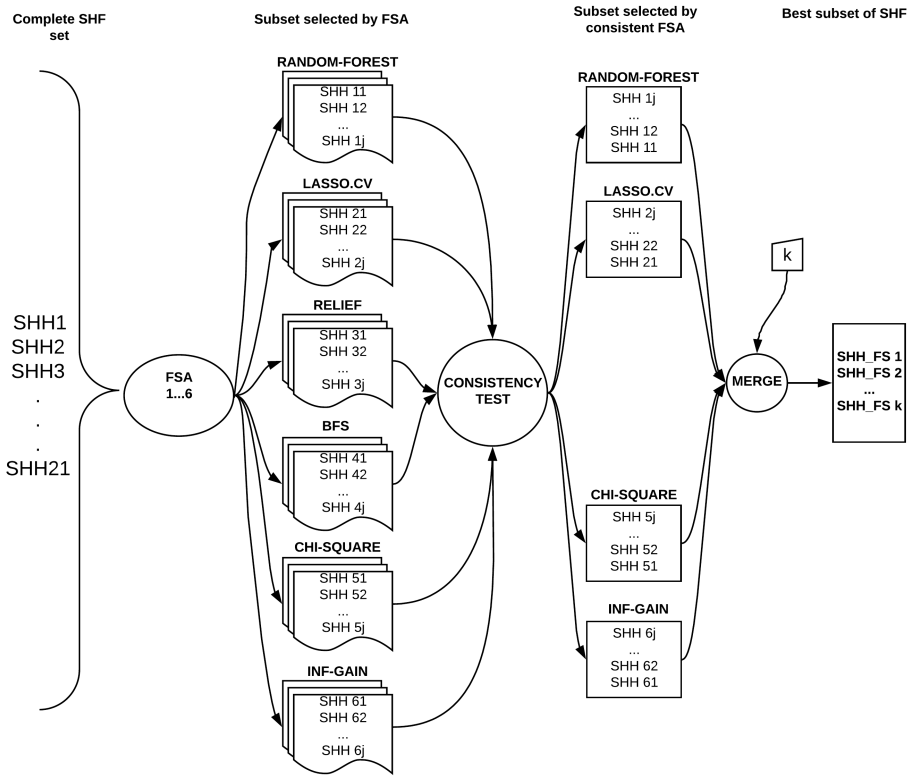


Figure 4.5: Merge-rank Method for feature selection.

Each process in the method of Merge-rank is explained in the next itmes:

1. **Feature Selection (1..6):** Each one of six methods had as data input, the entire set of features to generate an optimal subset $(1, 2..., j)$ according with its own selection criterion.

2. **Consistency Test:** The Merge-rank tests consistency of each method by executing multiple times the algorithm and comparing the resulting subsets. Relief and BSF methods did not pass the consistency test due to differs in each execution of the algorithm, thus they were discarded for the final process. Table 4.2 shows the weights that consistent algorithms

assigned to each SHF after perform FS on the complete dataset. The 0.00 value indicates that the algorithm assigned a very low value, while (—) means that the algorithm assigned no value to the factor. We show in this table the values in the scale that each algorithm uses to calculate weights.

Table 4.2: Assigned Weight per Algorithm.

Factors	Feature Selection Methods			
	Random Forest	Chi Square	CV.LASSO	Information Gain
NAP	—	0.00	—	0.00
SLEEP-TIME	0.57	0.25	0.15	0.03
WUP-TIME	0.29	0.00	—	0.00
EV-EXERCISE	—	—	—	—
MO-EXERCISE	—	—	—	—
AF-EXERCISE	—	—	—	—
STAY-IN-BED	—	—	—	—
TOBACCO	—	—	—	—
ALCOHOL	—	—	—	—
CAFFEINE	—	—	—	—
INT-ACT-BS	—	—	0.01	—
STRESS-BS	1.24	0.39	0.49	0.08
N-SLEEP-BU	—	—	—	—
UNC-BED	0.24	—	0.07	—
LUMINOSITY	—	—	—	—
NOISY	—	—	—	—
COLD	—	—	—	—
HEAT	—	—	—	—
INT-WK-BS	0.46	—	—	—
WORRIES-BS	0.62	0.31	0.16	0.05
DINNER	—	—	—	—

3. **Ordering:** In this step the algorithm ordered in descending way the factors of Table 4.2 according with their weights, posteriorly the algorithm extracts the first k factor of each set which we shown in Table 4.3.

4. **Merging:** As final task, the algorithm merge each subset and it selects the factors repeated in each list to conform the final subset of features to train the models. Fig. 4.6 shows the consistency of four algorithms in determining the weights for each selected factor.

Table 4.3: Best set of SHF per Feature Selection Algorithm.

Random Forest		**Chi Square**		**CV.LASSO**		**Information Gain**	
Factor	**W**	**Factor**	**W**	**Factor**	**W**	**Factor**	**W**
SBS	1.24	SBS	0.39	SBS	0.49	SBS	0.08
ST	0.57	WBS	0.31	WBS	0.16	WBS	0.05
WBS	0.54	ST	0.25	ST	0.15	ST	0.03
IABS	0.44	NA	0.00	UB	0.07	NA	0.00
WT	0.29	WT	0.00	IABS	0.01	WT	0.00
UB	0.24	—	—	—	—	—	—

IABS: Intellectual Activity Before Sleep, NA: NAP, SBS: Stress Before Sleep, ST: Sleep Time, UB: Uncomfortable Bed, WBS: Worries Before Sleep, WT: Wake up Time.

Table 4.4: Sleep Hygiene Factors selected after applying the merging process.

Factor	**RF**	χ^2	**CV.LASSO**	**Entropy**
STRESS-BS (SH12)	1.78	2.67	1.44	1.65
WORRIES-BS (SH20)	0.78	2.13	0.46	1.02
SLEEP-TIME (SH2)	0.82	1.67	0.44	0.65

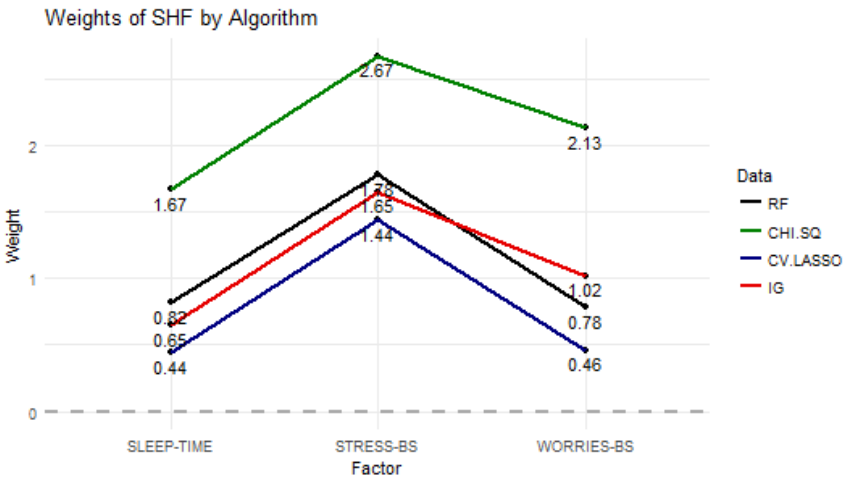


Figure 4.6: Comparative plot of results per algorithm.

4.2 FS Evaluation

We evaluated results of feature selection process by comparing accuracy to predict sleep quality through three different MLA and their variants. We use as data source the dataset of three selected features and the complete dataset of 21 features. To chose the algorithms we use to evaluate the selection of the SHF, we reviewed literature and selected the most used MLA for predicting and classifying (See Table 4.5). After, we performed a process to select three of them based in criteria described below.

Table 4.5: MLA most used in reviewed literature.

MLA	Type	Math	AOL
R-linear	Supervised	$J(\theta) = \frac{1}{2}m \sum_{i=1}^{m} \left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2, \frac{\partial y}{\partial x}$	Yes
R-logistic	Supervised	$J(\theta) = -\frac{1}{m}\sum_{i}^{m}\left[y^{(i)} \log h_\theta\left(x^{(i)}\right) + \left(1 - y^{(i)}\right) \log\left(1 - h_\theta\left(x^{(i)}\right)\right)\right]$	Yes
ANN	Supervised	$z = \sum_{x=1}^{n} w_i x_i + w_b b, a = \sigma(z) = \frac{1}{1+e^{-z}}$	Yes
K-means	No supervised	$min\left(d = \sqrt{(cx_{1\ldots n} - x_{1\ldots n})^2 + (cy_{1\ldots n} - y_{1\ldots n})^2}\right)$	No
D-Trees	Supervised	$H(t) = -\sum_{i=1}^{l} P(t=i)\log_2(P(t=i))$ or $G(t,D) = 1 - \sum_{l \in levels(t)} P(t=l)^2$	No
KNN	Supervised	$M_k(q) = \underset{l \in levels(t)}{\arg\max} \sum_{i=1}^{k} \frac{1}{dist(q,d_i)^2} * \delta(t_i, l)$	No
SVM	Supervised	$J(\theta) = \min_\theta C \sum_{i=1}^{m} [y^{(i)} cost_1(\theta^T x^{(i)}) + (1-y^{(i)}) cost_0(\theta^T x^{(i)})] + \frac{1}{2}\sum_{j}^{n}\theta_j^2$	Yes

Algorithms fitting the following criteria were chosen: 1) The algorithm was designed to train models in supervised approach; 2) The math behind of the algorithm is not stochastic and can be explained; 3) The algorithm could be used to train models in '*online mode*', it means, models are trained as data arrives.

From Table 4.5, AOL is the column that most contributes to discriminate the algorithms. Four algorithms meet this criterion, however, linear regression has been discarded because its utility is limited to problems where target feature is continuous. Based on the above, algorithms chosen for the evaluation are: Logistic Regression (LR), SVM and Artificial Neural Network (ANN).

The process carried out to evaluated the efficiency of FS is described below.

1. We composed two subsets of the data to train the models and compare the prediction

accuracy. The first set (**F21**) was composed of the 21 SHF (**SH1**, **SH2** ... **SH21**) as predictive features and SQ as target feature. The second set (**F3**) was composed of the three selected factors (SLEEP-TIME (**SH2**), STRESS-BS (**SH12**) and WORRIES-BS (**SH20**)) by the Merge-rank algorithm as predictive features and SQ as target feature. Each predictive feature can take values between 0 and 4, while the target feature is defined on the PSQI scale criterion as follows:

$$SQ = \begin{cases} 0 \leq sum(SQTT) \leq 5 & \Rightarrow \quad 0 \ (good\ sleep\ quality.) \\ 6 \leq sum(SQTT) \leq 21 & \Rightarrow \quad 1 \ (poor\ sleep\ quality.) \end{cases}$$

Table 4.6: MLA and its variants generating eight predictive models.

MLA	Variant of algorithm	Abrreviation
Logistic Regression	Linear polynomial	LR-DG-1
	Two degree polynomial	LR-DG-2
	Three degree polynomial	LR-DG-3
ANN	One hidden layer three neurons	NN-3
	One hidden layer four neurons	NN-4
SVM	Linear kernel	SVM-LN
	Radial kernel	SVM-RD
	Sigmoid kernel	SVM-SIG

2. We use the variants of three MLA in Table 4.6 and cross-validation technique[3] to train eight models to predict SQ with **F21** as a data source. We repeated ten times the process, and after, we calculated the mean and standard deviation per trained model (See Table 4.7).

[3]Cross Validation Technique divides the dataset into three parts: Training(60% - 80% of the data), Validation(10% - 20% of data) and Test(10% - 20% of data). The algorithm iterates by training the model with training subset, validating the prediction accuracy in the validation subset and adjusting the parameters to optimize the cost function. When criterion of '*stop*' is reached, the algorithm test the prediction accuracy of the model using the test subset. Since this subset has not been part of the training, the accuracy prediction of resulting model is close to what the model will have when predicting new instances. (See Fig. 4.7)

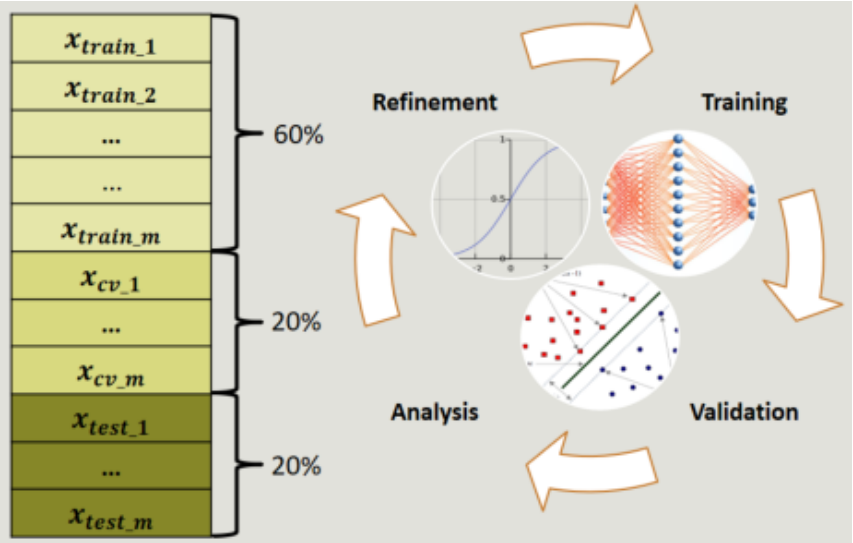


Figure 4.7: Cross Validation Technique in the process of models' training.

Table 4.7: Results of ten trainings of predictive models using F21.

MLA	**Runs**										μ	σ
	1	2	3	4	5	6	7	8	9	10		
LR-DG-1	0.62	0.69	0.67	0.67	0.64	0.6	0.67	0.64	0.64	0.67	0.65	0.03
LR-DG-2	0.71	0.73	0.56	0.62	0.62	0.62	0.80	0.69	0.73	0.71	0.68	0.07
LR-DG-3	0.44	0.40	0.64	0.64	0.64	0.62	0.42	0.58	0.73	0.71	0.58	0.12
NN-3	0.62	0.64	0.56	0.58	0.56	0.76	0.64	0.56	0.44	0.71	0.61	0.09
NN-4	0.51	0.56	0.58	0.51	0.64	0.62	0.53	0.47	0.69	0.64	0.57	0.07
SVM-LN	0.64	0.73	0.73	0.56	0.76	0.56	0.58	0.67	0.73	0.71	0.67	0.08
SVM-RD	0.58	0.53	0.60	0.53	0.64	0.47	0.73	0.69	0.58	0.71	0.61	0.09
SVM-SIG	0.51	0.67	0.69	0.58	0.73	0.62	0.56	0.58	0.56	0.60	0.61	0.07

3. We perform the step above using as data source the **F3** dataset (See Table 4.8).

Table 4.8: Results of ten trainings of predictive models using F3.

MLA	**Runs**										μ	σ
	1	2	3	4	5	6	7	8	9	10		
LR-DG-1	0.67	0.76	0.73	0.69	0.73	0.73	0.56	0.64	0.76	0.80	0.71	0.07
LR-DG-2	0.67	0.69	0.76	0.73	0.76	0.73	0.73	0.67	0.73	0.80	0.73	0.04
LR-DG-3	0.71	0.73	0.73	0.76	0.80	0.78	0.73	0.76	0.73	0.64	0.74	0.04
NN-3	0.58	0.60	0.62	0.71	0.60	0.56	0.62	0.56	0.69	0.60	0.61	0.05
NN-4	0.73	0.58	0.64	0.53	0.58	0.67	0.76	0.67	0.56	0.69	0.64	0.08
SVM-LN	0.73	0.76	0.67	0.69	0.78	0.67	0.71	0.69	0.76	0.69	0.72	0.04
SVM-RD	0.73	0.64	0.84	0.67	0.73	0.71	0.60	0.71	0.69	0.67	0.70	0.06
SVM-SIG	0.69	0.73	0.62	0.62	0.64	0.60	0.67	0.67	0.62	0.69	0.66	0.04

4. We compare the prediction accuracy for each pair of models training with **F21** and **F3**. We discussed the results in following section.

