Treball de Fi de Grau

Grau en Enginyeria en Tecnologies Industrials

Development of Wake-Up Optimization Software Based on Heart Rate and Sleep Cycle

REPORT

Author: Berta Leach Maristany
Supervisor: Antoni Grau

Call: July 2025

ETSEIB

Escola Tècnica Superior
d'Enginyeria Industrial de Barcelona

UPC

Resum

Aquest projecte presenta el desenvolupament d'una aplicació de despertador intel·ligent per a rellotges Garmin, dissenyada per despertar l'usuari en el moment òptim del seu cicle de son, dins d'una finestra horària preestablerta, per tal de minimitzar la sensació de fatiga o desorientació en despertar-se.

El projecte inclou una investigació exhaustiva sobre els factors fisiològics que contribueixen a un despertar més efectiu. Es farà una anàlisi de dades com la freqüència cardíaca, la variabilitat de la freqüència cardíaca i el moviment, recollides pels sensors del rellotge durant diverses nits en diferents persones.

S'entrenarà un model d'aprenentatge automàtic LSTM (Long Short-Term Memory) per a distingir entre les fases de son lleuger i no lleuger (REM o profund). El model es refinarà de manera iterativa per maximitzar la precisió evitant el sobreajustament.

Un cop identificat el patró òptim per despertar-se, es desenvoluparà una aplicació en Monkey C pels rellotges Garmin Forerunner 55 que processi les dades dels sensors i activi l'alarma dins la finestra de 30 minuts escollida per l'usuari, en el moment més adequat.

Resumen

Este proyecto presenta el desarrollo de una aplicación de despertador inteligente para relojes Garmin, diseñada para despertar al usuario en el momento óptimo de su ciclo de sueño, dentro de un intervalo de tiempo preestablecido, con el fin de minimizar la sensación de fatiga o desorientación al despertar.

El proyecto incluye una investigación exhaustiva sobre los factores fisiológicos que contribuyen a un despertar más efectivo. Se analizarán datos como la frecuencia cardíaca, la variabilidad de la frecuencia cardíaca y el movimiento, recogidos por los sensores del reloj durante varias noches en diferentes personas.

Se entrenará un modelo de aprendizaje automático LSTM (Long Short-Term Memory) para distinguir entre las fases de sueño ligero y no ligero (REM o profundo). El modelo se refinará de forma iterativa para maximizar la precisión evitando el sobreajuste.

Una vez identificado el patrón óptimo para despertar, se desarrollará una aplicación en Monkey C para los relojes Garmin Forerunner 55 que procese los datos de los sensores y active la alarma dentro de la ventana de 30 minutos elegida por el usuario, en el momento más adecuado.

Abstract

This project presents the development of a smart alarm application for Garmin smartwatches, designed to wake the user at the optimal point in their sleep cycle, within a preset time window, to minimize feelings of fatigue or grogginess upon waking.

The work includes an in-depth investigation into the physiological factors that contribute to a more effective wake-up experience. Data such as heart rate, heart rate variability, and movement—collected from the watch's sensors over several nights—will be analyzed across multiple individuals.

An LSTM (Long Short-Term Memory) machine learning model will be trained to distinguish between light and non-light (REM or deep) sleep phases. The model will be refined iteratively to maximize accuracy while avoiding overfitting.

Once the optimal wake-up pattern is identified, a Monkey C application for the Garmin Forerunner 55 will be developed to process sensor input and trigger the alarm during the user's chosen 30-minute window at the most suitable moment.

Contents

Abbreviations and Symbols

Adam: Adaptive Moment Estimation

AI: Artificial Intelligence

API: Application Programming Interface

APP: Application

CNMC: Comisión Nacional de los Mercados y la Competencia (National Commission of Markets and Competition)

CO2: Carbon Dioxide

CSV: Comma-Separated Values

ECG: Electrocardiogram

EEG: Electroencephalogram

EMG: Electromyogram

EOG: Electrooculogram

FIT: File Interchange Type (commonly refers to Garmin's Flexible and Interoperable Data Transfer format in fitness tracking)

HR: Heart Rate

HRV: Heart Rate Variability

IBI: Inter-Beat Interval

IDE: Integrated Development Environment

LED: Light-Emitting Diode

LSTM: Long Short-Term Memory

NREM: Non-Rapid Eye Movement

PPG: Photoplethysmography

PSG: Polysomnography

ReLU: Rectified Linear Unit

REM: Rapid Eye Movement

RMSSD: Root Mean Square of Successive Differences

RNN: Recurrent Neural Network

RR: Inter-beat Interval

SDG: Sustainable Development Goal

SDNN: Standard Deviation of Normal-to-Normal Intervals

SDK: Software Development Kit

SMOTE: Synthetic Minority Oversampling Technique

VO2 MAX: Maximum Volume of Oxygen Consumption

List of figures

List of tables

1. Preface

Almost one year ago, two of my friends and I, all industrial engineers, bought our first Garmin smartwatches. We found them fascinating, especially in today's data-driven world. Simply wearing a watch all day could provide so many insights into your health: your resting heart rate, stress levels, VO2 max, and more. Among all the insights, the ones related to sleep piqued our curiosity the most. Every morning, we'd check how we slept—how many sleep cycles we had, how the sleep stages were distributed in each cycle, how our heart rate varied during the night, and how it differed across nights. For instance, we noticed that training intensely for a marathon significantly lowered our resting heart rates, while being ill prevented us from resting well, as our heart rates would steadily increase throughout the night.

Our curiosity didn't stop there. That same summer, we spent the month of July studying Garmin and sleep. My friends and I built a platform that used Garmin data to provide even more detailed sleep information than what Garmin offered. We invented a scoring system to measure sleep quality and created various visualizations to display the data in more informative ways. I learned a lot during this process, particularly as I focused on embedded software development. My ultimate goal was to develop the Garmin sleep alarm that I'm presenting in this thesis. However, starting with little knowledge about Garmin and programming, one summer wasn't enough to achieve that. Still, I learned how to build apps for Garmin and even developed a simple pink watch face. Beyond the technical skills, I enjoyed the process and discovered what I truly love about engineering: the ability to learn and create anything.

After that summer, I was left with an itch. I knew I'd return to the project someday, and what better time than now, as the final step to becoming a real engineer?

2. Introduction

2.1. Motivation

"Prevention is better than cure", a principle I have always believed in. In past decades, society was unaware of the dangers of smoking, excessive sugar, or alcohol consumption. Today, despite strong evidence, many still overlook the long-term consequences of unhealthy habits until medical intervention becomes necessary. However, it is often simple, natural routines, such as regular exercise and quality sleep, that contribute most to a longer, healthier life.

In recent years, there has been a growing shift toward healthier lifestyles. Technologies such as fitness tracking apps (e.g., Strava) and wearable devices (e.g., Garmin smartwatches) have transformed health practices into measurable and even social experiences. They have made physical activity and sleep monitoring more accessible and engaging.

With this project, I aim to further highlight the importance of sleep, an often overlooked yet fundamental pillar of health. Specifically, I will focus on the waking phase: designing a system that supports a more natural, biologically-aligned awakening by identifying the optimal moment to start the day. This contributes not only to improved well-being but also to better daily performance and long-term health.

2.2. Scope

During the course of this thesis, I encountered several limitations, which I hope to overcome in future iterations of this work. Nevertheless, I developed the best possible version of the system within the constraints of the available resources.

One major limitation was the inability to access the Garmin API, which prevented real-time transmission of sensor data to a high-capacity external device. This restricted my ability to implement a machine learning model that could continuously learn from the user's sleep patterns and personalize the waking experience over time. To address this, I simplified the approach and designed a lightweight algorithm capable of running locally on the smartwatch, taking into account its limited processing power and storage.

Another significant constraint involved the detection of sleep phases. Ideally, accurate identification of sleep stages requires brain wave data (EEG), typically collected in a laboratory setting. Without access to such facilities, I relied on *Garmin Forerunner 55's* built-in sleep stage detection algorithm. According to available data, this algorithm achieves an overall accuracy of 69.7%, with a sensitivity (proportion of actual positive cases) of 95.8% for detecting sleep and a specificity (proportion of actual negative cases) of 73.4% for detecting wakefulness [1]. While this level of accuracy introduces a margin of error, it remains the most feasible solution given the available sensors, which do not include direct brain activity measurement.

2.3. Objectives

The primary goal of this project is to develop a smartwatch alarm application that wakes users at the ideal point in their sleep cycle, using physiological data (heart rate, heart rate variability and movement) to ensure adaptability to individual users. To achieve this, the following specific objectives will be addressed:

- Understand sleep cycles by analyzing data from Garmin sensors (accelerometer, heart rate sensor) and identify patterns using machine learning models.
- Conduct a theoretical analysis of sleep physiology and existing research to identify the ideal sleep stage for waking.
- Design a machine learning-based algorithm to predict the optimal wake-up time, subsequently optimized for the limited processing capacity of a smartwatch.
- Develop an intuitive and user-friendly wake-up alarm application for Garmin devices using Monkey C.
- Publish the application on the Connect IQ platform, making it accessible for all Garmin users to download and use.
- Ultimately, contribute to a healthier society by promoting better sleep quality through accessible, personalized wake-up solutions.

3. Theory or Theoretical background

3.1. Theoretical Foundations

3.1.1. Introduction to Sleep

Sleep is a complex and dynamic biological process essential for maintaining physical and mental health. It is characterized by a reversible state of reduced responsiveness to external stimuli, accompanied by distinct physiological changes such as alterations in brain wave activity, breathing, heart rate, and body temperature.

Sleep regulation is governed by two primary mechanisms: the circadian rhythm and the sleep pressure. The circadian rhythm, an intrinsic biological process, is observed in organisms with lifespans exceeding seven days. This rhythm operates on an approximately 24-hour cycle and functions independently of solar cycles. However, exposure to sunlight facilitates synchronization of the circadian rhythm to a precise 24-hour period, recalibrating the internal biological clock daily.

The circadian rhythm oscillates independently of sleep history, maintaining its cycle regardless of sleep duration or deprivation. This independence underlies the phenomenon of jetlag, where transmeridian travel disrupts alignment between the internal circadian rhythm and the external environment. Consequently, individuals experience difficulty remaining awake during periods when the circadian rhythm promotes sleep, or difficulty sleeping when the rhythm promotes wakefulness, irrespective of prior wakefulness duration.

The circadian rhythm fluctuations occur at distinct times of the day depending of the person, giving rise to three primary chronotypes: morning types, evening types, and intermediate types. Morning types are characterized by earlier peak alertness, while evening types display later peak alertness. Societal structures often disadvantage evening types, whose circadian rhythms predispose them to sleep onset between 1:00 AM and 2:00 AM and awakening between 9:00 AM and 10:00 AM. When required to commence work at 8:00 AM, evening types may experience cognitive impairment akin to a sleep-like state. Furthermore, chronic misalignment with their circadian rhythm increases the likelihood of developing mental and physical health disorders.

Sleep pressure is determined by the duration of wakefulness. A neurochemical, adenosine, accumulates in the brain over time, incrementally increasing the propensity for sleep. This signal can be temporarily suppressed by caffeine, the most widely consumed psychoactive substance. While caffeine masks the perception of adenosine-induced sleepiness, the underlying adenosine levels continue to rise. Consequently, when caffeine's effects dissipate, the accumulated sleep pressure results in heightened feelings of fatigue. Adenosine levels are reduced exclusively through sleep.

The interplay between sleep pressure and the circadian rhythm governs levels of alertness and fatigue. The urge to sleep is influenced by the balance between adenosine accumulation and the oscillations of the circadian rhythm. As illustrated in Figure 1, which depicts an all-night wakefulness scenario, the urge to sleep is notably greater at 3:00 AM compared to midday, despite longer wakefulness duration at the latter time point. This counterintuitive finding highlights the dominant influence of circadian rhythm fluctuations on sleep propensity.

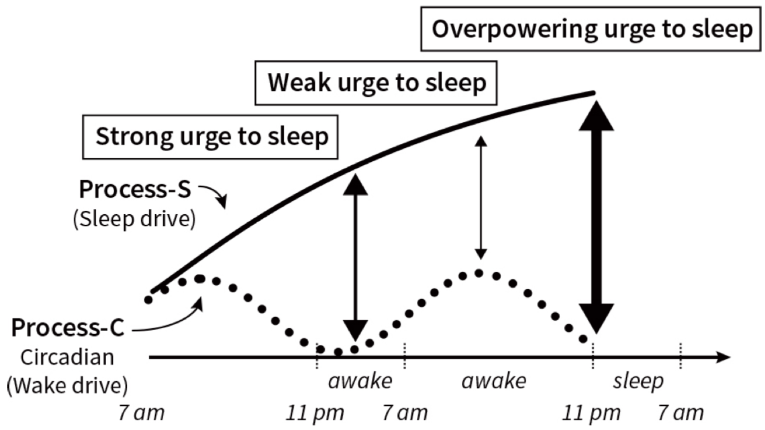


Figure 1 Circadian Rhythm and Sleep Pressure in Sleep deprivation. [Source [2]]

Despite its apparent simplicity, sleep encompasses various stages and mechanisms that contribute to its restorative functions. Understanding these processes is crucial, as sleep affects numerous aspects of health, including cognitive performance, emotional regulation, and immune function. The subsequent sections of this thesis will delve deeper into the stages of sleep, physiological indicators, and the impact of sleep on overall health.

3.1.2. Sleep Stages

Sleep is a dynamic process composed by multiple stages that cycle throughout the night, each characterized by distinct physiological and neurological patterns. These stages are broadly categorized into non-rapid eye movement (NREM) sleep and rapid eye movement (REM) sleep. A typical night's sleep involves four to six cycles, each lasting approximately 70 to 120 minutes, with the composition of these cycles evolving as the night progresses (see Figure 2).

