Machine Learning
Data Analytics

Friedrich-Alexander-Universität
Technische Fakultät

Machine Learning based Sleep Staging from Contactless Vital Sign Radars

Master's Thesis in Medical engineering

submitted
by

Carlos Herrera Krebber

born 12.12.1996 in Nuremberg

Written at

Machine Learning and Data Analytics Lab
Department Artificial Intelligence in Biomedical Engineering
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU)

in Cooperation with

Medicine 1
University Hospital FAU Erlangen-Nuremberg

Advisors:	Prof. Dr. Bjoern Eskofier, Daniel Krauß, M. Sc.
Started:	01.01.2024
Finished:	01.08.2024

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.
Die Richtlinien des Lehrstuhls für Bachelor- und Masterarbeiten habe ich gelesen und anerkannt, insbesondere die Regelung des Nutzungsrechts.

Erlangen, den 1. August 2024

Übersicht

Schlaf ist eine lebenswichtige Funktion des Körpers, notwendig für die ordnungsgemäße Funktion des Gehirns und ein physiologisch und psychologisch wichtiger Faktor für das Wohlbefinden des Menschen, der sogenannte gute Schlaf. Es hat sich gezeigt, dass Schlafprobleme mit einer Vielzahl von neurodegenerativen Erkrankungen in Verbindung stehen. Eine davon ist die Parkinson-Krankheit, die nach Angaben der Weltgesundheitsorganisation in mehr als 8 Millionen Fällen diagnostiziert wurde, wobei die Zahl der Patienten steigt. Sie kann in motorische und nicht-motorische Symptome unterteilt werden. Eine Art von Schlafstörung, die so genannte REM-Schlaf-Verhaltensstörung (rapid-eye-moment), wird als frühes nicht-motorisches Symptom identifiziert. Eine verlässliche Schlafüberwachung ist daher unerlässlich, um diese Symptome zu erkennen. Der Goldstandard der Schlafanalyse ist die Polysomographie. Dabei handelt es sich um ein System von Sensoren zur Messung von Vitaldaten, das gute und leicht interpretierbare Ergebnisse für die Schlafstadienbestimmung liefert. Allerdings ist es zeitaufwändig, es am Patienten anzubringen und zu analysieren. Vor allem aber behindert es den Schlaf sehr stark und ist für eine Langzeitüberwachung nicht geeignet. Vielversprechende Alternativen dazu sind tragbare oder kontaktlose Geräte. Ein gängiges kontaktloses Gerät ist ein Radarsensor. In dieser Arbeit haben wir ein $61\,Ghz$- continous wave Radarsystem zur Erfassung von Vitaldaten und Deep Learning (DL) Algorithmen zur Klassifizierung von Schlafstadien verwendet, um eine Lösung für dieses Problem zu bieten. Die im Zuge dieser Abschlussarbeit durchgeführte Arbeit umfasst die Datenerfassung, die Merkmalsextraktion, die Aufbereitung und die Klassifizierung der Schlafstadien mithilfe von DL-Netzwerken. Bei den untersuchten DL-Algorithmen handelte es sich um ein convolutional neural network (CNN), ein long-term short-term memory cell (LSTM) und eine time convolutional network (TCN) Architektur. Während die Herzschlag-Extraktion mit einem F1-Score von $0,803$ gut abschnitt, erfassten wir $97,07\,\%$ der auftretenden Herzschläge. Jedoch war die folgenden Klassifizierungen der Schlafphasen nicht zufriedenstellend, wobei die besten Ergebnisse für eine einfache Wake/Sleep-Klassifizierung erzielt wurden. Wir erreichten eine balanced accuracy von $0,631$ und einen F1-Score von $0,732$ für eine CNN-Architektur. Der Matthews-Korrelationskoeffizient (MCC) erreicht jedoch nur $0,203$. Diese Ergebnisse zeigen, dass es möglich ist, die Herzfrequenz für die Schlafanalyse zu verwenden, aber wir benötigen zusätzliche Daten, wie z.B. Atemwellen oder Körperbewegungen, um die Vorhersage der Schlafphase zu verbessern.

Abstract

Sleep is a vital function of the body, necessary for the proper function of the brain, and a physiological and psychologically important factor for the well-being of humans, the so-called good night of sleep. It has been shown that sleep issues are related to a wide variety of neurodegenerative pathologies. One of these is Parkinson's Disease which, according to the World Health Organisation, has been diagnosed in more than 8 million cases, with an increasing number of patients. It can be divided into motor and non-motor symptoms. A type of sleep disorder, so-called rapid-eye-moment (REM) sleep behaviour disorder (RBD), is identified as an early occurring non-motor symptom. Reliable sleep monitoring is, therefore, essential to detect these symptoms. The gold standard sleep analysis the polysomography. A system of sensors measuring vital signals, producing good and easily interpretable results for sleep staging. However, it is time-consuming to mount to the patient and to analyse. Most importantly, it is very obstructive to sleep and not feasible for longitudinal monitoring. Promising alternatives to this are wearable or contactless devices. One common contactless device is a radar sensor. In this work, we used a $61\,Ghz$ continuous wave radar system for capturing vital signs and deep learning (DL) algorithms for sleep stage classification to offer a solution to this problem. The work performed within this thesis contains data collection, feature extraction, preparation, and sleep stage classification using DL networks. The inspected DL algorithms were a convolutional neural network (CNN), a long-term short-term memory cell (LSTM) and a time convolutional network (TCN) architecture. While the heartbeat extraction performed well with an F1 score of 0.803, we found $97,07\,\%$ of the occurring heartbeats. The following sleep stage classifications were not satisfactory, with the best results for a simple Wake/sleep classification. We achieved a balanced accuracy of 0.631 and an F1 score of 0.732 for a CNN architecture. However, the Matthews correlational coefficient (MCC) only reaches 0.203. These results demonstrate the possibility of using the heart rate for sleep staging, but we need additional features, such as respiratory waves or bodymovements, to enhance sleep stage prediction.

Contents

Chapter 1

Introduction

Parkinson's disease (PD) is a neurodegenerative disorder. It is the second most common neurological disorder after Alzheimer's disease [Tol21; Aya23; Lam22]. Worldwide, the World Health Organization (WHO) reported more than 8 million cases in 2019 [WHO24a], with men more likely to develop the disease than women by a ratio of approximately 3:2 [Tol21]. Unfortunately, the number of cases, as well as disability and death due to PD, is increasing rapidly [WHO24a]. In 1990, there were an estimated 2.5 million cases of PD [Aya23]. As age is considered to be the main risk factor for PD, the increase in PD cases may be due to an increased life expectancy of more than 30 years in the recent decades [WHO24b]. Other mentioned factors in the literature include excessive caffeine intake, smoking and exposure to environmental toxins, as well as depression, stress, genetics, trauma, head injury and inheritance. Although several modulating effects have been found, the cause of PD remains unknown, and no cure or prevention has yet been found [Lam22; Aya23; Tol21]. Early diagnosis is desirable not only to slow the progression of PD but also to improve the quality of life for patients, as for now only symptoms can be treated.

Still, there is a relatively good understanding of what happens during the progression of the disease. Physiologically, the dopamine-producing cells in the substantia nigra, a part of the midbrain, degenerate or die. This leads to dysregulation in the basal ganglia pathway, which is essential for controlling movement. Dopamine is particularly important in this part of the brain for smooth and coordinated control of muscle and, therefore, movement [Aya23; Bei14; Bar09; Spr13].

Symptoms of PD only occur when there is a significant amount of cell damage [Aya23; Bei14; Sve16]. These symptoms can be divided into motor and non-motor symptoms. The motor symptoms are the most visible part of the disease. Motor symptoms include tremors, muscle rigidity and postural problems. The non-motor symptoms are diverse and include various sleep

disorders, hallucinations, pain, anxiety, depression, different degrees of cognitive impairment and dementia [Sch17; Tol21; Wei22; Par23; Sve16]. However, both motor and non-motor symptoms overlap with other neurodegenerative diseases, such as Huntington's and Alzheimer's, and increase in severity over time, making clinical diagnosis complex [Lam22; Wat10]. This results in a moderately high error rate of about 15 - 24 % [Tol21]. The clinical routine for diagnosis usually starts with the onset of motor symptoms, however, these symptoms develop relatively late in the progression of the disease [Wei22; Tol21].

Despite the difficulty of clinical diagnosis, non-motor symptoms are known to occur up to 10 years before the development of motor symptoms and recognition of clinical manifestation. It is reported that 80% of PD patients have sleep disturbances either caused by non-motor symptoms or as a result of motor symptoms [Zuz20; Gro20; Sve16]. One such sleep disorder is rapid eye movement (REM) sleep behaviour disorder (RBD). RBD is a parasomnia in which the normal paralysis of REM sleep is lost in such a way that patients appear to act out their dreams [Pos19]. In a meta study by Figorill et al., it is stated that 0.5–1% of the general population over the age of 60 have RBD. Furthermore, isolated RBD (iRBD) is widely recognised as a prodromal manifestation of alpha synucleinopathies, a type of neurodegenerative diseases including PD and dementia, present when the neurodegeneration progress has already begun but cardinal symptoms of the disease have not yet manifested. Approximately 90% of patients with iRBD receive a clinical diagnosis of an alpha-synucleinopathy disease at 15 years of follow-up, namely Lewy body dementia (LBD) in about 45%, PD in 45%, and multiple system atrophy (MSA) in 5% [Fig23].

This implies the possibility of using sleep behaviour disorders for early diagnosis. It is additionally reported that patients with the occurrence of RBD as early non-motor symptoms have a higher likelihood of developing a mild cognitive impairment. This is a condition which greatly deteriorates the progression of PD, increasing its morbidity and rapidly impacts the quality of life [Wat10; Mag21]. Additionally, a higher risk for dysfunctions of the autonomic nervous system, and an increased risk of cognitive decline were shown. RBD also leads to a more severe development of motor dysfunctions [Gro20]. REM sleep is important for procedural memory, motor skill processes and long-term memory as well as emotional memory especially fear conditioning [Sak23; Erl24; Ler21]. However, contradicting findings have been reported stating that REM sleep is not particularly important for motor skill and procedural memory consolidation [Sar17; Con23].

The gold standard for sleep analysis is polysomnography (PSG). A set of sensors and electrodes capture the vital signals of the body and brainwaves during sleep, allowing for accurate detection of different sleep stages and objective assessment of sleep quality. Sleep is divided into two main phases: slow wave (SW) sleep and REM sleep [Pat10; Bar21]. However, PSG is very invasive

and disruptive to sleep. Furthermore, the sensor placement, supervision of the patient during the night, and manual data analysis require trained personnel and a large time expenditure [Hir16]. Therefore, it is poorly suited for long-term monitoring. A reliable, easy-to-use, and especially unobtrusive long-term sleep monitoring method could allow early diagnosis of PD and improve sleep disorders detection. It would also benefit the understanding of sleep and sleep disorders in general, improving overall sleep quality and, therefore, the quality of life [Bar21].

One established method for sleep monitoring is the use of wearable sensors; Miller et al. evaluated six commonly used devices (Apple Watch, Garmin, Polar, Oura (Gen.2), WHOOP (3.0), Somfit). They found an overall high agreement (>90 %) with PSG for sleep-wake and heart rates classification but lack adequate performance for multi-sleep staging and wake states within sleep, only reaching a 50 to 65 % agreement with the PSG data. It is further stated that devices working with heart rate better detect the wake phases within the sleep [Mil22]. Since wearables still require body contact, they can also be obstructive during sleep. A promising alternative is sleep staging with radar sensors, a contactless method ensuring unobstructed sleep to observe normal sleep behaviour and facilitating long-term monitoring [Kag16; Hon18; Pat23]. The radar sensors measure vital signs contactless and use machine learning algorithms, e.g., neural networks for data analysis. In recent studies, the accuracy of radar-based phase sleep staging has been improved from around 57% by Kagawa et al. [Kag16] to 91% by Park et al. [Par24] in two stage prediction. In recent years for multi-sleep staging, reported accuracies ranged from 76 to 89% [Lee24] and 76 to 85% [Par24], while back in 2016 accuracies ranged from 34 to 57% [Kag16]. The reason for this performance increase is the advanced processing of vital signs with deep learning algorithms. Park et al. used a convolutional neural network (CNN) [Par24] and Lee et al. utilised an attention-based long short-term memory (LSTM) model [Lee24].

In this work, we used a $61\,GHz$ continuous wave (CW) radar system to extract vital signs with focus on the heart rate, additionally extracting body movement or respiratory waves. We calculated heart rate variability (HRV) features to use for sleep stage classification with a CNN. The data was recorded during an overnight sleep study at our laboratory using the aforementioned radar system and PSG as ground truth. Using this dataset, the following research questions were addressed:

- How accurate is overnight sleep staging using non-invasive vital signs compared to the gold standard PSG?
- How well do the extracted biosignal features perform with different machine learning algorithms?

Here, the CNN is compared against other deep learning (DL) networks namely time convolutional neural network (TCNN) and LSTM.

This thesis is structured as follows: Chapter 2 covers the medical background of sleep and basics of PSG as well as the sleep disorders in PD. In Chapter 3 the technical background, the basics of the employed radar system and DL algorithms, are explained. Chapter 4 describes the current state of the art in radar sleep analysis. Chapter 5 contains the applied methodology, including a description of the data acquisition, an explanation of the conducted study, the data handling and preparation for the CNN and the network architecture itself. Chapter 6 presents the results of the work and compares them with other machine learning and deep learning networks. Chapter 7 discusses the results. Finally, Chapter 8 offers a conclusion and outlook to possible use cases of the proposed algorithm.

Chapter 2

Medical Background

To emphasise on the medical background and motivation behind this work, we will discuss the different phases of sleep in Section 2.1, highlight the importance of sleep in Section 2.2, talk about the gold standard PSG for sleep analysis in Section 2.3. Besides, we will take a short look at the sleep disorders during Parkinson's disease in Section 2.4.

2.1 Human sleep behaviour

Humans experience different sleep stages or sleep phases during the night. Typically, human sleep consists of one to eight cycles per night [Le 20]. Between these cycles, there commonly are short periods of wakefulness. A typical cycle consists of four to five stages, depending on whether the wake state during the night is regarded as a sleep stage, although wakefulness between a cycle does not always occur. These sleep stages are namely N1 (from wakefulness to sleep), N2 (light sleep) and N3 (deep sleep, slow wave (SW) sleep or delta sleep) and REM sleep (preparation for the return of consciousness) and are associated with specific brain waves. Sleep spindles, a type of fast oscillation in the range of 11 to 16 Hz, occur mainly during N2 and N3. While REM sleep is apart from the eye movements, also associated with the suppression of muscle tonus. Other sleep classification approaches distinguish between three stages (wakefulness, non rapid eye movement (NREM) and REM sleep) or simply two stages (wakefulness and sleep) [Boo17; Bar21; Lew21; Le 20].

During sleep, the heart rate is generally lower than during daytime activity, and the heart rate variability is also more stable. The heart rate reaches its lowest point during SW sleep; as the NREM sleep is associated with the parasympathetic nervous system, the body temperature and blood pressure also decrease. While the REM sleep is associated with the sympathetic nervous

system, the heart rate is rising as well as the body temperature and blood pressure [Bro12; Sej22; Mal18].

2.2 Importance of sleep for psychological and physical health

Recent studies found that in healthy human, all sleep phases and their interactions contribute to normal cognitive functions such as memory consolidation and skill learning [Sak23; Sar17; Bar21].

Sleep is essential for maintaining cognition and healthy brain functions, and NREM and REM sleep has restorative functions, such as removal of metabolic waste products. Disruption of sleep interferes with normal restorative functions, causing a widespread range of psychological and physical malfunctions. These include respiratory and cardiovascular problems, but also changes in emotional reactivity and reduced cognitive performance leading to impairments in attention, memory, and decision making [Lew21; Bro12]. It is to state that the quality of sleep has a high impact on the well-being of humans [Sco21; Fab21; Lew21; Le 20]. Prolonged bad sleep quality can lead to the development and worsening of neurodegenerative diseases [Fig23; Ben19]. Further, diet choices and physical activity impact sleep quality; a bad diet negatively affects sleep quality, while healthy food and physical activity, in general, are positive for sleep quality [Sej22]. Martin et al. reported that during REM sleep dream reports are on average, longer, more vivid, bizarre, emotional and story-like compared to those collected after NREM [Mar20]. However, the exact relationship between sleep phases and their significance for sleep has yet to be fully understood. Nevertheless, during a regular night in healthy humans, there is more NREM sleep at the beginning of the night and more REM sleep at the end of the night [Bar21]. In humans, the stages of sleep are, to some extent, intertwined as the duration of the different stages over a cycle and the night varies depending on the total duration and length of previous sleep stages. The occurrence of the specific sleep stages, how long and often they appear, seems to have a strong influence on the quality and overall duration of sleep, as well as on its effectiveness, i.e. how well rested one feels [Le 20; Lew21; Bar21; Bar23], but also their effect on the cognitive abilities of humans [Scu15].