4.2.1 Results

Tables 4.9 and 4.10 show the means of accuracy in prediction by models and their variants, as well as the training time for each model and subset of data.

Table 4.9: Results for F21

MLA	Accuracy	Time
LR-DG-1	0.65	0.33
LR-DG-2	0.68	0.42
LR-DG-3	0.58	2.05
NN-3	0.61	0.17
NN-4	0.57	0.20
SVM-LN	0.67	1.15
SVM-RD	0.61	1.07
SVM-SIG	0.61	0.97

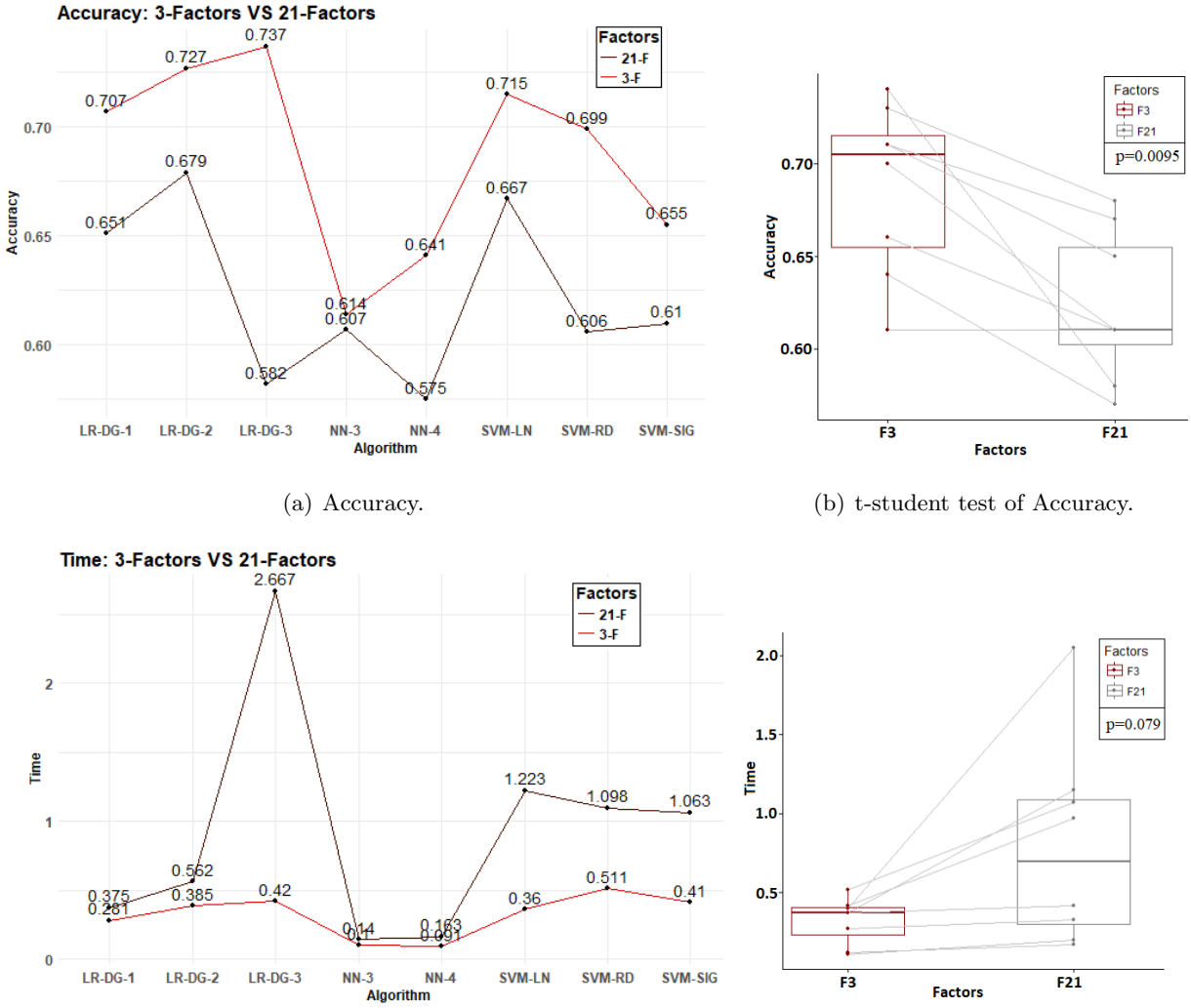
Table 4.10: Results for F3

MLA	Accuracy	Time
LR-DG-1	0.71	0.27
LR-DG-2	0.73	0.37
LR-DG-3	0.74	0.40
NN-3	0.61	0.12
NN-4	0.64	0.11
SVM-LN	0.71	0.37
SVM-RD	0.70	0.52
SVM-SIG	0.66	0.42

We perform normality and homogeneity of variances tests before we compare the means. Table 4.11 shows that all sets of means came from normal data. The Barttlet tests corroborate variance homogeneity between means of accuracy ($p = 0.7381$) and not variance homogeneity in means of time ($p = 0.0008$). We use t-student test to compare difference between means of accuracy and Wilcox test for means of time.

Table 4.11: Results of normality tests on data of precision and time

Data	p.value	Conclusion
Means of accuracy F21	0.4073	Strong evidence to not refute Ho
Means of accuracy F3	0.4241	
Means of time F21	0.1826	
Means of time F3	0.3142	
Ho : Data come from normal sample.	Significance Criterion: (p.value=0.05)	



(a) Accuracy.

(b) t-student test of Accuracy.

(c) Time.

(d) Wilcox test of Time.

Figure 4.8: Comparison of accuracy and time for dimensionality reduction through FS.

Figs. 4.8 (a) and 4.8 (b) show the means comparison of accuracy in prediction for the ten runs we executed to each subset of data. We can see in Figure (a) that the eight models we trained using **F3** report higher accuracy of prediction than their counterparts we trained with the **F21**. We tested Ho: *Difference between means is equal to 0* and rejected (p = 0.009) as we show in Fig. (b). We have strong evidence to say that means of accuracy in models differs when they are training using the two two datasets.

We perform a similar analysis with the time of training. We show in Figs. 4.8 (c) and 4.8 (d) that models we trained with **F3** spend less time than those we trained with **F21**. Nevertheless, the p-value (p=0.160) does not provide evidence to reject the Wilcox null hypothesis: "*Means of time are from the same population*". The time we spend in training models with three features is not statistically significant compared to the time we used for training models with 21 features. We suppose that this is due to the size of the datasets which does not present a challenge to the processor used to train the models.

4.3 Algorithm Selection

4.3.1 Accuracy assessment

We used the results in Table 4.8 to evaluate the performance of each algorithm in their variants with the purpose of selecting the best algorithm to train a model with this dataset.

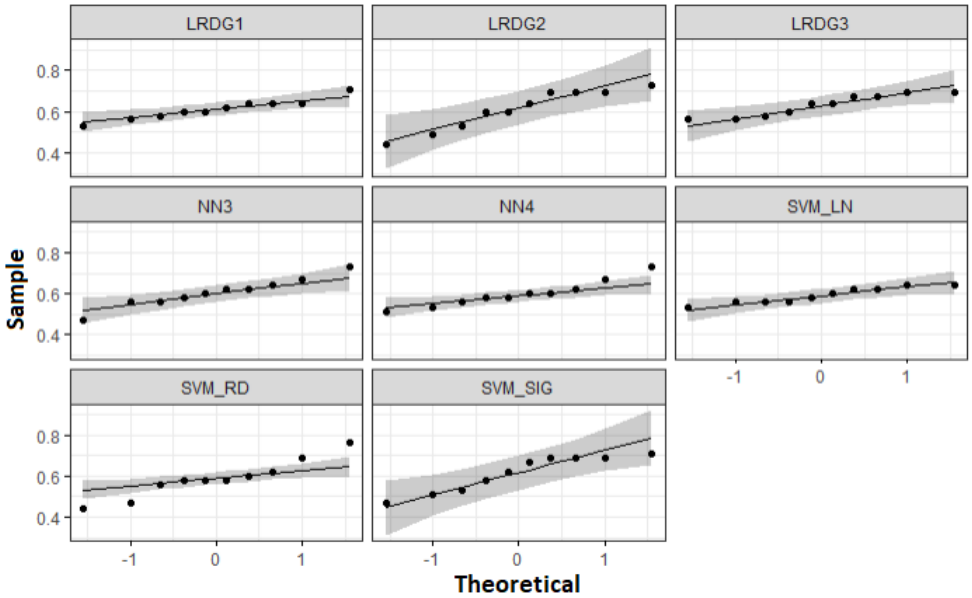


Figure 4.9: Normal Q-Q plots of prediction accuracy for each MLA.

Before performing the statistical test, we verify the nature of the data (parametric/non-parametric) to guarantee the quality of the results. Fig. 4.9 shows that all samples are from normality data and a Bartlett test gives us evidence of variance homogeneity ($p = 0.09285$)

among all samples as we can see in the following result:

```
Bartlett test of homogeneity of variances
    data:  accalg$ACC and accalg$ALG
    Bartlett's K-squared = 12.243, df = 7, p-value = 0.09285
```

After we corroborated the parametric nature of the data, we perform Analysis Of Variance (ANOVA):

```
Analysis of Variance Table
Response: accalg$ACC
               Df  Sum Sq   Mean Sq    F value   Pr(>F)
    accalg$ALG   7  0.01341  0.0019164   0.3693   0.9173
    Residuals   72  0.37366  0.0051897
```

A p-value (p=0.9173) strongly support the evidence of no significant differences among the means of the samples. The accuracy in prediction is not a relevant criterion to choose the algorithm that should be used in generation of predicting models for this data.

4.3.2 Assessment of Time

Similarly to accuracy, we verify the nature of data by selecting a parametric or non parametric method to compare the means. Visually, some plots in Fig. 4.10 shown points out of normality threshold, thus we performed the Shapiro-Wilk test to contrast Ho:"*Data come from normal data*" at 0.05 of significance.

Table 4.12: Normality test on training time samples.

	LR-DG-1	**LR-DG-2**	**NN-3**	**NN-4**	**SVM-LN**	**SVM-SIG**
p-value	0.0002	0.0303	0.0001	0.0012	0.0064	0.024

Table 4.12 shows six out of eight tests indicating no normality in the samples ($p < 0.05$). Due to this results, we use the non parametric Kruskal-Wallis test to proof Ho:"*Difference among all means of samples are equal to 0*". We found the following:

```
Kruskal-Wallis rank sum test

data:  Time by MLA

Kruskal-Wallis chi-squared = 64.359296, df = 7, p-value = 0.02E-9
```

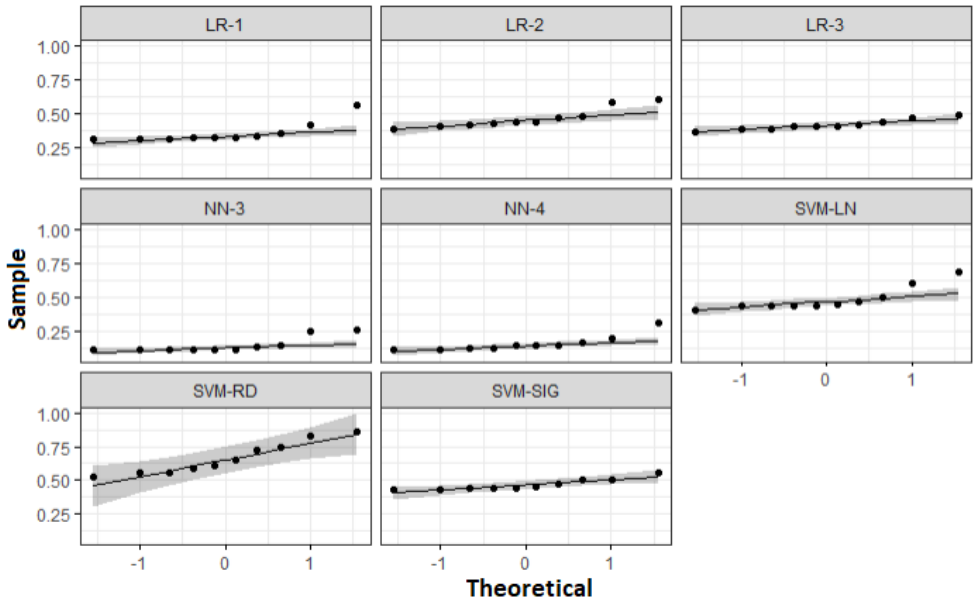


Figure 4.10: Normal Q-Q plots of execution time for each MLA.

$p - value < 0.05$ gives us strong evidence to refute the null hypothesis. There is significant difference among means of the executing time. In order to find the pairs of samples with significant differences we perform a Parwise-Wilcox test obtaining the results shown in Table 4.13.

Table 4.13: Pairwise comparisons using the Wilcoxon rank sum test.

	LR-1	**LR-2**	**LR-3**	**NN-3**	**NN-4**	**SVM-LN**	**SVM-RD**
LR-2	1.15E-02	-	-	-	-	-	-
LR-3	5.75E-02	0.345	-	-	-	-	-
NN-3	9.92E-05	9.92E-05	9.92E-05	-	-	-	-
NN-4	1.70E-04	9.92E-05	9.92E-05	0.837	-	-	-
SVM-LN	6.79E-03	1.00	0.169	9.92E-05	9.92E-05	-	-
SVM-RD	5.55E-04	4.98E-03	9.92E-05	9.92E-05	9.92E-05	8.71E-03	-
SVM-SIG	8.71E-03	1.00	0.574	9.92E-05	9.92E-05	1.00	1.70E-04

Except the pairs converging in gray cells (LR-2 vs LR-3; LR-2 vs SVM-LN; LR-2 vs SVM-

SIG; LR-3 vs SVM-LN; LR-3 vs SVM-SIG; NN-3 vs NN-4; SVM-LN vs SVM-SIG) all others pairs have significant differences. Box-plot of Fig.4.11 shows that NN3 and NN4 ranked first having the best time (there is no significant difference between them) to training the model. Third place is for RL-1 which have significant differences with the remaining algorithms.

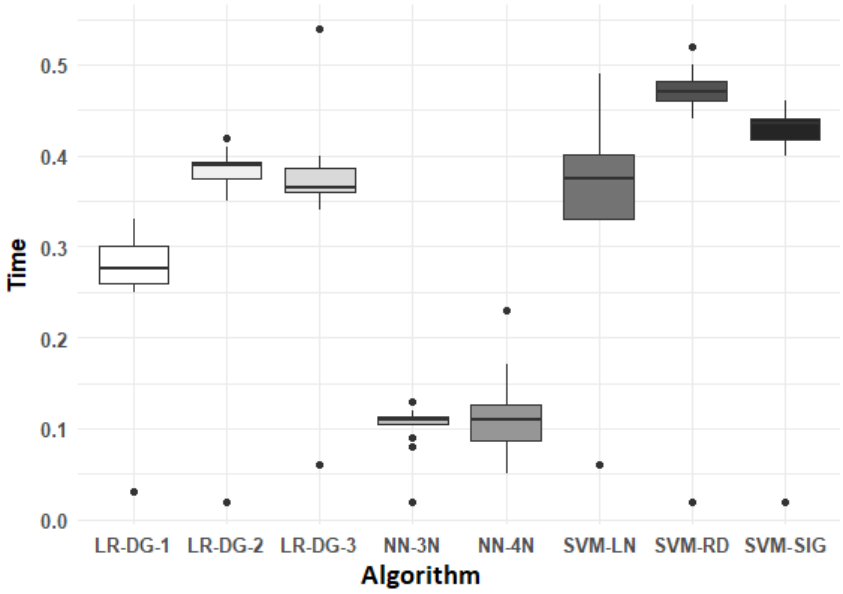


Figure 4.11: Training time per algorithm using 3F as data source.