NREM sleep consists of three stages:

- Light Sleep (Stage N1): This initial stage marks the transition from wakefulness to sleep and lasts about 7 minutes. It is characterized by a decrease in muscle activity and slow eye movements. Individuals in this stage can be awakened easily.
- Light Sleep (Stage N2): This stage lasts approximately 10 to 25 minutes in the initial cycle and lengthens in subsequent cycles. It is marked by a further decrease in heart rate, respiration rate and body temperature as well as relaxed muscles. It is still light sleep, but the sleeper is less likely to be awakened.
- Deep Sleep (Stage N3): It is the most restorative stage, crucial for tissue repair and growth hormone release. There is no eye movement or muscle activity. Stage N3 occurs predominantly in the beginning of the night and becomes shorter in later cycles. At this stage, waking up becomes harder, and if you do, you're likely to feel disoriented.

Following NREM sleep, the cycle progresses to:

- REM Sleep: This stage is characterized by rapid eye movements under your eyelids, increased brain activity resembling wakefulness, and muscle paralysis. REM sleep is associated with vivid dreaming and plays a vital role in memory consolidation. The first REM period typically occurs after about 90 minutes of sleep and lengthens with during the night's cycles.

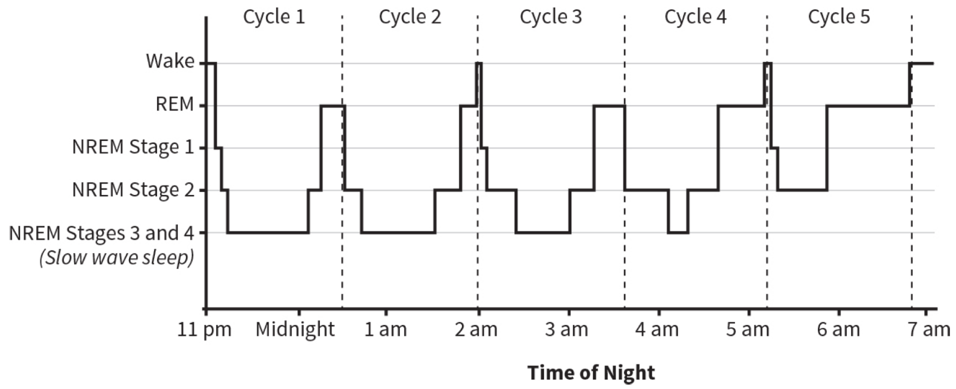


Figure 2 Sleep Stages architecture during the night. [Source [2]]

Garmin's sleep tracking technology utilizes data from heart rate and movement sensors to estimate time spent in each sleep stage. This information provides users with insights into their sleep, helping them understand and improve their sleep quality.

3.1.3. Physiological Indicators of Sleep

Physiological indicators reflect bodily changes across sleep stages, enabling sleep monitoring. Wearable devices like Garmin smartwatches track some indicators effectively, while others require specialized equipment, making them harder to monitor in real-world settings. As Garmin notes, “Sleep stages are identified through a combination of heart rate, heart rate variability, and body movement data”[3], but those are not the most accurate indicators to detect sleep, and sleep stages.

A study was conducted in 2019 by Garmin along with the university of Kansas and other scientists [1], where it was compared the Garmin vívosmart 3 with the results of an in-laboratory polysomnography (PSG), stated as the reference 100% accurate sleep tracker. The results showed a 69.7% of accuracy with the following Confusion Matrix in Figure 3:

	True Deep	True Light	True REM	True Awake
Predicted Deep	68.9%	12.0%	1.4%	1.9%
Predicted Light	29.1%	68.6%	26.4%	14.9%
Predicted REM	0.9%	13.7%	69.8%	9.7%
Predicted Awake	1.1%	5.7%	2.3%	73.4%

Figure 3 Confusion matrix between real and predicted Sleep Stages. [Source [1]]

It is also observed that Garmin combines N1 and N2 into a single Light sleep stage for simplicity. Accordingly, this study adopts the same assumption from this point forward.

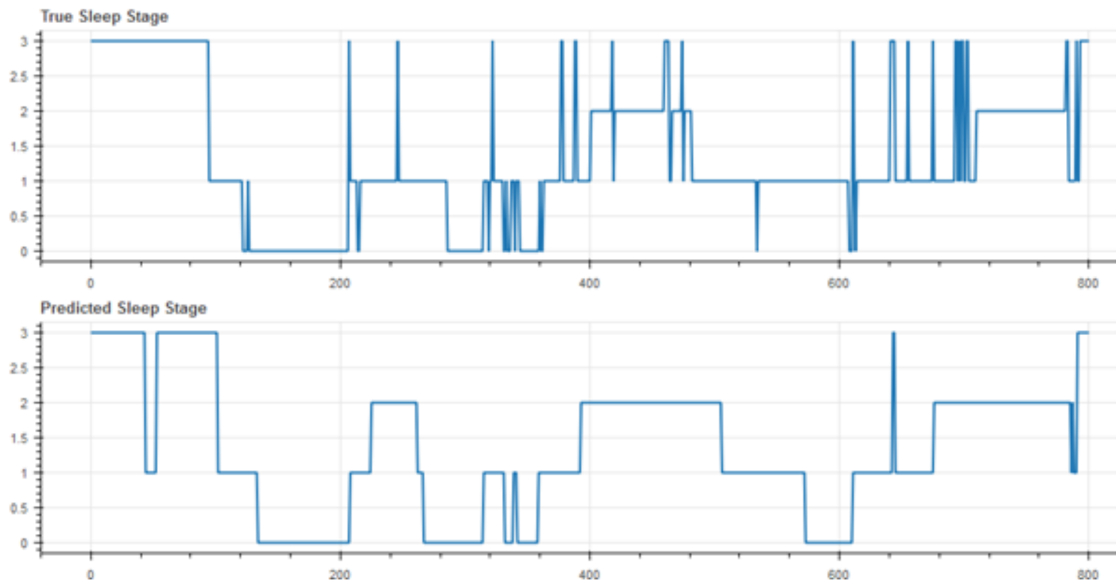


Figure 4 Median result from Garmin Sleep Study. [Source [1]]

Figure 4 illustrates the behavior of the median result of this study in comparison with the reference PSG-based data. The Garmin watch did not accurately detect all short awakenings during this individual's night and identified three sleep cycles (three distinct REM stages), whereas only two were present.

Numerous physiological indicators can provide insights into sleep patterns. Table 1 presents a selection of these indicators along with their corresponding trackability.

Table 1 Physiological indicators of sleep and their trackability. [Source: own elaboration]

	Description	**Trackable by Wearables**
Heart Rate	Measures heart beats per minute [bpm].	Yes
Heart Rate Variability (HRV)	Measures variation of time [ms] between heartbeats.	Yes
Movement	Detects physical activity via x, y and z accelerometers.	Yes
Respiratory Rate	Measures breaths per minute. Helps detect disturbances like sleep apnea.	Yes
Blood Oxygen Levels	Measures oxygen saturation with SpO2 sensors.	Yes (some advanced models)
Body Temperature	Core temperature drops around 1°C to initiate sleep, rises upon waking.	No
Blood Pressure	Blood pressure reduces in non-REM, varies in REM based on dream activity.	No
Other Physiological Activity	Kidney function slows, growth hormone release increases during sleep.	No (requires lab tests)

All these indicators reflect how the body responds to different stages of brain activity. As previously mentioned, brain activity, measured through electroencephalography (EEG), is visualized in polysomnography (PSG). Sleep stages are primarily identified based on the amplitude and frequency of EEG waveforms. During light sleep, EEG recordings show low-amplitude, mixed-frequency waves, often accompanied by characteristic features such as sleep spindles (brief bursts of high-frequency activity) and K-complexes (distinctive biphasic waves). In deep sleep, or slow-wave sleep, EEG activity is dominated by high-amplitude, low-frequency delta waves. REM sleep, on the other hand, presents EEG patterns that closely resemble wakefulness, low amplitude and high frequency, making it difficult to identify based on EEG alone. Therefore, other physiological markers are essential for detecting REM.

In addition to EEG, PSG monitors other physiological functions, including eye movements via electrooculography (EOG), which is the most reliable method for detecting REM sleep, and muscle activity through electromyography (EMG), which helps identify the muscle atonia characteristic of REM sleep. Polysomnography also typically includes electrocardiography (ECG) to track heart rhythm.

This thesis focuses on the development of an alarm system for Garmin devices, which do not track brain activity. From this point forward, the analysis will concentrate on heart rate (HR), heart rate variability (HRV), and movement, acknowledging a minimum expected error margin of 30%, as indicated by the previously referenced Garmin study[1].

Heart rate and heart rate variability are measured by using optical photoplethysmography (PPG). PPG emits a light into the skin and detects changes in the intensity of the reflected light. These changes occur as blood volume in the capillaries shifts with each heartbeat. The raw data extracted by this sensor are beat to beat intervals (RR) in milliseconds [ms], meaning the time between heart beats. To calculate HR and HRV they are used the following formulas:

$$HR = \frac{60000}{RR}\ [bpm]$$ *Equation 1*

HRV can be calculated using two common formulas: RMSSD (Root Mean Square of Successive Differences), which is the standard method used by Garmin devices, and SDNN (Standard Deviation of RR intervals):

$$HRV_{RMSSD} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N-1}(RR_{i+1} - RR_i)^2}\ [ms]$$ *Equation 2*

$$HRV_{SDNN} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(RR_{i+1} - \overline{RR})^2}\ [ms]$$ *Equation 3*

Where N denotes the number of interbeat intervals used in the calculation, and RR represents the interbeat intervals in milliseconds (ms).

Lastly, movement is measured using an accelerometer, which records acceleration along the x, y, and z axes as defined by the orientation of the device. The output values are expressed in milli-g units, where 1 g represents the acceleration due to gravity (approximately 9.81 m/s²), and 1 milli-g equals one-thousandth of that (0.00981 m/s²). This allows for precise detection of even small movements.

Table 2 How do HR, HRV and movement change between sleep stages. [Source: own elaboration]

	Light Sleep	Deep Sleep	REM Sleep	Awake
Heart Rate	Gradually decreases as parasympathetic activity increases.	Lowest levels due to dominant parasympathetic tone.	Variable and may increase, resembling wakefulness.	Highest levels, influenced by physical and mental activities.
HRV	Increases compared to wakefulness, reflecting enhanced parasympathetic activity.	Peaks, indicating maximal parasympathetic dominance and restorative processes.	Decreases compared to deep sleep due to increased sympathetic activity.	Lowest, due to higher sympathetic nervous system activity.
Movement	Moderate movement; body may shift positions.	Minimal to no movement; body remains mostly still.	Very little to no movement due to muscle atonia, except for possible brief twitches.	Frequent and often intense movements, including voluntary activities.

Table 2 illustrates how different sleep stages can be identified using physiological signals measurable by Garmin smartwatches. Notably, heart rate and heart rate variability (HRV) are closely linked to the balance between sympathetic and parasympathetic nervous system activity. The sympathetic nervous system increases physiological alertness and prepares the body for responsiveness, while the parasympathetic nervous system promotes relaxation and supports essential restorative functions such as digestion, urination, and defecation when the body is in a safe, resting state.

3.1.4. Impact of Sleep on Health

Sleep is a fundamental biological process that influences nearly every aspect of human health. Adequate and quality sleep is essential for optimal cognitive function, emotional well-being, physical health, and safety.

- **Cognitive Function**

Sleep plays a critical role in cognitive processes, including memory consolidation, learning, attention, and decision-making. During sleep, particularly during Deep sleep and REM stages, the brain processes and integrates new information, strengthening neural connections. Sleep deprivation impairs these functions, leading to decreased alertness, slower reaction times, and reduced problem-solving abilities.

- **Emotional Well-being**

Adequate sleep is vital for emotional regulation and mental health. Insufficient sleep can increase negative emotional responses to stressors and decrease positive emotions, heightening the risk of mood disorders such as depression, anxiety or even suicidal ideation.

- **Physical Health**

Sleep is integral to physical health, supporting tissue repair, immune function, and metabolic regulation. Chronic sleep deprivation disrupts hormonal balances and metabolic processes, increasing the risk of obesity, type 2 diabetes and cardiovascular diseases.

- **Safety and Performance**

Sleep significantly affects alertness and performance. Sleep deprivation impairs cognitive and motor functions, leading to increased risk of accidents and errors. Drowsy driving, for instance, is a major safety concern, contributing to numerous road accidents annually.

3.2. State of the Art

The concept of wearable technology dates back centuries, with early innovations laying the groundwork for today's sophisticated devices. One of the earliest known ideas resembling a wearable fitness tracker was conceived by Leonardo da Vinci in the 15th century. Da Vinci sketched designs for a mechanical pedometer intended for military applications, aiming to measure the distance soldiers marched. While it's uncertain whether he built a working model, his vision highlighted the potential of wearable devices for tracking physical activity.

By the late 16th century, mechanical pedometers had been developed, utilizing mechanisms similar to contemporary watches. These devices gained popularity among European aristocrats as tools for measuring walking distances. In 1780, Swiss watchmaker Abraham-Louis Perrelet created a self-winding watch that also functioned as a pedometer, measuring steps and distance walked. This innovation was introduced to the United States by Thomas Jefferson, who acquired a French pedometer during his time in Europe.

The era of wearable fitness tracking began in the 20th century. In 1965, a Japanese professor, Dr. Yoshiro Hatano, developed the "Manpo-kei", which translates to "10,000 steps meter." Hatano's research suggested that walking 10,000 steps daily could help maintain a healthy lifestyle, a very groundbreaking concept for that time, that remains influential in fitness tracking today.

The 1970s and 1980s witnessed significant advancements in wearable technology. In 1977, Professor Seppo Säynäjäkangas invented the first battery-operated wireless heart rate monitor to aid the Finnish National Cross-Country Ski Team. This led to the founding of Polar Electro, a sports equipment manufacturer which released, in 1982, the world's first commercially available wireless heart rate monitor (the Sport Tester PE 2000). This device allowed athletes to monitor their heart rates in real-time during training, revolutionizing endurance sports and personal fitness tracking.