Barbato states that short sleep durations have a higher percentage of SW sleep, and with an increase in total sleep time, the percentage and duration of REM sleep increased; long sleep periods were perceived of better quality [Bar21]. In particular, this indicates that REM sleep is an important measure of sleep quality, the so-called good night of sleep [Bar21].

Baena et al. found that with age, there is a decrease of sleep spindle slow wave coupling during NREM sleep, which negatively affects the problem-solving skill of humans [Bae24]. Their

findings support that the quantity and quality of sleep are significantly reduced with age and that age-related changes in spindles and slow waves during NREM sleep are associated with a reduced benefit of sleep on sleep-dependent memory processes and consolidation. However, the degree of the preservation of coupling with age correlates with the extent of problem-solving skill consolidation during sleep [Bae24].

It has been shown that sleep is essential for memory and motor skill consolidation [Sti05]; however, which sleep stage is mainly responsible for this phenomenon remains unclear. On the one hand, Sara states that the REM sleep is not particularly important for procedural memory and motor skill consolidation and the processes happening to equal parts in the SW sleep and REM sleep. Instead, the REM sleep is important for the emotional memory and here, especially the fear conditioning [Sar17]. Lerner et al. found that REM sleep impaired recall of the original fear-related memories, but it improved the ability to generalise these memories to novel situations that emphasised the discrimination between threat and safety signals, increasing the sensitivity to the most relevant stimuli previously associated with fear [Ler21]. On the other hand, Sakai reported in a more recent paper that the REM sleep is the most important sleep stage for consolidating new memory in the long term memory as well as the learning of new skills as the long-term synaptic changes happen during REM sleep, and that SW sleep and sleep spindles are responsible for declarative learning as it does not require large amount of synaptic changes [Sak23].

Erlacher et al. also show that after complex motor skill training, the REM sleep density is increased [Erl24]. And Almeida-Filho et al. showed in a rat study that phasic REM sleep supports spatial learning [Alm21], Conessa et al. found that a daytime nap after physical practice, motor imagery, or action observation supports the skill consolidation, and its generalizability toward the inter-manual transfer of skill after action observation. They found that a temporal cluster organisation of sleep spindles underlies motor memory consolidation. Moreover, they state that their findings may have practical impacts on the development of non-physical rehabilitation interventions for patients having to remaster skills following peripherical or brain injury [Con23].

In conclusion it can be said that, the sleep dynamic is complex and needs further research. Depending on age, whether sleeping in the day or nighttime, sleep duration, and what actions and food are taken before sleep affect the sleep structure, restorative effects, and quality of sleep.

2.3 Gold standard of sleep analysis - PSG

The gold standard for sleep analysis is polysomnography (PSG), which has several benefits but also considerable drawbacks. It is used to objectively evaluate sleep, document and investigate

vital functions during sleep by simultaneously and continuously recording neurophysiological signals and vital signs of the human body, such as the brain waves, the heart rate or respiratory waves [Blo97; Jaf10; Bou19]. It enables the study of human sleep behaviour and sleep structure up to a neuronal level by incorporating the assessment of brain waves via the electroencephalography (EEG). Additional sleep-quantifying questionnaires contribute to a more objective sleep research. An PSG is usually performed in a sleep lab with constant monitoring of sleep. Mounting the sensor usually requires between 30 to 60 minutes [Blo97].

Figure 2.1 shows a participant mounted to PSG and depicts the multiple sensors that are placed on the participant. It highlights a major drawback of PSG, which is its obtrusive character that can have a disturbing effect on an individual's sleep. Additionally, a sleep lab poses an unfamiliar environment, which further reduces sleep quality.

Figure 2.1: Example of PSG sensors and placement, the circles indicate electrodes at the shin. The placed sensors are pairs of electromyography (EMG) electrodes on each forearm and shin (the blue circles) as well as on the chin to record jaw movements, electrocardiography (ECG) electrodes for the heartbeat, electrooculography (EOG) electrodes diagonally on the side of the eyes to pick up the eye movements, an EEG cap to measure the brain waves, two straps one for the abdomen and one for the thorax respiration, a nasal tube to measure respiratory pressure and a finger pin for the oxygen level and heartbeat. Finally, there is a microphone at the chin to measure noises such as apnoea during the night and video recording to allow monitoring of the participants from afar.

2.4 Sleep disorders in Parkinson's disease

Sleep disturbances affect around 80 % of all PD patients and are therefore a common non motor symptom. The percentage is increasing with the progression of the disease. The symptoms range from excessive daytime sleepiness, the difficulty to remain alert and wake during the day leading to "sleep attacks", to nocturnal sleep disorders. These nighttime sleep disorders include:

- Insomnia, which is the difficulty initiating, maintaining, consolidating sleep or generating an overall good sleep quality, despite satisfying opportunities for sleep and resulting in daytime impairment.
- RBD a loss of muscle tonus suppression during REM sleep.
- Restless leg syndrome, which is the urge to move the legs usually associated with leg discomfort with a worsening condition during nighttime.
- Obstructive sleep apnoea is the repeated reduction of the airflow during the night due to anatomical obstruction of the airway. The sleep apnoea can be caused by PD itself or the medication.
- General circadian rhythm disorders, which are chronic or recurrent sleep disturbances due to alteration of the circadian system or a misalignment between the endogenous circadian rhythm and socially determined sleep-wake schedules.

These symptoms can occur on their own but also follow after the motor symptoms or medication [Sch17; Zuz20; Mag21; Gro20].

As mentioned in Chapter 1 RBD is especially useful for early diagnosis of PD as it occurs up to 15 years before the development of the disease. Recently, Lahlou found that a lower sleep spindle density during the N3 sleep stage was associated with a worsening of declarative memory consolidation as well a reduced temporal clustering of the spindles for PD patients, hinting to be one reason for worsening cognitive function in the progress of the disease [Lah24].

Chapter 3

Technical Background

In this chapter, the fundamental principles of continuous wave radar will be highlighted in Section 3.1. Further examples for a $61\,GHz$ system will be given that emphasises its suitability for vital sign measurement. Further, the basics of the employed deep learning algorithm for the sleep analysis will be explained with a focus on the CNN in Section 3.2 and Section 3.3.

3.1 Fundamentals of continuous wave radar

Radar sensors are of interest to healthcare professionals as they pose a promising non-invasive method for measuring vital signs [Pat23; Wan14; Yav16; Muñ17].

The wide range of possible applications demonstrates the diagnostic power of contactless measured vital signs. It can be used in nearly all medical fields of healthcare, ranging from elderly care (e.g., fall detection) [Han21] to telemedicine and telemonitoring technique with its ability to monitor general physiological signals, such as respiration, pulse wave propagation, heart sounds, heartbeats or body movements [Ang20; Yav16]. More specific proposed applications are in sleep analysis [Kag16; Hon18], disaster medicine (life detection) [Che00], or cancer therapy [Gu12] but also uses in dermatology are possible, e.g., for patients with compromised skin [Bor16]. Further, radar can be used as a communication method to transmit data from on or in-body sensors to medical personnel for monitoring or diagnostic purposes, one example being the endoscopic capsules [Wan12].

Radar is a sensor technique, more precisely, a localisation technique using electromagnetic waves to determine the angle, velocity and distance of objects relative to the origin of the wave, the antenna [Rah19]. To measure the distance of a target, the time of flight method or round trip time of flight method, as the start and end point are the same, is commonly used for sensor systems

utilising wave propagation.

In radar systems, the electromagnetic wave is sent with an antenna in the direction of the target, which scatters and reflects the wave. Some of the scattered waves are received back by the antenna. Equation 3.1 shows the distance calculation omitted in this way, where c is the wave propagation speed in the medium, and τ is the round trip time of flight [Kra24; Rah19].

$$d = \frac{1}{2}c\tau \tag{3.1}$$

In air, c is the speed of light, $c \approx 3 * 10^8\, ms^{-1}$, when we assume a $20\, cm$ distance to the target, the round trip travel time calculates with $\tau = 2d/c$ to $1,33\, ns$. Further, as equation 3.2 shows, the propagation speed c is embedded in the carrier frequency f_c of the electromagnetic wave that defines the frequency band of the radar and its wavelength.

$$f_c = \frac{c}{\lambda},\ \lambda = \frac{c}{f_c} \tag{3.2}$$

The wavelength is important as it determines the radar system's accuracy, resolution, and applicability. Note that c is determined by the medium and not always constant. The maximum unambiguous range of a radar system is defined as

$$d_{max} = \frac{c}{2f_c} = \frac{\lambda}{2} \tag{3.3}$$

For example, a carrier frequency of $61\, GHz$ with a wavelength of $4,918\, mm$ would lead to $2.459\, mm$ as the maximum unambiguous range. This requires a high sample frequency of the radar to capture movements bigger than $2.459\, mm$, e.g., chest or limb movements [Kra24].

The radar regulation of the International Telecommunication Union (ITU) states approved carrier frequencies for industrial, scientific and medical (ISM) applications. They are in frequency bands from $6.765\, MHz - 246\, GHz$ and shown in Tab. 3.1.

Table 3.1: ITU radio regulation approved frequency bands for ISM applications and centre frequencies [ITU20]

Frequency band	centre frequency
6765 - 6795 kHz	6780 kHz
433.05 - 434.79 MHz	433.92 MHz
61 - 61.5 GHz	61.25 GHz
122 - 123 GHz	122.5 GHz
244 - 246 GHz	245 GHz

In medical application common frequencies are $5.8\,GHz$ [Li18], $24\,GHz$ [Kag16; Han21] and $61\,GHz$ [Wen23]. One, in literature, frequently used radar mode is the CW radar [Wan14; Kag16; Han21; Pat23; Kra24; Yav16; Lee24; Li18], but frequency modulated continuous wave radar [Wan14] or ultra wideband radar [Par24; Yan18] were also utilised in the scope of biomedical research.

Continuous wave (CW) radar or Doppler radar works based on the Doppler principle. Next to pulse-based radar, it is one of the two general methods to employ radar sensors. It has the advantage that it is usually smaller and simpler in design, as well as lighter than pulse-based radar. Further, it transmits lower power and detects at a shorter range than pulse radar [Rah19]. However, noise leaking from transmitter to receiver and its low unambiguous target detection range are its most prominent limitations [Rah19; Kra24].

The Doppler frequency f_d of the continuous wave radar is defined as the difference between the transmitted signal s_{tx} and the receiving signal s_{rx}. The frequency f_d can be expressed as

$$f_d = f_{rx} - f_{tx}, \text{ with } f_{tx} = f_c \text{ or } f_d = 2\frac{v_r}{\lambda} = 2\frac{v_r f_c}{c} \tag{3.4}$$

where v_r is the radial velocity of the target moving from the antenna, and f_c is the transmitted frequency or carrier frequency. The general Doppler frequency can then be denoted as

$$f_d = \pm 2\pi f_c \frac{2\dot{d}}{c} \tag{3.5}$$

with the plus sign for approaching and the minus sign for receding targets, $\dot{d}$ is the travel distance of the target [Rah19].

An CW radar system can be realised with a homodyne receiver, which is depicted in Fig. 3.1. Considering a voltage-controlled oscillator in the homodyne receiver the transmitting signal s_{tx} can be written as

$$s_{tx}(t) = A_{tx}\cos(\omega_c + \varphi_0) \tag{3.6}$$

with the amplitude A_{tx} and the angular frequency $\omega_c = 2\pi f_c$. While the receiving signal is

$$s_{rx}(t) = A_{rx}\cos(\omega_c(t-\tau) + \varphi_0 + \varphi_d) \tag{3.7}$$

which is generally the same as the transmitted signal but with a time delay of τ, an amplitude of $A_{rx} = A_{tx}C$, with C being the scattering property of the material lowers A_{rx}. Further s_{rx} has a phase shift of φ_d influenced by the offset of the radar hardware and the phase shift properties of

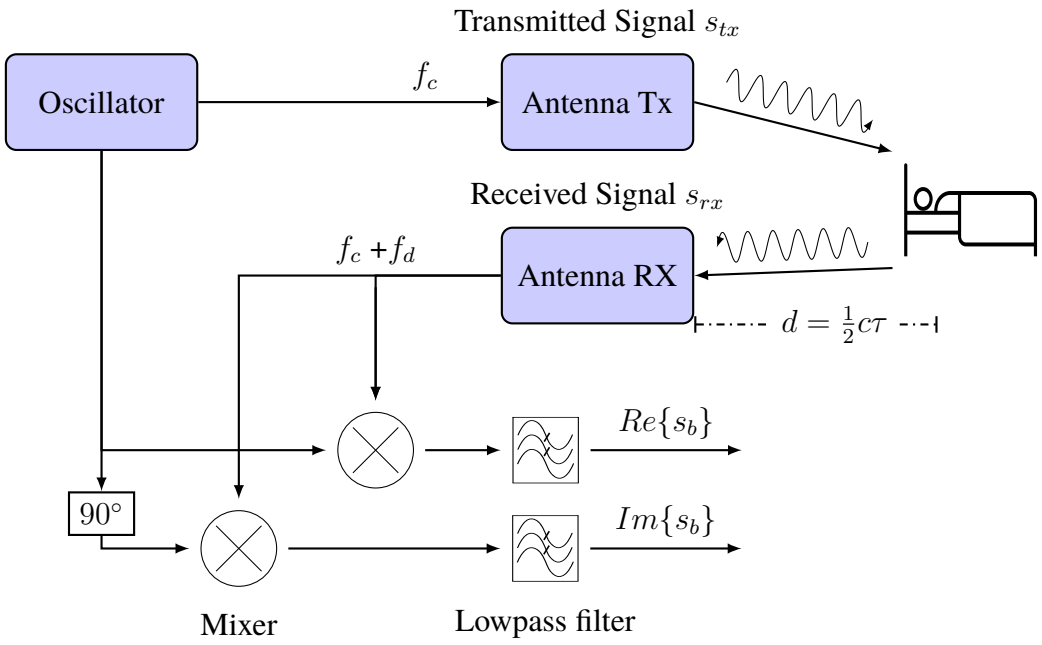

Figure 3.1: Bock diagram of a CW radar system with homodyne receiver, and an oscillator, quadrature mixer and low pass filter.

the target material [Alb24; Kra24; Rah19].

After utilising a quadrature mixer to mix the receiving signal and the transmitted signals together, obtaining in-phase and quadrature (I/Q) components, the retrieved signal is the baseband $s_b(t)$. After applying a low pass filter to remove unwanted signal components at twice the carrier frequency, it can be used to calculate the radial velocity of the target [Alb24; Kra24; Rah19; Muñ17].
The baseband signal s_b can be denoted as

$$\begin{aligned} s_b(t) &= Re\{s_b(t)\} + jIm\{s_b(t)\} \\ &= \frac{1}{2}A_{tx}A_{rx}{\cdot}exp(j(\omega_c\tau{-}\varphi_0)) \\ &= A_b{\cdot}exp(j(\varphi_b)) \end{aligned} \tag{3.8}$$

Where the phase $\varphi_b = \omega_c\tau{-}\varphi_0$ of the s_b is proportional to the delay time τ which is the round trip time of equation 3.1 and A_b is proportional to the power of the signal [Alb24]. This shows that the distance d can be derived from the baseband signal. However, the ambiguity of the phase by 2π and φ_0 has to be considered. It shows that a distance change of half a wavelength rotates the phase by 2π, which also corresponds with equation 3.3. Further, it can be derived that for a frequency

of $60\,GHz$ and, therefore, a wavelength of $5\,mm$, a phase change of 1° corresponds with a $7\,\mu m$ distance change. This enables the detection of micro-movements from the skin, e.g., the pulse wave or heartbeat [Kra24]. Therefore, a CW radar system with $60\,GHz$ and $61\,GHz$ as well is suitable to detect vital signs.

3.2 Fundamentals of convolutional neural networks (CNN)

A CNN is a classical DL algorithm originally developed to analyse local features in images but can also be used on time series; in this case, it is also called 1D convolutional neural networks (1D-CNN) [Zha17; Liu19].

The idea behind CNNs is to recognise patterns in the input data by processing several neighbouring data points at a time. This is implemented by convolving several kernels with trainable weights with the input in each layer and, therefore, enhancing distinct features of the original input. To reduce dimensionality, the convolution layer results are passed through pooling layers in between. Finally, a fully connected layer classifies the convolution layer output [Zha17]. In Fig. 3.2, a basic structure of a CNN is depicted.