Table 4.14 summarizes the analysis of results for evaluations on accuracy and time. The first two rows show comparisons between two sets of data, one of 21 SHF and other of three SHF selected by the FS process (21F vs 3F). The last two rows are the results of comparing the accuracy of SQ prediction and the training time among MLA variants (8 MLA-V).

Table 4.14: Summary of accuracy and time assessments.

Metrics	**Groups**	**Test**	**p-value**	**Conclusion**	**strength**
Accuracy	21F vs 3F	t-student	0.009	Refute Ho	Strong
Time	21F vs 3F	Wilcox	0.160	Does not refute Ho	Strong
Accuracy	8 MLA-V	ANOVA	0.9173	Does not refute Ho	Very Strong
Time	8 MLA-V	Kruskal-Wallis	0.02E-9	Refute Ho	Very Strong
Ho:Difference beteewn means are equal to 0.			Significance Level: 0.05.		

4.4 Summary

Table 4.15: Summary of comparisons of MLA

Rank	**Accuracy**	**Time of training**	**Feasibility of online implementation**
First	There is no evidence of	NN-3, NN4	LR-1
Second	significant difference in	LR-1	LR-2, LR-3
Third	Accuracy for any of all	LR-2, LR-3, SVM-LN,	NN-3, NN-4
	MLA variants.	SVM-SIG	
Fourth		SVM-RD	SVM-LN, SVM-RD, SVM-SIG

After we performed all proofs and we analyzed the results, we synthesized the information in Table 4.15. As we can see, all algorithms have the same performance in accuracy when results are statistically compared to a level of significance of 0.95. The ANN have the best performance in time of training, placing in second place the Linear Logistic Regression. On the other hand, Logistic Regression is the most feasible algorithm to implement in the On-line approach. We prioritizes '*accuracy in prediction*' and '*feasibility of on-line implementation*' over '*time of training*'. The time of training is not a priority since the system will train the model one time a week, which does not represent high computational cost. In this scenario LR-1 is the most appropriate algorithm to train a model for predicting SQ based on the critical SHF of the population studied. In the next chapter, we tested through a prospective study if the SHF selected from the SP could generalize the data obtained from the personal observations of the volunteers.

Chapter 5

Adaptive and Predictive Model

Chapter 4 covers the first part of the model development to predict SQ in SP. We perform feature selection to find the most relevant SHF and we verify that the selected factors predict better than the complete set of factors. Likewise, we selected the appropriate algorithm to train a model to predict SQ in SP. This Chapter covers the second part which consist of fitting a model to predicts SQ in a personal approach. First, we test if the SHF selected in the cross-sectional study characterize at least with the same level of accuracy, personal data from subjects participating in the longitudinal study. The second purpose is to improve the SQ prediction through a personalized model compared with the maximum accuracy achieved in the cross-sectional study (73%). A third objective is to determine the number of weeks of training that a model requires reaching SQ predictions with low Mean Square Error (MSE) and at least a 90% of variance explained (R^2).

5.1 Study design

We recruited three volunteers to participate in the longitudinal study, two women and one man. We refer them **VF1**, **VF2** and **VM** for privacy reasons. We explain them the reason for the study, making emphasis in the truthfulness of the data they provided. The process to carry out the longitudinal study consisted of:

- Application of a questionnaire to each volunteer. We used the **PSQI-SHI**(see Section 3.1.1)

questionnaire to collect information on SQ and SH from the volunteers. The questionnaire also ask for information about the suitability of volunteers to be part of the longitudinal study, for example, people suffering some chronic illness that affects sleep do not be part of the study.

- Functioning explanation of the electronic device to score sleep quality. We use Beddit 3 Sleep Monitoring, an unnoticeable device to place under the sheet to monitor the sleep when the user lie down on the bed. Beddit technology rests on a scientific principle of ballistocardiography (BCG), an unobtrusive technique for measuring the mechanical activity of the heart, lungs and other body functions. Beddit adapts its measures to each person through implement embedded MLA [1]. The volunteers were given the manual and each installed the mobile application that connects to the device via Bluetooth to shows the results. The electronic device monitor subjects and estimates SQ by collecting and analyzing sleep-related data such as sleep time, heart rate, breath rate, body and room temperature, movement and snoring. The estimation scale is from 0 to 100, placing the threshold of good and poor quality at 75 points.

- Instruction in filling a digital form to collect data during the days of study. Volunteers used the device and filled the form for 30 days (time that PSQI uses to estimate sleep quality in clinicians environments). They captured the data for the 21 SHF (See 3.6) by dichotomous items (Yes/No). Likewise, they introduced daily the estimation of **SQ** given by the device in the 0-100 scale.

- Preparation of the data for the analysis, which involved cleaning, extrapolation and transformation (see 5.3).

- Generation of the model in two steps: 1) Designing a feature selection algorithm and 2) Training a model through the linear regression algorithm Elastic-Net [10] using the SHF. Each instance consists of 21 predictors and one target features. The predictor features take values between 0 and 3 such as in the previous model. The target feature is **SQ** provided by the electronic device on a scale of 0-100, different to the target feature on

[1] https://www.apple.com/shop/product/MR9P2LL/A/beddit-3-sleep-monitor.

the previous data where SQ is dichotomous as PSQI suggest (*'Good/Poor'*). Given the target feature type, it was modeled using Elastic-Net (See Equations 5.1 and 5.2), which generates linear models by adjusting the coefficients through an efficient regularization. Depending on the purpose of modeling, the algorithm allows us to use regularization with penalty $l_1(\alpha = 1)$ transforming into the LASSO algorithm [148], $l_2(\alpha = 0)$ becoming in Ridge regression [151], or use a mixture of both algorithms by selecting an intermediate value for alpha ($0 \leq \alpha \leq 1$).

$$\min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} R_\lambda(\beta_0,\beta) = \min_{(\beta_0,\beta)\in\mathbb{R}^{p+1}} \left[\frac{1}{2N} \sum_{i=1}^{N} (y_i - \beta_0 - x_i^T\beta)^2 + \lambda P_\alpha(\beta) \right] \tag{5.1}$$

where

$$P_\alpha(\beta) = (1-\alpha)\frac{1}{2}\|\beta\|_{l_2}^2 + \alpha\|\beta\| l_1 = \sum_{j=1}^{p} \left[\frac{1}{2}(1-\alpha)\beta_j^2 + \alpha\left|\beta_j\right| \right] \tag{5.2}$$

- Application of a questionnaire to the volunteers. We collected data on the volunteers experience when participating in the study. We include questions to identify the participant's level of knowledge on SH and SQ before and after the study. The questionnaire inquires about the perception that participants have on the SHF they believe influence their SQ.
- Confrontation of the study results against the volunteers perception.

5.2 Results of PSQI-SHI application

We collected subjective information of SH amd SQ from the volunteers through the **PSQI-SHI**questionnaire. Likewise, we collect information on the health and chronic conditions that could influence the quality of sleep. Based on Tables 3.7 and 3.5 we computed the SQ score of the questionnaire and we compared it with the SQ estimated by the electronic device in the 30 days of monitoring. Table 3.7 describes the scale to classify SH, in a scale of 84 points divided in five categories. Interpretation: *'Very good (0-16)'*, *'Good (17-33)'*, *'Regular (34-50)'*, *'Poor (51-65)'*, *'Very Poor(66-84)'*. The PSQI scale described in Table 3.5 is in a scale of 21 points with two levels *'Good (0-5)'* and *'Poor (6-21)'*. On the other hand, the electronic device reports

its results on a 0 to 100 scale with two levels, *'Good(76-100)'* and *'Poor (0-75)'*.

Table 5.1: Comparison of PSQI-SHI vs Longitudinal Study.

	SHI		**PSQI**		**VP**	**L. Study**		
	Score	**SH**	**Score**	**SQ**	**SQ**	**Score**	**Rate GD/PD**	**SQ**
VF1	19	Good	2	Good	Good	76±19	1.5	Good
VF2	30	Good	8	Poor	Good	80±23	4.5	Good
VM	25	Good	3	Good	Good	82±13	1.9	Good

VP: Volunteer Perception.

The summary presented in Table 5.1 shows that results scores of two volunteers (**VF1** and **VM**), estimated by the electronic device during the study, coincides with qualification resulting from PSQI and also with the participant's own perception of their SQ. Since we compared subjective against objective data, the coincidence we found is interesting; more when **VF1** and **VM** reported having different levels of knowledge and awareness about sleep quality and its health influence, before starting the study. **VF1** reported *'high knowledge'* and *'high awareness'* while **VM** expressed *'little knowledge'* and *'little awareness'* about it. PSQI scores of **VF2**(*'Poor'*) do not match the results of the study (*'Good'*) but matches with the participant's personal perception, who said having *'average knowledge'* about their quality of sleep and *'very high awareness'* about the impact of sleep on health.

With respect to SH, the three participants came up with scores that place them in the *'Good'* classification, even though they considered their level of knowledge on this subject different before participating in the study. **VF1** said to have *'high knowledge'*, **VF2** replied that she had *'average knowledge'* and **VM** said that his knowledge of the subject was *'null'*.

Fig. 5.1 shows densities of the target feature (**SQ**) on the data collected during the study. **VM** had most concentrated data around the mean ($\sigma = 13$), however, its variability is good. **VF1** and **VF2** had $\sigma = 19$ and $\sigma = 23$. This means that the target feature meet the requirements to generate reliable data by extrapolation to train models using MLA. The predictor features ("***SH1, SH2, ..., SH21***") will be discussed in the next section where we explain the process we performed to extrapolate and transform the data.

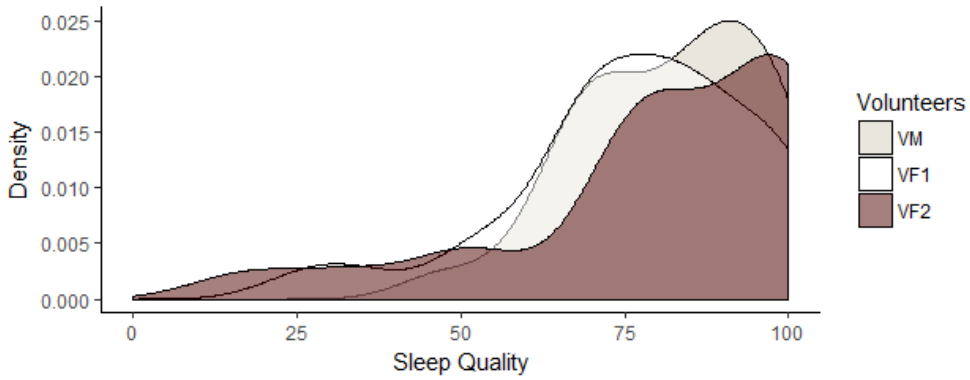


Figure 5.1: Density plot of data collected during longitudinal study.

5.3 The dataset

The dataset consists of an identifier composed of the date and time when the users captured the record, the participant name, 21 dichotomous predictor features (*'Yes/No'*) and a target feature taking values in a 0-100 range. We extrapolate the data to have enough records for training the models. More, we transformed the data to give it the format used by PSQI to acquire information and score SQ. Table 5.2 shows how data is stored when the volunteers captured them[2]. The predictor features are text (*'Yes/No'*) and the target feature is a number between 0 and 100.

Table 5.2: Data as captured by participants.

Days	**Predictive features**						**Target Feature**
	SH1	**SH2**			**SH20**	**SH21**	**SQ**
1	No	Yes	Yes	Yes	Yes	Yes	57
2	No	Yes	Yes	No	Yes	No	91
3	No	No	No	No	Yes	No	95
				...			
27	Yes	No	No	No	Yes	No	94
28	Yes	Yes	No	No	Yes	Yes	81
29	No	No	Yes	No	Yes	Yes	78
30	Yes	Yes	Yes	Yes	No	Yes	75

[2]From here we separate the data for each participant to carried out the following process

We converted the predictive features corresponding to 30 to numeric and we extrapolated the data by the binomial distribution (5.3) to reach 1456 records equivalent to 208 weeks. Table 5.3 shows the data for SH after we extrapolate them while we calculate the column **SQ** in the next step.

$$P(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x} \tag{5.3}$$

Table 5.3: Extrapolated data by Binomial Distribution.

Days	**Predictive features**						**Target Feature**
	SH1	**SH2**			**SH20**	**SH21**	**SQ**
1	1	1	1	1	1	1	
2	1	1	1	2	2	1	
3	1	1	1	2	1	1	
.					.		
.					.		
1453	2	2	2	1	2	1	
1454	1	1	1	2	1	1	
1455	2	1	2	1	2	2	
1456	2	2	2	2	1	2	

To calculate SQ, we use the algorithm KNN of the package FNN [5] (see algorithm 1). We identify in the original dataset the k nearest neighbors for each record we generated by the extrapolation. After, we calculated the bounded mean of SQ in the vicinity and we assigned the new value of the SQ feature in the record of the new dataset.

In the last step we format the data as PSQI works. PSQI inquires on habits observed during 30 days by taking the week as a reference. For example, a question in PSQI is: "*During the past 30 days, how many times per week did you drink coffee within 4 hours before going to the bed?*" An answer can be one of the following: *'Never (0)'*, *'Once (1)'*, *'twice (2)'*, *'three times (3)'*, *'four or more times (4)'*. With previous criteria, we divided the new dataset in blocks of seven records. For each block, we generate a record by counting the number of times the participant answered *'Yes'* in each predictive feature. We assigned to the record the target feature SQ by computing the average of SQ in the block (see Algorithm 2).

Algorithm 1 Computing SQ by KNN.

procedure CSQbyKNN($dsOriginal, dsExtrapoledData, k$) ▷ k must be odd.
  for all $recordindsExtrapoledData$ **do**
    $nn \leftarrow knn(dsOriginal, record, k = k)$
    $SQN \leftarrow dsOriginal[nn.indexes, SQ]$
    if $k > 1$ **then**
      $m \leftarrow (k - 3)/2$
      $SQNS[record] \leftarrow$ **mean**($SQN[$ **order**$(SQN)[(m + 1) : (m + 3)]]$)
    else
      $SQNS[record] \leftarrow SQN$
    end if
  end for
  return $SQNS$
end procedure

Algorithm 2 Transforming data at PSQI format.

procedure GcatData($dsOfDays$)
  $nOfFactors \leftarrow$ **ncol**$(dsOfDays) - 1$ ▷ gets the number of SHF
  $k \leftarrow 1$
  while $k \leq$ **nrow**$(dsOfDays)$ **do** ▷ go through all records
    $week \leftarrow dsOfDays[i : (i + 6), 1 : nOfFactors]$
    $weeks[k]_{1...m} \leftarrow ifelse\left(\sum_{i=k}^{i+6} X_i <= 4, \sum_{i=k}^{i+6} X_i, 4\right)$
    $SQave[k] \leftarrow$ **mean**$(dsOfDays[i : (i + 6), (nOfFactors + 1)])$
    $k \leftarrow k + 1$
    $i \leftarrow i + 7$
  end while
  return $cbind(weeks, SQave)$
end procedure

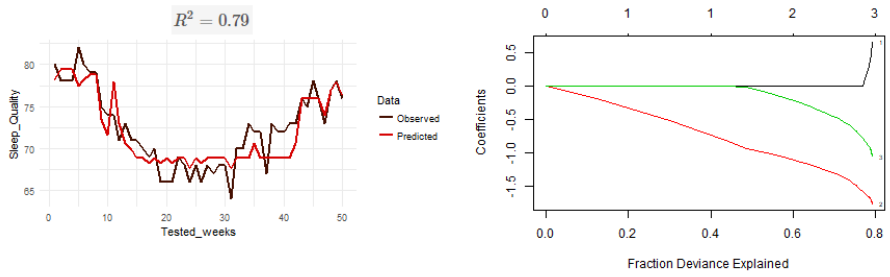
After this transformation, the dataset locks like Table 5.4 shows. All predictor features are categorical in five levels (0 to 4), and the target feature is numeric in a 0-100 scale.