Simultaneously, the development of digital watches with added functionalities marked the emergence of smartwatches. In 1972, Pulsar introduced the Time Computer Calculator, a watch that combined timekeeping with a calculator function, using a red LED digital display. This innovation paved the way for more sophisticated wearable devices. The 1980s and 1990s saw the proliferation of digital watches with various features, including alarms, calculators, and data storage. In 1983, Seiko released the Data 2000, a watch capable of storing memos and appointments, marking a significant step toward the multifunctional smartwatches we know today. These developments laid the foundation for integrating health and fitness tracking into wearable devices.

As smartwatches and fitness bands became more capable, developers began expanding their functionality beyond daytime activity tracking. By the early 2000s, a growing interest in understanding and improving sleep patterns prompted the integration of sleep tracking features. The first wave of devices capable of analyzing sleep typically relied on actigraphy, sensors that monitored wrist movements to infer rest and wake periods. One of the pioneering devices in this space was the Actiwatch, developed in 1999 by Mini Mitter Co. (later acquired by Philips Respironics). Though originally intended for clinical use, actigraphy technology would become the basis for later consumer-grade wearables.

Around the late 2000s, commercial interest in sleep tracking grew rapidly. The original Fitbit, released in 2009, was among the first consumer devices to combine daytime activity monitoring with sleep tracking features in 2013. By wearing it at night, users could receive a basic overview of their sleep patterns. However, these insights were still relatively simplistic, focusing on sleep duration and movement rather than detailed sleep stages.

It wasn’t long before manufacturers recognized the potential of not only tracking sleep but optimizing the user’s waking experience. One of the first consumer products to incorporate a “smart alarm” was the aXbo Sleep Phase Alarm Clock, released in 2006 by Austrian company aXbo. Unlike traditional alarms that wake users at a fixed time regardless of their sleep state, the aXbo used actigraphy-based sensors in a wearable wristband to detect sleep phases and determine the best moment to wake someone within a pre-set time window. This was a significant milestone in combining sleep science with practical consumer technology.

In the following years, several companies adopted and expanded on this principle. Notably, the Fitbit Versa, released in the 2018, integrated smart alarm functionality into their devices, and it was from then on, they adopted this functionality in almost every Fitbit wristband. These smart alarms operated on similar principles, which are identifying the user's light sleep phase within a preset interval to reduce grogginess upon waking. Fitbit's large user base and intuitive app interface helped bring smart alarms into mainstream awareness.

Garmin, one of the leaders in multisport wearables, has also embraced the smart alarm concept. While earlier models offered basic sleep tracking, more recent devices like the Vívoactive 6 series (released in 2025) include advanced sleep stage detection, nap tracking, and smart wake functionalities. Garmin's implementation prioritizes user-defined wake windows and uses accelerometer and heart rate data to optimize the wake-up moment.

Today, the ecosystem of sleep-focused wearables has expanded significantly. The Oura Ring, first released in 2015, stands out as one of the best wearables for sleep tracking, though it lacks a smart alarm. The Pavlok Shock Clock 3 uses vibrations, sounds, and mild electrical stimuli to train consistent waking habits. Meanwhile, apps like AutoSleep for the Apple Watch continue to innovate in smart alarm functionality through refined sleep analysis software. The Withings Sleep Mat, released in 2018, presented a non-wearable alternative. Placed under the mattress, it used pressure sensors to detect movements, breathing patterns, and heart rate, offering detailed sleep analysis. Additionally, smartphone applications like Sleep Cycle offer smart alarm features by analyzing movement and sound using the phone's sensors. However, their accuracy in detecting specific sleep stages tends to be lower compared to dedicated wearable devices.

In summary, the journey from simple pedometers and heart rate monitors to sophisticated sleep-aware wearable devices highlights the rapid evolution of consumer health technology. Smart alarms represent a meaningful intersection of sleep science and daily habit formation, offering a glimpse into the future of personalized wellness. As wearable technology becomes more affordable and sensor accuracy improves, it’s likely that smart alarms and adaptive wake systems will become standard features, not only in premium devices, but across the full spectrum of consumer wearables.

Figure 5 illustrates the evolution of wellness tracker wearables, tracing their development from the 15th century to the present day.

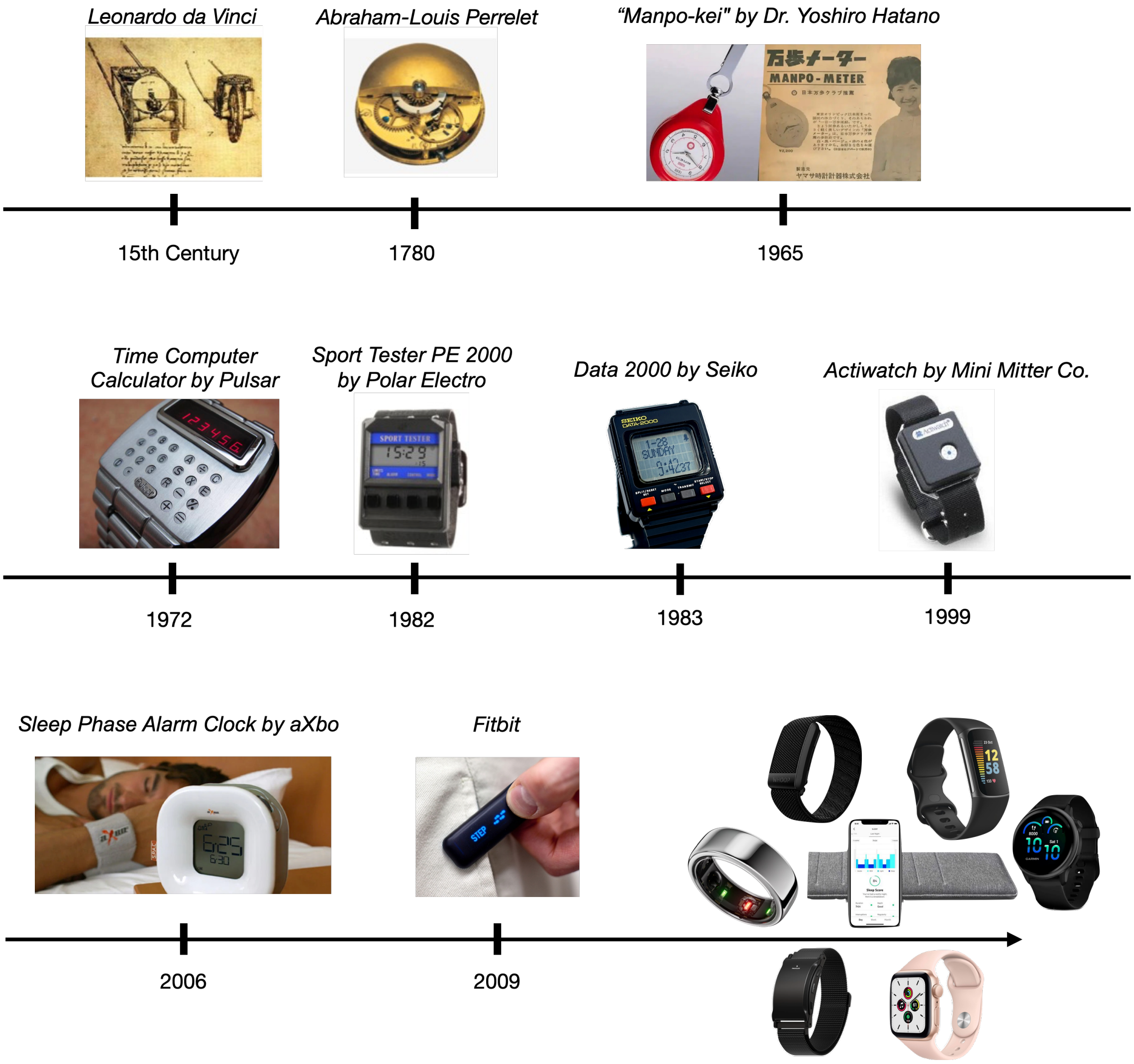


Figure 5 The evolution of wellness tracker wearables [Source: own elaboration, Source of pictures: Google]

Beyond wearables, innovative non-wearable solutions have transformed the sleep tech landscape, specially the Eight Sleep's Pod series. The Eight Sleep Pod is a smart mattress cover that goes beyond tracking sleep metrics. It actively enhances sleep quality by adapting the sleep environment through dynamic temperature control, adjustable bed elevation, and curated surround sound for relaxation. The temperature feature, which adjusts throughout the night to optimize comfort for each sleep phase, can be customized independently for each side of the bed, making it ideal for couples with different preferences. Eight Sleep reports benefits like increased time in deep sleep, reduced resting heart rate, and improved heart rate variability. However, its high cost makes it a premium choice, limiting accessibility.

4. Development Process of the Garmin Smart Wake up Alarm

The development of the Garmin app, designed to wake users during light sleep phases to minimize morning fatigue, required a multifaceted approach that integrated algorithm design, backend data access, and frontend user interface development. This project was structured around two parallel tracks: (1) designing an algorithm to determine optimal wake-up times based on sensor data from the Garmin Forerunner 55, and (2) implementing a functional application to access real-time sensor data and present it through an intuitive user interface. These tracks were interdependent, as the algorithm's design relied on understanding the specific data that could be reliably extracted from the watch's sensors, while the app's development ensured that the data could be processed and displayed effectively.

The following sections detail the development process, beginning with the setup of the development environment. This is followed by dataset collection and postprocessing, the selection and refinement of a machine learning model, the creation of a simplified version of the model adapted to the smartwatch's capacity constraints, and finally, the development of the Garmin app in Monkey C and integration of the algorithm.

4.1. Setting Up the Development Environment

4.1.1. Garmin's Connect IQ Platform and Developer Portal

To begin development, the Garmin's Connect IQ platform was used, a robust ecosystem designed to enable developers to create custom applications, watch faces, widgets, and data fields for Garmin wearable devices. The Connect IQ platform, introduced in 2015, allows third-party developers to extend the functionality of Garmin devices by leveraging the proprietary Monkey C programming language. The Garmin Developer Portal[1] serves as the central hub for accessing resources, documentation, and tools necessary for app development. The portal provides comprehensive guides, including:

- **API Documentation**: Detailed references for accessing device features such as sensors, user profiles, and communication modules.
- **SDK Downloads**: Tools for compiling, testing, and deploying applications.
- **Programmer's Guide**: Tutorials and best practices for writing Monkey C code, managing resources, and adhering to user experience guidelines for designing intuitive, device-appropriate interfaces.

[1] https://developer.garmin.com/connect-iq/overview/

The Connect IQ platform is tailored to accommodate the resource-constrained nature of Garmin wearables, which prioritize battery life and performance over computational power. Monkey C, the language used for Connect IQ development, is an object-oriented, dynamically typed language inspired by Java, JavaScript, Python, PHP, and Ruby. It is designed to be lightweight and efficient, using reference counting for memory management instead of garbage collection, which reduces memory overhead on the devices.

4.1.2. Software Development Kit (SDK) Setup

A Software Development Kit (SDK) is a collection of tools, libraries, and documentation designed to facilitate application development for a specific platform or framework. The Connect IQ SDK includes the Monkey C compiler (that transforms Monkey C into bytecode), a device simulator (any existing Garmin model can be simulated), and libraries for interacting with Garmin device hardware and APIs. The SDK is essential for compiling Monkey C code into bytecode that runs on the Connect IQ virtual machine, ensuring compatibility with the Garmin wearables.

4.1.3. Choosing the Development Environment: Cursor

For coding, Cursor was selected, a modern Integrated Development Environment (IDE) based on Visual Studio Code and enhanced with artificial intelligence capabilities aimed at improving developer productivity. Cursor is increasingly adopted by the developer community due to its AI-assisted code suggestions, real-time debugging support, and natural language-based code generation, which collectively streamline the software development workflow. Cursor offers a lightweight, customizable interface with integrated support for Monkey C through the official Garmin Connect IQ extension.

Following the installation of the Monkey C extension, Cursor effectively functioned as a comprehensive Connect IQ development environment, providing:

- Syntax highlighting and autocompletion for Monkey C.
- Build integration for compiling apps directly from the IDE.
- Debugging tools for testing apps in the Connect IQ simulator.
- Project creation wizards to set up new Connect IQ projects.

Using the extension's project creation wizard, a watch app project targeting the Garmin Forerunner 55 was generated. The wizard automatically produced a structured project directory that included:

- A manifest.xml file, defining app configuration and device compatibility.

- A monkey.jungle file, specifying build configurations
- A primary Monkey C source file (source/App.mc), containing the app's core logic.

4.2. Auxiliary App Development: Accessing Sensor Data

The first milestone was to create a simple Connect IQ app to access and display real-time sensor data from the Garmin Forerunner 55. This step was crucial to understand the data available for the wake-up algorithm and to validate the app's ability to interact with the device's sensors.

The app focuses on two sensors: the heart rate monitor, providing heart rate (HR) and heart rate variability (HRV) data via interbeat intervals (IBIs); and the accelerometer, capturing movement data. The raw data from these sensors required processing to be usable for the algorithm. The heart rate monitor provides IBIs, which are the time intervals (in milliseconds) between consecutive heartbeats. The *Toybox.Sensor.HeartRateData.heartBeatIntervals* API returns an array of up to ten IBI values per second, though during sleep, typically only one to three values are recorded due to lower heart rates. The accelerometer measures linear acceleration across three axes (x, y, z) relative to the Earth's gravitational field. The *Toybox.Sensor.AccelerometerData.x, .y, and .z APIs return arrays of 25 values per second* for each axis, measured in milli-g (where 1000 milli-g equals 1 g), with positive or negative values.

In order to record full-night data for subsequent analysis and algorithm development, a separate application was developed to access and store the required information using a different set of commands than those employed previously. This was necessary to ensure the data could be written to a FIT file for later download. The process began with the creation of an activity session. Once the session was initiated, the FIT file automatically recorded heart rate (HR) at one-second intervals, along with other parameters. To capture specific fields, such as interbeat intervals (IBIs) and accelerometer data, explicit configuration was required.