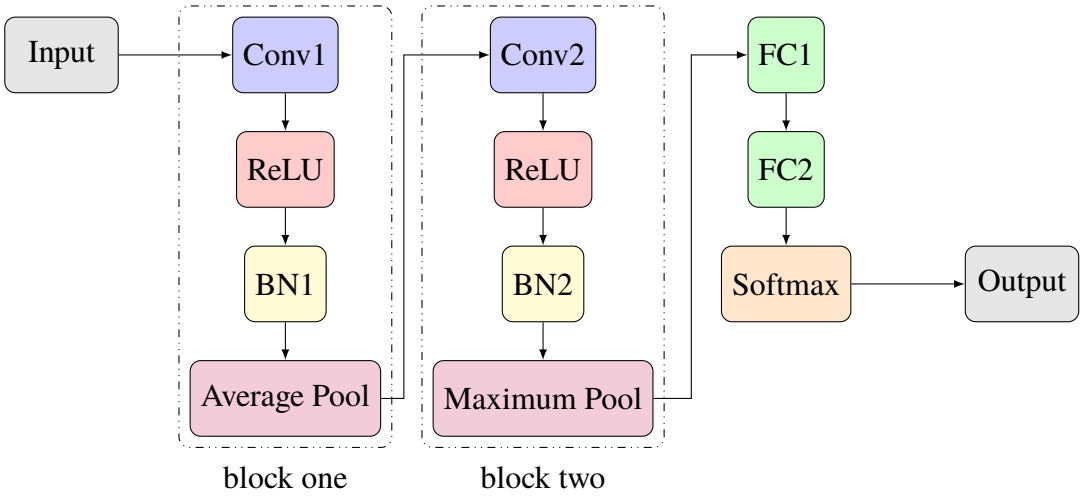

Figure 3.2: Example of a two Bock CNN with two convolutional layers (conv1, conv2), each followed by a rectified linear unit function (ReLU) activation function, an batch norm (BN) layer (BN1, BN2) and a pooling layer. Lastly, there are two fully connected (FC) layers and a softmax function for classification.

A 1D convolution can be written as shown in equation 3.9, where k is the kernel size utilised for the convolution and $\star$ is the valid cross-correlation operator behaving similar to the convolution operator $*$, N is the batch size, C denotes the number of channels, L is the length of the signal

sequence for $i, j \in L$. In the simplest case, the output value of the layer with input size (N, C_{in}, L) has a size of (N, C_{out}, L_{out}) and is

$$\text{out}(N_i, C_{out_j}) = \text{bias}(C_{out_j}) + \sum_{k=0}^{C_{in}-1} \text{weight}(C_{out_j}, k) \star \text{input}(N_i, k) \tag{3.9}$$

L_{out} is calculated with

$$L_{out} = \lfloor \frac{L_{in} + 2 \times \text{padding} - \text{dilation} \times (k-1) - 1}{\text{stride}} + 1 \rfloor \tag{3.10}$$

Here, stride, padding and dilation are hyperparameters characterising the discrete convolution.

- Stride controls the step size by which the kernel should be shifted.
- Padding controls the amount of overlap the convolving kernel over the border of the input data.
- Dilation controls the spacing between the kernel points.

The inclusion of a batch normalisation layer is not always needed. However, it helps to reduce internal covariate shift. In doing so it, accelerates the training of deep neural networks by normalising the mean and variance of the layer input [Ser15]. It further helps the gradient flow through the network by reducing dependency on the initial values or scale of the parameters and can be stated as

$$y = \frac{x - E[x]}{\sqrt[2]{Var[Var[x] + \epsilon}} * \gamma + \beta \tag{3.11}$$

The mean and standard deviation are calculated per dimension over the mini-batches, and γ and β are learnable parameter vectors of size C (where C is the number of features or channels of the input). By default, the elements of γ are set to 1, and the elements of β are set to 0.

The pooling layer is a reduction of input features. The example shown in Fig. 3.2 uses average pooling in the first convolution block and max pooling in the second block. Figure 3.3 shows the working principle of a pooling layer where, in the case of average pooling, the average is used for the next layer, and in the case of max pooling, the maximum value is used.

The fully connected layers are a linear transformation that is applied after a flattening of the last convolutional block, they classify the features globally, and the last FC layer has to be the size of the class numbers that are to predict. The final prediction probability of the classes is usually

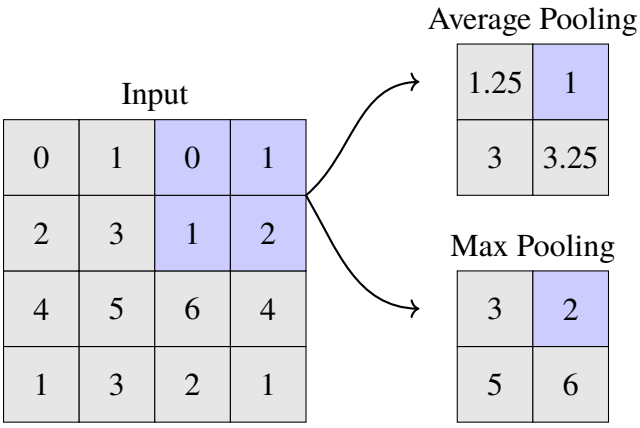

Figure 3.3: Pooling layer example for average and max pooling, with a 2x2 kernel

calculated with a softmax function. The softmax function denotes as

$$Softmax(x_i) = \frac{exp(x_i)}{\sum_j exp(x_j)} \tag{3.12}$$

The equations 3.9, 3.10, 3.11 and 3.12 can be found in the Pytorch documentation [Pyt24].

3.3 Other machine learning algorithms

The Section will give an overview of the DL , which are used to compare the results of the CNN with. Generally machine learning (ML) is allowing the development of algorithms that enable machines to make predictions or identify patterns based on data. It is usually differed between classic ML and DL algorithms. ML algorithms are often simpler in structure than DL approaches, which are commonly called neural networks as they are inspired by the neuron structure in the human brain. DL networks are more general and inherently able to learn features on their own but thus also less understandable, as it can be hard to trace how exactly the results are produced. They are, therefore, often referred to as black box models, which can be challenging in medical applications, where transparency on the decision-making is very important. While classical ML approaches commonly require explicit pre-processing and expert feature engineering, the decision process is more interpretable [Kra24].
Deep learning algorithms used:

- LSTM (long short-term memory): A type of recurrent neural network especially designed to overcome the exploding/vanishing gradient problems that typically arise when learning long-term dependencies. It usually consists of 3 gates: a forget gate, which states how much

information from the previous step is forgotten; an input gate, which imports the information of the current step; and an output gate, which updates information passed to the next step. [Hoc97].

- TCNN/TCN (time convolutional neural network/ time convolutional network): A type of convolutional network that usually has no leaking from future to past, casual convolution, and maps the input to an output of the same length [Bai18]. It is suitable for long sequential data, e.g., language, music or traffic [Lin21] and sensor data from wearables [Ing21]. It uses casual convolutions, dilated convolutions and residual connections to enable a larger receptive field [Lin21].

Chapter 4

Related Work

This Chapter will focus on related research involving radar systems for sleep staging. There are various ways to utilise the gathered radar signals, either by directly feeding them into the classification algorithm [Lee24] or by extracting the vital signs from the radar signal and feeding these into the classification. There is often used a different variety of combinations for vital signs ranging from all possible signals to only one selected few, e.g., heartbeats, respiration and body movements [Kag16; Hon18] to only respiration [Par24] or respiration and body movements [Lau20; Tof20].

Lee et al. feed the raw radar data into the sleep prediction algorithm without previous feature extraction. They used two $60\ GHz$ frequency modulated continuous wave radar sensors, one above and one below the bed, to predict obstructive sleep apnea. They utilising an attention-based bidirectional LSTM consisting of four LSTM and attention layers. Their best results were achieved using both radar sensors and classifying three sleep stages (Wake, NREM and REM) with an overall accuracy of 85.2 % and a four-stage classification (Wake, light, deep and REM sleep) with an overall accuracy of 80.3 % [Lee24].

Kagawa et al. used two $24\,GHz$ microwave radar sensors placed under the mattress to extract the heart rate (HR), heart rate variability (HRV), body movements, and respiratory signals during sleep. They fed these features in a canonical discriminant analysis for the sleep staging resulting in the following accuracies: 66.4 % for identifying Wake and sleep, 57.1 % for three stages (Wake, REM, and NREM) and 34 % for four stages (Wake, light, deep and REM sleep) [Kag16].

Hong et al. as well used heart rate, respiration and body movements extracted by a low-power CW Doppler radar with a frequency of $2.4\ GHz$, with the radar positioned above the bed. They were comparing boost tree and bagged decision tree with a subspace K-nearest neighbour algorithm; the last one outperforms the other classifiers with a highest accuracy of up to 86.6 % for four

sleep stage classification (Wake, light, deep and REM sleep) with class accuracies of 75.8 % (light), 89.1 % (deep), 87.0 % (REM) and 82.0 % (Wake), boosted trees achieving 86.4 % highest accuracy and bagged trees had a highest accuracy of 85.7 % [Hon18].

Park et al. utilised ultra wideband radar connected to a smartphone. They applied a 1D-CNN to extract the respiratory features from the resulting Doppler map. Finally, a four-layer multi-head attention transformer block was used to classify the features, finding that epoch-by-epoch comparison between the predicted and expert annotated four sleep stages (Wake, light, deep and REM sleep) resulted in 76 % accuracy and Cohen's kappa of 0.64, with class-wise sensitivities of 50.3 % (deep), 76.4 % (light), 88.2 % (REM), and 80.7 % (Wake). For three classes, they achieved an accuracy of 85.8 % and Cohen's kappa of 0.735 with a sensitivity of 86.3 % (NREM), 85.4 % (REM), and 83.1 % (Wake) [Par24].

Lauteslager et al. used a Circadia (Circadia Technologies Ltd., London, United Kingdom) Contactless Breathing Monitor C100 that utilises pulsed ultra wideband radar to monitor the respiration and body movement of patients during the night as a possible home monitoring device. In their study, the C100 outperforms the medical grade wrist-worn inertia measurement unit (IMU) based monitoring devices by Philips (Philips, Amsterdam, Netherlands) or Fitbit (Google Fitbit, Mountain View, USA), having a true positive rate of 75 %, 59.9 %, 74,8 % and 57,1 % for deep, light, REM and Wake classification [Lau20].

Toften et al. performing a clinical sleep study using Somnofy (Vitalthings AS, Trondheim, Norway) an impulse radio ultra-wideband radar sensor with a center frequency of $7.3\,GHz$, a bandwith of $1.4\,GHz$ and sampling rate of $23.3\,GHz$, and the Somnofy sleep staging algorithm version 1.0, which uses non-causal temporal neural networks like a TCNN and LSTM recurrent neural networks (RNN) that are fed with respiration and movement data from the radar [Tof20]. Toften et al. achieved high Epoch-by-epoch accuracy in a clinical validation study for young adults. With 76 % accuracy and a Cohen's kappa of 0.63 against a PSG accuracy of 88 % and Cohen's kappa of 0.82. The sleep stage differentiation for Somnofy was 0.75 for N1/N2, 0.74 for N3 and 0.78 for REM, and PSG scorers had an agreement of 0.83 for N1/N2, 0.92 for N3, 0.96 for REM and 0.91 for Wake [Tof20].

However, besides these, there are multiple approaches for ML-based sleep staging that utilise a broad range of input signals and features. For instance, non contactless measured raw EEG data geather with during PSG recording was used for CNN-based sleep staging [Zhu20; Mas24; Par23] or random forest (RF)-based sleep classification [Coo19]. In the scope of PD, Parajuli et al. used a CNN to detect mild cognitive impairment from PD patients with EEG signals, achieving high differential accuracy between normal and impaired cognition [Par23]. There is also literature

using wearable systems to detect vital signs for the sleep stages classification mainly by employing HRV through the sleep phases with CNN classification [Hab23; Wan22; Mal18].

Chapter 5

Methods

The chapter describes the performed sleep study in Section 5.1 and further contains a description of the dataset used in this work. In Section 5.2, the feature extraction from the collected data will be covered, while Section 5.3 highlights the feature preprocessing for the CNN. Section 5.4 states the training pipeline and Section 5.5 describes the models (CNN, LSTM,TCNN). The hyperparameter search values are shown in Section 5.6.

The implementation of this work was done in Python using the *Scikit-learn* and *NeuroKit* packages [Ped11; Mak21] for feature extraction, the *PyTorch* package [Pyt24] for model implementation.

5.1 Data acquisition

To acquire sleep data, a study was conducted in cooperation with the Department of Neurology of the University Hospital Erlangen (Medicine 1). The control group was recorded in the Machine Learning and Data Analytics Lab (MaD Lab).

The overnight recording included a PSG as ground truth, performed with a Somno medic (SOMNOmedics AG, Randersackerl, Germany) *SOMNO HD eco* system whose components are illustrated in Fig. 5.1. Further, the experimental setup included four radar sensors to measure the vital signals, arranged as shown in Fig. 5.2. The sensors are placed in an array over the width of the mattress under the bed in the chest area. The study used CW radar sensors developed within the ”Empatho-Kinaesthetic Sensory Systems” (Empkins) collaborative research center at the Technical University Hamburg (TUH) in the chair of Prof. Dr. Kölpin, the institute of High-Frequency Technology. They are utilising a carrier frequency of $61\,GHz$ and sampling rate of $1953.125\,Hz$. Their structure and working principle are described in detail in [Alb24].

Figure 5.1: An overview of the different components of the utilised *SOMNO HD eco* PSG system: [1] somno hub, [2] leg EMG electrodes, [3] arm and chin EMG electrodes, [4] ECG electrodes, [5] EOG electrodes, [6] Hub for EEG and EOG, [7] EEG electrode cap, [8] night camera, [9] nose tubes, [10] abdomen and torso straps, [11] finger pin, [12] microphone

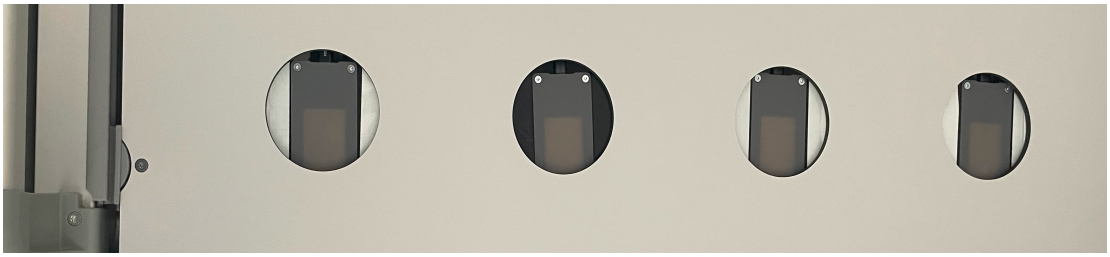

Figure 5.2: Radar sensor placement under the bed, from left to right sensor 1,2,3 and 4

As ground truth labels, we used Wake, N1, N2, N3, and REM, acquired from the automatic *Somno eco* labelling algorithm. For every $30s$ epoch, one label is assigned by the Somnos algorithm, following the guidelines of the American Academy of Sleep Medicine (AASM) for sleep stage classification.

Prior to the data recording, the participants were asked to fill out four self-assessment questionnaires regarding sleep quality:

1. The Pittsburgh sleep quality index (PSQI), 26 items
2. Stop-Bang to asses for obstructive sleep apnoea, seven items
3. RBD Screening questionnaire, 13 items

4. SF-12 Health Survey to overall evaluate the physic and psychic state of the participants, 12 items

Resulting in a total of 58 questions, example prints of the used questionnaires are included in Appendix A.

5.1.1 Dataset description

In the control group, data from 44 participants was collected, with 16 male and 28 female participants. Of the total 44 participants, three did not give permission to use their demographic data. Further, 15 participants were excluded from the study data due to technical issues related to the recording of the ground truth or related to the radar sensors.
Of those, in seven cases there were the following issues during the recording of ground truth data during the night:

- In three instances EEG signals were incomplete due to the disconnection of cables.
- For two instances poor EEG electrode connection occurred.
- In one instance the EOG electrode detached.
- For one instance the video signal was lost, preventing the determination of sleep termination.
- For one instance the automatic ground truth labelling was absent.

The recorded radar data was insufficient a total of six times:

- In two instances due to unexpected technical issues of the recording laptop.
- For four instances due to missing synchronization signals.

Lastly, one participant withdrew from the study during the night.

After excluding the erroneous recordings from the dataset, 29 participants remained. The statistical characteristics of the remaining cohort are listed in table 5.1, while table 5.2 shows the statistical description of the sleep durations.

Table 5.1: Statistical description of the control group demographics after excluding bad data

	Age [years]	Weight [kg]	Height [cm]		Gender
Mean	38,51	70,46	172,18	Female	17
Std	16,16	12,06	10,40	Male	12
Min	18	49	152		
Max	77	105	197		

Table 5.2: Statistical description of the recorded sleep durations

	Time [hh:mm:ss]
Total Time	223:42:54
Mean	07:42:54
Std	00:55:41
Min	05:50:30
Max	09:15:30

5.2 Feature extraction

In the initial step, the heartbeats were extracted from the recorded radar data using the I and Q channels. The radar data was synchronized with the ECG data from the PSG to ensure the time ranges matched. The ECG signal was resampled to the radar sample rate of $1953.125\,Hz$. For synchronization the four radars nodes with each other the *emrad-io*[1] package was utilised. A previously developed and validated LSTM-based model, deployed in the *empkins-micro* package[2] was applied after the synchronisation to predict the heartbeats. This model provides a probability function indcation the likelihood of the heartbeats occurring in the radar signals. The bi-directional LSTM is applied individually to the recorded streams of each radar, after which they are summarised to dampen noise in the LSTM output and for a more stable beat prediction. An example of the individual LSTM probability for each radar is shown in Fig. 5.3, illustrating not every radar node captures each heartbeat consistently.