Table 5.4: Example of data as used to train the model.

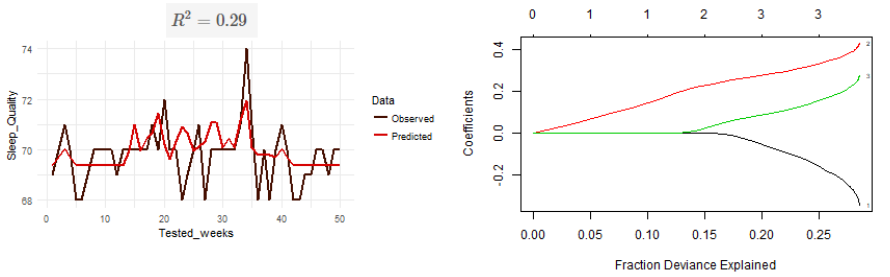
Weeks	**Predictive features**						**Target Feature**
	SH1	**SH2**			**SH20**	**SH21**	**SQ**
1	2	0	1	0	2	0	96
2	1	0	4	0	0	4	89
3	2	0	0	2	3	0	90
				...			
				...			
205	3	2	4	1	0	1	88
206	1	1	1	4	2	1	87
207	4	1	1	1	3	2	76
208	0	4	1	4	4	0	60

5.4 Feature selection

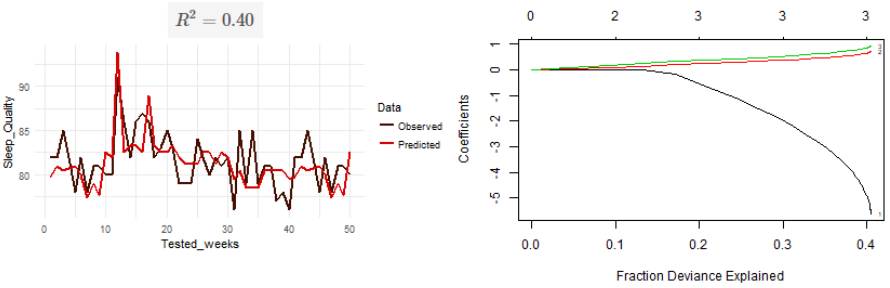
In Chapter 4, we use FSA to identify the most influential factors on the SQ of the SP. We found that trained models with the three identified factors achieve better prediction accuracy than models trained with the complete set of 21 SHF. Yet, even when predictions where better, the highest was around of 0.73; low for a predictive model. We train models through Elastic-Net using the data of each volunteer and the SHF (**SH2**, **SH12** and **SH20**) selected in the analysis of the data of the SP. Figures 5.2(a) and 5.2(c) show that the models for **VF2** and **VM** explain 29% and 40% of the variance, while Figure 5.2(a) shows the highest percentage of the variance explained ($R^2 = 79\%$) for **VF1**. Since a regression model is coherent if explaining at least the 70% of the variance and highly reliable if it explains from 95% onward [152] (For this study, we set the lower threshold to 90%); we can say that the selected factors from the SP do not generate models explaining satisfactorily the variance of the data from personal monitoring.



(a) Data of VF1.



(b) Data of VF2.



(c) Data of VM.

Figure 5.2: Fraction deviance explained by models using 3F of volunteers data.

Due to the previous results, we include into the method of producing models a feature selection process. For this purpose, we test three FSA for linear regression given as data input 208 records including the 21 dataset as predictors and SQ as the target feature. We show the results in Table 5.5.

No algorithm reaches the 90% of variance explained. The algorithm producing the highest

Table 5.5: Results of models generated by SHF selected by three FSA

	Random Forest		**LASSO**		**XGBOOST**	
	SHF	R^2	**SHF**	R^2	**SHF**	R^2
VF1	1,4,10	85%	13,1,6,20,10, 14,7,11,12	79%	7,6,20,1,10	76%
VF2	21,5,6,20,2	81%	1,5,3,14,18, 21,6,15,2	70%	1,14,5,9,7	83%
VM	19,18,21,2,20, 13,4,12,5	60%	2,3,18,13,12, 7,5,20,6	29%	12,3,21,5,20, 10	70%

percentage is Random Forest, which however does not reach the criterion. Furthermore, there is the disadvantage of not having a metric to determine the number of factors to choose from when differences among their weights are very small. To give a solution to these drawbacks we design the algorithm 3 with the following main characteristics:

1. To combine results of XGBOSST, RF and LASSO through ‘merge—expand‘to select a set of critical factors. ‘merge‘picks only factors selected by all algorithms, while ‘expand‘chose all factors selected by algorithms.

2. To implement the “Best-Search”technique in forward or backward mode in order to choose the optimal set of factors within the list of selected factors. The stop criterion is reaching at least 90% of the data variability after training a model through Elastic-Net.

3. Use recursion to add factors in a second or third iteration if the result is unsatisfactory. From the second iteration onwards the algorithm is applied on the set of factors not selected as critical in the previous iterations.

Algorithm 3 fsXLR.

procedure FSXLR($workDS, try, ve, ...$) ▷ other parameters are listed below due their extension.
  $cardnonzero \leftarrow$ **AllFSCardGreat1**($workDS$) ▷ Selects variables with cardinality greater than 1.
  $listofselfeat \leftarrow$ **SelFMultAlgFS**($workDS[cardnonzero], list\_of\_alg$) ▷ Returns a list of all features selected by algorithms involved in selection (LASSO, Random-Forest, XGBOOST).
  $ordfeatbyweight \leftarrow$ **OrderFS**($listofselfeat, selecttype$)
  $bestfeatsearch \leftarrow$ **BruteForceFS**($workDS, ordfeatbyweight, FBmode, R^2$) ▷ Returns a list of sets of features searched by Best-Search in FBMode(Forward—Backward).
  if *replace* **then** ▷ Build the vector of features to do the new selection if necessary.
    $lstofnonsel \leftarrow !(cardnonzero$ **in** $bestfeatsearch)$
  else
    $lstofnonsel \leftarrow !(cardnonzero$ **in** $orderfeatbyweight)$
  end if
  if $(ve < R^2)$ **and** $(try <= maxoftries)$ **and** $length(cardnonzero) > 2$ **then** ▷ Verifies the stop criteria, if all criterion are TRUE, fsXLR is invoked recursively.
    $dfSelFeat \leftarrow$ **fsXLR**($workDS[lstofnonsel], try + 1, ve, ...$)
  else
    return $bestfeatsearch$
  end if
end procedure

Table 5.6: Description of Algorithm 3

fsXLR		Recursive Feature Selection Algorithm using Bagging and Best-Search
Description		The algorithm selects a set of relevant features using LASSO, RF and XGBOOST. Searches best set of features through Best-Search using as stop criterion the determination coefficient (R^2). Returns a list of best regressions for provided data.
Usage		fscombolaha(workDS, selectType, ve, maxOfTrys, FBMode, R^2, ...). Searches best set of features through Best-Search using as stop criterion the determination coefficient (R^2). Returns a list of best regressions for provided data. Returns a list of factors classified by importance.
Arguments	**default**	**Description**
workDS		Data set containing all features and data concerning to a subject,
previousSelectedF	null	(Vector[String]) In recursion applies, this parameter brings to new invocation the feature selected in previous rounds.
explain	0.1	(Variable[double]) Is the threshold to select as relevant a feature of list provided by SelFMultAlgFS
selectType	expand	(Variable-Factor[expand—merge]) Indicates the type of selection to do. 'merge' includes repeated features in all sets. 'expand' includes all features in all sets.

Continued on next page

Table 5.6 – *Continued from previous page*

Arguments	**default**	**Description**
ve	0.0	(Variable[double]) Explained Variance, this parameter keeps the best R^2 in Best-Search algorithm. It sets up its value every time that a set of features exceeds the old greatest value. It is the most important stop criterion for Best-Search and recursion. At the beginning it is established to zero.
maxOfTries	3	(Variable[integer]) Establishes the maximum number of tries to reach R^2
FBMode	backward	(Variable-Factor[forward—backward]) Determines if Best-Search begins search with the biggest (backward) or the smallest (forward) set of features.
minOfFeatures	3	(Variable[integer]) Minimum of features to begin (FBMode=Forward) or stop (FBMode=Backward) the search.
startRsquare	.70	(Variable[double]) Establishes the minimum value of R^2 to add a regression to the list of regressions.
squareR	.90	(Variable[double]) Explained Variance, is the threshold to stop Best-Search and Recursion, works together whit *ve*.
maxOfFeatures	7	(Variable[integer]) Maximum number of features to chose by algorithms.
listOfAlgorithms	null	(Vector[string]) Contains the name of algorithms to use in the selection. If null, the algorithm uses: c("LASSO","RF","XGBOOST").

Continued on next page

Table 5.6 – *Continued from previous page*

Arguments	default	Description
replace	TRUE	(Variable[boolean]) Establishes if the selected features by the algorithms but not selected by Best-Search must be part of the new set of data in the next round of recursion.
best_regressions	null	(List) Keeps the list of regressions reaching criterion in each round.

5.5 Results

To verify the efficiency of the algorithm we run fsXLR using XGBOOST, LASSO and RF separately for the data of the three volunteers, then, we run the algorithm using the three algorithms together in *'merge'* and *'expand'* mode. We generate models with the selected factors through the Elastic-Net ($\alpha = 0.3$) to evaluate the performance of fsXLR by comparing the percentage of the variance explained shown in Table 5.7 against the results of the Table 5.5.

Table 5.7: Results of feature selection through fsXLR.

	VF1			VF2			VF3		
	SHF	R^2	**MSE**	**SHF**	R^2	**MSE**	**SHF**	R^2	**MSE**
LASSO	4,6,12, 18,20	96%	4.54	5,11,16	97%	0.18	1,5,19	93%	2.22
R-Forest	1,7,10	93%	3.66	6,14,15, 16	97%	3.92	1,3,17	97%	0.12
XGBOOST	6,7,20	94%	2.96	3,6,15	96%	1.12	5,12,19	90%	3.28
LRX-EX	4,6,12, 18,20	96%	4.54	3,11,13, 16,18,19	93%	7.1	1,7,21	95%	0.72
LRX-ME	6,7,20	94%	2.96	2,7,15, 16	96%	2.58	7,13,21	90%	0.3
minOfFeatures:3, maxOfFeatures:7, maxOfTries:2, FBMode:Backward, replace:F									

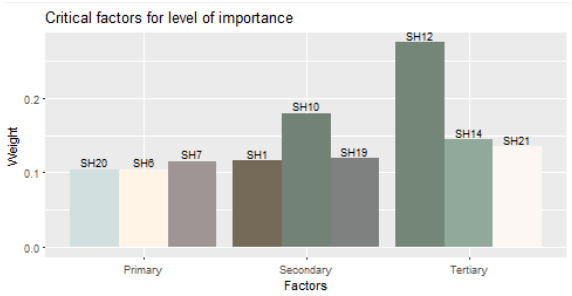
We conclude that we can obtain better sets of factors when we use the algorithm fsXLR than if we use any of the three algorithms out of fsXLR to select the relevant SHF. We show in Table

5.7 that each test reaches an explained variance higher than 90%, while in Table 5.5 the highest variance explained is 85%. This is reasonable since the algorithm fsXLR adds the search for the best combinations of factors to explain the variability through the Best-Search algorithm. As we shown in Table 5.7 the MSE is relatively small in any case in terms of what it represents in practical terms. For example, if we generate SQ randomly, MSE for **VF2** is 471 while the largest MSE in this table is 7.1. On average, the greater fraction of variance explained is achieved using the RF algorithm and the smaller MSE is obtained when Bagging is applied over the three algorithms (**LA-RF-XG-ME**).

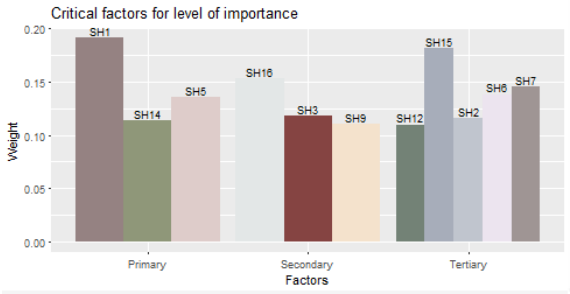
The feature selection algorithm fulfills its objective by delivering a set of factors that explain the variance of the data by at least 90%. The prediction accuracy is also high if we consider the MSE relatively low for all cases. In the next section we compare the results of the algorithm in the selection of relevant SHF using the response of participants. This is important because the ultimate goal of the hypothetical system is to provide information so that people interested in improving their SQ have a more objective reference on those factors that may affect it. In a real scenario, the system informs to the users their SQ and the most influential factors to estimate it. The users receive the information and they provide feedback to confirm or deny the prediction of the model. Through users' feedback, the model adjusted its parameters to generate better predictions as the weeks pass. Finally, the system converges with the users perception, and they are aware of the SHF that most influence their SQ.

5.5.1 Volunteer perception

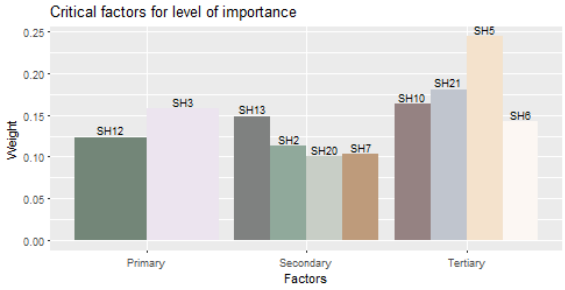
The algorithm fsXLR besides regressions with the better combinations of SHF generates a classification of factors in order of relevance. We show in Figure 5.3 the classification of the factors. To **VF1** and **VF2** the relevant factors are classified into two groups: primary and secondary, while for **VM** a third group was added with tertiary factors. The number of groups formed depends on the iterations that algorithm needed to reach the 90% of the variance explained.



(a) Factors for VF1.



(b) Factors for VF2.



(c) Factors for VM.

Figure 5.3: Critical individual SHF classified in relevance order.

In the final section of the questionnaire, we ask the participants to classify the SHF from the highest to the lowest weight of influence in their sleep quality. They should mark ten factors from 1 to 10 being 1 the highest weight and 10 the least. We ranked the SHF as follows: primary from 1 to 3, secondary from 4 to 6, and tertiary from 7 to 10. We carry out a concordance analysis using Cohen's Kappa to compare the selection of the system against the classification of the volunteers. First, we considered four levels of classification: primary, secondary, tertiary and non-relevant factors; second, we compare selected vs non-selected factors independently of the level in which the system and users classify them. We show results in Tables 5.8, 5.9 and 5.10.