4.3. Night data analysis

The FIT files generated by the Garmin device are in a binary format and require conversion to a human-readable structure for analysis. To facilitate this, the FitCSVTool.jar utility included in the Connect IQ SDK was used to convert the FIT files into CSV format. The resulting CSV files contained raw sensor data but exhibited inconsistencies and redundant entries that necessitated further cleaning. While heart rate (HR) values were automatically derived from interbeat intervals (IBIs) by the activity recorder, both IBIs and accelerometer data required additional manual processing in Excel in order to compute heart rate variability (HRV) and movement-related metrics. These calculations where afterwards automated with python.

4.3.1. Data Cleaning

The FIT file records IBIs as arrays of ten values per second, separated by vertical bars (|), with a default value of 65535 indicating no data. (See an example in Table 3)

Table 3 Raw Interbeat intervals data. [Source: own elaboration]

btb_intervals	1244|65535|65535|65535|65535|65535|65535|65535|65535|65535	ms
btb_intervals	995|65535|65535|65535|65535|65535|65535|65535|65535|65535	ms
btb_intervals	1033|65535|65535|65535|65535|65535|65535|65535|65535|65535	ms
btb_intervals	1033|65535|65535|65535|65535|65535|65535|65535|65535|65535	ms
btb_intervals	892|916|65535|65535|65535|65535|65535|65535|65535|65535	ms

Analysis of multiple nights of data revealed that, during sleep, typically one or two interbeat interval (IBI) values were recorded per second. Values equal to 65535, which indicate invalid readings, were filtered out, and valid IBIs were extracted into separate columns for further analysis. Repeated IBI values were also observed, as illustrated in the third and fourth rows of the example above; such repetitions could occur multiple times throughout the dataset. To address this issue, duplicate rows were removed, resulting in a cleaned dataset as shown below in Table 4.

Table 4 Cleaned Interbeat intervals data. [Source: own elaboration]

btb_intervals	1244		ms
btb_intervals	995		ms
btb_intervals	1033		ms
btb_intervals			ms
btb_intervals	892	916	ms

Accelerometer data required comparatively less preprocessing. Each row in the file contained calibrated accelerometer readings for the x, y, and z axes. For each second, 25 values per axis were recorded and presented as sequences separated by vertical bars (|). (See an example in Table 5)

Table 5 Raw accelerometer data. [Source: own elaboration]

calibrated_accel_x	-622.0|-629.0|-596.0|-685.0|-565.0|-621.0|-640.0|-838.0|-604.0|-724.0|-736.0|-760.0|-755.0|-755.0|-739.0|-779.0|-752.0|-755.0|-766.0|-731.0|-773.0|-780.0|-763.0|-753.0|-746.0	g	calibrated_accel_y	…

To simplify the analysis, the average of the 25 accelerometer values per second was computed for each axis using Excel, resulting in a single representative value per axis per second.

In order to merge the three sensor datasets—heart rate (HR), interbeat intervals (IBIs), and accelerometer data—into a continuous timeline, the raw data were examined. As explicit timestamp cells were absent, the timestamp of the first recorded second was extracted from the filename, with subsequent data assumed to follow at one-second intervals.

Upon inspection, irregularities were identified in the structure of the raw data. Typically, each second included three rows: one containing HR and IBI data, one with accelerometer data, and one with additional fields not relevant to the present analysis. However, in certain cases, multiple rows of the same data type were recorded within a single second. To maintain temporal consistency, only the first instance of each data type per second was retained, and duplicates were discarded. This cleaning process ensured the generation of a coherent and synchronized dataset suitable for algorithm development.

4.3.2. Training dataset creation

The heart rate (HR) variable was already structured appropriately for analysis. For heart rate variability (HRV), the two columns of interbeat intervals (IBIs) were used in conjunction with Equation 2. As noted in Garmin's documentation, HRV is calculated "based on analysis of 5-minute time windows"[4], which may indicate either a single HRV value every 5 minutes or a rolling calculation with a higher sampling frequency. To perform these calculations, given their computational complexity, an Excel macro was developed.

This macro processed the most recent 300 rows (300 seconds) of IBI data, merging the two IBI columns into a single continuous array. Due to the variable number of IBI entries per second (with some rows containing two values and others being empty), the resulting array could contain more or fewer than 300 values. The final number of values in the array was treated as N in Equation 2.

Movement was defined as the difference between consecutive acceleration values along the same axis. To derive a single scalar value representing overall movement, considering that acceleration is a vector, the magnitude was calculated using the following equation:

$$MOV = \sqrt{\overrightarrow{MOV}_X^2 + \overrightarrow{MOV}_Y^2 + \overrightarrow{MOV}_Z^2} \qquad \textit{Equation 4}$$

For an average 8-hour night, the dataset consisted of 28,800 values per variable. However, depending on the model, a larger dataset does not necessarily improve performance, as excessive data can introduce noise and increase the risk of overfitting. To mitigate this, data reduction was performed by computing averages over 1-minute or 5-minute intervals, as detailed in the following section. This approach also supported the development of the device algorithm, which operates under the computational constraints of the Garmin Forerunner 55.

For the target variable, sleep stages, data were extracted from Garmin Connect, the user platform that provides detailed insights into activities and wellness, including minute-by-minute sleep stage summaries. Analysis showed that using 1-minute averages for the input variables (HR, HRV, and movement) improved the model's accuracy compared to 5-minute averages. Moreover, since Garmin Connect defines sleep stages at 1-minute intervals, the 1-minute averaging aligned better with the reference data. Therefore, 1-minute intervals were selected for the final dataset to optimize both model performance and compatibility with the target variable's temporal resolution.

4.3.3. Choosing the experiment samples of data

To evaluate the generalizability of the model and reduce potential bias toward individual characteristics, sensor data from three distinct subjects were selected and analyzed independently. Subjects were chosen to introduce variation in both sex and age within a narrow range, reflecting the primary demographic of typical Garmin Forerunner 55 users. All selected individuals fell within the age range of 20 to 23 years, which aligns with the core market segment for this device.

Each subject contributed multiple nights of sleep data, with recordings randomly selected across different days of the week and weeks, thereby introducing natural variability in sleep patterns and durations. The inclusion of nights with different lengths further ensured that the model would be exposed to a range of real-world sleep conditions.

Inclusion criteria for participation required subjects to:

1. Personally own a Garmin Forerunner 55 device.
2. Have already used Garmin Connect sleep tracking for at least a week.

The characteristics of the selected subjects and the corresponding data samples are summarized in Table 6.

Table 6 Description of the experiment subjects and its data. [Source: own elaboration]

	Age	Sex	Number of nights	Number of samples
Subject 1	21	Female	6	2798
Subject 2	23	Male	1	331
Subject 3	20	Male	1	533

4.3.4. Exploratory visual analysis

To investigate potential relationships between the input variables—heart rate (HR), heart rate variability (HRV), and movement (MOV)—and the output sleep stages—Deep (D), Light (L), REM (R), and Awake (A)—, an initial visual inspection of the data was conducted. Visual analysis serves as a critical preliminary step, as it can reveal patterns and trends not immediately evident through numerical methods and inform the direction of the analysis.

Figures 6, 7, 8 present the sleep stages as shaded columns, overlaid with the input variables plotted as continuous lines. Contrary to theoretical expectations, the visualizations showed some inconsistencies. For instance, while the literature generally describes REM sleep as a state of near paralysis accompanied by low HRV, Figure 6 reveals notable movement during REM periods. Figure 7 also shows movement during REM, although the pattern appears more sporadic, isolated spikes amid prolonged periods of stillness, potentially aligning more closely with theoretical assumptions.

In Figure 7, a particularly noteworthy segment appears between 4:00 AM and 5:00 AM, where high HRV, absence of movement, and a drop in HR suggest a possible deep sleep episode. However, the model fails to identify this correctly and classifies the period as light sleep. On the other hand, in Figure 8, a clear peak in HRV is observed accompanied by no movement and low HR at the beginning of the night, and it is correctly classified as deep sleep.

Heart rate, in general, did not display any clear visual correlation with the sleep stages in the plots. This absence of obvious visual patterns does not preclude the existence of underlying relationships that could be uncovered through machine learning techniques.

At this stage, we are using Garmin's sleep model as a reference. However, its accuracy appears limited and may not serve as a reliable ground truth.

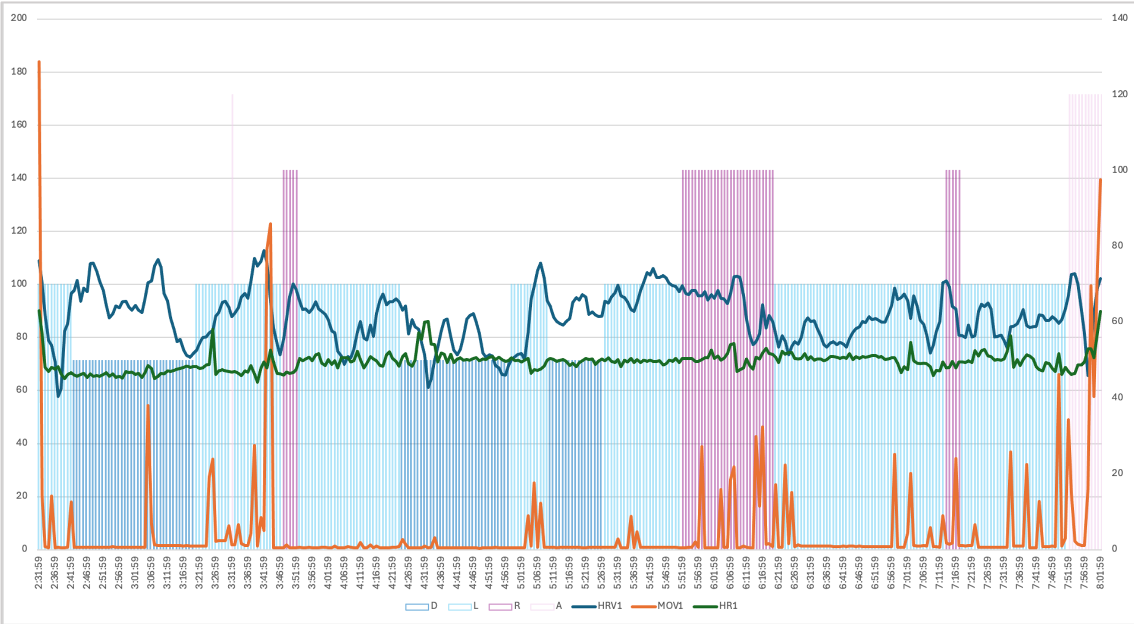


Figure 6 Subject 2's 01/06/2025 night analysis. [Source: own elaboration]

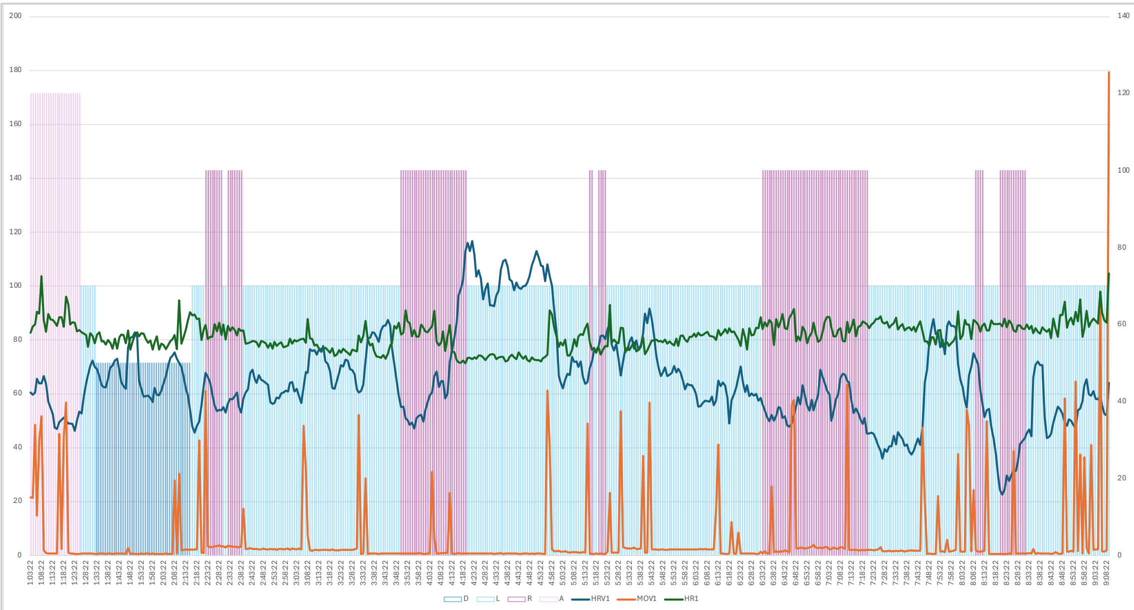


Figure 7 Subject 1's 01/06/2025 night analysis. [Source: own elaboration]

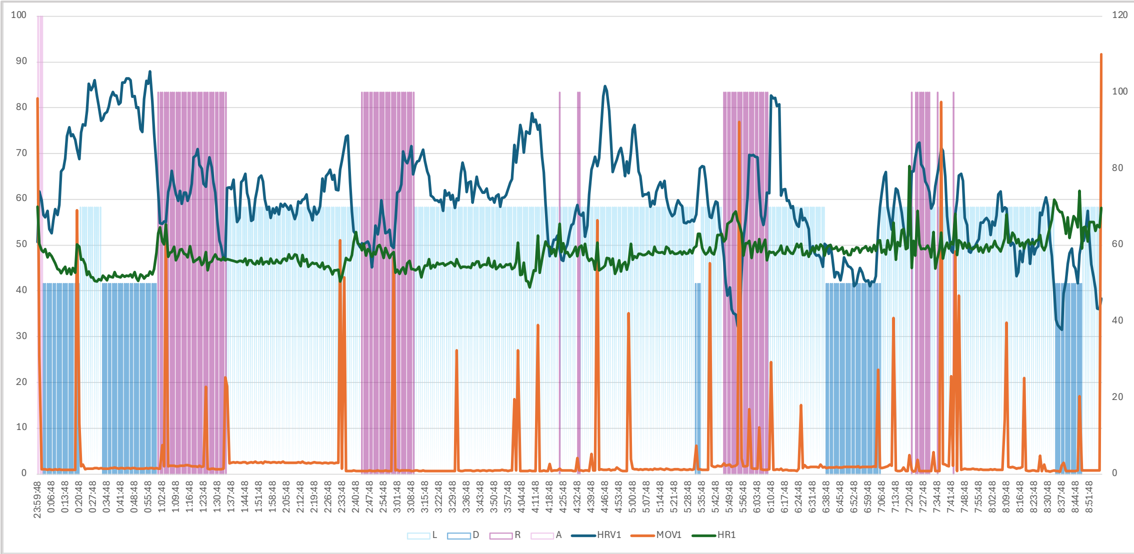


Figure 8 Subject 1's 23/05/2025 night analysis. [Source: own elaboration]

Before attributing inconsistencies solely to Garmin's model, it is necessary to first consider potential limitations within the present analysis. An initial observation includes the presence of significant movement during REM periods, which contradicts the patterns described in existing literature. This anomaly may indicate a hardware-related issue, particularly inaccuracies in the accelerometer.