The radar data was fed into the LSTM in $30s$ batches. The final heartbeats were extracted using peak detection, employing the $\mathrm{find\_peaks}$ method from the Python package *SciPy* [Vir20]. As hyperparameters for this peak detection function, the height parameter was set to 0.24, and the distance to $0.3 *$ radar sample frequency respectively. These values were chosen after a grid search containing different parameters, shown in Section 6.1.

Fig. 5.4 shows an example of the summarized LSTM probability outputs and ground truth. For comparison two height thresholds (0.32 and $0.0.26$) and the respectively found heartbeats

[1] https://github.com/empkins/empkins-io/tree/main/empkins_io/sync
[2] https://github.com/empkins/empkins-micro/tree/main/empkins_micro/emrad

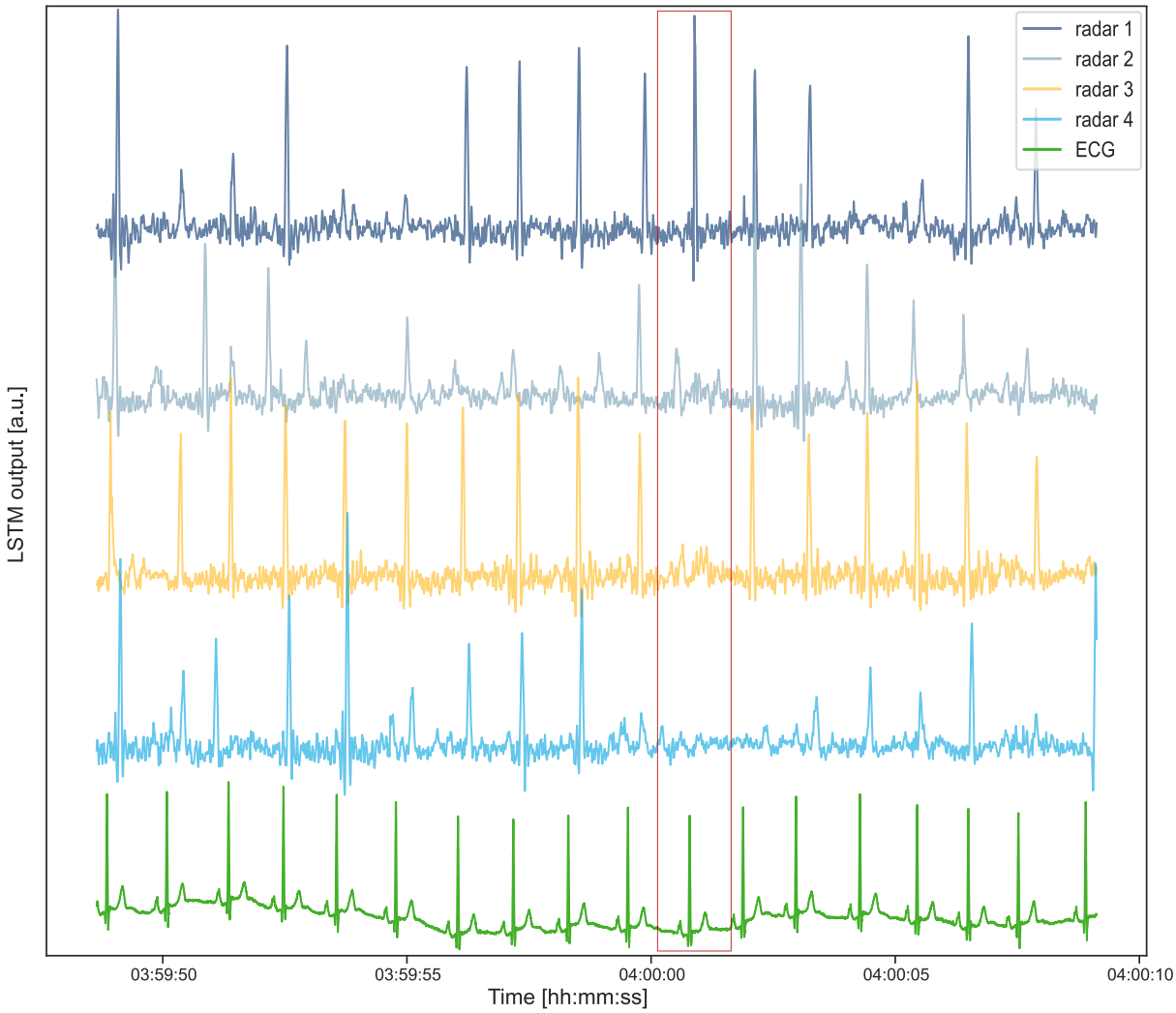

Figure 5.3: Example of the LSTM probability output for each radar sensor 1 (blue), 2 (light blue), 3 (yellow), and 4 (turquoise) and the reference ECG signal (green); e.g., radar 3 captures almost every heartbeat expect one shortly after four o clock (red rectangle), but radar 1 captures the heartbeat at this position.

in the LSTM probability are depicted. It shows a lower threshold will lead to more correctly detected heartbeats, but also results in more false positives. The performed grid search contained the following pipeline: The true heartbeats were extracted using the *BioPsyKit* package [Ric21]. Using the corresponding time stamps of the raw ECG and beats position to compare them with the beat positions calculated with the find_beats method from LSTM heartbeat prediction. Tested were the values $0.24, 0.26, 0.28, 0.30, 0.32$ for a selection of 5 participants (VP: 2,6,10,28, and 42), from the beginning, middle and end of the control group recordings.

The obtained heartbeats are used to calculate the HRV features to predict the sleep stages. The

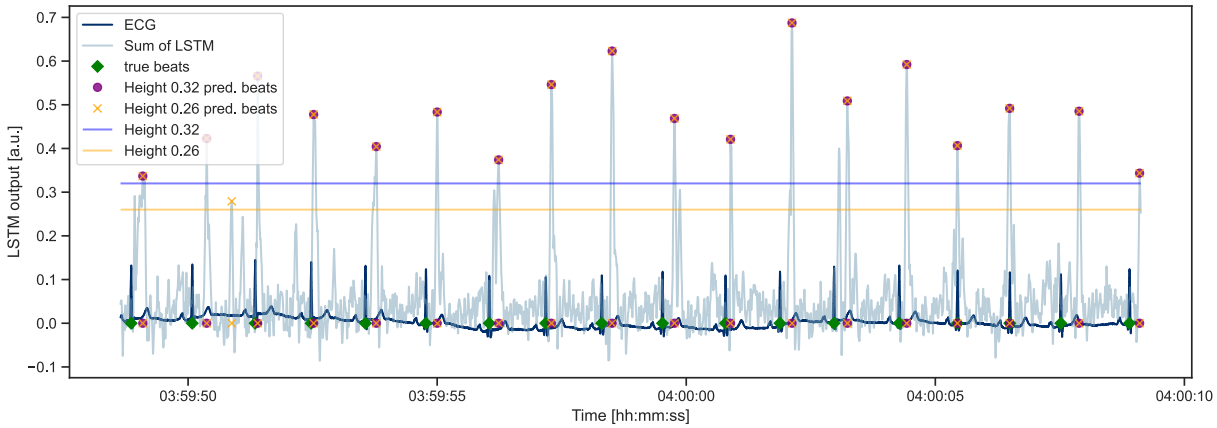

Figure 5.4: $20s$ example of the LSTM output in (light blue) with the found heartbeats in (purple, orange circle) for the heights 0.32 and 0.26 from the find_peaks against the ground truth ECG in (dark blue), the true beats are marked with green diamonds. The predicted heartbeats are shown in both the LSTM output and ECG for easier comparison. Lastly, the blue and orange lines represent the height value for included beats.

HRV features were calculated using the hrv method from the Python package *NeuroKit2* [Mak21; Pha21] a biosignal processing package, calculating a range of time-domain, frequency-domain, and non-linear HRV features. The hrv method uses the heartbeat position in the time frame and their sampling rate to calculate the HRV indices by utilising the beat-to-beat (RR) intervals. The extracted heartbeats are windowed with a sliding window of different sizes with a $30s$ step to match the labelling interval of the ground truth labels of the sleep stages. For each window, the HRV features were calculated with the hrv method, which produces all available features in *NeuroKit*. Depending on the length of the input window, some features are more reliable than others. However, in the *NeuroKit* documentation paper, Pham et al. recommend the use of time and frequency domain features for shorter window sizes as they produce more reliable HRV indices compared to the non-linear features, which usually require longer window sizes (up to 24 h) [Pha21]. Possible features from the time-domain, frequency-domain, and non-linear indices are detailed in the package documentation [Mak21].

The window sizes of $60s$, $180s$ and $360s$ were tested, resulting in three different numbers of feature channels from all 94 features that can be calculated by the hrv method, as the windows are too short for some of the time- and frequency-domain HRV features as well as non-linear indices provided by the hrv method. The window size of $60s$ resulted in 74 usable features, the $180s$ window resulted in 84 features and the $360s$ window resulted in 88 total features.

Listed in the table 5.3 are the possible time- and frequency-domain indices.

Table 5.3: Feature table with all possible HRV features in the time- and frequency-domain. The marked features are used for training the DL networks. The features marked with an x occur in all three window sizes, those marked with y only in the window sizes $180s$ and $360s$.

Feature	Description – Time-Domain	
MeanNN	The mean of the RR intervals.	x
SDNN	The standard deviation of the RR intervals.	x
SDANN1, SDANN2, SDANN5	The standard deviation of average RR intervals extracted from n-minute segments of time series data (1, 2 and 5 by default). Note that these indices require a minimal duration of signal to be computed (3, 6 and 15 minutes, respectively) and will be silently skipped if the data provided is too short.	
SDNNI1, SDNNI2, SDNNI5	The mean of the standard deviations of RR intervals extracted from n-minute segments of time series data (1, 2 and 5 by default). Note that these indices require a minimal duration of signal to be computed (3, 6 and 15 minutes respectively) and will be silently skipped if the data provided is too short.	
RMSSD	The square root of the mean of the squared successive differences between adjacent RR intervals. It is equivalent (although on another scale) to SD1, and therefore it is redundant to report correlations with both.	
SDSD	The standard deviation of the successive differences between RR intervals.	x
CVNN	The standard deviation of the RR intervals (SDNN) divided by the mean of the RR intervals (MeanNN).	
CVSD	The root mean square of successive differences (RMSSD) divided by the mean of the RR intervals (MeanNN).	
MedianNN	The median of the RR intervals.	
MadNN	The median absolute deviation of the RR intervals.	
MCVNN	The median absolute deviation of the RR intervals (MadNN) divided by the median of the RR intervals (MedianNN).	
IQRNN	The interquartile range (IQR) of the RR intervals.	
SDRMSSD	SDNN / RMSSD, a time-domain equivalent for the low Frequency-to-High Frequency (LF/HF) Ratio.	
Prc20NN	The 20th percentile of the RR intervals.	
Prc80NN	The 80th percentile of the RR intervals.	

Table 5.3: continued

Feature	Description – Time-Domain	
pNN50	The proportion of RR intervals greater than 50ms, out of the total number of RR intervals.	
pNN20	The proportion of RR intervals greater than 20ms, out of the total number of RR intervals.	
MinNN	The minimum of the RR intervals.	
MaxNN	The maximum of the RR intervals.	
TINN	A geometrical parameter of the HRV, or more specifically, the baseline width of the RR intervals distribution obtained by triangular interpolation, where the error of least squares determines the triangle. It is an approximation of the RR interval distribution.	x
HTI	The HRV triangular index, measuring the total number of RR intervals divided by the height of the RR intervals histogram.	
Feature	**Description – Frequency-Domain**	
ULF	The spectral power of ultra-low frequencies (by default, .0 to .0033 Hz). Very long signals are required for this indice to be extracted; otherwise, it will return NaN.	
VLF	The spectral power of very low frequencies (by default, .0033 to .04 Hz).	y
LF	The spectral power of low frequencies (by default, .04 to .15 Hz).	x
HF	The spectral power of high frequencies (by default, .15 to .4 Hz).	x
VHF	The spectral power of very high frequencies (by default, .4 to .5 Hz).	x
TP	The total spectral power.	x
LFHF	The ratio obtained by dividing the low-frequency power by the high-frequency power.	
LFn	The normalised low-frequency, obtained by dividing the low-frequency power by the total power.	
HFn	The normalised high-frequency, obtained by dividing the low-frequency power by the total power.	
LnHF	The log-transformed HF.	

The features marked in the table 5.3 are basic features that are contained in all window sizes and were used for the DL networks to train, except the VLF which only occurs in window size $180s$ and $360s$. A detailed table with five exemplary entries of all features for the $60s$ window size

is shown in Appendix B.

5.3 Feature preparation

In the first data preparation step the label and HRV features were aligned to the same $30s$ epoch. The PSG labels were converted into numerical values for the classification purposes: Wake, N1, N2, N3 and REM were mapped to 0,1,2,3, and 4 for five-stage classification, three and two-stage/ binary classification accordingly, while binary or 2 stage classification just compares between wake and sleep epochs. Lastly, the time points labelled as artefacts in the ground truth labels were dropped in both the HRV features and the ground truth.

Additionally, the HRV features which range from 10^3 to 10^{-2}, were scaled to prevent gradient problems for the CNN and other DL algorithms. The scaling was performed utilizing the $\mathrm{StandardScaler}$ method from the *Scikit-learn* package [Ped11] and applied in a way to scale the dataset based on the selected test data. These steps are performed for each subject individually, except the scaling, which was performed over the whole dataset.

After scaling a further sliding window was placed over the cleaned data to transform the data accommodating the CNN input shape. The middle label of this window is located in the centre of the sliding window, with the exception of the end points; here the middle label slides to the start and end points. Depending on the architecture, different window sizes between five and 15 minutes were tested, resulting in containing 10, 15, 20, and 30 samples were used. Lastly, the data was transformed into a torch tensor and reshaped form $[N, L, C]$ into $[N, C, L]$. N is the total recorded time for subject (recording length of data or batch size), C are the number of HRV feature channels used, and L is the sample size of the sliding window (CNN window). Here, L is the convolving dimension.

5.4 Training and testing

For the model training, the dataset was split into a train and test set using a $80/20$ split. The data split was performed between subjects to prevent data leaking from training to testing.

Further, the training data was split with a three-fold group k-fold cross-validation into a testing and validation set, ensuring that a subject does not appear in both the validation and training set simultaneously. During the training, the data was randomly loaded from the subjects in the training set, while during validation the data was loaded sequentially into the model, to mimic the real world data scenario. As a loss function, the weighted $\mathrm{Cross\ entropy\ loss}$ was employed, which

eliminates the need to apply a softmax function to the network output and one-hot-encoding of the labels as the *torch* method handles this automatically. For binary classification, a weighted $\mathrm{BCEWithLogitsLoss}$ was applied, which applies a sigmoid and binary cross entropy loss on the logits of the network. The chosen loss functions are commonly used in ML and reported good results for the applied use case. The class weights were established by dividing the most frequent occurring class with the number of other classes to address the class imbalance.

For scoring the validation and test performance and loss, the *Scikit-learn* package was employed using the $\mathrm{matthews\_corrcoef}$ method as performance values [Ped11]. The Matthews correlation coefficient (MCC) score is used in ML as a measure of the quality for binary and multi-class classifications. It takes into account true positives and false positives as well as the true negatives and false negatives. Generally, it is regarded as a balanced measure that can be used even if the classes are of very different sizes. The MCC, in essence, is a value between -1 and $+1$, where $+1$ represents a perfect prediction, 0 an average random prediction and -1 represents an inverse prediction [Ped11]. Further scores calculated with the *sk_metrics* library from *Scikit* package were accuracy, balanced accuracy, precision, recall, F1 score, cohen's kappa, and specificity. Additionally the confusion matrix for the tested classes were calculated.

The early stopping criteria could be reached after at least 60 epochs with either no improvement of the loss, a performance of 0.0 or a negative development of the loss of 0.03. For testing and validation, the data was loaded sequentially into the ML algorithms.

For the CNN, the different sliding window sizes of 10, 15, 20 and 30 were included as hyperparameters in model. The different window sizes for the calculated HRV features were tested for the proposed CNN architectures in Section 5.5.1, as well for the LSTM and TCNN model. The models were tested for three different classification granualities of sleep stages after the AASM classification binary (Wake/sleep), three stage (Wake/REM/NREM) and five stage (Wake/N1/N2/N3/REM), and the three diffrent HRV features window sizes ($60s$, $180s$, and $360s$).

5.5 Model description

This Section describes the general model structures for the DL algorithms CNN, LSTM, and TCNN. It describes each model on an simple parameter configuration.

5.5.1 1D-CNN

This Subsection introduces the network architecture developed within this work. A 1D-CNN was tested with two-block CNN with a layer structure as shown in Section 3.2.

Figure 5.5 shows the network architecture used in this work. This baseline model was proposed as the work of Zhao et al. and showed promising results [Zha17]. They stated a good accuracy and good signal-to-noise ratio for 1D-CNN's. The model used in this work has the addition of a batch norm layer to have a more stable training, and a Dropout layer as a measurement against overfitting and for improved generalisation of the network. The Dropout layer as well reduces dependency on specific neurons by randomly dropping a percentage of them during the training.