Table 5.8: Concordance VF1 vs System

Levels	VF1	System
Primary	SH12, SH21,SH20	SH20,SH6,SH7
Secondary	SH4,SH3,SH2	SH1,SH10,SH19
Tertiary	SH1,SH16,SH15,SH7	SH12,SH14,SH21
Cohen's Kappa = 0.0215, p= 0.87 (Four levels)		
Cohen's Kappa = 0.137, p=0.52 (Two levels)		

Table 5.9: Concordance VF2 vs System

Levels	**VF2**	**System**
Primary	SH12,SH13,SH16	SH1,SH14,SH5
Secondary	SH20,SH19,SH14	SH16,SH3,SH9
Tertiary	SH1,SH3,SH4,SH2	SH12,SH15,SH2,SH6,SH7
Cohen's Kappa = 0.003, p=0.98 (Four levels)		
Cohen's Kappa = 0.236, p=0.279 (Two levels)		

Table 5.10: Concordance VM vs System

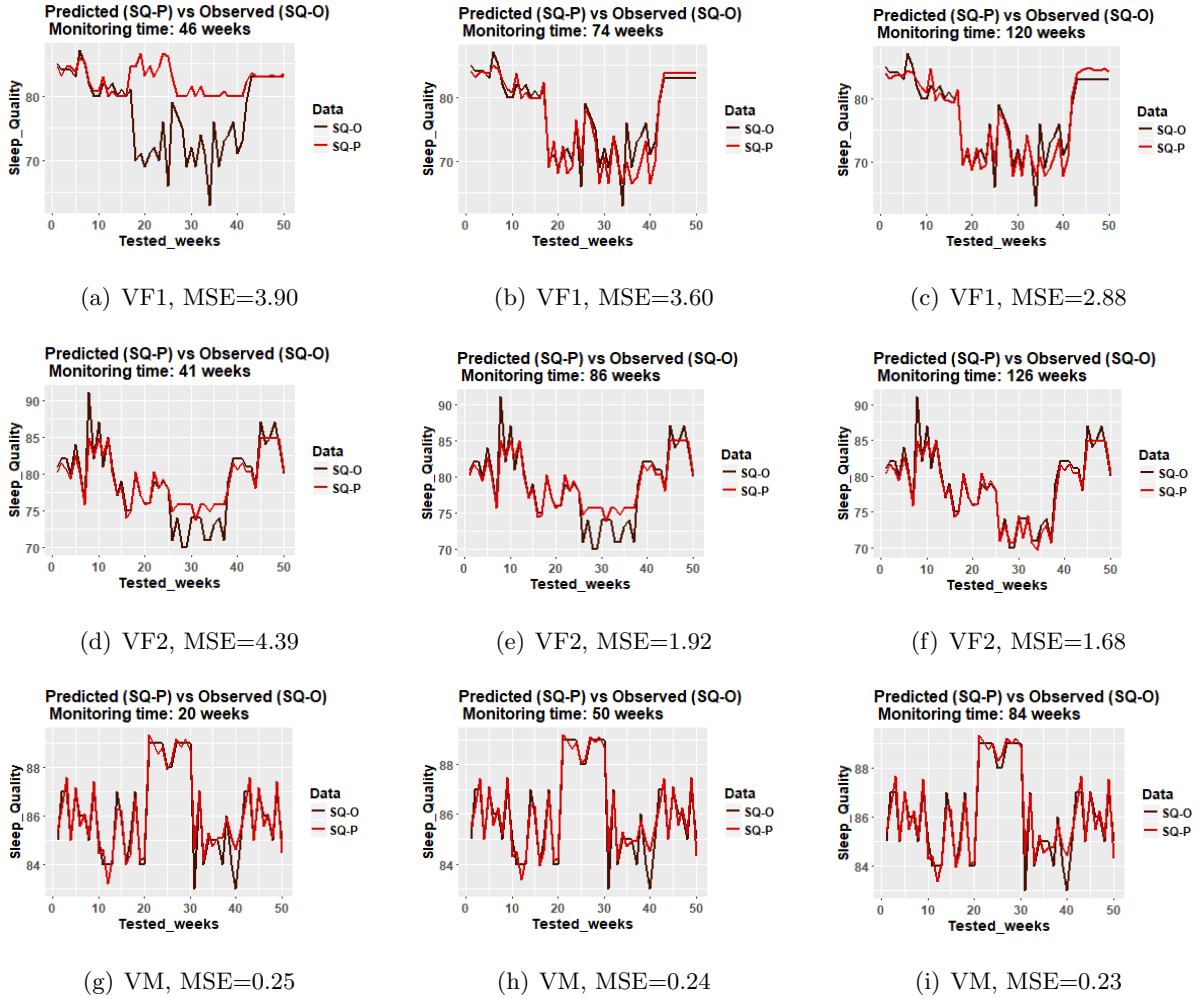
Levels	**VM**	**System**
Primary	SH7,SH12,SH2	SH12,SH3
Secondary	SH3,SH11,SH19	SH13,SH2,SH20,SH7
Tertiary	SH20,SH10,SH4,SH13	SH10,SH21,SH5,SH6
Cohen's Kappa = 0.192, p=0.15 (Four levels)		
Cohen's Kappa = 0.427, p=0.05 (Two levels)		

We proof through the coefficient of determination (R^2) that the algorithm has efficiency in the selection of the relevant factors. The results allow to explain the variability of the data and make predictions close to the SQ generated by the device during the longitudinal study. On

the other hand, we can observe a poor concordance in almost all cases, except for VM in two levels where we found a p-value in the threshold of the significance level (p=0.05) and a Kappa coefficient close to 0.5, which reveals a medium level of concordance. It suggests us that people participating in the study does not be aware of the SHF that may affect their SQ.

5.6 Convergence

We perform tests to identify the number of records we need for generate efficient models. Each record in the dataset resulting from the extrapolation explained in Section 5.3 represents a week of monitoring; thus we used the records of this dataset to simulate training of the model during a period of time $10 \leq t \leq 208$. We trained the model from 10 instances onward and we save three events: the time to begin producing satisfactory results ($R^2 \geq 90$), the time for model convergence, and finally an intermediate time between the two previous marks. We show the results for each volunteer in Fig. 5.4 and we summarized it in Table 5.11.



(a) VF1, MSE=3.90 (b) VF1, MSE=3.60 (c) VF1, MSE=2.88

(d) VF2, MSE=4.39 (e) VF2, MSE=1.92 (f) VF2, MSE=1.68

(g) VM, MSE=0.25 (h) VM, MSE=0.24 (i) VM, MSE=0.23

Figure 5.4: Simulation of model's convergence over time per volunteer.

Table 5.11: Simulation results of model's convergence over time per volunteer.

	VF1	**VF2**	**VM**
Start week ($R^2 \geq 90$)	46	41	20
MSE start week	3.90	4.39	0.25
Intermediate week	74	86	50
MSE Intermediate	3.60	1.92	0.24
Convergence week	120	126	84
MSE convergence	2.88	1.68	0.23

Table 5.11 shows that models take a minimum of 20 weeks to produce satisfactory results and a maximum of 46 weeks. This time in terms of therapies that involve changing behavior patterns in humans is not out of the range. The following scenario is considered: from week zero to the first week with satisfactory results, the user of the model becomes aware of his quality of sleep by reading the prediction of the model daily and entering SHF. From the initial week of satisfactory predictions onward, the users begins to be aware of the most relevant factors, which can lead them to decide improve their habits. If users arrive at the week of convergence, enough time will have passed for any change favoring their SH.

5.7 Chapter conclusions

1. The most relevant SHF to the SP, generate models with low explanation of the variance of the personal tracking data to the SH and SQ of the volunteers during the longitudinal study. We conclude that it is necessary to train personalized models to achieve more efficiency in the prediction of sleep quality.

2. The feature selection algorithms LASSO, RF and XGBOOST select sets of relevant factors explaining between 60% and 85% of the variability of the data, which does not guarantee good predictions when training models. Thus, a robust algorithm was proposed applying techniques such as Bagging and BFS.

3. The FS algorithm fsXLR increases the cost of processing time, but guarantefsXLRes sets of factors that explain the variability of data in high percentages ($> 90\%$). In scenarios such as these where models must be trained one time per week and the time of training is in the order of minutes, the use of fsXLR is justified.

4. There are different combinations of factors explaining satisfactory the variance of the data with a low MSE. These are no conclusive, but suggestive results. In a real scenario, the users of the system could identify a set of factors statistically linked with their SQ. As time passes and the users feed the system for new conjectures the model makes better predictions, and the users learn more about their SH and SQ. This empower the users with

information on specific changes they need to make in their behavior to improve their SH.

5. The convergence times of the model are in line with the therapy times when talking about treatments involving behavioral patterns.

6. In order to improve the performance of the prototype, it is necessary to monitor individuals for a long time (months) in such a way that the data to select factors and train models are observed and not simulated data.

Chapter 6

Summary

The approach to the specific objectives of this study generated questions that guided the development of the research. In this chapter we conclude the work by answering the research questions. We discuss the findings we found by following the strategies planned to to answer each question.

6.1 Answers to research questions

6.1.1 is it feasible to investigate if SH influence to some extent on SQ?

Quality of sleep is a topic that has taken relevance in the last decades, every time there is more evidence showing the association between the level of rest and well being of the human being. The biological and physiological processes that occur during this state of unconsciousness are largely responsible for the cellular restoration in practically all the organs of our body. Figure 6.1 shows some areas of impact of sleep quality on the individual and how they correspond to some dimensions of the wellbeing framework proposed by academics at Harvard University. At cognitive level, repercussions of poor sleep in the short term are manifested in problems of concentration and memory while in the long term could causes degenerative diseases such as Parkinson's and Alzheimer's. In physical dimension, one of the most affected organs by a poor quality of sleep is the heart. A prolonged sleep deprivation or a sleep disorder not treated properly, such as OSA, can lead to cardiovascular problems that put life at high risk. The repercussions of bad sleep cover also the emotional plane because it produces changes in

the mood, and recurrence in these changes can lead to depression. In fact there is no patient diagnosed with depression who does not have a sleep disorder.

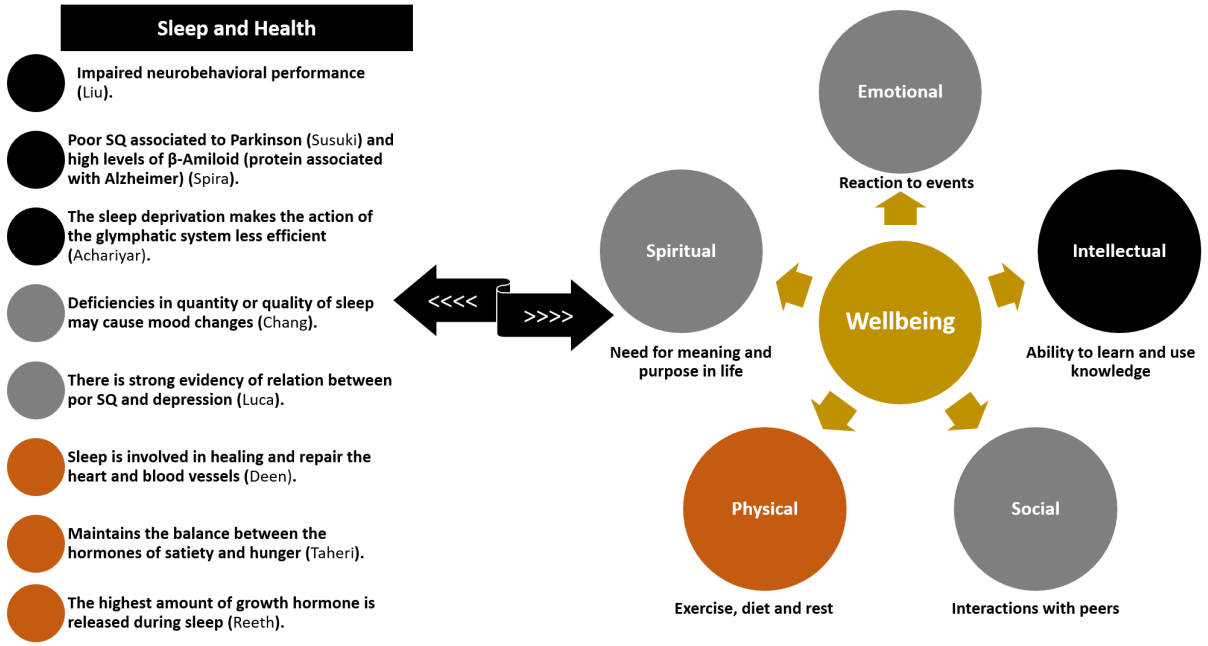


Figure 6.1: Associations Between wellbeing and Sleep Quaity. (Liu [153], Susuki [55], Spira [54], Achariyar [154], Chang [62], Luca [89], Deen [128], Taheri [57], Reeth [90])

On the other hand, Sleep Hygiene is a concept introduced by Kleitman in the first half of the 20th century and taken up again 50 years later by Hauri. The concept encompasses a set of behaviors and elements of the environment[1] that may favor or impair the level of quality of sleep. The Kleitman hypothesis has been questioned and debated. Several studies have been done to investigate whether SHF both individually and collectively have an impact on SQ.

Researchers have been conducted studies in healthy people and people with insomnia to seek the correlation between the perception of sleep quality and the SH without obtaining conclusive results. However, there are studies showing a direct connection between a SHF and physiological indicators of SQ. For example, exercise near bedtime keeps the heart rate and body temperature higher than normal for several hours, which is not favorable for the quality of sleep. A similar effect occurs with an acute stress event near to the bedtime. Furthermore, researchers have been

[1] called Sleep Hygiene Factors in this study

found evidence of the coincidence of sunset and the start of melatonin production (called the sleep hormone) and how it is blocked to some extend with the artificial light. Especially the blue light regardless of its intensity.

So far we can hypothesize about the idea that SHF have physiological repercussions on the quality of sleep. However, imperceptible to the people who claims to sleep with the same level of quality regardless of their lifestyle practices. This coupled with the finding over the incidence of environmental factors in the generation of marks at the genetic level, put before us many challenges and research questions. We can say that it is feasible to do research in this field with the certainty of contributing evidence in favor or against that the sleep hygiene behaviors influence the sleep quality.

6.1.2 How has technology evolved to study sleep and its disorders?

We found that a test performed for the first time in the 1960s and turned into a laboratory test by Stanford University a decade later became the gold standard for monitoring sleep and diagnosing various disorders clinically. This test allowed researchers to more accurately classify the sleep into stages and observe the physiological changes that occur in each one. Until today, it is the most used technological tool in sleep studies because of the amount of accurate information that it throws. However, being a laboratory test has the disadvantage of being invasive and it can influence the results in some patients whose sleep is susceptible to non natural environments. In addition, it does not allow to physicians the clinical monitoring of patients whose problem is difficult to diagnose in a single night of study.

Actigraphy has been proposed as an alternative and less invasive technique that physicians can practice in the natural sleep environment. It does not provide as much information as PSG; however it allows the patient to be followed for long periods of time, which makes it especially useful in certain populations, for example older adults and children. In the last decade, the manufacture of portable electronic devices of relatively low cost and sufficient processing and memory capacities to perform moderately complex processes has proliferated. This has enhanced the ability to design new methods to explore the sleep. Technology has been developed to monitor and diagnose sleep disorders as well as applications that implement therapies with the purpose

of helping users sleep better. Within this technology there are home PSG; applications that use the sensors of the mobile phones or smart watches to function as actighraphs; and systems using various sensors placed in the bedroom environment to observe sleep parameters and to send information to computers.

We classify the technology by its maturity. We consider technology clinically approved as mature while other technologies designed for research or trade purposes was classified as emerging technology. There is a debate among the various international organizations about the turning point to decide whether a technology designed to intervene health aspects should be clinically evaluated before use it in an uncontrolled environment. The consensus today is that technology that does not put at risk the health of users can be allowed without regulation, such is the case of all applications aimed at changes in the lifestyle. Among them, there are the applications for nutrition, physical exercise and sleep. This position is debatable, because only those immediate and drastic risks are being considered, however, there may be systems or applications that intervene with the intention of modifying behaviors that may impact on the health of the user in the long term. If the impact is positive, at a good time, but what about those changes that may affect health in the long term?

6.1.3 How should the emerging technology related to the SQ be evaluated?

The analysis of 18 applications designed for some aspects related to sleep from any of their approaches, reveals that there are no standardized protocols to evaluate technology for health when they do not have the intention of entering the clinical field. Each researcher proposes goals to be evaluated at the end of the study and themselves design strategies to conduct the assessment. The level of rigor in the evaluation methods vary, which makes the comparison of technological developments complex. After reviewing the literature, we concluded that it is necessary to build a reference framework to evaluate this type of technological development, which should include two areas of evaluation: efficiency and user perception.

Efficiency evaluation: The results must be tested by statistical methods to contrast the proposed hypothesis in a given level of significance. Unlike clinical results that require a level of significance of 99% or higher, the researchers must evaluate the efficiency of this kind

of studies between 90% to 95% of significance. The evaluation must adhere to existing methodologies that have proven reliability or, the evaluation processes should be described clearly and in detail.