Accelerometers are known to exhibit bias and drift due to integration errors. Even when stationary, these sensors may not register zero acceleration because of inherent imperfections. Such bias can accumulate when acceleration is integrated to estimate velocity or position, ultimately resulting in drift. To address this, many devices implement a Kalman filter to estimate the true state of the system and compensate for sensor bias.

In this study, integration to calculate velocity or position is not performed. Instead, movement is derived from the difference between consecutive accelerometer readings, a method that inherently minimizes the influence of constant bias.

The possibility of a temporal misalignment between movement data and sleep stage classifications was also considered. For instance, Garmin may correctly identify stillness during REM, while the recorded movement data could be shifted temporally. Nevertheless, this hypothesis can be discarded based on, for example, Figure 6, where the start and end of the sleep period are clearly marked by a significant decrease and increase in movement and heart rate, corresponding with the expected awake states. This suggests that the data is well-aligned with minimal temporal offset.

Additionally, the heart rate and movement data used in this analysis correspond closely to the graphical data displayed in Garmin's own application (Garmin Connect). Although movement is smoothed and aggregated in Garmin's interface, the general patterns shown in Figure 9 are consistent to the corresponding data of study in Figure 8, reinforcing the reliability of the input data used in this study.

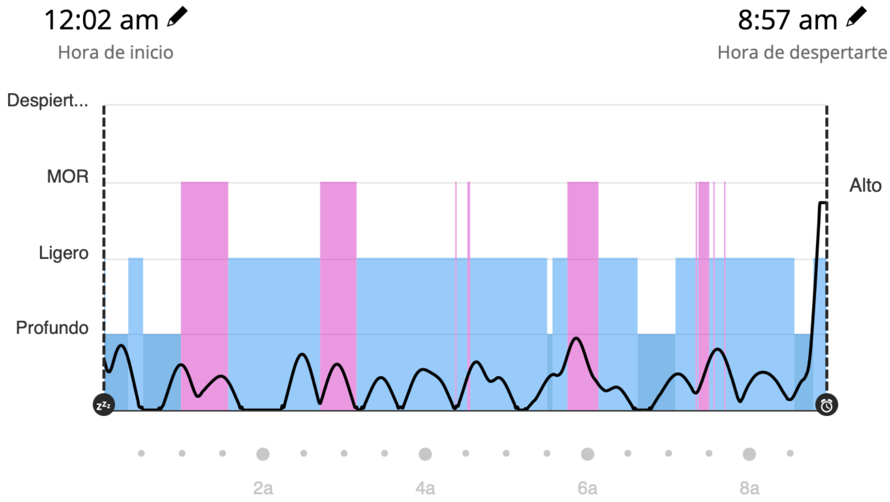


Figure 9 Sleep stage display in Garmin connect for Subject 1's 23/05/2025 night. [Source: Personal Garmin Connect platform]

Based on the preceding observations, the reliability of Garmin's sleep stage classification is called into question. Even Figure 2, which presents a standard night's sleep architecture, does not match the stage distributions provided by Garmin's model for any of the subjects studied.

As previously noted in this thesis, Garmin devices, particularly the Forerunner 55 model, are not known for their excellence in sleep tracking. Garmin has traditionally focused on activity tracking and has integrated sleep analysis primarily as a feature to monitor recovery for athletes. The hardware capabilities of the Forerunner 55 are relatively limited in this context.

Therefore, while this study continues to use Garmin's sleep stage outputs as a reference, it does so with caution. The limitations of the Forerunner's hardware and its generalized sleep tracking algorithm suggest that the results may not be as accurate as those obtained using higher-end Garmin models or devices specifically designed for sleep analysis.

4.3.5. LSTM model decision

A machine learning model was developed to identify suitable moments based on the processed sensor data. Among various machine learning approaches considered, a Long Short-Term Memory (LSTM) model was selected due to the sequential structure and substantial volume of the dataset. The LSTM architecture was deemed the most appropriate for this application owing to its capacity to learn temporal patterns in long, sequential time-series data.

LSTMs are a specialized type of recurrent neural network (RNN) designed to model and learn from sequential data, particularly when dealing with long-term dependencies. Unlike traditional feedforward neural networks, which treat inputs independently, LSTMs maintain a memory cell that can store information across many time steps, making them well-suited for time-series analysis. In the context of this project, the dataset consists of sequential measurements recorded minute by minute, where patterns in sleep stages (light, deep, or REM sleep) emerge over extended periods. The LSTM's ability to capture temporal relationships and retain context from earlier data points was critical for accurately detecting light sleep phases based on these patterns.

Alternative models, including traditional Recurrent Neural Networks (RNNs) and decision trees, were considered. However, standard RNNs are prone to vanishing gradient issues, which reduce their effectiveness when applied to long sequences such as an 8-hour sleep dataset. While decision tree-based models like Random Forests are effective for static data, they lack the capability to capture the temporal dependencies inherent in time-series data.

The LSTM model was designed as a categorical classifier to predict discrete sleep stages (light, deep, REM, or awake) for each time window, with the objective of identifying light sleep as the optimal period for triggering the wake-up alarm. To determine whether the user was in light sleep, the output was explored in a simplified binary classification: 1 for light sleep (indicating a suitable time to wake up) and 0 for non-light sleep (indicating the user should continue sleeping). This version will lead to a simplified yet more accurate model.

4.3.6. LSTM Model Optimization and Performance Evaluation

The LSTM model development begins with the loading and concatenation of data from the previously cleaned Excel files, representing sleep data collected over various nights from a single individual. This approach focuses on learning personalized patterns to create a tailored model. Should the model be later tested in other individuals to validate it.

To mitigate the impact of outliers that could skew prediction results, all variables are capped at the 99th percentile. Features are organized into an input matrix "X", and the binary target variable is stored in an array "y". Subsequently, features are normalized using the *MinMaxScaler*, which transforms values to a range of [0, 1] to ensure consistency across different scales.

During data consolidation, an imbalance between the two target classes may arise, where one class is disproportionately represented. Such imbalances could bias the model toward predicting the majority class. To address this, the Synthetic Minority Oversampling Technique (SMOTE) is employed, generating synthetic samples for the minority class based on existing data points to achieve a balanced dataset.

Following preprocessing, sequences are defined for the LSTM model by selecting the number of time steps used for predictions. This parameter can be optimized later to enhance performance. Given that the data is recorded at one-minute intervals, an initial sequence length of 15 - 30 minutes is selected to capture temporal trends preceding each prediction.

As is standard in machine learning, the dataset is split into a training set (80%) and a testing set (20%). This ratio strikes a balance: 80% gives the model enough data to learn meaningful patterns, while 20% is typically large enough to reliably evaluate how well the model generalizes to unseen data. Using too little for training can lead to underfitting, while too little for testing can result in unreliable performance estimates.

The model is constructed as a Recurrent Neural Network (RNN) utilizing Long Short-Term Memory (LSTM) layers. Initial parameters are set to a standard default configuration, which will be iteratively refined to maximize accuracy. As shown in Figure 10, the model architecture comprises the following components:

- Input Layer: Specifies the input shape, which includes the number of time steps and the number of features (three: HR, HRV, and MOV).
- LSTM Layer: Includes 64 memory units (neurons) as a starting configuration. A bidirectional LSTM, which considers both past and future values, was not adopted, as real-time sleep stage classification requires predictions based solely on past and current data.
- Dropout Layer: Applies a dropout rate of 40%, meaning 40% of the neurons are randomly ignored during training. This helps reduce overfitting by preventing the model from relying too heavily on specific neurons or noise in the data.
- Dense Layers: The first dense layer includes 16 units and uses the ReLU (Rectified Linear Unit) activation function. The final output layer uses a sigmoid activation function, which outputs a probability between 0 and 1 for binary classification (this makes possible to see how confident the model is about its decision, the closer it gets to 0 or 1, the more confident of the decision it is). Both activation functions are illustrated in Figure 11.

These parameters are subject to adjustment based on model performance during experimentation and iteration.

```
model = Sequential([
    Input(shape=(time_steps, X_train.shape[2])),
    LSTM(64, return_sequences=True),
    Dropout(0.4),
    LSTM(32),
    Dropout(0.4),
    Dense(16, activation='relu'),
    Dense(1, activation='sigmoid')
])
```

Figure 10 Section of the LSTM training model code. [Source: own elaboration]

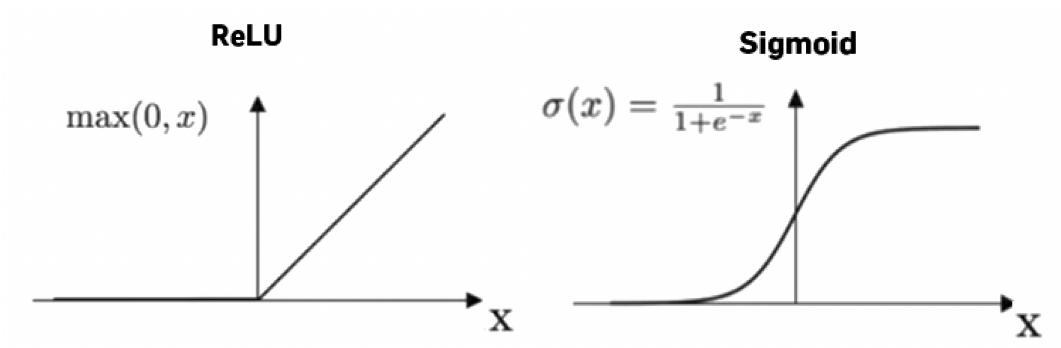


Figure 11 ReLu and Sigmoid Activation Functions. [Source [6]]

The model is compiled using the Adam optimizer (Adaptive Moment Estimation), which dynamically adjusts the learning rate during training, starting at a default value of 0.001. This rate can be modified, reduced to mitigate underfitting by slowing training or increased to accelerate convergence. The loss function is defined as binary cross-entropy, which quantifies the discrepancy between predicted and true labels, penalizing confident incorrect predictions more severely. Accuracy, defined as the proportion of correct predictions, is selected as the primary metric for optimization.

The model uses 2,960 sequences (80% of the full dataset of 3,676) for training. A validation split of 0.2 is applied to this training set, leaving 2,368 sequences for actual training and 592 sequences for validation. The model will be trained over 50 epochs, meaning it will make 50 complete passes through the entire training dataset, learning from each pass and updating its internal weights accordingly.

With a batch size of 32, the number of batches per epoch is computed as:

$$\frac{2368}{32} \approx 74\ batches$$

To enhance training efficiency and prevent overfitting, the following strategies are implemented:

- If no improvement is observed after five consecutive epochs, the learning rate is reduced by 50%.
- If no improvement occurs after ten consecutive epochs, training is stopped, and the model reverts to the best-performing version based on validation accuracy.

These training parameters are standard defaults, known to perform reliably across various objectives and model architectures.

Training was conducted to evaluate the model's capacity to predict sleep stages. Through iterative experimentation, a sequence length of 30 minutes (corresponding to 30 time steps) was determined to provide optimal performance. To identify the best model configuration, training was initially performed using data from a single subject, selected based on the availability of the largest number of samples. After reaching an optimal model for this subject, its generalizability was assessed by testing on data from additional subjects to verify consistent learning performance.

Using data from Subject 1 collected over six nights (2,798 samples), an initial test accuracy of 85.05% was obtained. As illustrated in Figure 12, the training and validation accuracies exhibit similar trends, suggesting the absence of overfitting, which typically presents as increasing training accuracy alongside decreasing validation accuracy. Moreover, the accuracy curves indicate that additional training epochs beyond the 42nd epoch do not yield further performance improvements, suggesting convergence to an optimal solution.

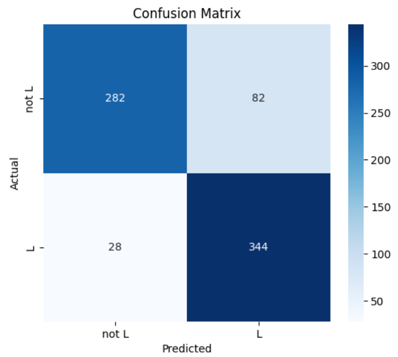
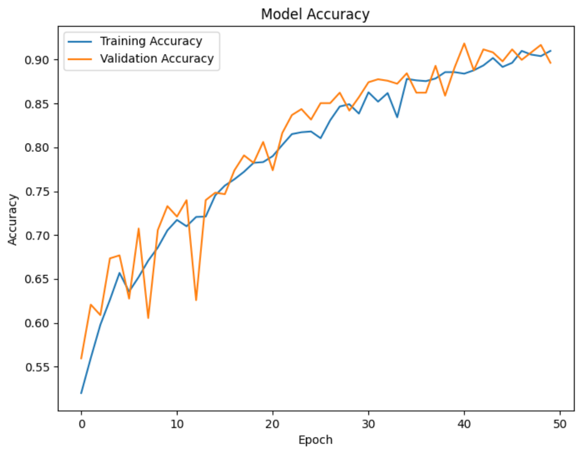


Figure 12 Training and Validation Accuracies throughout epochs (1st iteration). [Source: own elaboration]

Figure 13 Confusion matrix of the test (1st iteration). [Source: own elaboration]

The 85.05% accuracy is very high staring point, from where the model can be pushed by little changes to try to increase it through iterations. With the time steps fixed at 30, the first and second LSTM layers, initially set to 64 and 32 units, respectively, are adjusted to 128 and 64 units to enable the model to capture more complex sequential patterns. This modification will add more parameters to the model and could risks overfitting but is evaluated empirically. The test accuracy improved by 3.39% to 86.68%. Figure 14 confirms that both training and validation accuracies continue to rise, indicating no overfitting.