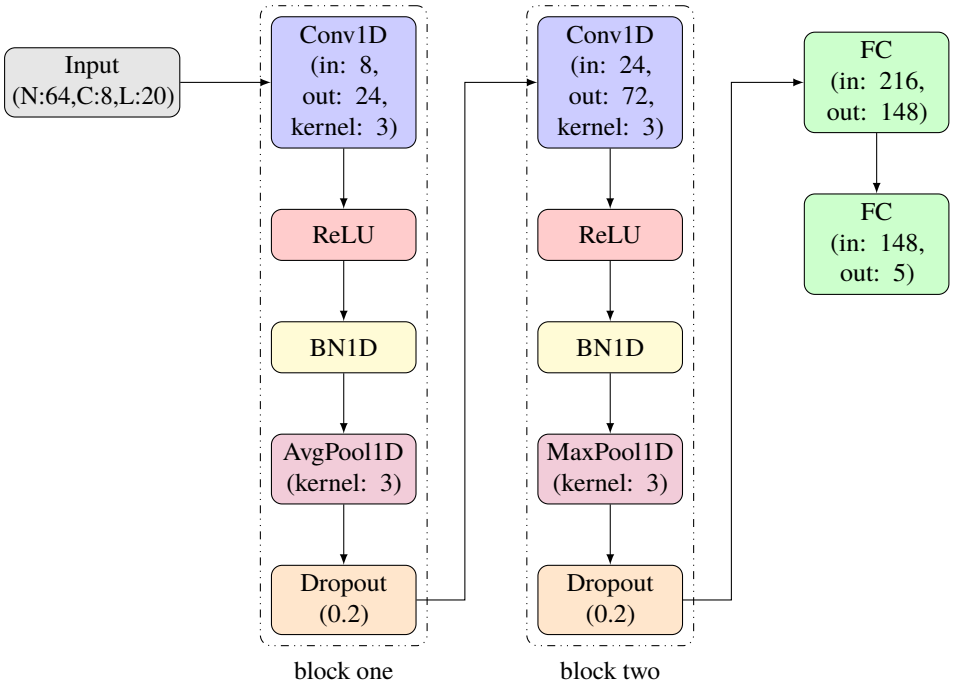

Figure 5.5: Example of a two-block CNN model architecture showing the input and output sizes (in, out) of the layers and the kernel size of each convolution, pooling, dropout and FC layer. The general input size of the model is determined by N: the batch size, C: the number of HRV features and L: the size of the CNN window, in this case, 20. The final FC layer has the class numbers as output, here 5.

Further, a model with four convolution blocks was tested; the general structure is shown in Fig. 5.6. While the architecture is generally the similar to the previously stated two block 1D-CNN, the output layers from block 2 and 3 with an are combined with each other to preserve dimensions for the fourth block of the CNN. The final FC layers has the same size as the smaller network.

5.5.2 LSTM

The provided LSTM network architecture, is shown in Figure 5.7. The network consist of a memory cell accepting the data sequence, the cell state (cn) and hidden state (hn), which are

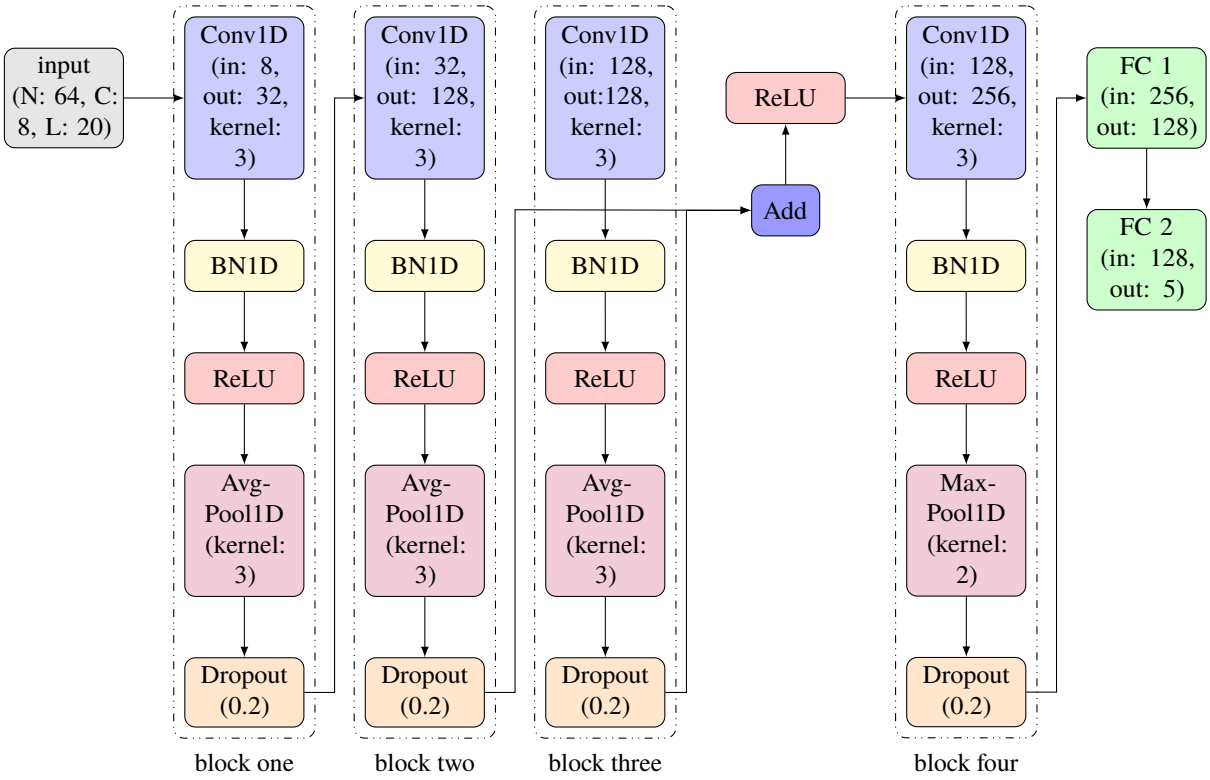

Figure 5.6: Structure of a four-block CNN, following with the input and output sizes (in, out) and the kernel size of each convolution layer, pooling layer and dropout values. Shown is an example input, with a batch size (N) of 64, a feature channel size (C) of 8, and a length of sliding window (L) of 20, a kernel size of 3 for the average and 2 for the max pooling layer, the dropout value is 20 %, and the class number is 5.

updated after each layer of the LSTM. The final hidden state output is then passed through a ReLU function to introduce non-linearities and two dense layers for the final classification of the sleep phases.

5.5.3 TCNN/ TCN

The provided TCNN architecture is shown in Fig. 5.8. It shows the model structure for a temporal block. Each block contains two sequences of one 1D-CNN layer, a chomp layer for trimming the input tensor to remove a number of elements from the last dimension followed by a ReLU activation and dropout layer. The output of the last sequence is added with a downsampled input to the temporal block itself and fed into a ReLU activation function to be the input for to the next temporal block or a FC layer for the final classification of the sleep stages. The number of layers is determined by the channel size and hidden number.

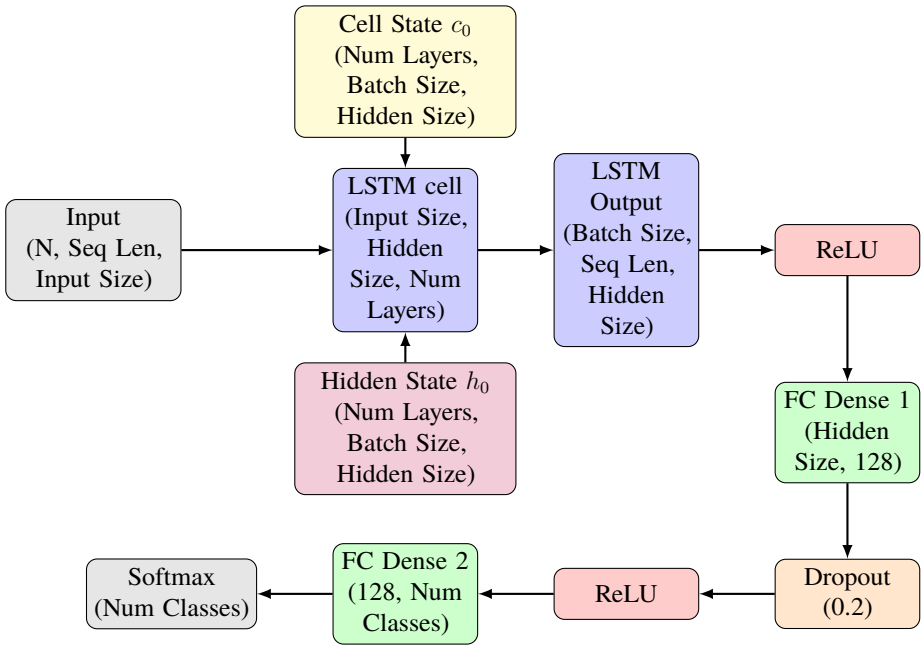

Figure 5.7: Structure of the provided LSTM model, for one LSTM layer, hidden state (hn) is and cell state are initialised with $h0$, $c0$ the states passing through the cell and updated within. For final class prediction the LSTM output is fed into two dense layers (FC Dense 1,2) between two ReLU and dropout with a value of 0.2. Finally, a softmax function gives the final class probabilities.

5.6 Hyperparameter search

A grid search was employed to find the best set of hyperparameters for the respective models. The best set of hyperparamter was chosen Pareto front between from the lowest mean and standard deviation (Std) validation loss. This gives a range of values that have the lowest mean and Std loss helping to chose a value minimising the two criteria. The best model was finally taken from the best performing fold of the best parameter set.

For the CNN, the hyperparameter search was performed with the *sweep* library[3] containing in the *wandb* package [Bie20]. The search contained the different sliding window sizes for the CNN time sample, the batch size, the kernel size of both the layer and batch norm, the kernel application time and the learning rate. For each HRV sliding window ($60s$, $180s$, and $360s$) and class (binary, three-stage, and five-stage) a new grid search was performed, resulting in 9 sets of hyperparameter for each model.

The hyperparameter search for the LSTM and temporal convolutional network (TCN) was performed with *optuna* library containing in *tcpc* package [Küd23]. For these models the best

[3] https://docs.wandb.ai/ref/python/sweep

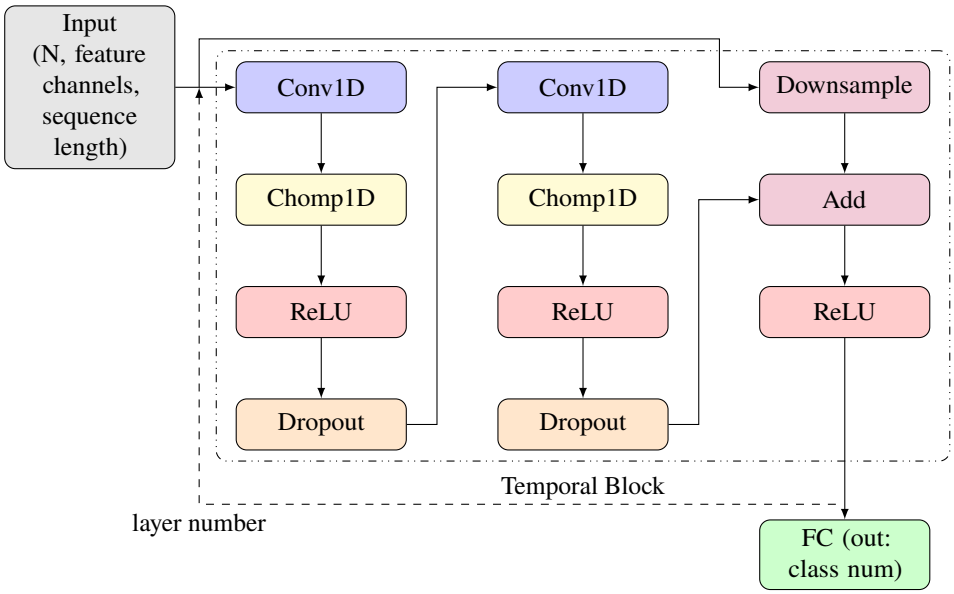

Figure 5.8: Structure of the TCNN model with one temporal block and a FC layer for final classification. The number of temporal blocks is dependent on the channel number. Each temporal block contains two 1DCNN layers (Conv1D), followed by a trimming layer (Chomp1D), an ReLU activation and a dropout layer. The output of the second convolution block is added with a downsampled input of the temporal block and passed on to an ReLU and finally into the FC layer for the final prediction.

parameter set was chosen after the best MCC performance in the during validation. The LSTM, the hyperparameter search included different hidden layer sizes, number of layers, sequence length, learning rates and batch sizes. While the TCNN, search included parameters for the channel number, the hidden layers, kernel and batch size, dropout values and learning rate.

The table 5.4 list the possible combination sets of hyperparameters used for the models.

Table 5.4: Hyerparameter sets for the models employed in this work (CNN, LSTM, and TCNN)

Model	Parameters	Values
two block CNN	window ccn	10, 15, 20,
	max pooling kernel size	2, 3
	average pooling kernel size	2, 3
	kern application	3, 4, 5
	learning rate	0.001, 0.0005
four block CNN	window ccn	15, 20, 30
	max pooling kernel size	2, 3
	average pooling kernel size	2,3
	kern application	3, 4,5
	learning rate	0.001, 0.0005
LSTM	hidden size	range(4, 600) step = 4
	number layers	1, 5
	Sequence length	21, 51, 101
	learning rate	range(0.0001, 0.005)
	batch size	32, 64, 128, 256, 512
TCN	number channels	2, 6
	hidden layers	8, 512, 32
	kernel size	2, 5
	dropout	range(0.1, 0.5), step=0.1
	learning rate	0.0001, 1
	batch size	64, 128, 256

Chapter 6

Results

This chapter provides the results of the methods mentioned tests, following the structure of the previous chapter. First, we see the grid search results for the optimal threshold to classify the heartbeats based on the probability distribution of the bi-directional LSTM in Section 6.1. Section 6.2 shows the results of the sleep staging for the models, the found hyperparameter sets for the respective class granularities and different window sizes that were used to compute the HRV. Section 6.3 compares the CNN, TCN and LSTM results with each other.

6.1 Heartbeat prediction

Five different values $(0.24, 0.26, 0.28, 0.30, 0.32)$ were tested to find an optimal parameter for the heartbeat classification. For the decision of a good threshold, the f1 score was used, representing the harmonic mean between precision and recall. It ranges from one to zero, with one being the best score, and it can be calculated with the formula

$$f1 = \frac{2tp}{(2tp) + fp + fn} \tag{6.1}$$

The precision is the ratio between $tp/(tp + fp)$, where tp is the number of true positives and fp is the number of false positive values. It represents the classification performance with respect to the wrong classification of extra heartbeats. The recall is the ratio between $tp/(tp + fn)$, where fn is the number of false negative labels. It is the ability to find all heartbeats within the signal [sci24]. The overall scores are shown in Tab. 6.1, while the individual scores for the selected participants are in Appendix C.

The best overall F1 score reveals a threshold of 0.24 as the best performing parameter for the

Table 6.1: Overall Comparison of F1, precision and recall scores for different height thresholds, Tt: Total True ECG Beats, Tpb: Total predicted beats, Tp: True positives, Fn: False negatives, Fp: False positives

Threshold	Tt	Tpb	Tp	Fp	Fn	F1	Precision	Recall
0.24	146175	141902	115798	26104	30377	0.803	0.816	0.792
0.26	146175	137904	113349	24555	32826	0.798	0.821	0.775
0.28	146175	133449	110312	23137	35863	0.789	0.826	0.754
0.30	146175	128589	106721	21868	39454	0.776	0.829	0.730
0.32	146175	123108	102448	20660	43727	0.760	0.832	0.700

heartbeat classification. This involves a trade-off between detecting more overall heartbeats and increasing the number of false positive beats. This reduces precision, opposing the detection of fewer heartbeats and false positives, resulting in improved precision. However, this trade-off is balanced by an increase in recall, as more otherwise false negative beats are found with a lower threshold. The threshold of 0.24 has a f1 score of $80,39\,\%$ and results in $97,07\,\%$ of all heartbeats of which $81,60\,\%$ are correctly detected and $18,39\,\%$ are falsely detected beats. In comparison, the f1 score for a threshold level of 0.32 is $76\,\%$ and results in only $84,21\,\%$ of all beats with a higher rate of true positive heartbeats, $83,21\,\%$, and false positive beats of $16,78\,\%$.

An inspection of the individual subjects reveals that the probability function of the different radar nodes often does not return a peak at the same time point but with a slight offset around the actual heartbeat, depending on where the subject is lying in bed. This leads to broader and lower peaks when simply summating the individual heartbeat probabilities. As a result, a lower threshold has to be set to find more heartbeats. However, this increases the chance of capturing false heartbeats.