Usability evaluation: The technological developments should incorporate the evaluation of the users' perception of ease of use, the intuitive of the interaction, the level of intrusiveness and the effectiveness in fulfilling the purpose of the designers. In emergent technology, this evaluation is as important or more than the evaluation of the efficiency. Through this you can learn much of what users and third parties expect from the system or application, hence, you can determine its success or failure when you think in release them as a product. In addition, many researchers have the medium or long term purpose of having their technological developments approved in the clinical field; The usability evaluation can guide them to improve their design in order to obtain better dividends when they reach the stage of approving the clinical evaluation protocols.

6.1.4 is there a set of factors characterizing the SP so that we can train more efficient models than if we train them using the complete set of SHF?

A cross-sectional study was conducted in SP as described in Section 3.1. After conducting the study, a robust algorithm for the FS was implemented and the prevailing factors were found in SP. To corroborate that these set of SHF characterize SP better than the complete set of SHF we trained 16 models using variants of three MLA and the cross-validation technique. Eight models were trained with the selected SHF and eight with the complete set of SHF. Means of accuracy in prediction were compared through t-student obtaining a significant difference ($t = 3.0108$, $p = 0.0095$) by contrasting Ho = "*True difference in means is equal to 0*". Models trained with the three selected SHF (**SH12** (Stress before sleep), **SH2** (Regular schedule to get up) and **SH20**(Attending worries before sleep)) showed significantly better efficiency when predicting SQ.

At this point it is important to note that even though the accuracy of the prediction was better when training the models using the data of the selected factors, the average prediction level is low as shown in Table 6.1. These results could be due to two reasons mainly: 1) The nature of the phenomenon studied. It is difficult to generalize the quality of sleep because each

Table 6.1: Descriptive statistics of accuracy prediction of SQ with 3F and 21 SHF

	Min.	**1st Q.**	**Median**	**Mean**	**3rd Q**	**Max**	**Dev.**
3 SHF	0.610	0.655	0.705	0.687	0.715	0.740	0.040
21 SHF	0.570	0.602	0.610	0.622	0.655	0.680	0.045

individual has different preferences, needs and perceptions of their reality regarding the quality of their sleep. 2) The constructs to obtain information from the SP regarding the subject being studied, even when validated, maintain risks inherent to their nature. The questionnaires are subject to the memory, objectivity and sometimes to the mood of the respondent.

6.1.5 Which MLA is the most appropriate algorithm to model this problem in terms of prediction accuracy and processing time?

To answer this question we used the results of accuracy in predicting the SQ and the execution time that each MLA variant took to train models using the information of the selected factors as data sources.

Accuracy: After verifying normality and variance homogeneity in the data, we used ANOVA to test Ho: "*The difference between the means of the samples is equal to zero*". The results show that there is no statistically significant difference between the samples (F = 0.369, p=0.91). The precision in prediction is not a criterion to opt for an algorithm or another when generating predictive models with the analyzed data. This result gives us the freedom to use other criteria to select the algorithm, for example, the execution time or the ease of implementation in different platforms and languages.

Execution time: The data concerning to the execution time did not fulfill the assumption of normality, therefore we use non-parametric test Kruskall-Wallis to test Ho:"*The difference between the means of the samples is equal to zero*". The results show strong evidence (χ^2=57.8, $p = 4.11E - 11$) of a significant difference between at least a couple of samples. We use compare the means by pairs to identify the algorithms that better performs in time. We use the Wilcox test and we show the results in a crossed table and a box plot

(see Section4.3.2). The Artificial Neural Networks reached the best performance while the Linear Logistic Regression took the second place. We found difference statistically significant among these algorithms and the other five variants of MLA.

The criteria for choosing an algorithm are three, and we listed them below in order of importance: 1) Accuracy in the prediction of sleep quality, the primary objective of the model. 2) The feasibility and simplicity of the implementation of the algorithm in different platforms, for example, web and mobile applications for smart phones and smart watches. 3) The execution time. The time the third criterion because to this system the time between predictions is in the term of weeks. Since we did not found significant differences among the algorithms for the accuracy criterion, we use the second one. Of the three algorithms and their variants, the regression of degree one is the most feasible algorithm to implement in different platforms. It is a light algorithm simple to reproduce in any programming language.

6.1.6 To what extent do the SHF obtained from the SP characterizes the SQ of a PIMEV observed individually?

To answer this question a longitudinal study was conducted, where three volunteers measured their sleep quality for 30 days using an electronic device that observed physiological parameters and objective data on their sleep patterns. The same number of days the volunteers captured data assessing their sleep hygiene through the 21 SHF. We use the data corresponding to the factors selected in the previous stage to carried out an extrapolation process in the aim of reach enough data to train the model. One of the three models explains more than 70% of the variance of the data while for the remaining two, the coefficient of determination R^2 was below to 50%. The answer to the research question indicates that we cannot use the selected factors characterizing SP to explain the data obtained from the observation of one subject of the same population.

These results evince that a MLA based system for estimating sleep quality must learn from personal data instead of data from a sample of the population. Such that the factors identified by the model as relevant are those SHF that best describe the subject.

6.1.7 Is it possible to identify a subset of SHF producing personalized models to predict SQ by reaching at least 90% of efficiency?

We designed a prototype of the system to generate customized models to estimate SQ from personal data. We extrapolated the data corresponding to the 21 SHF and the SQ score from the Beddit device for each volunteer. After this process, we train a model per volunteer through a Linear Regression algorithm. Similar to the test with three factors the results were not satisfactory. We did not get any model explaining more than 73% of the variance of the data. We optimize the training by using FSA to eliminate confounding and correlated factors. This process improves the predictions; however, the coefficient of determination to measure the explanation of the variance of the data did not reach more than 85% in the best model as we showed in Table 6.2

Table 6.2: R^2 of models by selecting features through three different algorithms

	Random Forest		**LASSO**		**XGBOOST**	
	SHF	R^2	**SHF**	R^2	**SHF**	R^2
VF1	1,4,10	85%	13,1,6,20,10, 14,7,11,12	79%	7,6,20,1,10	76%
VF2	21,5,6,20,2	81%	1,5,3,14,18, 21,6,15,2	70%	1,14,5,9,7	83%
VM	19,18,21,2,20, 13,4,12,5	60%	2,3,18,13,12, 7,5,20,6	29%	12,3,21,5,20, 10	70%

The algorithm `'fsXLR'` (see 3) was designed by adding robustness to the selection of features by integrating '*bagging*' and '*best-search*' in a recursive algorithm that uses three feature selection algorithms (XGBOOST, LASSO and RF). The algorithm extracts the relevant factors classified as primary, secondary and tertiary. In addition, it finds the best subset of those factors that would explain a high percentage of the variance of the data of each subject (See 5.4).

In all the cases of the subjects that participated in the longitudinal study, we found that there is more than one combination of factors that explain the variance of the data in high percentages (above 90 %). This led us to validate the prototype through the construction of hypothetical datasets, for example, profiles such as VM2 were generated whose quality of sleep depends on

four factors: Caffeine, time to get up, stress before sleeping and worries before sleeping. When we run the process to identify the critical factors to this subject, the algorithm selected exactly the four factors above and it generated only a regression involving this SHF. This model explained the 0.94% of the variance of the data.

Since the prototype generates models that can predict efficiently when there is a set of SHF that directly affects SQ, we could infer that for the study subjects who participated in this research there is no dominant set of SHF influencing the SQ. We need to observe SH and SQ in a longer time to determine if there is a really dominant set of factors as in the hypothetical cases.

6.1.8 How much observation time is it necessary?

We perform tests using the extrapolated data as giving information (from 30 days of observation to 48 months). The sections 5.5 and 5.6 show the process in detail. After we determined the set of SHF that explains the highest percentage of the variance with a quantity of data corresponding to 208 weeks, we carried out tests to determine the number of records we need to produce efficient models (each record equals one week). The results varied for each subject: the model of **VF1** needed 46, **VF2** needed 41 and **VM** needed 20 weeks to start producing satisfactory results. From these results we can infer that there is no specific time, or a certain time range for models to predict efficiently SQ from SH. The time that the model takes to stabilize depends on how regular are the patterns read in the data. For data from subjects with regular patterns the convergence times are shorter than in those whom parameters are irregular.

6.2 Conclusions and future work

In this work we tested the hypothesis that it is possible to predict SQ from monitoring SHF. We chose a population of people that work more intellectual than physically to find if there were SHF that would characterize their SQ. We conducted this finding through a cross-sectional study by analyzing the data through FSA. We identified three SHF that produce the best predictions: *Inconsistency in getting up* (SH2), *stress before going to sleep* (SH12) and *addressing concerns*

before going to sleep (SH20).

We use eight variants of MLAs to contrast the predictive efficiency by the complete set of SHF (21F) and do so by the subset of the three selected SHF (3F). Models trained by 3F are more efficient at predicting than those trained by 21F (p = 0.0095). Using 3F on average the algorithms predict with 60% efficiency (min = 44%, max = 73%, $\sigma = 0.068$). The most consistent algorithm (LR-DG-1) predicts an average of 61% (min = 53%, max = 71%, $\sigma = 0.047$). We validated the prediction efficiency by training regression models using the data from observations made individually to three subjects of the gls sp by a 30-day longitudinal study. Only in one of the cases does the variance explanation rate ($R^2 = 79\%$) exceed the highest prediction percentage of SP (73%). In the data of the two remaining subjects the explanation of the variance is very low ($R^2 = 40\% and 29\%$). The factors that characterize the SP can not be generalized to particular cases to achieve at least the same prediction efficiency.

We developed a process to generate models that explain in at least 90% the variance of the data of the longitudinal study participants. The process includes the development of a variable selection algorithm (fsXLR) exploiting the power of three algorithms (XGBOOST, LASSO, RF) and two techniques (Bagging, Best-Search) that have been widely used for feature selection purposes. The tests show that fsXLR selects factors that generate more efficient models (> 90%) than those generated by the factors selected with any of the three algorithms ($R^2_{\max(XGBOOST)} = 83\%$, $R^2_{\max(LASSO)} = 79\%$, $R^2_{\max(RandomForest)} = 85\%$) used individually outside of the fsXLR environment. On the other hand, convergence tests of the models show that a minimum of 20 (VM) and a maximum of 46 (VF1) observation weeks are needed to start producing efficient models.

We found that the study participants have a perception of their level of SQ and SH comparable to the results of the PSQI-SH questionnaire and the estimates made by the Beddit electronic device. However, the results of the study of concordance between their opinions about the most influential SHF in their sleep quality and those found by the fsXLR algorithm is low. Considering four levels of classification $Kappa_{\max} = 0.19, p = 0.15$ and considering two levels $Kappa_{\max} = 0.42, p = 0.05$, both for VM. Cohen's Kappa for VF1 and VF2 is below 0.25 at any of the rating levels ($p > 0.26$ in all cases). This shows that the study subjects had a misperception

about the SHF that most influence their SQ.

Our work from the perspective of computer science contributes to research on SQ and SH. However, the results require verification through more extensive studies, with a larger population and conducted under clinical protocols by an interdisciplinary team.

Bibliography

[1] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [En línea]. Disponible: https://www.R-project.org/

[2] RStudio Team, *RStudio: Integrated Development Environment for R*, RStudio, Inc., Boston, MA, 2016. [En línea]. Disponible: http://www.rstudio.com/

[3] H. Wickham, R. Francois, L. Henry, and K. Mller, *dplyr: A Grammar of Data Manipulation*, 2017, r package version 0.7.4. [En línea]. Disponible: https://CRAN.R-project.org/package=dplyr

[4] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch, *e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien*, 2017, r package version 1.6-8. [En línea]. Disponible: https://CRAN.R-project.org/package=e1071

[5] A. Beygelzimer, S. Kakadet, J. Langford, S. Arya, D. Mount, and S. Li, *FNN: Fast Nearest Neighbor Search Algorithms and Applications*, 2013, r package version 1.1. [En línea]. Disponible: https://CRAN.R-project.org/package=FNN

[6] Microsoft and S. Weston, *foreach: Provides Foreach Looping Construct for R*, 2017, r package version 1.4.4. [En línea]. Disponible: https://CRAN.R-project.org/package=foreach

[7] P. Romanski and L. Kotthoff, *FSelector: Selecting Attributes*, 2016, r package version 0.21. [En línea]. Disponible: https://CRAN.R-project.org/package=FSelector

[8] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. [En línea]. Disponible: http://ggplot2.org

[9] A. Kassambara, *ggpubr: 'ggplot2' Based Publication Ready Plots*, 2017, r package version 0.1.6. [En línea]. Disponible: https://CRAN.R-project.org/package=ggpubr

[10] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software, Articles*, vol. 33, no. 1, pp. 1–22, 2010. [En línea]. Disponible: https://www.jstatsoft.org/v033/i01

[11] G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables, *gplots: Various R Programming Tools for Plotting Data*, 2016, r package version 3.0.1. [En línea]. Disponible: https://CRAN.R-project.org/package=gplots

[12] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017. [En línea]. Disponible: https://www.R-project.org/

[13] S. M. Bache and H. Wickham, *magrittr: A Forward-Pipe Operator for R*, 2014, r package version 1.5. [En línea]. Disponible: https://CRAN.R-project.org/package=magrittr

[14] D. Bates and M. Maechler, *Matrix: Sparse and Dense Matrix Classes and Methods*, 2018, r package version 1.2-14. [En línea]. Disponible: https://CRAN.R-project.org/package=Matrix

[15] S. van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–67, 2011. [En línea]. Disponible: http://www.jstatsoft.org/v45/i03/

[16] S. Fritsch and F. Guenther, *neuralnet: Training of Neural Networks*, 2016, r package version 1.33. [En línea]. Disponible: https://CRAN.R-project.org/package=neuralnet

[17] O. Dag, A. Dolgun, and N. Konar, "onewaytests: An r package for one-way tests in independent groups designs." *The R Journal*, 2018.

[18] L. Komsta, *outliers: Tests for outliers*, 2011, r package version 0.14. [En línea]. Disponible: https://CRAN.R-project.org/package=outliers

[19] C. Strobl, A.-L. Boulesteix, A. Zeileis, and T. Hothorn, "Bias in random forest variable importance measures: Illustrations, sources and a solution," *BMC Bioinformatics*, vol. 8, no. 1, p. 25, Jan 2007. [En línea]. Disponible: https://doi.org/10.1186/1471-2105-8-25

[20] T. Hothorn, P. Bhlmann, S. Dudoit, A. Molinaro, and M. J. Van Der Laan, "Survival ensembles," *Biostatistics*, vol. 7, no. 3, pp. 355–373, 2006. [En línea]. Disponible: +http://dx.doi.org/10.1093/biostatistics/kxj011

[21] C. Strobl, A.-L. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis, "Conditional variable importance for random forests," *BMC Bioinformatics*, vol. 9, no. 1, p. 307, Jul 2008. [En línea]. Disponible: https://doi.org/10.1186/1471-2105-9-307

[22] C. Sievert, C. Parmer, T. Hocking, S. Chamberlain, K. Ram, M. Corvellec, and P. Despouy, *plotly: Create Interactive Web Graphics via 'plotly.js'*, 2017, r package version 4.7.1. [En línea]. Disponible: https://CRAN.R-project.org/package=plotly

[23] X. Robin, N. Turck, A. Hainard, N. Tiberti, F. Lisacek, J.-C. Sanchez, and M. Mller, "proc: an open-source package for r and s+ to analyze and compare roc curves," *BMC Bioinformatics*, vol. 12, p. 77, 2011.