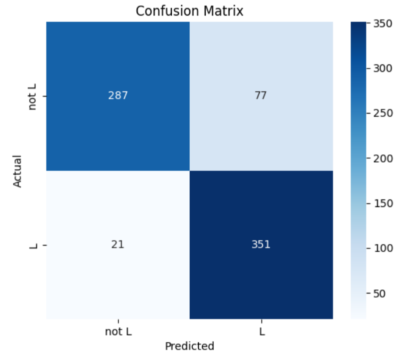
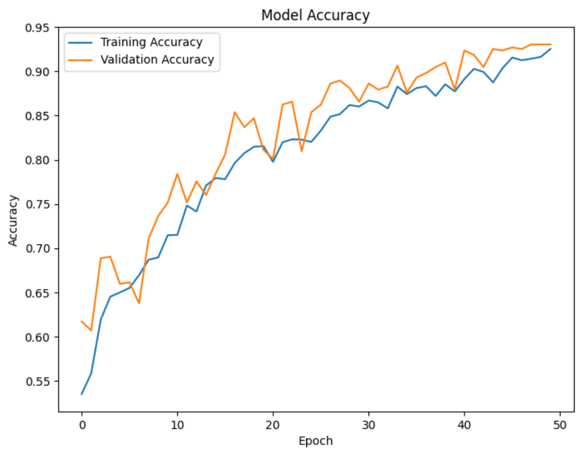


Figure 14 Training and Validation Accuracies throughout epochs (2nd iteration). [Source: own elaboration]

Figure 15 Confusion matrix of the test (2nd iteration). [Source: own elaboration]

To further enhance performance, a third LSTM layer with 32 units with a 40% dropout rate is introduced. This modification increased the test accuracy by 4.49% to 91.17%. As shown in Figure 16, no overfitting is observed, since the validation accuracy is constantly higher than the training accuracy and ends in the last epoch with an accuracy similar to the one obtained in the test. A comparison between Figures 15 and 17 reveals a substantial reduction in the number of false positives, specifically instances where light sleep was incorrectly predicted. Minimizing false positives is particularly important in this context, as erroneous identification of light sleep may result in premature or untimely user awakenings. In contrast, false negatives, failing to detect light sleep episodes, are considered less detrimental to the overall objective.



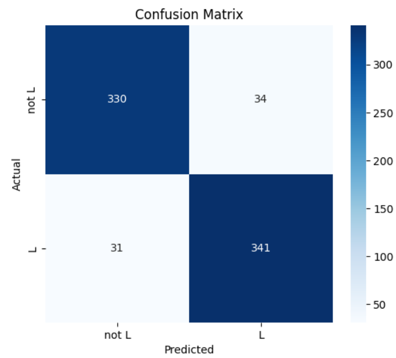
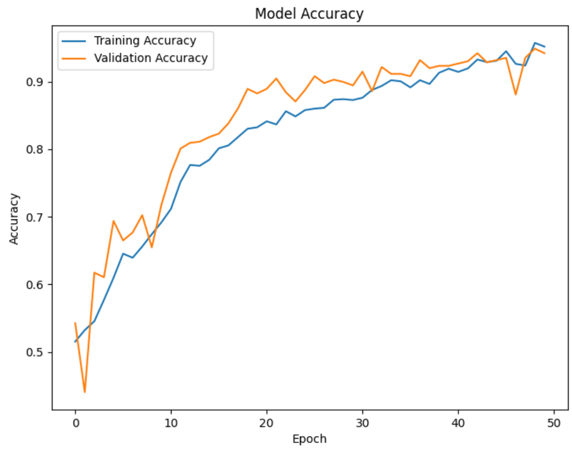


Figure 16 Training and Validation Accuracies throughout epochs (3rd iteration). [Source: own elaboration]

Figure 17 Confusion matrix of the test (3rd iteration). [Source: own elaboration]

Encouraged by these results, the LSTM layer units are doubled again to 256, 128, and 64, respectively. This adjustment improved the test accuracy by 0.95% to 92.12%. However, Figure 18 reveals that the validation accuracy curve begins to flatten, and validation accuracy starts to go below training accuracy, suggesting the onset of overfitting. To test this hypothesis, the units are doubled again (512, 256, 128), resulting in a reduced test accuracy of 62.50% since the training was stopped in the 12th epoch due to an early stop caused by no improvement. This confirms that the configuration with 256, 128, and 64 units represents an optimal balance, yielding a test accuracy of 92.12%.



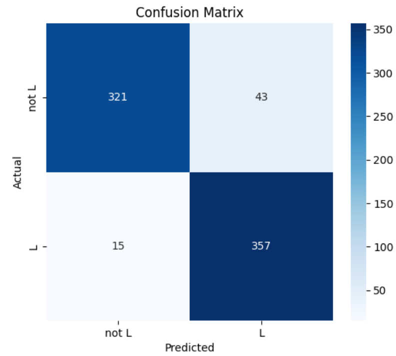
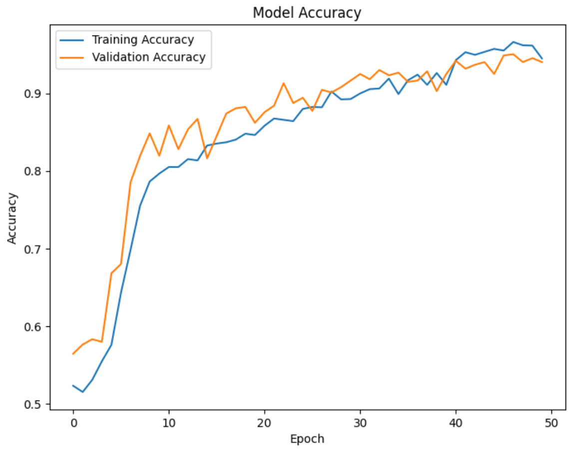


Figure 18 Training and Validation Accuracies throughout epochs (4th iteration). [Source: own elaboration]

Figure 19 Confusion matrix of the test (4th iteration). [Source: own elaboration]

Another parameter explored was the batch size, defined as the number of training samples processed before each weight update. The default batch size of 32 was reduced to 16, resulting in noisier gradient estimates that can potentially improve model generalization. This smaller batch size increased the number of weight updates per epoch from 74 to 147, consequently extending the training duration. The adjustment led to an increase in test accuracy to 92.39%, representing a 0.27% improvement over the previous model. As illustrated in Figure 21, this modification also helped balance false negatives and false positives, notably reducing false positives, which are critical in this application.



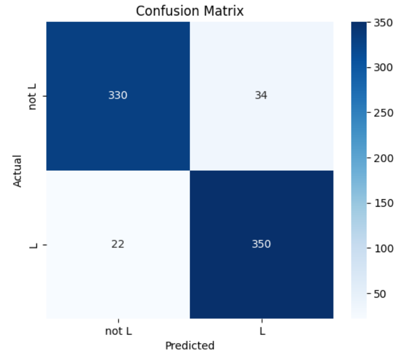


Figure 20 Training and Validation Accuracies throughout epochs (6th iteration). [Source: own elaboration]

Figure 21 Confusion matrix of the test (6th iteration). [Source: own elaboration]

The learning rate was subsequently examined by reducing the default Adam optimizer rate from 0.001 to 0.0001 to evaluate whether slower convergence would enhance performance. This adjustment, however, led to a substantial decrease in test accuracy to 77.17%, suggesting underfitting likely caused by insufficiently large weight updates.

Consequently, the learning rate was restored to its default value, and the next parameter was explored for optimization. The dropout rate was increased from 0.4 to 0.5 in an effort to further reduce overfitting by decreasing neuron co-dependencies. As depicted in Figure 20, slight overfitting is observed in later epochs, since validation accuracy lags slightly behind training accuracy, indicating that a higher dropout rate may be beneficial. Nonetheless, the increased dropout rate resulted in a reduced test accuracy of 67.93%, which is notably lower than the 92.39% accuracy achieved in the sixth iteration.

Achieving 92.39% accuracy is considered strong for this model type, given the substantial data complexity and inherent noise. While random classification accuracy for this binary task is 50%, improving upon this baseline by 42.39% represents a significant gain. Additionally, it is important to note that this model utilizes only past and present data, whereas the Garmin algorithm benefits from full-night data access, which likely contributes to its performance advantage.

The full code for the LSTM model in python can be found in annex 1.

4.4. Simplified model for Connect IQ App

The Garmin Forerunner 55 lacks sufficient computational resources and battery capacity to support a resource-intensive Long Short-Term Memory (LSTM) model for continuous sleep stage classification throughout the night without external device connectivity. Consequently, a simplified model was developed to operate natively on the device, prioritizing computational efficiency while maintaining acceptable classification accuracy. Previous analyses (see Section [4.3.4]) concluded that Garmin Connect's proprietary sleep stage classifications are unreliable. Therefore, this model was designed based on established sleep research literature, ensuring a theoretically robust foundation independent of Garmin's outputs.

4.4.1. Theoretical foundations for the output decision

Sleep stages are characterized by distinct physiological patterns, as described in section 3.1.3. From those patterns, with a Garmin watch can be measured only heart rate variability (HRV), movement (MOV), and heart rate (HR). And they follow the following relationships:

- Light Sleep (L): Elevated HRV, moderate movement, and reduced HR compared to wakefulness.
- REM Sleep (R): Reduced HRV, minimal movement due to paralysis, and elevated HR due to heightened brain activity.
- Deep Sleep (D): Highest HRV, minimal movement, and significantly reduced HR.
- Awake (A): Low HRV, high movement, and elevated HR.

Sleep cycles, averaging 90 minutes in duration, exhibit dynamic patterns where REM sleep durations increase and Deep Sleep durations decrease as the night progresses. Although individual physiological baselines vary, these patterns provide a reliable framework for sleep stage classification. HR was deemed less important for distinguishing sleep stages due to its variability across contexts. Thus, the model prioritizes HRV and movement as primary features, as they already offer a clear differentiation between sleep stages.

The objective was to develop a binary classification model to distinguish "Light Sleep or Awake" (L/A) from "REM sleep or Deep sleep" (R/D). This binary output is tailored to identify optimal wake-up times during Light Sleep, as waking during REM or Deep Sleep can lead to sleep inertia and reduced alertness.

4.4.2. Data and Preprocessing

Physiological data already cleaned and preprocessed for the training of the lstm model was again used for this case. But some further preprocessing was made, with the idea of simplifying the model's calculations and make it adaptable to any Garmin user.

- To account for the accelerometer's limited precision, a margin of 5 units in the movement variable was used to indicate no movement.
- To normalize HRV and account for inter-individual variability, a normalized HRV was computed as the ratio of the current HRV to the baseline HRV, defined as the average of nightly mean HRV values over the previous seven days.

A synthetic dataset was constructed by labeling sleep epochs based on literature-derived patterns. The 90-minute sleep cycle structure was considered during labeling to ensure temporal alignment with expected REM and Deep Sleep distributions.

4.4.3. Model Development

A decision tree was selected over a quadratic expression because it is easier to interpret, simpler to compute, and well-suited for binary classification problems where clear feature thresholds are helpful. A decision tree is a model that makes predictions by asking a series of yes/no questions about the input features. At each step, it splits the data based on one feature and a specific threshold, creating branches that lead to further questions or final decisions. This structure is similar to a flowchart and makes it easy to follow how the model arrives at a prediction. The decision tree was initially trained using the Scikit-learn library in Python with the following parameters:

```
DecisionTreeClassifier(random_state=42, max_depth=2, min_samples_leaf=5,
                       class_weight={0: 2, 1: 1})
```

The maximum depth was modified during iterations but later limited to 2 to ensure simplicity and prevent overfitting. The class weight prioritized higher penalties for misclassifying REM or Deep sleep as Light sleep or Awake satges, as false positives (waking during REM or Deep Sleep) are less desirable for the wake-up application.

The initial decision tree was trained on the synthetic dataset, but the machine-learned structure did not fully utilize both HRV and movement for all possible outputs. To address this, the decision tree was manually refined by adjusting thresholds and restructuring the logic to ensure both variables influenced every classification outcome. This refinement was performed in Microsoft Excel for easy visualization and testing, where thresholds were iteratively tuned to maximize alignment with the literature-based labels.

The final decision tree structure is as follows:

```
|--- Movement > 5
    |--- class: Wake up
|--- Movement <= 5
    |--- HRV normalized > 1.1
        |--- class: Don't wake up
    |--- HRV normalized <0.7
        |--- class: Don't wake up
    |--- else
        |--- class: Wake up
```

The decision tree's performance was evaluated by comparing its classifications against the literature-based labels in the synthetic dataset. Accuracy was calculated as the proportion of correctly classified epochs (number of correct responses divided by the total number of samples). The model achieved an average accuracy of 83.71%, indicating robust performance for a simplified approach. The class-weighted training and manual threshold tuning contributed to this high accuracy, particularly in prioritizing correct identification of non-light sleep to avoid suboptimal wake-up times.

Limitations include the reliance on a synthetic dataset, as validation against polysomnography (PSG) data was not feasible within the study's scope. Future work could incorporate PSG validation or integrate time-based adjustments to account for sleep cycle dynamics, such as increasing REM duration later in the night. Nevertheless, this model demonstrates the feasibility of theory-driven, lightweight sleep stage classification for optimizing wake-up timing on low-cost wearables.