6.2 Hyperparameter search results

This Section presents the hyperparameter search of the different models(two block 1D-CNN, four block 1D-CNN, LSTM and TCN); shown are the testing metrics and confusion matrix for each class granularity after the AASM sleep stage classification binary is Wake/sleep, three-stage is Wake/NREM/REM, and five stage is Wake/N1/N2/N3/REM sleep, and different window size to calculate the HRV ($60s$, $180s$, and $360s$).

6.2.1 Two block 1D-CNN grid search

Performance Analysis

The grid search results for two block 1D-CNN are summarised in Table 6.2. The binary classification reveals the best overall performance of HRV window sizes. The highest MCC score of 0.203 and a Std of 0.073 is found for a HRV window size of $60s$. The corresponding F1 score is 0.732 with a variance of 0.1. This indicates a good balance between precision and recall. For the higher class granularity, the lowest variance in the metrics is observed for the $60s$ window size. As a general trend, we see a slight decrease in all metrics for higher window sizes and an increase in the standard deviation. For three-stage and five-stage, the $180s$ HRV window size shows slightly better MCC performance compared to the $60s$ window size. The $60s$ and $180s$ window sizes performed better across all metrics than the 360-second window size, particularly for three-stage and five-stage sleep classifications.

Confusion matrix

To better understand the metrics, the figures 6.1, 6.2 and 6.3 are picturing the corresponding confusion matrix for the different HRV window sizes and classes. First, considering the binary stage confusion matrix in Figure 6.1, we see the moderate ability to distinguish between Wake and sleep across all window sizes. For the $180s$ window size, the model finds slightly more sleep stages than for the other window size. Wake stage predicting is slightly decreasing for higher window sizes.

The three-stage confusion matrix in Figure 6.2 consistently performs well in classifying NREM. However, it shows an overall difficulty in distinguishing between Wake and NREM, as well as between NREM and REM. Across all window sizes, REM sleep has the lowest true positive prediction. In comparison, the highest true positive prediction of NREM and REM sleep occurs for a window size of $60s$. $180s$ had the highest true positive prediction of the Wake state.

The five-stage confusion matrix is shown in Figure 6.3. While it reveals a higher true prediction of the Wake state, the miss classification as Wake significantly increases for larger window sizes. The N1 stage is consistently poorly detected but also has low false negatives. The N2 stage has the highest true positive prediction for a $60s$ window size and a high value for the other window sizes. However, the model behaves similarly to the three-stage classification, with difficulties in distinguishing between Wake and N2 and N2 and REM. We see that the model predicts the majority classes Wake and N2 well, but misclassifies the minority classes N1, N3, and REM repeatedly as N2. For larger HRV window sizes, increasing difficulty in differentiating between Wake and

Table 6.2: Two block CNN mean metrics of test results and standard deviation for all classes and HRV windows

Class	HRV Window	$60s$	$180s$	$360s$
Wake/Sleep	Accuracy	0.641 ± 0.088	0.637 ± 0.033	0.541 ± 0.063
	Bal. Accuracy	0.631 ± 0.061	0.628 ± 0.065	0.580 ± 0.087
	Precision	0.868 ± 0.109	0.864 ± 0.115	0.848 ± 0.132
	Recall	0.651 ± 0.135	0.651 ± 0.047	0.551 ± 0.113
	F1	0.732 ± 0.100	0.737 ± 0.047	0.652 ± 0.051
	Kappa	0.167 ± 0.057	0.163 ± 0.081	0.089 ± 0.085
	Specificity	0.612 ± 0.204	0.605 ± 0.170	0.608 ± 0.231
	MCC	$\mathbf{0.203 \pm 0.073}$	$\mathbf{0.196 \pm 0.096}$	$\mathbf{0.122 \pm 0.123}$
Wake/NREM/REM	Accuracy	0.477 ± 0.062	0.478 ± 0.088	0.434 ± 0.055
	Bal. Accuracy	0.401 ± 0.053	0.424 ± 0.101	0.379 ± 0.065
	Precision	0.613 ± 0.115	0.609 ± 0.136	0.577 ± 0.104
	Recall	0.477 ± 0.062	0.478 ± 0.088	0.434 ± 0.055
	F1	0.508 ± 0.082	0.503 ± 0.122	0.468 ± 0.076
	Kappa	0.092 ± 0.043	0.087 ± 0.045	0.045 ± 0.041
	Specificity	0.489 ± 0.146	0.482 ± 0.122	0.460 ± 0.150
	MCC	0.103 ± 0.044	$\mathbf{0.105 \pm 0.062}$	0.049 ± 0.052
Wake/N1/N2/N3/REM	Accuracy	0.312 ± 0.066	0.284 ± 0.069	0.264 ± 0.066
	Bal. Accuracy	0.261 ± 0.020	0.255 ± 0.026	0.221 ± 0.041
	Precision	0.442 ± 0.083	0.481 ± 0.148	0.418 ± 0.116
	Recall	0.312 ± 0.066	0.284 ± 0.069	0.264 ± 0.066
	F1	0.338 ± 0.061	0.315 ± 0.077	0.290 ± 0.079
	Kappa	0.075 ± 0.020	0.079 ± 0.040	0.041 ± 0.041
	Specificity	0.640 ± 0.084	0.646 ± 0.080	0.634 ± 0.082
	MCC	0.084 ± 0.025	$\mathbf{0.092 \pm 0.048}$	0.046 ± 0.045

REM starts to show frequently misclassifying Wake for REM. The highest REM prediction can be observed for the $180s$ window size.

Hyperparmeter sets

The chosen hyperparameter sets for the two block 1D-CNN are shown in Table 6.3. The class granularities and HRV window sizes are stated. The sets were chosen from the Pareto front between the mean loss and Std mean loss for all folds. To maximise performance, the set with the lowest mean loss was picked. Despite the higher variability of the loss, we see a low loss for the binary state during the training in accordance with the test results. In contrast to the test results, the variance decreases for the higher HRV window size. This is observable in all class grains. The relatively high variability of the losses indicates a high dependency on the training validation split. We can also see that all classes have a virtually equal set with slight differences. For the sliding

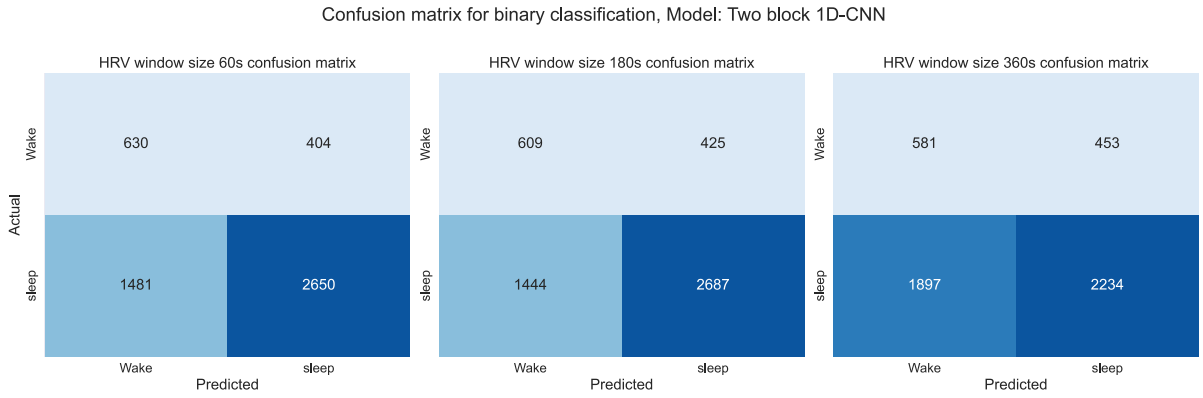

Figure 6.1: Two block 1D-CNN confusion matrix for all HRV window sizes and Wake/sleep classification

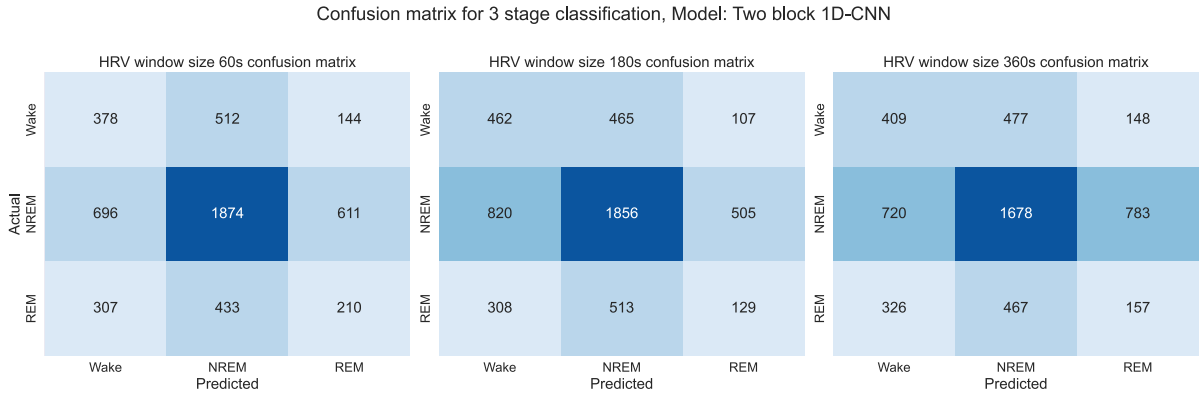

Figure 6.2: Two block 1D-CNN confusion matrix for all HRV window sizes and Wake/NREM/REM classification

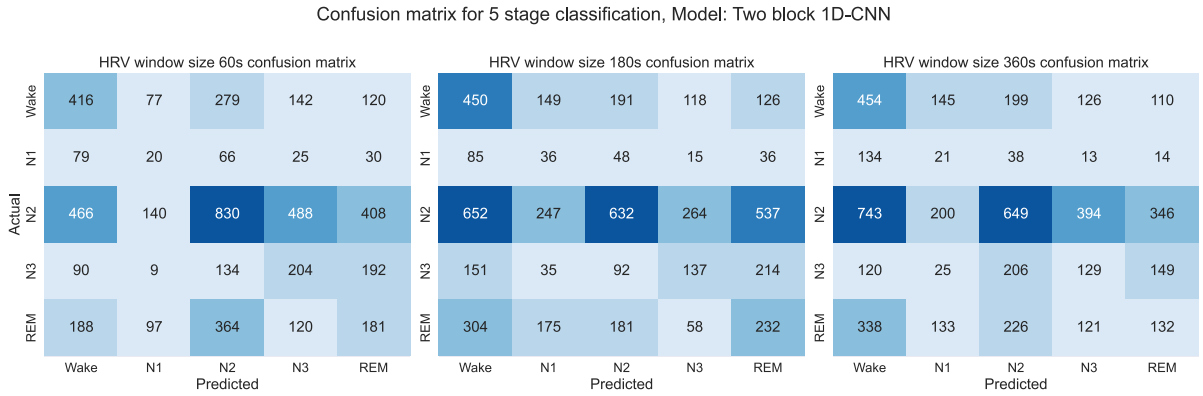

Figure 6.3: Two block 1D-CNN confusion matrix for all HRV window sizes and Wake/N1/N2/N3/REM classification

Table 6.3: Best hyperparameter set combination for two block 1D-CNN, with the acronyms Max pooling (Maxp), Average pooling(Avgp)

Class number	binary			3stage			five stage		
HRV windows:	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$
Parameters:									
Window ccn	20	20	20	10	15	15	20	20	15
Maxp. kernel	3	3	3	2	3	3	2	3	3
Avgp. kernel	2	3	3	2	2	3	2	3	3
Kern application	3	5	4	5	5	5	4	5	5
Learning rate	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005	0.0005
Mean val loss	0.3499	0.3596	0.3523	0.6851	0.711	0.6754	1.07	1.068	1.075
Std val loss	0.03989	0.02916	0.0267	0.1264	0.1031	0.09823	0.1627	0.06738	0.1212

CNN window size, a higher window size dominates overall classes and HRV window sizes, and a low learning rate occurs in all sets. Also, a higher value for the kernel application is predominant. This hints that a larger network is preferable to a smaller one as the kernel application increases the weights and biases of the model. Higher values for the CNN window size and kernel application seem desirable to lower the loss. The kernel size for the max pooling layer appears to be ideal at three. At the same time, for the window sizes of $180s$ and 360, a higher average pooling kernel of three seems beneficial, and for the $60s$ window, a lower size is indicated to improve the loss.

6.2.2 Four block 1D-CNN grid search

Performance Analysis

The grid search results for the four block 1D-CNN are presented in Table 6.4. The Table lists the metrics for the different class granularities and HRV window sizes. The best performing classification is the binary class. We achieve the highest MCC of 0.194 and the relative lowest Std of 0.072 for the HRV window size of $360s$. However, for three- and five-stage classification, the model performs better for a window size of $60s$, showing overall higher metrics than for a window size of $180s$ and $360s$.

Confusion matrix

The figures 6.4, 6.5, and 6.6 picturing the corresponding confusion matrix for the class grains and the HRV windows. The Figure 6.4 shows the binary confusion matrix. The prediction for true Wake states variates through the window sizes, with a decrease during the $180s$. The highest true predicted Wake states are seen for a $60s$ window size. However, the frequent true predicted sleep

Table 6.4: Four block CNN mean metric of test results and standard deviation for all classes and HRV windows

Class	HRV Window	$60s$	$180s$	$360s$
Wake/Sleep	Accuracy	0.558 ± 0.082	0.600 ± 0.026	0.597 ± 0.045
	Bal. Accuracy	0.631 ± 0.061	0.611 ± 0.064	0.629 ± 0.049
	Precision	0.874 ± 0.134	0.855 ± 0.127	0.867 ± 0.110
	Recall	0.517 ± 0.134	0.607 ± 0.038	0.606 ± 0.092
	F1	0.638 ± 0.124	0.704 ± 0.044	0.702 ± 0.032
	Kappa	0.122 ± 0.057	0.130 ± 0.060	0.152 ± 0.072
	Specificity	0.712 ± 0.168	0.614 ± 0.161	0.652 ± 0.169
	MCC	$\mathbf{0.173 \pm 0.079}$	$\mathbf{0.164 \pm 0.083}$	$\mathbf{0.194 \pm 0.072}$
Wake/NREM/REM	Accuracy	0.483 ± 0.083	0.494 ± 0.049	0.381 ± 0.069
	Bal. Accuracy	0.432 ± 0.078	0.430 ± 0.092	0.410 ± 0.069
	Precision	0.633 ± 0.129	0.623 ± 0.117	0.594 ± 0.115
	Recall	0.483 ± 0.083	0.494 ± 0.049	0.381 ± 0.069
	F1	0.509 ± 0.105	0.520 ± 0.086	0.410 ± 0.077
	Kappa	0.118 ± 0.049	0.101 ± 0.082	0.059 ± 0.049
	Specificity	0.502 ± 0.136	0.488 ± 0.145	0.468 ± 0.142
	MCC	$\mathbf{0.145 \pm 0.064}$	0.119 ± 0.093	0.074 ± 0.060
Wake/N1/N2/N3/REM	Accuracy	0.332 ± 0.035	0.281 ± 0.081	0.275 ± 0.095
	Bal. Accuracy	0.311 ± 0.097	0.240 ± 0.047	0.246 ± 0.054
	Precision	0.479 ± 0.133	0.448 ± 0.128	0.472 ± 0.165
	Recall	0.332 ± 0.035	0.281 ± 0.081	0.275 ± 0.095
	F1	0.356 ± 0.061	0.304 ± 0.088	0.287 ± 0.111
	Kappa	0.107 ± 0.035	0.053 ± 0.028	0.062 ± 0.041
	Specificity	0.660 ± 0.081	0.639 ± 0.076	0.643 ± 0.080
	MCC	$\mathbf{0.121 \pm 0.047}$	0.061 ± 0.031	0.074 ± 0.054

stages are found for the window size of $180s$ and with the highest misclassification for the $60s$ window.

In the three-stage classification matrix, in Figure 6.5, we see the highest true NREM sleep prediction at the $180s$ window size and the lowest for $360s$. As for the classification of Wake and REM sleep, we see similar to the two block CNN a difficulty distinguishing Wake and NREM likewise NREM and REM from each other. The highest true predictions for REM and NREM are found for the window size of $180s$. While the highest true classification for the Wake state occurs for the window size of $360s$. For this window size, the NREM sleep is also frequently misclassified as Wake or REM, showing a corresponding behaviour with the three-stage classification from the two block CNN.

The five-stage confusion matrix in Fig 6.6 shows overall window sizes a consistent true prediction of the Wake state with a slight increase. The misclassification shifts from N3 and REM

to N1 and N2 across the higher window sizes, and it remains significant. N2 has relatively high true positive prediction overall window sizes and high misclassification rates. N1, N3 and REM are consistently poor detected over the HRV window sizes. This classes show a high misclassification rate in the other stages.