[24] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002. [En línea]. Disponible: http://CRAN.R-project.org/doc/Rnews/

[25] H. Wickham, "Reshaping data with the reshape package," *Journal of Statistical Software*, vol. 21, no. 12, 2007. [En línea]. Disponible: http://www.jstatsoft.org/v21/i12/paper

[26] J. VanDerWal, L. Falconi, S. Januchowski, L. Shoo, and C. Storlie, *SDMTools: Species Distribution Modelling Tools: Tools for processing data associated with species distribution modelling exercises*, 2014, r package version 1.1-221. [En línea]. Disponible: https://CRAN.R-project.org/package=SDMTools

[27] W. Chang, J. Cheng, J. Allaire, Y. Xie, and J. McPherson, *shiny: Web Application Framework for R*, 2017, r package version 1.0.5. [En línea]. Disponible: https://CRAN.R-project.org/package=shiny

[28] W. Chang, *shinythemes: Themes for Shiny*, 2016, r package version 1.1.1. [En línea]. Disponible: https://CRAN.R-project.org/package=shinythemes

[29] A. Kowarik and M. Templ, "Imputation with the R package VIM," *Journal of Statistical Software*, vol. 74, no. 7, pp. 1–16, 2016.

[30] T. Chen, T. He, M. Benesty, V. Khotilovich, and Y. Tang, *xgboost: Extreme Gradient Boosting*, 2017, r package version 0.6-4. [En línea]. Disponible: https://CRAN.R-project.org/package=xgboost

[31] D. J. Buysse, "Sleep health: Can we define it? does it matter?" *Sleep*, vol. 37, no. 1, pp. 9–17, 2014.

[32] H. Liu and H. Motoda, *Feature selection for knowledge discovery and data mining*. Springer, Boston, MA, 1998.

[33] M. Hirshkowitz, K. Whiton, S. M. Albert, C. Alessi, O. Bruni, L. DonCarlos, N. Hazen, J. Herman, E. S. Katz, L. Kheirandish-Gozal, D. N. Neubauer, A. E. ODonnell, M. Ohayon, J. Peever, R. Rawding, R. C. Sachdeva, B. Setters, M. V. Vitiello, J. C. Ware, and P. J. A. Hillard, "National sleep foundations sleep time duration recommendations: methodology and results summary," *Sleep Health*, vol. 1, no. 1, pp. 40 – 43, 2015.

[34] D. Evans, "Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions," *Journal of clinical nursing*, vol. 12, no. 1, pp. 77–84, 2003.

[35] D. Pati, "A framework for evaluating evidence in evidence-based design," *HERD: Health Environments Research & Design Journal*, vol. 4, no. 3, pp. 50–71, 2011.

[36] P. Miller and A. Jones-Harris, "The evidence-based hierarchy: Is it time for change? a suggested alternative," *Journal of Manipulative and Physiological Therapeutics*, vol. 28, no. 6, pp. 453 – 457, 2005. [En línea]. Disponible: http://www.sciencedirect.com/science/article/pii/S0161475405001715

[37] J. Tang, S. Alelyani, and H. Liu, "Feature Selection for Classification: A Review," *Data Classification: Algorithms and Applications*, pp. 37–64, 2014.

[38] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: A survey of the state of the art and research challenges," *CoRR*, vol. abs/1306.2356, 2013. [En línea]. Disponible: http://arxiv.org/abs/1306.2356

[39] M. Botterman, "Internet of Things : an early reality of the Future Internet," *For the European Commision Information Society and Media Directorate General, Networked Enterprise & RFID Unit-D4*, 2009, czech Republic.

[40] WHO, "Global status report on noncommunicable diseases 2014," Geneva, Tech. Rep., 2014. [En línea]. Disponible: http://www.who.int/nmh/publications/ncd-status-report-2014/en/

[41] ODCE, "Obesity Update 2017," Tech. Rep., 2017, information accessed in October 2017. [En línea]. Disponible: www.oecd.org/health/obesity-update.htm

[42] F. Lucivero and B. Prainsack, "The lifestylisation of healthcare? 'Consumer genomics' and mobile health as technologies for healthy lifestyle," *Applied and Translational Genomics*, vol. 4, pp. 44–49, 2015.

[43] S. Luo, H. Xia, Y. Gao, J. S. Jin, and R. Athauda, "Smart Fridges with Multimedia Capability for Better Nutrition and Health," *2008 International Symposium on Ubiquitous Multimedia Computing*, pp. 39–44, oct 2008.

[44] H. Badawi and A. E. Saddik, "Towards a Context-Aware Biofeedback Activity Recommendation Mobile Application for Healthy Lifestyle," *Procedia Computer Science*, vol. 21, no. 613, pp. 382–389, jan 2013.

[45] B. Silva, I. Lopes, J. Rodrigues, and P. Ray, "Sapofitness: A mobile health application for dietary evaluation," in *e-Health Networking Applications and Services (Healthcom), 2011 13th IEEE International Conference on*, June 2011, pp. 375–380.

[46] T. Chang, "Food fight: A social diet network mobile application," Master's thesis, EECS Department, University of California, Berkeley, May 2012. [En línea]. Disponible: http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-133.html

[47] E. Kyriacou, C. Pattichis, M. Pattichis, A. Jossif, L. Paraskeva, A. Konstantinides, and D. Vogiatzis, "An m-health monitoring system for children with suspected arrhythmias," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007, pp. 1794–1797.

[48] M. Vazquez-briseno, M. Diaz-arce, and E. Jimenez-garcia, "mhealth platform and architectures to provide nutritional guidance to children," *International Journal of Interactive Mobile Technologies*, vol. 7, no. 4, pp. 15–20, 2013.

[49] S. Youm, G. Lee, S. Park, and W. Zhu, "Development of remote healthcare system for measuring and promoting healthy lifestyle," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2828–2834, Mar. 2011.

[50] M. J. Hasselberg, J. McMahon, and K. Parker, "The validity, reliability, and utility of the iButton for measurement of body temperature circadian rhythms in sleep/wake research." *Sleep medicine*, vol. 14, no. 1, pp. 5–11, jan 2013.

[51] T. Hao, G. Xing, and G. Zhou, "isleep: Unobtrusive sleep quality monitoring using smartphones," in *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '13. New York, NY, USA: ACM, 2013, pp. 4:1–4:14.

[52] T. Scherini, P. Melo, T. van Craenendonck, W. Zou, and M. Kaptein, "Enhancing the sleeping quality of partners living apart," in *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, ser. DIS '10. New York, NY, USA: ACM, 2010, pp. 171–174.

[53] M. Sekine, T. Tatsuse, N. Cable, T. Chandola, and M. Marmot, "U-shaped associations between time in bed and the physical and mental functioning of Japanese civil servants: The roles of work, family, behavioral and sleep quality characteristics," *Sleep Medicine*, vol. 15, no. 9, pp. 1122–1131, 2014.

[54] A. P. Spira, Y. An, and S. M. Resnick, "Self-reported sleep and b-amyloid deposition in community older adults–reply," *JAMA Neurol*, vol. 71, no. 5, pp. 651–652, May 2014, 24818679[pmid].

[55] K. Suzuki, M. Miyamoto, T. Miyamoto, M. Iwanami, and K. Hirata, "Sleep Disturbances Associated with Parkinson's Disease," *Parkinson's Disease*, vol. 2011, pp. 1–10, 2011.

[56] S. R. Patel and F. B. Hu, "Short sleep duration and weight gain: A systematic review," *Obesity*, vol. 16, no. 3, pp. 643–653, 2008.

[57] S. Taheri, L. Lin, D. Austin, T. Young, and E. Mignot, "Short sleep duration is associated with reduced leptin, elevated ghrelin, and increased body mass index," *PLoS Med*, vol. 1, no. 3, p. e62, Dec 2004, 15602591[pmid].

[58] B.-Q. Zhu, X.-M. Li, D. Wang, and X.-F. Yu, "Sleep quality and its impact on glycaemic control in patients with type 2 diabetes mellitus," *International Journal of Nursing Sciences*, vol. 1, no. 3, pp. 260–265, 2014.

[59] S. C. Markt, A. Grotta, O. Nyren, H.-O. Adami, L. A. Mucci, U. A. Valdimarsdottir, P. Stattin, R. Bellocco, and Y. T. Lagerros, "Insufficient sleep and risk of prostate cancer in a large swedish cohort," *Sleep*, vol. 38, no. 9, pp. 1405–1410, Sep 2015, sp-00795-14[PII].

[60] X. Zhang, E. L. Giovannucci, K. Wu, X. Gao, F. Hu, S. Ogino, E. S. Schernhammer, C. S. Fuchs, S. Redline, W. C. Willett, and J. Ma, "Associations of self-reported sleep duration and snoring with colorectal cancer risk in men and women," *Sleep*, vol. 36, no. 5, pp. 681–688, May 2013.

[61] S. J. Bolitho, S. L. Naismith, S. M. W. Rajaratnam, R. R. Grunstein, J. R. Hodges, Z. Terpening, N. Rogers, and S. J. G. Lewis, "Disturbances in melatonin secretion and circadian sleep-wake regulation in Parkinson disease." *Sleep medicine*, vol. 15, no. 3, pp. 342–7, mar 2014.

[62] K. Chang, S. Son, Y. Lee, J. Back, K. Lee, S. Lee, Y. Chung, K. Lim, J. Noh, H. Kim, S. Koh, H. Roh, M. Park, J. Kim, and C. Hong, "Perceived sleep quality is associated with depression in a korean elderly population," *Archives of gerontology and geriatrics*, vol. 59, no. 2, pp. 468–473, 2014.

[63] Y. ABE, K. MISHIMA, Y. KANEITA, L. LI, T. OHIDA, T. NISHIKAWA, and M. UCHIYAMA, "Stress coping behaviors and sleep hygiene practices in a sample of japanese adults with insomnia," *Sleep and Biological Rhythms*, vol. 9, no. 1, pp. 35–45, 2011.

[64] S. de Biase, G. Milioli, A. Grassi, S. Lorenzut, L. Parrino, and G. L. Gigli, *Sleep Hygiene*. Milano: Springer Milan, 2014, pp. 289–295.

[65] L. Genzel, M. C. W. Kroes, M. Dresler, and F. P. Battaglia, "Light sleep versus slow wave sleep in memory consolidation: A question of global versus local processes?" *Trends in Neurosciences*, vol. 37, no. 1, pp. 10–19, 2014.

[66] Z. Xia and D. Storm, "Role of circadian rhythm and rem sleep for memory consolidation," *Neuroscience Research*, vol. 118, pp. 13 – 20, 2017, cutting-edge Approaches to Unwrapping the Mysteries of Sleep. [En línea]. Disponible: http://www.sciencedirect.com/science/article/pii/S0168010217302377

[67] B. Portha, A. Fournier, M. D. Ah Kioon, V. Mezger, and J. Movassat, "Early environmental factors, alteration of epigenetic marks and metabolic disease susceptibility," *Biochimie*, vol. 97, no. 1, pp. 1–15, 2014.

[68] V. A. Rozanov, "Epigenetics: Stress and behavior," *Neurophysiology*, vol. 44, no. 4, pp. 332–350, 2012.

[69] R. Jaenisch and A. Bird, "Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals," *Nature Genetics*, vol. 33, no. 3S, pp. 245–254, 2003.

[70] M. B. First, R. L. Spitzer, M. Gibbon, J. B. Williams *et al.*, "Structured clinical interview for dsm-iv-tr axis i disorders, research version, patient edition," SCID-I/P, Tech. Rep., 2002.

[71] D. J. Buysse, C. F. R. Reynolds, T. H. Monk, S. R. Berman, and D. J. Kupfer, "The Pittsburgh Sleep Quality Index: A New Instrument for Psychiatric Practice and Research," pp. 193–213, 1989. [En línea]. Disponible: http://www.ncbi.nlm.nih.gov/pubmed/2748771

[72] L. K. Pilz, L. K. Keller, D. Lenssen, and T. Roenneberg, "Time to rethink sleep quality: Psqi scores reflect sleep quality on workdays," *Sleep*, p. zsy029, 2018.

[73] L. Meltzer, H. Montgomery-Downs, S. Insana, and C. Walsh, "Use of actigraphy for assessment in pediatric sleep research," *Sleep Medicine Reviews*, vol. 16, no. 5, pp. 463–475, 2012.

[74] A. Sadeh, "The role and validity of actigraphy in sleep medicine: An update," *Sleep Medicine Reviews*, vol. 15, no. 4, pp. 259–267, 2011.

[75] E. K. Choe, S. Consolvo, N. F. Watson, and J. A. Kientz, "Opportunities for computing technologies to support healthy sleep behaviors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 3053–3062. [En línea]. Disponible: http://doi.acm.org/10.1145/1978942.1979395

[76] M. Kay, E. K. Choe, J. Shepherd, B. Greenstein, N. Watson, S. Consolvo, and J. A. Kientz, "Lullaby: A capture & access system for understanding the sleep environment," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, ser. UbiComp '12. New York, NY, USA: ACM, 2012, pp. 226–234.

[77] J. Behar, A. Roebuck, J. S. Domingos, E. Gederi, and G. D. Clifford, "A review of current sleep screening applications for smartphones," *Physiological Measurement*, vol. 34, no. 7, p. R29, 2013. [En línea]. Disponible: http://stacks.iop.org/0967-3334/34/i=7/a=R29

[78] D. Posner and P. R. Gehrman, "Sleep hygiene," in *Behavioral Treatments for Sleep Disorders*, B. Perlis, Michael and Aloia, Mark and Kuhn, Ed. San Diego: Academic Press, 2011, ch. Sleep Hygi, pp. 31–43.

[79] B. I. Voinescu and A. Szentagotai-Tatar, "Sleep hygiene awareness: its relation to sleep quality and diurnal preference," *J Mol Psychiatry*, vol. 3, no. 1, p. 1, Jan 2015.

[80] B. Bjorvatn, E. Fiske, and S. Pallesen, "A self-help book is better than sleep hygiene advice for insomnia: A randomized controlled comparative study," *Scandinavian Journal of Psychology*, vol. 52, no. 6, pp. 580–585, 2011.

[81] C. A. Brick, D. L. Seely, and T. M. Palermo, "Association between sleep hygiene and sleep quality in medical students," *Behav Sleep Med*, vol. 8, no. 2, pp. 113–121, 2010.

[82] Y. ABE, K. MISHIMA, Y. KANEITA, L. LI, T. OHIDA, T. NISHIKAWA, and M. UCHIYAMA, "Stress coping behaviors and sleep hygiene practices in a sample of japanese adults with insomnia," *Sleep and Biological Rhythms*, vol. 9, no. 1, pp. 35–45, 2011.

[83] J. S. Peltz and R. D. Rogge, "The indirect effects of sleep hygiene and environmental factors on depressive symptoms in college students," *Sleep Health*, vol. 2, no. 2, pp. 159 – 166, 2016.

[84] M. C. Teodorescu, A. Y. Avidan, M. Teodorescu, J. J. Harrington, A. O. Artar, C. R. Davies, and R. D. Chervin, "Sleep medicine content of major medical textbooks continues to be underrepresented," *Sleep Medicine*, vol. 8, no. 3, pp. 271–276, 2018/01/26 2007.

[85] J. S. Bauer, S. Consolvo, B. Greenstein, J. Schooler, E. Wu, N. F. Watson, and J. Kientz, "Shuteye: Encouraging awareness of healthy sleep recommendations with a mobile, peripheral display," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '12. New York, NY, USA: ACM, 2012, pp. 1401–1410.

[86] T. Shamah-Levi, L. Cuevas-Nasu, J. Dommarco-Rivera, and M. Hernandez-Avila, "Encuesta Nacional de Salud y Nutrición de Medio Camino 2016." Instituto Nacional de Salud Pública, Tech. Rep., 2016. [En línea]. Disponible: http://promocion.salud.gob.mx/dgps/descargas1/doctos{_}2016/ensanut{_}mc{_}2016-310oct.pdf

[87] L. A. Irish, C. E. Kline, H. E. Gunn, D. J. Buysse, and M. H. Hall, "The role of sleep hygiene in promoting public health: A review of empirical evidence," *Sleep Med Rev*, vol. 22, pp. 23–36, Aug 2015.

[88] P. H. Finan, P. J. Quartana, and M. T. Smith, "The effects of sleep continuity disruption on positive mood and sleep architecture in healthy adults," *Sleep*, vol. 38, no. 11, pp. 1735–1742, Nov 2015, sp-00003-15[PII].