4.4.4. Garmin App design and functionalities

Once the algorithm is defined, it is implemented into the final alarm application. As with any alarm system, the user is required to set an alarm time. In this case, the user sets a latest acceptable wake-up time, although the alarm may trigger up to 30 minutes earlier if determined optimal by the algorithm.

The alarm-setting interface is illustrated in Figure 22. Upon launching the application, a default wake-up time is displayed. This time is retained from night to night, allowing users with a consistent wake-up schedule to avoid reconfiguration.

Time selection is performed in two steps. Initially, the hour value is highlighted (indicated in blue in Figure 22, Step 1) and can be adjusted using the up and down buttons on the watch.

Confirmation is performed via the start/stop button, after which the minute selection is enabled (highlighted in blue in Figure 22, Step 2). Once both values are confirmed, the alarm is set, and a different screen is displayed (Figure 22, Step 3). This screen shows the current time during the night, along with the status "Wake up" or "Don't wake up", indicating the algorithm's real-time decision.

Additionally, during the alarm-setting process, it is possible to access further configuration options. By holding the up button, a settings menu appears. One of the configurable parameters is the alarm feedback mode, which allows selection between sound and vibration or vibration only (Figure 22, Steps 4, 5, and 6).

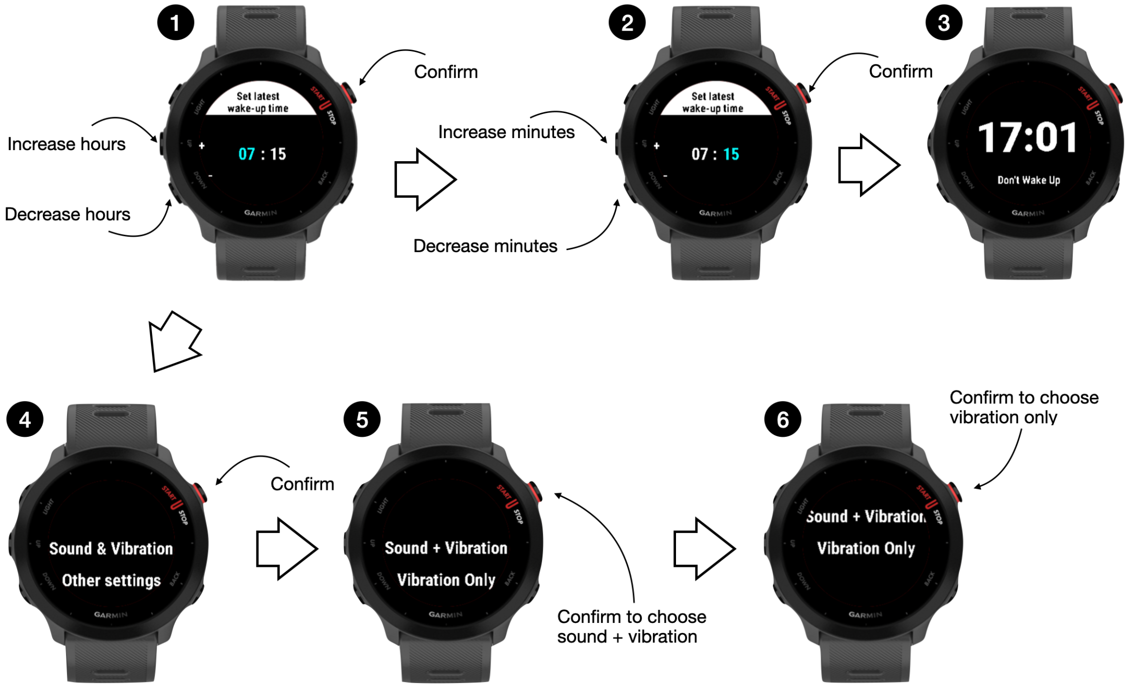


Figure 22 Alarm configuration steps. [Source: own elaboration]

Relevant parts of the Garmin wake-up alarm code in Monkey C can be found in annex 2.

5. Results and Validation

The final LSTM model achieved a 92.34% accuracy on a personalized dataset comprising six nights of sleep data from Subject 1 (female, age 21). To assess its generalizability across different individuals and sexes, the model was further validated using data from Subject 2 (male, age 20) and Subject 3 (male, age 23).

The model was retrained with one night of sleep data from each of Subjects 2 and 3, yielding satisfactory results. For Subject 2, the model attained an accuracy of 85.92%, representing a 35.92% improvement over a random binary baseline (50%). For Subject 3, the model achieved a higher accuracy of 90.14%, improving by 40.14% over the same baseline. Although these accuracies are lower than that of Subject 1, the difference is likely attributable to the smaller dataset sizes for Subjects 2 and 3 (one night each) compared to Subject 1's six nights. These results demonstrate the model's ability to adapt to new users, including across sexes, but suggest that performance may improve with larger datasets. Further testing with more diverse subjects and extended data collection would enhance robustness and confirm generalizability.

The application was successfully tested and performed as intended. It accurately read sensor data (heart rate, heart rate variability, and movement) from the Garmin smartwatch, processing these metrics minute-by-minute to determine optimal wake-up times. The alarm functionality worked reliably, triggering the alarm precisely when the algorithm identified an appropriate light sleep phase within the user's specified wake-up window, and stopping it via the corresponding button, with options for sound and vibration or vibration only, as configured by the user.

However, further research is needed to evaluate the application's effectiveness in reducing morning grogginess (inertia) and improving sleep quality perception. Long-term studies with a larger and more diverse participant pool are essential to validate its performance across a broader range of individuals, including variations in age, sex, and sleep patterns, and to quantify its impact on wake-up experience and overall well-being.

6. Planning

The project was structured into four main phases: the literature review, the development of the Garmin application using Monkey C, the data collection and preprocessing, and finally, the training and evaluation of the LSTM model along with its simplified version. Each of these phases is described in greater detail and scheduled day by day in annex 3.

7. Economic assessment

This section presents the estimated cost of developing the present work, as if it had been carried out in a professional setting. The assessment considers the time invested, operational expenses, experimental resources, equipment depreciation, and potential licensing costs. All values are expressed in euros (€).

1. Labor Cost

According to the university, the workload for this Final Project is equivalent to 360 hours. Assuming a cost of €15/hour, the total labor cost is:

$$360\text{ hours} \times 15€/\text{hour} = 5{,}400€$$

A breakdown by task (estimated):

- Literature review and research: 60 h
- Data collection and preprocessing: 50 h
- Algorithm development and training: 60 h
- App interface design and implementation: 100 h
- Testing and validation: 30 h
- Writing and documentation: 60 h

2. Operational Expenses

These include basic utilities, office supplies, and indirect usage costs during the project development. (See Table 7)

Table 7 Operational expenses breakdown. [Source: own elaboration]

Item	Estimated Net Cost (€)	Estimated Gross Cost (€)
Electricity, water, heating	60 €	50 €
Internet	70 €	58 €
Miscellaneous	20 €	17 €
Total	**155 €**	**125 €**

3. **Experimental Expenses**

The expenses used in the experimentation part are broken down in Table 8.

Table 8 Experimental expenses breakdown. [Source: own elaboration]

Item	Net Cost (€)	Gross cost
Garmin Forerunner 55 smartwatch	180 € x 3 units	446.28 €
Cursor subscription	20 $/month x 5 months = 17.30 € x 5	71.49 €
Grok subscription	39 $/month x 5 months = 25.95 € x 5	107.23 €
Total	**756.25 €**	**625 €**

4. Depreciation of Equipment

Laptop used during the project, estimated at €1,000 with a useful life of 5 years:

→ Annual depreciation: $€1{,}000 \;/\; 5 \;=\; €200$

→ Proportional use for 5 months: $€200 \text{ x } 5 \;/\; 12 \sim €80$

Smartphone used for testing, estimated at €600 with a 3-year life:

→ Annual depreciation: $€600 \;/\; 3 \;=\; €200$

→ Proportional use for 5 months: $€200 \text{ x } 5 \;/\; 12 \;\sim €80$

(See depreciation expenses in Table 9)

Table 9 Equipment depreciation expenses breakdown. [Source: own elaboration]

Equipment	Depreciation (€)
Laptop	80 €
Smartphone	80 €
Total	160 €

4. Total Cost (Without Tax)

The total cost of the project before taxes is broken down in Table 10.

Table 10 Subtotal expenses breakdown before taxes. [Source: own elaboration]

Category	Total (€)
Labor	5,400 €
Operational expenses	125 €
Experimental expenses	625 €
Equipment depreciation	160 €
Subtotal	**6,310€**

6. Including Taxes

Assuming a 21% VAT rate:

$$\text{VAT: } 6{,}310€ \times 0.21 = 1{,}325.10€$$

$$\text{Total including VAT: } 6{,}310€ + 1{,}325.10€ = 7{,}635.1€$$

The total cost of the work amounts to 6,310€, which increases to 7,635.10€ after applying the corresponding VAT (21%).

The total estimated cost of over 7,500 € reflects a reasonable valuation for the work involved, considering it includes deep research, dataset and LSTM model development, and smartwatch integration. This cost would be higher in a commercial or industrial setting, especially if advanced features such as LSTM model integration with external API data were fully implemented and licensed.

Given the technical depth of the work and the multi-disciplinary nature of the tasks,the cost can be considered moderate and justified.

8. Environmental assessment

The growing use of artificial intelligence (AI) raises concerns about its environmental impact, particularly through energy consumption and carbon dioxide (CO_2) emissions. While cloud-based AI systems rely on large data centers that consume significant electricity and water, on-device AI models running on smartphones also contribute to environmental footprints. This section estimates the CO_2 emissions of the Long Short-Term Memory (LSTM) model developed in this thesis for Garmin smartwatches, which is executed on smartphones via the Garmin API, assuming adoption by a large user base.

Evaluating the environmental impact of AI models is essential to understand their carbon footprint, especially for applications used daily by many users. Efforts should focus on optimizing these models to reduce energy use and emissions. In the Garmin Connect IQ platform, which allows users to download applications and watch faces for Garmin smartwatches, popular apps can reach up to 1 million downloads. A conservative estimate of 500,000 downloads is assumed for this model, meaning 500,000 smartphones would run the model each night. This scale requires careful consideration of environmental consequences.

The model runs for an average of 8 hours each night, performing one inference per minute, resulting in:

$$8\,h \cdot \frac{60\,min}{h} \cdot \frac{1\,inference}{min} = 480\,infereces\ per\ night$$

The model operates on users' own smartphones. And is defined by the following parameters which are used to estimate the emissions:

- Number of Parameters: 30,369, reflecting the model's computational complexity.
- Inference Time: 70 ms per inference, measured on a personal computer and assumed applicable to smartphones, as modern devices have processors optimized for lightweight AI tasks.
- Power Draw: 2.5 W during active CPU inference, based on typical smartphone CPU power consumption during intensive tasks.
- Emission Factor: 0.283 kg CO_2e/kWh, as reported by the *Comisión Nacional de los Mercados y la Competencia* (CNMC)[5].

The total compute time per night per subject is:

$$480\,inferences \times 0.07\,sec = 33.6\,sec = 0.009\overline{3}\,h$$

The energy consumed per night per subject is:

$$Energy = 2.5\,W \times 0.009\overline{3}\,h = 0.02\overline{3}\,Wh = 0.00002\overline{3}\,kWh$$

The CO_2 emissions per night per subject are:

$$0.00002\overline{3}\,kWh\;\times\;0.283\,kg\,CO_2/kWh\;=\;0.00000660\overline{3}\,kg\,CO_2 =\;6.60\overline{3}\,mg\,CO_2e$$

The annual energy consumption for 500,000 users is:

$$0.00002\overline{3}\,kWh \cdot 500{,}000\,users \cdot 365\,days = 4{,}258.\overline{3}\,kWh$$

The annual CO_2 emissions for 500,000 users are:

$$0.00000660\overline{3}\,kg\,CO_2e \cdot 500{,}000\,users \cdot 365\,days = 1{,}205.1\,kg\,CO_2e$$

As a reference, a single ChatGPT query consumes approximately 0.3 Wh, which is more than the total overnight energy consumption of this application for 10 users. Over the course of a year, the total CO_2e emissions generated by the app are roughly equivalent to those produced by a single gasoline-powered car. In essence, widespread use of the app would be like adding one more petrol car to the planet.

The environmental impact of the application is relatively modest; however, its energy consumption contributes to the broader cumulative footprint of global digital technologies. To enhance sustainability, future iterations should focus on optimizing energy efficiency, minimizing computational requirements while maintaining functionality, and aligning with efforts to reduce the overall carbon intensity of digital infrastructure.

9. Social and gender equality assessment

The application is designed to be inclusive across gender and social groups. The algorithm, based on a long short-term memory (LSTM) model, adapts to individual user data, ensuring functionality across diverse populations, including men, women, and individuals with non-traditional gender identities. Testing was conducted on a small but balanced sample (one woman, two men, with comparable nights studied), confirming the algorithm's adaptability across sexes. Since the app focuses on clinical metrics (HR, HRV, and movement), it avoids gender-specific biases in its core functionality.

However, the algorithm's performance relies on the accuracy of Garmin's sleep stage detection, which provides input data for the LSTM model. A 2019 Garmin Health study conducted with the University of Kansas Medical Center (KUMC) validated Garmin's Advanced Sleep Monitoring, reporting an overall accuracy of 69.7% for sleep stage classification using photoplethysmography (PPG) and actigraphy data. The study included 67 participants (47 men, 20 women, aged approximately 35.6 ± 8.3 and 35.9 ± 10.5, respectively) with no known sleep-related conditions. While the study was conducted in real-world conditions, the participant pool was skewed toward men, raising potential concerns about gender-specific biases in Garmin's algorithm, particularly for women whose sleep cycles may vary due to hormonal fluctuations.

Moreover, the application's reliance on Garmin smartwatch technology introduces potential socioeconomic barriers. Garmin devices are not universally affordable, which may limit access for lower-income individuals or vulnerable groups, such as immigrants or those with limited resources. To mitigate this, the app was developed to function on lower-end Garmin models, expanding accessibility compared to similar functionalities typically requiring higher-end devices.