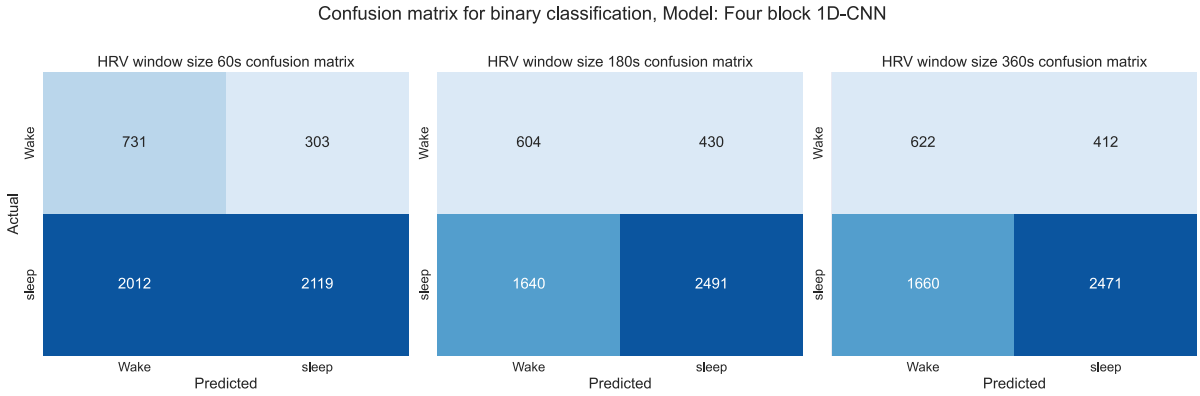

Figure 6.4: Four block 1D-CNN confusion matrix for all HRV window sizes and Wake/sleep classification

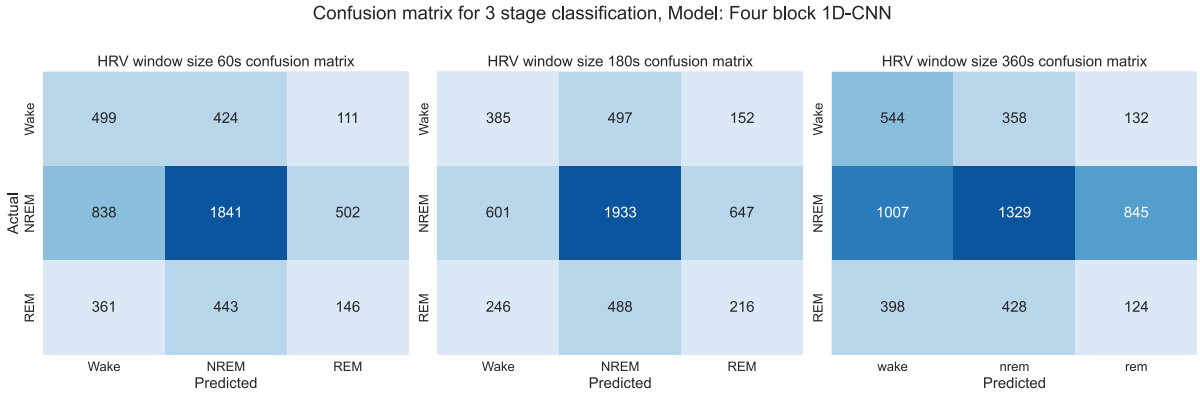

Figure 6.5: Four block 1D-CNN confusion matrix for all HRV window sizes and Wake/NREM/REM classification

Hyperparameter sets

Table 6.5 shows the best hyperparameter set from the four block 1D-CNN gird search for the class and window combinations. At inspection of the sets, we find that a high sliding window size is present in all sets; further, the kernel size of three is dominant for the max and average pooling layers. A higher learning rate seems to produce lower losses. Lastly the kernel application reduced

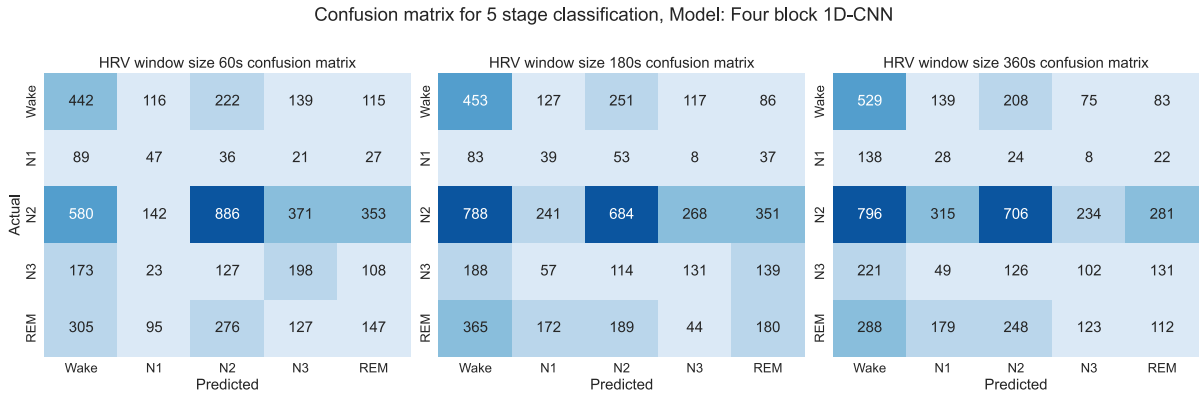

Figure 6.6: Four block 1D-CNN confusion matrix for all HRV window sizes and Wake/N1/N2/N3/REM classification

with a higher class granularity and HRV window size. Further, the Std overall classes are relatively high.

Table 6.5: Best hyperparameter set combination for four block 1D-CNN, with the acronyms max pooling (Maxp), average pooling (Avgp)

Class number	binary			three-stage			five-stage		
HRV windows:	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$
Parameters:									
Window ccn	30	15	30	30	30	30	30	30	30
Maxp. kernel	3	3	3	2	3	3	3	3	2
Avgp. kernel	3	2	3	3	3	3	3	3	3
Kern application	5	3	4	4	3	3	3	3	2
Learning rate	0.0005	0.001	0.0005	0.0005	0.001	0.001	0.001	0.0005	0.001
Mean val loss	0.3473	0.3609	0.3524	0.7118	0.6931	0.6066	1.05	1.002	1.031
Std val loss	0.0416	0.0320	0.0231	0.0693	0.0810	0.0537	0.1123	0.0742	0.1169

6.2.3 LSTM grid search

Performance analysis

After performing the optuna hyperparameter search on the LSTM we found the results shown in Table 6.6. The classification of binary, three-stage, five-stage are noted over the metrics. We see very bad performance metrics across all classes and window sizes, with the best MCC for a window size of $180s$ at five-stage classification 0.139 and a Std of 0.081 but a low F1 score of 0.301 with a high Std of ± 0.147. For two-stage classification, we have a very high specificity of 0.929 for a $60s$ window and 0.96 for a window size of $180s$. Notably we see a very slight negative

value for the $180s$ for binary and three stage classification and $60s$ for the five stage indicating an inverse prediction, further the for $360s$ wee see a value significant closer to zero, this also is present in the confusion matrix.

Table 6.6: LSTM results for all classes and HRV windows

Class	HRV Window	$60s$	$180s$	$360s$
Wake/Sleep	Accuracy	0.279 ± 0.136	0.204 ± 0.134	0.441 ± 0.220
	Bal. Accuracy	0.526 ± 0.048	0.498 ± 0.002	0.528 ± 0.053
	Precision	0.539 ± 0.435	0.359 ± 0.423	0.838 ± 0.123
	Recall	0.122 ± 0.133	0.036 ± 0.085	0.425 ± 0.321
	F1	0.191 ± 0.196	0.054 ± 0.123	0.486 ± 0.289
	Kappa	0.019 ± 0.031	-0.001 ± 0.003	0.049 ± 0.103
	Specificity	0.929 ± 0.058	0.960 ± 0.087	0.632 ± 0.299
	MCC	0.038 ± 0.068	-0.006 ± 0.019	$\mathbf{0.064 \pm 0.107}$
Wake/NREM/REM	Accuracy	0.388 ± 0.142	0.289 ± 0.164	0.333 ± 0.212
	Bal. Accuracy	0.411 ± 0.088	0.381 ± 0.081	0.421 ± 0.091
	Precision	0.607 ± 0.110	0.582 ± 0.176	0.569 ± 0.286
	Recall	0.388 ± 0.142	0.289 ± 0.164	0.333 ± 0.212
	F1	0.405 ± 0.118	0.262 ± 0.183	0.302 ± 0.237
	Kappa	0.082 ± 0.068	0.041 ± 0.066	0.101 ± 0.139
	Specificity	0.488 ± 0.159	0.475 ± 0.142	0.501 ± 0.163
	MCC	0.095 ± 0.072	0.064 ± 0.064	$\mathbf{0.132 \pm 0.140}$
Wake/N1/N2/N3/REM	Accuracy	0.395 ± 0.151	0.335 ± 0.137	0.184 ± 0.139
	Bal. Accuracy	0.184 ± 0.043	0.281 ± 0.044	0.156 ± 0.061
	Precision	0.299 ± 0.173	0.405 ± 0.177	0.135 ± 0.169
	Recall	0.395 ± 0.151	0.335 ± 0.137	0.184 ± 0.139
	F1	0.300 ± 0.121	0.301 ± 0.147	0.143 ± 0.153
	Kappa	-0.012 ± 0.067	0.115 ± 0.076	0.008 ± 0.048
	Specificity	0.593 ± 0.164	0.671 ± 0.102	0.631 ± 0.076
	MCC	-0.010 ± 0.097	$\mathbf{0.139 \pm 0.081}$	0.007 ± 0.060

Confusion matrix

The Figures 6.7, 6.8, and 6.9 show the corresponding confusion matrix for the classes and the HRV window sizes. When comparing the binary confusion matrix in Figure 6.7 with each other, we see the highest detection of actual sleep stages for the $360s$ window size. However, most sleep stages are still misclassified as Wake, while the true prediction of the Wake state decreases with increasing window size.

The three-stage confusion matrix in Figure 6.8 shows an improvement of the true positive rate for Wake state prediction across the window size. For the $60s$ window, we see the highest

true positive rate for NREM and REM sleep, while both are still considerably misclassified. The NREM and REM stages are consistently misclassified as Wake for higher window sizes, and the accurate prediction of the REM stage is very low across all window sizes.

For the five-stage classification depicted in Figure 6.9, we see the model has a high misclassification for all HRV window sizes. For $60s$, most Wake, N1 and N3 are not predicted at all and are misclassified as N2 or REM, with few correct predictions for N2 and REM. While for $180s$, some improvements for Wake and N2 prediction are to note. N1 and N3 remain challenging to predict correctly. Moreover, there is a high misclassification between Wake and N2. The window size $360s$ has improvements in predicting Wake and REM sleep but confuses N1, N2, and N3 to Wake or REM.

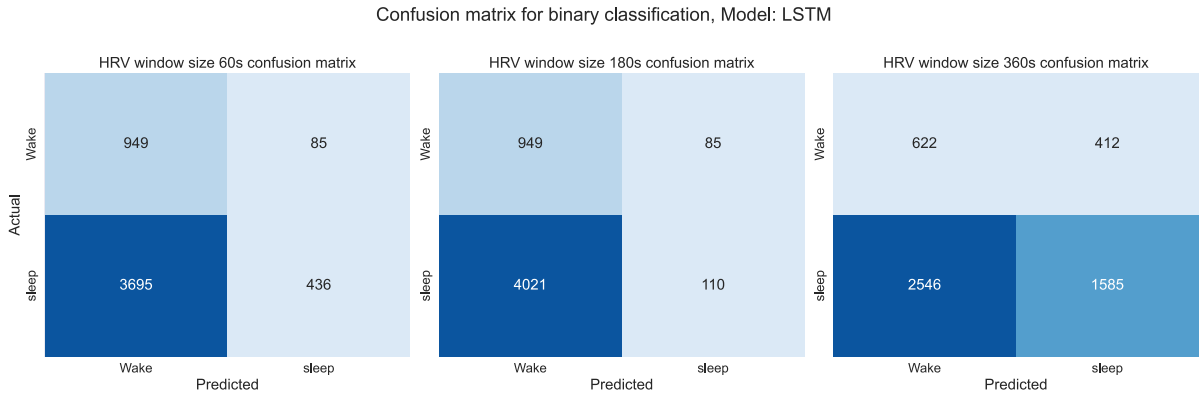

Figure 6.7: LSTM confusion matrix for all HRV window sizes and Wake/sleep classification

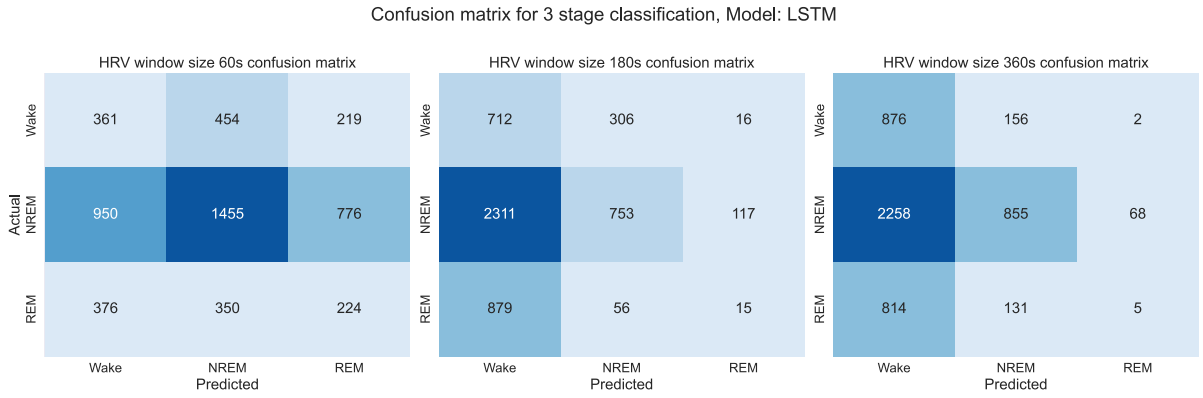

Figure 6.8: LSTM confusion matrix for all HRV window sizes and Wake/NREM/REM classification

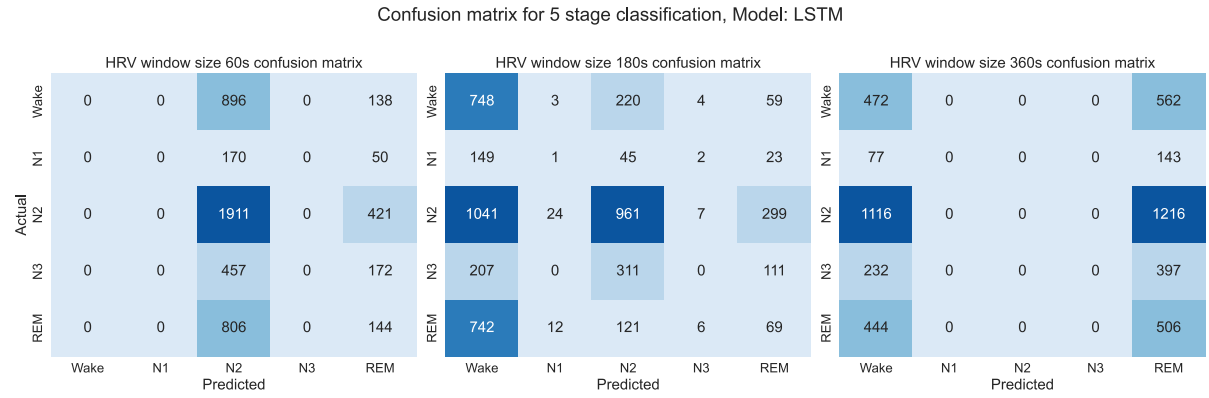

Figure 6.9: LSTM confusion matrix for all HRV window sizes and Wake/N1/N2/N3/REM classification

Hyperparameter sets

The found hyperparameter sets for the classes and windows are shown in table 6.7. We see from the different sequence lengths provided at testing that the test values' lower side is predominant. For the higher class granularity, a bigger hidden size is dominant. We see an increase in layer number the more classes the model has to distinguish. The learning rate is also on the lower end of the spectrum. While overall HRV window sizes, a high batch size is present except for the $180s$ window.

Table 6.7: Best hyperparamter set combination for LSTM

Class number	binary			three-stage			five-stage		
HRV windows:	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$
Parameters:									
Sequence length	101	51	21	21	51	51	51	51	101
Hidden size	420	68	80	296	436	528	440	48	592
Layer number	1	1	1	1	4	3	4	5	1
Learning rate	0.0001	0.0005	0.0009	0.0004	0.0001	0.0001	0.0020	0.0002	0.0003
Batch size	512	32	512	512	512	512	512	512	512
MCC val	0.1241	0.0195	0.2287	0.1088	0.1664	0.1454	0.0644	0.1057	0.0826

6.2.4 TCN grid search

Performance analysis

The test results for the TCN model are listed in Table 6.8; presented are the class stages and HRV window sizes over the metrics. We see the best MCC value of 0.113 ± 0.067 for three-stage

classification and a $360s$ window. For binary the highest MCC score of was 0.07 ± 0.069 reached for a $60s$ HRV window size and for five stage the $360s$ window produces a MCC of 0.099 ± 0.097. Overall, the HRV window size of $360s$ performs better than the lower sizes except for the binary classification.