[89] A. Luca, M. Luca, and C. Calandra, "Sleep disorders and depression: brief review of the literature, case report, and nonpharmacologic interventions for depression," *Clin Interv Aging*, vol. 8, pp. 1033–1039, Aug 2013, cia-8-1033[PII].

[90] O. V. Reeth, L. Weibel, K. Spiegel, R. Leproult, C. Dugovic, and S. Maccari, "Physiology of sleep (review)interactions between stress and sleep: from basic research to clinical situations," *Sleep Medicine Reviews*, vol. 4, no. 2, pp. 201 – 219, 2000.

[91] I. Alapin, C. S. Fichten, E. Libman, L. Creti, S. Bailes, and J. Wright, "How is good and poor sleep in older adults and college students related to daytime sleepiness, fatigue, and ability to concentrate?" *Journal of Psychosomatic Research*, vol. 49, no. 5, pp. 381 – 390, 2000.

[92] E. J. Stepanski and J. K. Wyatt, "Use of sleep hygiene in the treatment of insomnia," *Sleep Medicine Reviews*, vol. 7, no. 3, pp. 215–225, 2003.

[93] C.-M. Yang, S.-C. Lin, S.-C. Hsu, and C.-P. Cheng, "Maladaptive sleep hygiene practices in good sleepers and patients with insomnia," *Journal of Health Psychology*, vol. 15, no. 1, pp. 147–155, 2010, pMID: 20064894.

[94] L. A. Gellis and K. L. Lichstein, "Sleep hygiene practices of good and poor sleepers in the united states: An internet-based study," *Behavior Therapy*, vol. 40, no. 1, pp. 1–9, 2009.

[95] Y. Yamanaka, S. Hashimoto, N. N. Takasu, Y. Tanahashi, S.-y. Nishide, S. Honma, and K.-i. Honma, "Morning and evening physical exercise differentially regulate the autonomic nervous system during nocturnal sleep in humans," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 309, no. 9, pp. R1112–R1121, 2015.

[96] D. G. Davila, R. D. Hurt, K. P. Offord, C. D. Harris, and J. W. Shepard Jr, "Acute effects of transdermal nicotine on sleep architecture, snoring, and sleep-disordered breathing in nonsmokers." *American journal of respiratory and critical care medicine*, vol. 150, no. 2, pp. 469–474, 1994.

[97] D. Posner and P. R. Gehrman, "Sleep Hygiene," in *Behavioral Treatments for Sleep Disorders*, B. Perlis, Michael and Aloia, Mark and Kuhn, Ed. San Diego: Academic Press, 2011, ch. Sleep Hygi, pp. 31–43. [En línea]. Disponible: http://www.sciencedirect.com/science/article/pii/B9780123815224000031

[98] M. Carús-Cadavieco and I. de Andrés, "Adenosina y control homeostático del sueño. acciones en estructuras diana de los circuitos de vigilia y sueño," *Revista de Neurologia*, vol. 55, no. 7, pp. 413–420, 2012.

[99] C. Drake, T. Roehrs, J. Shambroom, and T. Roth, "Caffeine effects on sleep taken 0, 3, or 6 hours before going to bed," *J Clin Sleep Med*, vol. 9, no. 11, pp. 1195–1200, Nov 2013.

[100] R. Beun, "Persuasive strategies in mobile insomnia therapy: alignment, adaptation, and motivational support," *Personal and Ubiquitous Computing*, vol. 17, no. 6, pp. 1187–1195, 2013.

[101] A. Shechter, E. W. Kim, M.-P. St-Onge, and A. J. Westwood, "Blocking nocturnal blue light for insomnia: A randomized controlled trial," *Journal of Psychiatric Research*, vol. 96, pp. 196 – 202, 2018.

[102] S. A. Rahman, M. A. S. Hilaire, and S. W. Lockley, "The effects of spectral tuning of evening ambient light on melatonin suppression, alertness and sleep," *Physiology & Behavior*, vol. 177, pp. 221 – 229, 2017.

[103] J. Grnli, I. K. Byrkjedal, B. Bjorvatn, ystein Ndtvedt, B. Hamre, and S. Pallesen, "Reading from an ipad or from a book in bed: the impact on human sleep. a randomized controlled crossover trial," *Sleep Medicine*, vol. 21, pp. 86 – 92, 2016.

[104] D. Wentz and M. Wentz, *The Healthy Home: Simple Truths to Protect Your Family from Hidden Household Dangers*. Vanguard, 2011.

[105] A. KAKU, N. NISHINOUE, T. TAKANO, R. ETO, N. KATO, Y. ONO, and K. TANAKA, "Randomized Controlled Trial on the Effects of a Combined Sleep Hygiene Education and Behavioral Approach Program on Sleep Quality in Workers with Insomnia," *Industrial Health*, vol. 50, no. 1, pp. 52–59, 2012.

[106] C. Occhiuzzi and G. Marrocco, "The rfid technology for neurosciences: Feasibility of limbs' monitoring in sleep diseases," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 1, pp. 37–43, 2010.

[107] J. Behar, A. Roebuck, M. Shahid, J. Daly, A. Hallack, N. Palmius, J. Stradling, and G. Clifford, "Sleepap: An automated obstructive sleep apnoea screening application for

smartphones," *Biomedical and Health Informatics, IEEE Journal of*, vol. 19, no. 1, pp. 325–331, Jan 2015.

[108] L. Parra, S. Sendra, J. M. Jiménez, and J. Lloret, "Multimedia sensors embedded in smartphones for ambient assisted living and e-health," *Multimedia Tools and Applications*, pp. 1–27, 2015, article in Press.

[109] P. Klasnja, S. Consolvo, and W. Pratt, "How to evaluate technologies for health behavior change in hci research," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '11. New York, NY, USA: ACM, 2011, pp. 3063–3072.

[110] A. Shirazi, J. Clawson, Y. Hassanpour, M. Tourian, A. Schmidt, E. Chi, M. Borazio, and K. Van Laerhoven, "Already up? using mobile phones to track & share sleep behavior," *International Journal of Human Computer Studies*, vol. 71, no. 9, pp. 878–888, 2013.

[111] N. D. Lane, M. Lin, M. Mohammod, X. Yang, H. Lu, G. Cardone, S. Ali, A. Doryab, E. Berke, A. T. Campbell *et al.*, "Bewell: Sensing sleep, physical activities and social interactions to promote wellbeing," *Mobile Networks and Applications*, vol. 19, no. 3, pp. 345–359, 2014.

[112] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf wifi," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc '15. New York, NY, USA: ACM, 2015, pp. 267–276.

[113] C.-W. Jeong, S.-C. Joo, and Y. Jeong, "Sleeping situation monitoring system in ubiquitous environments," *Personal and Ubiquitous Computing*, vol. 17, no. 7, pp. 1357–1364, 2013.

[114] W.-H. Liao and J.-H. Kuo, "Sleep monitoring system in real bedroom environment using texture-based background modeling approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. 4, no. 1, pp. 57–66, 2013.

[115] S. Lee, L. Yan, T. Roh, S. Hong, and H.-J. Yoo, "The smart patches and wearable band (w-band) for comfortable sleep monitoring system," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, Aug 2011, pp. 6915–6918.

[116] V. Natale, D. Lger, M. Martoni, V. Bayon, and A. Erbacci, "The role of actigraphy in the assessment of primary insomnia: A retrospective study," *Sleep Medicine*, vol. 15, no. 1, pp. 111–115, 2014.

[117] M. Borazio and K. Van Laerhoven, "Combining wearable and environmental sensing into an unobtrusive tool for long-term sleep studies," in *Proceedings of the 2Nd ACM SIGHIT International Health Informatics Symposium*, ser. IHI '12. New York, NY, USA: ACM, 2012, pp. 71–80.

[118] A. Heinrich, V. Jeanne, and X. Zhao, "Lifestyle applications from sleep research," *Journal of Ambient Intelligence and Humanized Computing*, vol. 5, no. 6, pp. 829–842, 2014.

[119] V. Pejovic and M. Musolesi, "Anticipatory mobile computing: A survey of the state of the art and research challenges," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 47:1–47:29, Apr. 2015.

[120] F. Yousaf, S.-B. Hamida, and B. Ahmed, "Sleep ediary: A mobile health application for insomnia assessment," in *Middle East Conference on Biomedical Engineering, MECBME*, 2014, pp. 285–288.

[121] Z. Chen, M. Lin, F. Chen, N. D. Lane, G. Cardone, R. Wang, T. Li, Y. Chen, T. Choudhury, and A. T. Campbell, "Unobtrusive sleep monitoring using smartphones," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2013 7th International Conference on*. IEEE, 2013, pp. 145–152.

[122] Y. Bai, B. Xu, Y. Ma, G. Sun, and Y. Zhao, "Will you have a good sleep tonight?: sleep quality prediction with mobile phone," in *Proceedings of the 7th International Conference on Body Area Networks*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2012, pp. 124–130.

[123] S. Lawson, S. Jamison-Powell, A. Garbett, C. Linehan, E. Kucharczyk, S. Verbaan, D. A. Rowland, and K. Morgan, "Validating a mobile phone application for the everyday, unobtrusive, objective measurement of sleep," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '13. New York, NY, USA: ACM, 2013, pp. 2497–2506.

[124] E. K. Choe, "Design of persuasive technologies for healthy sleep behavior," in *Proceedings of the 13th International Conference on Ubiquitous Computing*, ser. UbiComp '11. New York, NY, USA: ACM, 2011, pp. 507–510.

[125] S. Kim, J. A. Kientz, S. N. Patel, and G. D. Abowd, *Are You Sleeping?: Sharing Portrayed Sleeping Status Within a Social Network*, ser. CSCW '08. New York, NY, USA: ACM, 2008, pp. 619–628. [En línea]. Disponible: http://libcon.rec.uabc.mx:3182/10.1145/1460563.1460660

[126] Y.-X. Chen, H.-C. Chen, L.-X. Chen, J.-W. Hu, C.-K. Shie, Y.-S. Lin, P. Borade, C.-C. Yeh, H.-H. Lin, S.-S. Chiang, Y.-C. Chen, W.-Z. Sun, and Y.-P. Hung, "Enhancing adherence to cognitive behavioral therapy for insomnia through machine and social persuasion," in *Proceedings - 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, GreenCom-iThings-CPSCom 2013*, 2013, pp. 750–757.

[127] D. Gartenberg, R. Thornton, M. Masood, D. Pfannenstiel, D. Taylor, and R. Parasuraman, "Collecting health-related data on the smart phone: Mental models, cost of collection, and perceived benefit of feedback," *Personal and Ubiquitous Computing*, vol. 17, no. 3, pp. 561–570, 2013.

[128] M. J. Deen, "Information and communications technologies for elderly ubiquitous healthcare in a smart home," *Personal and Ubiquitous Computing*, vol. 19, no. 3-4, pp. 573–599, 2015.

[129] V. Metsis, D. Kosmopoulos, V. Athitsos, and F. Makedon, "Non-invasive analysis of sleep patterns via multimodal sensor input," *Personal and Ubiquitous Computing*, vol. 18, no. 1, pp. 19–26, 2014.

[130] H. Ni, S. Wu, B. Abdulrazak, D. Zhang, X. Ma, and X. Zhou, "Non-intrusive sleep pattern recognition with ubiquitous sensing in elderly assistive environment," *Frontiers of Computer Science*, pp. 1–14, 2015.

[131] V. V. T. Van den Berg MH, Schoones JW, "Internet-based physical activity interventions: A systematic review of the literature," *Journal of Medical Internet Research*, vol. 9, no. 3, p. e26, 2007.

[132] C. Summerbell, E. Waters, L. Edmunds, S. Kelly, T. Brown, and K. Campbell, "Interventions for preventing obesity in children." *Cochrane database of systematic reviews (Online)*, no. 3, p. CD001871, 2005.

[133] J. Pérez, E. Iturbide, V. Olivares, M. Hidalgo, N. Almanza, and A. Martínez, *A Data Preparation Methodology in Data Mining Applied to Mortality Population Databases*. Cham: Springer International Publishing, 2015, pp. 1173–1182. [En línea]. Disponible: https://doi.org/10.1007/978-3-319-16486-1_116

[134] S. Zhang, C. Zhang, and Q. Yang, "Data preparation for data mining," *Applied Artificial Intelligence*, vol. 17, no. 5-6, pp. 375–381, 2003. [En línea]. Disponible: http://dx.doi.org/10.1080/713827180

[135] J. D. Kelleher, B. M. Namee, and A. D'Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics: algorithms, worked examples, and case studies*. London: The MIT Press, 2015, no. 1.

[136] D. F. Mastin, J. Bryson, and R. Corwyn, "Assessment of sleep hygiene using the sleep hygiene index," *Journal of Behavioral Medicine*, vol. 29, no. 3, pp. 223–227, 2006.

[137] S. de Biase, G. Milioli, A. Grassi, S. Lorenzut, L. Parrino, and G. L. Gigli, "Sleep hygiene," in *Sleepiness and Human Impact Assessment*. Springer, 2014, pp. 289–295.

[138] D. E. Beaton, C. Bombardier, F. Guillemin, and M. B. Ferraz, "Guidelines for the process of cross-cultural adaptation of self-report measures," *Spine*, vol. 25, no. 24, pp. 3186–3191, 2000.

[139] INEGI, "Principales resultados de la Encuesta Intercensal 2015: Estados Unidos Mexicanos / Instituto Nacional de Estadística y Geografía," INEGI, Tech. Rep. 304.601072, 2015. [En línea]. Disponible: http://www.beta.inegi.org.mx/contenidos/proyectos/enchogares/especiales/intercensal/2015/doc/eic2015{_}resultados.pdf

[140] R. E. Shiffler, "Maximum z scores and outliers," *The American Statistician*, vol. 42, no. 1, pp. 79–80, 1988. [En línea]. Disponible: http://www.tandfonline.com/doi/abs/10.1080/00031305.1988.10475530

[141] J. Miao and L. Niu, "A survey on feature selection," *Procedia Computer Science*, vol. 91, pp. 919–926, 2016.

[142] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[143] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[144] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.

[145] ——, *C4.5: Programs for Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1993.

[146] S. Ma and J. Huang, "Penalized feature selection and classification in bioinformatics," *Briefings in bioinformatics*, vol. 9, no. 5, pp. 392–403, 2008.

[147] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[148] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[149] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Aaai*, vol. 2, 1992, pp. 129–134.

[150] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1-2, pp. 273–324, 1997.

[151] A. E. Hoerl and R. W. Kennard, "Ridge regression 1980: Advances, algorithms, and applications," *American Journal of Mathematical and Management Sciences*, vol. 1, no. 1, pp. 5–83, 1981. [En línea]. Disponible: https://doi.org/10.1080/01966324.1981.10737061

[152] H. Pulido, R. de la Vara Salazar, A. Carrasco, and M. Sánchez, *Análisis y diseño de experimentos*. McGraw-Hill, 2008. [En línea]. Disponible: https://books.google.com.mx/books?id=FrB0QwAACAAJ

[153] X. Liu, Z. Zhao, and C. Jia, "Insomnia symptoms, behavioral/emotional problems, and suicidality among adolescents of insomniac and non-insomniac parents," *Psychiatry Research*, vol. 228, no. 3, pp. 797–802, 2015. [En línea]. Disponible: http://dx.doi.org/10.1016/j.psychres.2015.05.023

[154] T. M. Achariyar, B. Li, W. Peng, P. B. Verghese, Y. Shi, E. McConnell, A. Benraiss, T. Kasper, W. Song, T. Takana, D. M. Holtzman, M. Nedergaard, and R. Deane, "Glymphatic distribution of CSF-derived apoE into brain is isoform specific and suppressed during sleep deprivation," *Molecular Neurodegeneration*, vol. 11, no. 1, pp. 1–20, 2016. [En línea]. Disponible: http://dx.doi.org/10.1186/s13024-016-0138-8