Individuals with physical disabilities, particularly those with limited mobility (e.g., paralytics), may face challenges as well, as the algorithm heavily relies on movement to detect sleep stages. While this group is unlikely to purchase activity-tracking smartwatches, this limitation highlights a potential area for future improvement, such as integrating alternative metrics for sleep detection.

Team Composition, Language and Representation

The project team consisted of a female student and a male director, with testing conducted on one woman and two men. While the team is small, it reflects a reasonable gender balance. However, the limited sample size for testing (three participants) restricts broader conclusions about gender equity in the research process. A more diverse testing pool in future iterations would strengthen the project's inclusivity.

The project adheres to inclusive, non-sexist, and non-androcentric language in its documentation and user interface. No imagery or content promotes discriminatory ideas based on gender, race, culture, or socioeconomic status. The app's design focuses on functionality and accessibility, avoiding elements that could reinforce social hierarchies or exclusionary policies.

Sustainable Development Goals

The project aligns with three United Nations Sustainable Development Goals (SDGs), contributing to social and gender equity as well as health outcomes:

- **SDG 3:** Good Health and Well-Being

By enabling precise sleep stage detection and optimal wake-up timing, the application promotes better sleep management, which is critical for physical and mental health. Improved sleep quality can enhance cognitive function, emotional well-being, and overall quality of life. Future initiatives, such as public education campaigns on sleep health, could amplify the project's societal impact by raising awareness and promoting healthy sleep practices across diverse communities.

- **SDG 5:** Gender Equality

The application is designed to be gender-neutral, with the LSTM model adapting to individual physiological data regardless of gender. The algorithm was trained and tested on a balanced dataset (comparable nights studied for one woman and two men), ensuring equitable performance across sexes. This personalization minimizes potential biases, including those that might stem from Garmin's underlying technology, which may not fully account for gender-specific physiological variations.

- **SDG 10:** Reduced Inequalities

By enabling advanced sleep-tracking functionality on lower-end Garmin smartwatch models, the application broadens access to technology typically reserved for users of premium devices. This reduces economic barriers, making sleep health tools more accessible to individuals with limited financial resources. While the cost of Garmin devices still poses a challenge for some vulnerable groups, this development represents a step toward reducing technological disparities and fostering inclusivity in health monitoring.

10. CONCLUSIONS

This thesis successfully achieved its objectives by developing a functional beta prototype of a Garmin smartwatch application that tracks heart rate, heart rate variability, and movement to detect light sleep phases and trigger an optimal wake-up alarm. The prototype demonstrates robust basic functionalities, validated through testing across a small but diverse sample of users, achieving promising accuracies. However, as health technology is an ever-evolving field, there is significant potential for further refinement and enhancement of the application.

Future iterations of this project could focus on the following improvements:

- Reduce reliance on Garmin's sleep stage detection algorithm by developing a proprietary model trained directly on polysomnography (PSG) data. This would improve accuracy and address potential gender biases in Garmin's algorithm.

- Integrate the LSTM model into the alarm functionality via the Garmin API for seamless real-time processing.

- Expand the app's functionalities to enhance user convenience, such as enabling customizable alarms for different days of the week, adding a snooze option for users who would still struggle to wake up, and incorporating an option to trigger the alarm whenever Garmin's Body Battery metric reaches 100%, reflecting optimal energy levels.

- Improve user experience by transforming the app into a widget with automatic sleep detection, eliminating the need for manual activation each night and ensuring seamless operation upon sleep onset.

- Provide post-wakeup insights, such as sleep quality metrics, grogginess levels, or personalized recommendations, to enhance user engagement and quantify the alarm's impact on morning alertness.

- Validate the model by testing it across a diverse range of subjects and conduct in-depth research to identify more precise optimal wake-up times. Further refine wake-up timing by enabling user feedback, allowing users to report grogginess levels or confirm the alarm's effectiveness, thereby tailoring the model to individual physiological profiles.

- Promote and educate the public on the importance of sleep for overall health across diverse settings, from children in schools to adults in universities and workplaces, through targeted campaigns and integration of sleep health insights within the app to foster greater awareness and adoption of healthy sleep practices.

11. Acknowledgments

I wish to thank my thesis supervisor, Antoni Grau, for his guidance and helpful suggestions throughout the development of this project.

I am also immensely grateful to the participants who provided the datasets for training and validating the model. Their consistent availability, enthusiasm, and interest in the research outcomes were greatly appreciated.

Finally, I extend my thanks to my family and friends for their support and thought-provoking questions, which inspired critical improvements and enriched the iterations of this work.

12. Bibliography

References

[1] GARMIN. Garmin Health announces sleep study results. Garmin Blog, 22 April 2022. https://www.garmin.com/en-US/blog/health/garmin-health-announces-sleep-study-results/, [Accessed 11 May 2025].

[2] WALKER, Matthew. Why We Sleep: Unlocking the Power of Sleep and Dreams. Penguin Books, 2017.

[3] Garmin, "Sleep Tracking," [Online]. Available: https://www.garmin.com/es-ES/garmin-technology/health-science/sleep-tracking/. [Accessed: 13-May-2025].

[4] GARMIN. HRV Status – Garmin Technology. Garmin, n.d. https://www.garmin.com/en-US/garmin-technology/health-science/hrv-status/, [Accessed 11 May 2025].

[5] Generalitat de Catalunya, "Factors d'emissió associats a l'energia," Canvi Climàtic, 2025. [Online]. Available: https://canviclimatic.gencat.cat/en/actua/factors_demissio_associats_a_lenergia/index.html. [Accessed: 25-Jun-2025].

[6] Paperspace, "Activation functions," Machine Learning Wiki, [Online]. Available: https://machine-learning.paperspace.com/wiki/activation-function. [Accessed: 18-Jun-2025].

Additional bibliography

[1] Science of Sleep: What is Sleep? - Division of Sleep Medicine. Harvard Medical School. Available at: https://sleep.hms.harvard.edu/education-training/public-education/sleep-and-health-education-program/sleep-health-education-47 [Accessed 10 Mar 2025]

[2] Sleep homeostasis and the circadian clock. PubMed Central. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6584681/ [Accessed 13 May 2025].

[3] Sleep Foundation, "Stages of Sleep," [Online]. Available: https://www.sleepfoundation.org/stages-of-sleep. [Accessed: 13-May-2025].

[4] National Sleep Foundation, "What Are the Sleep Stages?" [Online]. Available: https://www.thensf.org/what-are-the-sleep-stages/. [Accessed: 13-May-2025].

[5] Neuroscientifically Challenged, "2-Minute Neuroscience: Stages of sleep," [Online]. Available: https://neuroscientificallychallenged.com/posts/2-minute-neuroscience-stages-of-sleep. [Accessed: 25-May-2025].

[6] [C. Armon, K. G. Johnson, A. Roy, and W. J. Nowack, "Polysomnography," Medscape, [Online]. Available: https://emedicine.medscape.com/article/1188764-overview. [Accessed: 25-May-2025].

[7] M. Radha, P. Fonseca, A. Moreau, M. Ross, A. Cerny, P. Anderer, X. Long, and R. M. Aarts, "Sleep stage classification from heart-rate variability using long short-term memory neural networks," Scientific Reports, vol. 9, Art. no. 14149, Oct. 2019. [Online]. Available: https://www.nature.com/articles/s41598-019-49703-y. [Accessed: 25-May-2025].

[8] E. Campo, "Wrist movement analysis for long-term home sleep monitoring," Expert Systems with Applications, vol. 198, Art. no. 116804, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417421013051. [Accessed: 25-May-2025].

[9] Cleveland Clinic, "Parasympathetic Nervous System (PSNS): What It Is & Function," [Online]. Available: https://my.clevelandclinic.org/health/body/23266-parasympathetic-nervous-system-psns. [Accessed: 25-May-2025].

[10] W. Wisner, "Can a Better Night's Sleep Boost Your Memory?," Verywell

Mind, Aug. 1, 2024. [Online]. Available: https://www.verywellmind.com/can-better-sleep-boost-memory-8668178. [Accessed: 25-May-2025].

[11] Columbia University Department of Psychiatry, "How Sleep Deprivation Impacts Mental Health," Mar. 16, 2022. [Online]. Available: https://www.columbiapsychiatry.org/news/how-sleep-deprivation-affects-your-mental-health. [Accessed: 25-May-2025]

[12] G. Medic, M. Wille, and M. E. H. Hemels, "Short- and long-term health consequences of sleep disruption," Nature and Science of Sleep, vol. 9, pp. 151–161, May 2017. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC5449130/. [Accessed: 25-May-2025].

[13] A. Rundo and M. Trenta, "Sleep disorders and cognitive impairment: A review," Sleep Medicine Reviews, vol. 35, pp. 63–72, Oct. 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847817306290. [Accessed: 25-May-2025].

[14] 2nd Breakfast, “The history of the heart rate monitor,” 2nd Breakfast, [Online]. Available: https://2ndbreakfast.audreywatters.com/the-history-of-the-heart-rate-monitor/. [Accessed: 18-Jun-2025].

[15] CrazyWatches, “Pulsar calculator Time Computer LED 1975,” CrazyWatches.pl, [Online]. Available: https://www.crazywatches.pl/pulsar-calculator-time-computer-led-1975. [Accessed: 18-Jun-2025].

[16] Fitbit, “Fitbit's Smart Wake explained,” Gadgets & Wearables, 09-Jul-2021. [Online]. Available: https://gadgetsandwearables.com/2021/07/09/fitbit-smart-wake/. [Accessed: 18-Jun-2025].

[17] Garmin Ltd., “Vívoactive 6 series owner’s manual – sleep & alarm,” Garmin, [Online]. Available: https://www8.garmin.com/manuals/webhelp/GUID-8C2C402F-55AC-431F-9CF2-1442B89CE149/EN-US/GUID-5F030BAF-DBBA-4201-A99C-5F3E1C16CC4D.html. [Accessed: 18-Jun-2025].

[18] MobiHealthNews, “Eight years of Fitbit news leading up to its planned IPO,” MobiHealthNews, 08-May-2015. [Online]. Available: https://www.mobihealthnews.com/43423/eight-years-of-fitbit-news-leading-up-to-its-planned-ipo. [Accessed: 18-Jun-2025].

[19] Oura Health Ltd., “History of Oura,” Oura Blog, 29-Mar-2022. [Online]. Available: https://ouraring.com/blog/es/history-of-oura/#:~:text=The%20first%2Dgeneration%20Oura%20Ring,and%20established%20the%20brand's%20credibility. [Accessed: 18-Jun-2025].

[20] Pavlok, “Shock Clock 3,” Pavlok, [Online]. Available: https://shop.pavlok.com/products/shock-clock-3?srsltid=AfmBOoqWmYUwnV4qtoVi1OoDojjPm5GwyY8ClLrl_eBzdAFBNpcQ1P01. [Accessed: 18-Jun-2025].

[21] Perrelet, “Nuestra historia,” Perrelet, [Online]. Available: https://perrelet.com/es-ES/nuestra-historia?srsltid=AfmBOordENPX8MacENfkonm868S4CgIVMw8etI0nl9UzVzV_4AcxHFMk. [Accessed: 18-Jun-2025].

[22] Richman Magazine, “Manpo-kei: 60 years of walking towards your health,” Richman Magazine, 2023. [Online]. Available: https://www.richmanmagazine.com/manpo-kei-60-years-of-walking-towards-your-health/. [Accessed: 18-Jun-2025].

[23] SleepCycle, “Smart alarm clock: how it works,” SleepCycle, [Online]. Available: https://sleepcycle.com/sleep-talk/smart-alarm-clock. [Accessed: 18-Jun-2025].

[24] Sleep Doctor, “Withings Sleep tracking mat review,” Sleep Doctor, [Online]. Available: https://sleepdoctor.com/pages/reviews/withings-sleep-tracking-mat-review. [Accessed: 18-Jun-2025].

[25] Tantsissa, “Smart alarm – AutoSleep Apple Watch app,” AutoSleep, [Online]. Available: https://autosleepapp.tantsissa.com/watch-use/smart-alarm#h.3ilen7obi34a. [Accessed: 18-Jun-2025].

[26] U.S. FDA, “K983533 – Actiwatch clinical sleep monitor,” FDA, [Online]. Available: https://www.accessdata.fda.gov/cdrh_docs/pdf/K983533.pdf. [Accessed: 18-Jun-2025].

[27] Wikipedia, “Pedometer,” Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/Pedometer. [Accessed: 18-Jun-2025].

[28] aXbo GmbH, “aXbo Sleep Phase Alarm Clock product launch press release,” aXbo GmbH, 11-Jun-2006. [Online]. Available: http://axbo.com/international/press/press_releases_EN/PR_aXbo_0611_product_launch_EN.pdf. [Accessed: 18-Jun-2025].

[29] PubMed Central, “The evolution of actigraphy for sleep,” PMC, [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7811468/. [Accessed: 18-Jun-2025].

[30] Google Store, “Trackers category,” Google Store, [Online]. Available:

https://store.google.com/category/trackers?hl=fr. [Accessed: 18-Jun-2025].

[31] Veproject1, "A brief history of wearable technology," YouTube, 10-Apr-2023. [Online]. Available: https://youtu.be/1qyE0AmpoBI. [Accessed: 18-Jun-2025].

[32] IEEE Xplore, "A Machine Learning Approach for Sleep Stage Classification Using Wearable Sensors," IEEE Xplore, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10565308. [Accessed: 24-Jun-2025].

[33] Eight Sleep, "How the Pod Improves HRV, Heart Rate, and Deep Sleep," Eight Sleep Blog, 2025. [Online]. Available: https://www.eightsleep.com/blog/hrv-hr-deep-sleep-improvement/. [Accessed: 24-Jun-2025].

[34] Eight Sleep, "Pod Cover," Eight Sleep EU, 2025. [Online]. Available: https://www.eightsleep.com/eu/pod-cover/. [Accessed: 24-Jun-2025].