Table 6.8: TCN results for all classes and HRV windows

Class	HRV Window	60s	180s	360s
Wake/Sleep	Accuracy	0.496 ± 0.049	0.297 ± 0.129	0.249 ± 0.118
	Bal. Accuracy	0.552 ± 0.051	0.520 ± 0.051	0.513 ± 0.010
	Precision	0.824 ± 0.137	0.817 ± 0.157	0.865 ± 0.151
	Recall	0.489 ± 0.094	0.200 ± 0.161	0.105 ± 0.099
	F1	0.601 ± 0.059	0.288 ± 0.144	0.168 ± 0.120
	Kappa	0.045 ± 0.044	0.018 ± 0.049	0.008 ± 0.008
	Specificity	0.615 ± 0.130	0.840 ± 0.156	0.920 ± 0.112
	MCC	$\mathbf{0.070 \pm 0.069}$	0.033 ± 0.100	0.040 ± 0.028
Wake/NREM/REM	Accuracy	0.509 ± 0.125	0.454 ± 0.142	0.357 ± 0.158
	Bal. Accuracy	0.381 ± 0.057	0.353 ± 0.017	0.419 ± 0.065
	Precision	0.495 ± 0.229	0.581 ± 0.106	0.634 ± 0.170
	Recall	0.509 ± 0.125	0.454 ± 0.142	0.357 ± 0.158
	F1	0.474 ± 0.170	0.442 ± 0.151	0.341 ± 0.184
	Kappa	0.023 ± 0.046	0.020 ± 0.031	0.078 ± 0.059
	Specificity	0.464 ± 0.148	0.457 ± 0.143	0.487 ± 0.144
	MCC	0.032 ± 0.063	0.018 ± 0.040	$\mathbf{0.113 \pm 0.067}$
Wake/N1/N2/N3/REM	Accuracy	0.344 ± 0.135	0.299 ± 0.132	0.350 ± 0.143
	Bal. Accuracy	0.237 ± 0.040	0.227 ± 0.035	0.253 ± 0.055
	Precision	0.463 ± 0.120	0.412 ± 0.123	0.468 ± 0.116
	Recall	0.344 ± 0.135	0.299 ± 0.132	0.350 ± 0.143
	F1	0.320 ± 0.118	0.286 ± 0.127	0.336 ± 0.121
	Kappa	0.059 ± 0.044	0.033 ± 0.055	0.082 ± 0.082
	Specificity	0.643 ± 0.063	0.636 ± 0.067	0.662 ± 0.074
	MCC	0.073 ± 0.058	0.040 ± 0.059	$\mathbf{0.099 \pm 0.097}$

Confusion matrix

The Figures 6.10, 6.11, and 6.12 picturing the confusion matrix for the class numbers and HRV window sizes. In figure 6.10, we see that for a higher HRV window size, we have an increasing number of correctly predicted Wake states but a significant decrease in the correct classification for the sleep stages.

For the three-stage classification, we observe an increase in correctly predicted Wake stages with an increase in the window size. Further, we notice an increasing misclassification between

NREM and Wake state for higher HRV windows. While the REM stage is poorly classified for all window sizes. For the window size of $60s$, it tends to be misclassified as NREM stage, while for the higher window size, this shifts to Wake. The NREM stage has the highest number of true positives but also a high number of false positives and false negatives.

The confusion matrix 6.12 shows that over all window sizes we have many Wake states falsely classified as N2. For the N1 stage, there are very few correct predictions, with a slightly higher correct prediction for the window size of $180s$; the N2 stage shows the highest correct prediction but is often misclassified for Wake, N3 or REM sleep. The correct prediction of N3 is increasing with a higher window size but tends to be misclassified for N2. REM sleep classification is increasing for $360s$. However, more often, many samples are misclassified as N1 or Wake. The model leans to the majority group.

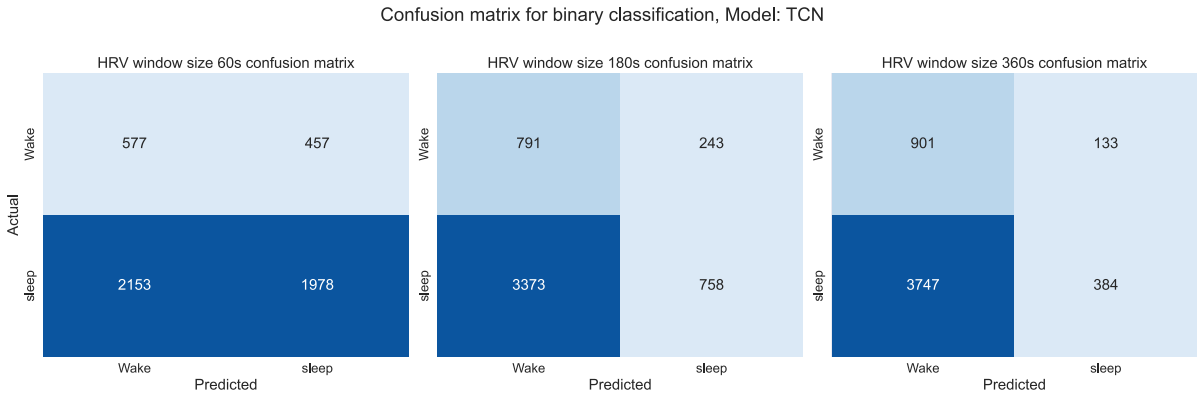

Figure 6.10: TCN confusion matrix for all HRV window sizes and Wake/sleep classification

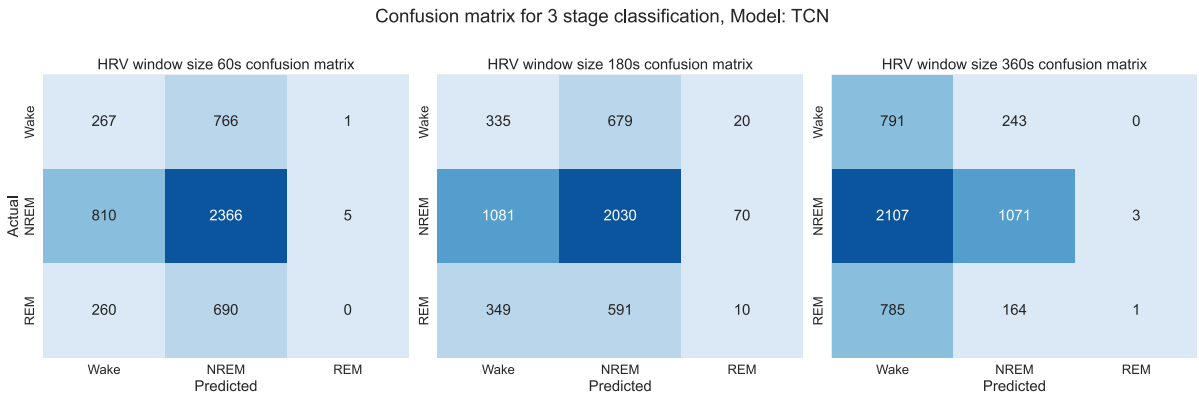

Figure 6.11: TCN confusion matrix for all HRV window sizes and Wake/NREM/REM classification

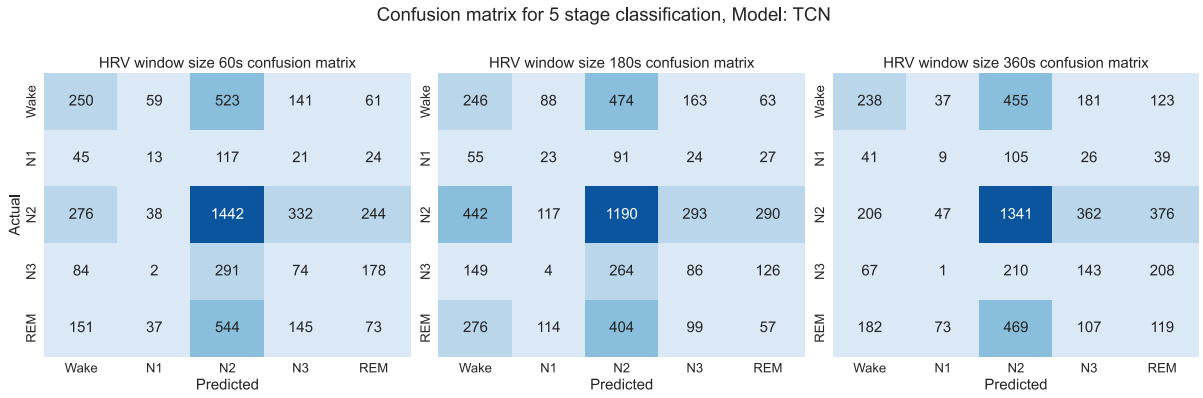

Figure 6.12: TCN confusion matrix for all HRV window sizes and Wake/N1/N2/N3/REM classification

Table 6.9: Best hyperparameter set combination for TCN

Class number	binary			three-stage			five-stage		
HRV windows:	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$	$60s$	$180s$	$360s$
Parameters:									
Channel number	5	6	3	6	4	5	5	4	2
Hidden number	200	328	424	488	232	488	72	8	168
Kernel size	4	5	4	3	5	3	3	2	4
Dropout	0.5	0.1	0.3	0.2	0.4	0.2	0.3	0.1	0.5
Learning rate	0.5552	0.6764	0.0308	0.7709	0.6028	0.9423	0.0018	0.0017	0.0009
Batch size	256	64	256	128	128	256	256	256	256
MCC val	0.0473	0.05928	0.07673	0.0448	0.0488	0.0958	0.0551	0.0897	0.1062

Hyperparemeter sets

The best performing hyperparameter set is shown in Table 6.9, the class numbers and HRV window combinations are listed over the hyperparameter sets. We notice that the channel number is decreasing with a higher HRV window size and class number. At the same time, the hidden number is decreasing for higher class grains but generally increasing for larger HRV window sizes. The dropout for the $60s$ window decreases with the higher class number while the value for $360s$ increases; contrasting for a $180s$ window size the dropout value peaks for three-stage classification. The value of the learning rate is higher for the lower class numbers while finding a minimal value for five-stage classification and a window size of $360s$. The batch size is dominant, with a high value over all classes and window sizes.

6.3 Model comparison

For the following Section comparison of the presented models, two block 1D-CNN, four block 1D-CNN, LSTM, and TCN performance metrics for the different class granularity and the HRV window size of $60s$, $180s$ and $360s$. Figure 6.13 shows the binary classification metrics, Figure 6.14 pictures the three stage metrics, and Figure 6.15 states the five stage metrics. As a general tend we observe in the Figures that the four block CNN has the best MCC performance across all HRV window and classes. We observe as well a general higher values in the metrics foe the CNN models with lower variability. However, the overall low Kappa and MCC sores while better than random guessing showing the need for improvements to achieve stronger performances.

Further we see an relative high variability for in the precision and recall metrics overall classes and HRV windows, specially for the LSTM. We can further see a decrease for the metrics the higher the numbers of sleep stages the models need to predict, but an increase in Specificity. In terms of recall and F1 score, this indicating it struggles to correctly identify positive instances. While the different HRV window size have a varying influence on the individual means of the metrics metrics, the range of the values is not as influences specially the MCC score is relatively stable for the classes granulites. For the CNN models the HRV window size has only a slight and varying influence on the metrics metric ranges. In contrast the HRV window size has a higher influence on th values for the LSTM and TCN.

Binary class comparison

For the binary classification in Figure 6.13 the CNN models both shows across all window sizes generally high values considering the balanced accuracy, precision, recall, and F1 scores showing the models are able to detect the wake and sleep state. Comparing to the CNNs, the LSTM only shows a precision in the same range for the window size of $360s$ with a high variability in recall, specificity and F1 score. It has however, the highest specificity of all models for the smaller window sizes. The TCN shows comparable precision and balanced accuracy values in all window sizes but with a significant lower recall and F1 for a bigger HRV window size than $60s$, for which in contrast the specificity are very high. For the CNN models we see the lowest variability in the metrics for the window size of $180s$. Here the two block CNN shows a higher MCC than the four block CNN.

Considering Figure 6.13 we observe across the HRV window sizes a similar trend in behaviour for the models. While the balanced accuracy shows that the models are able to that the models perform a an overall low score for the MCC score, but the four block CNN performs best across

window sizes followed by TCN and LSTM. It further has the lowest variance

Three stage comparison

For the three stage classification we see in figure 6.14 we see an assimilation for the metrics values over the window size, for the LSTM and TCN the variance is increasing with a higher window size. With considerably lower values for the balanced accuracy, precision, recall, F1 and specificity scores.

Five stage comparison

For the 5 stage classification in Figure 6.15 we see the LTSM model has the lowest performance for the window sizes of $60s$ and $360s$ while for the $180s$ window all models performs equality over the metrics except the LTSM has the highest MCC score. For the other window size the TCN, and CNN models are overall in the same range. The MCC and differing with the highest values for the four block CNN.

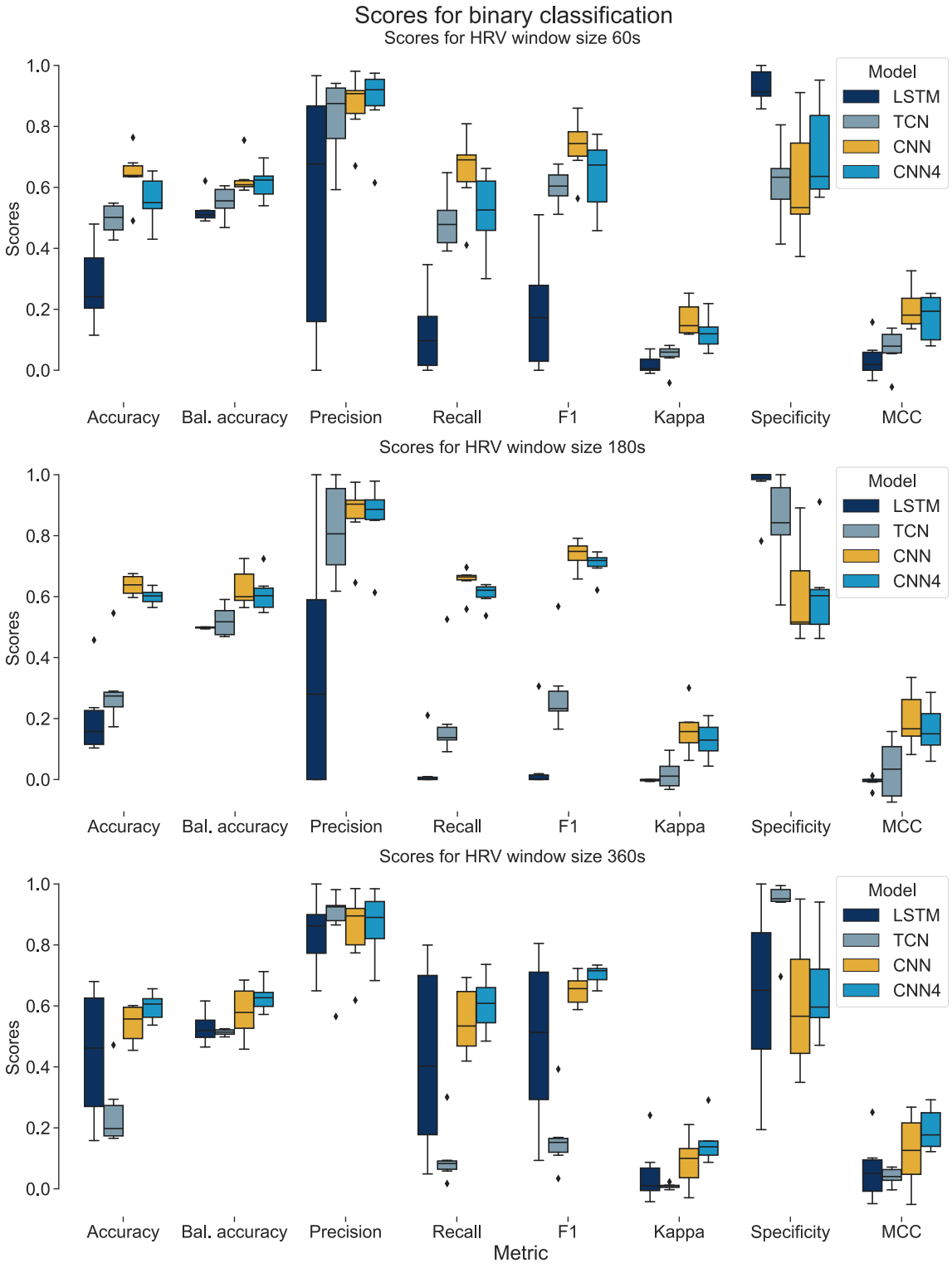

Figure 6.13: Comparing Boxplot for the binary classification (Wake/sleep) for the different HRV window sizes ($60s$, $180s$, and $360s$). The models (two block 1D-CNN (CNN), four block 1D-CNN (CNN4), LSTM and TCN) are pictured over the metrics (Accuracy, Balanced (Bal.) accuracy, Precision, Recall, F1, Kappa, Specificity and MCC)

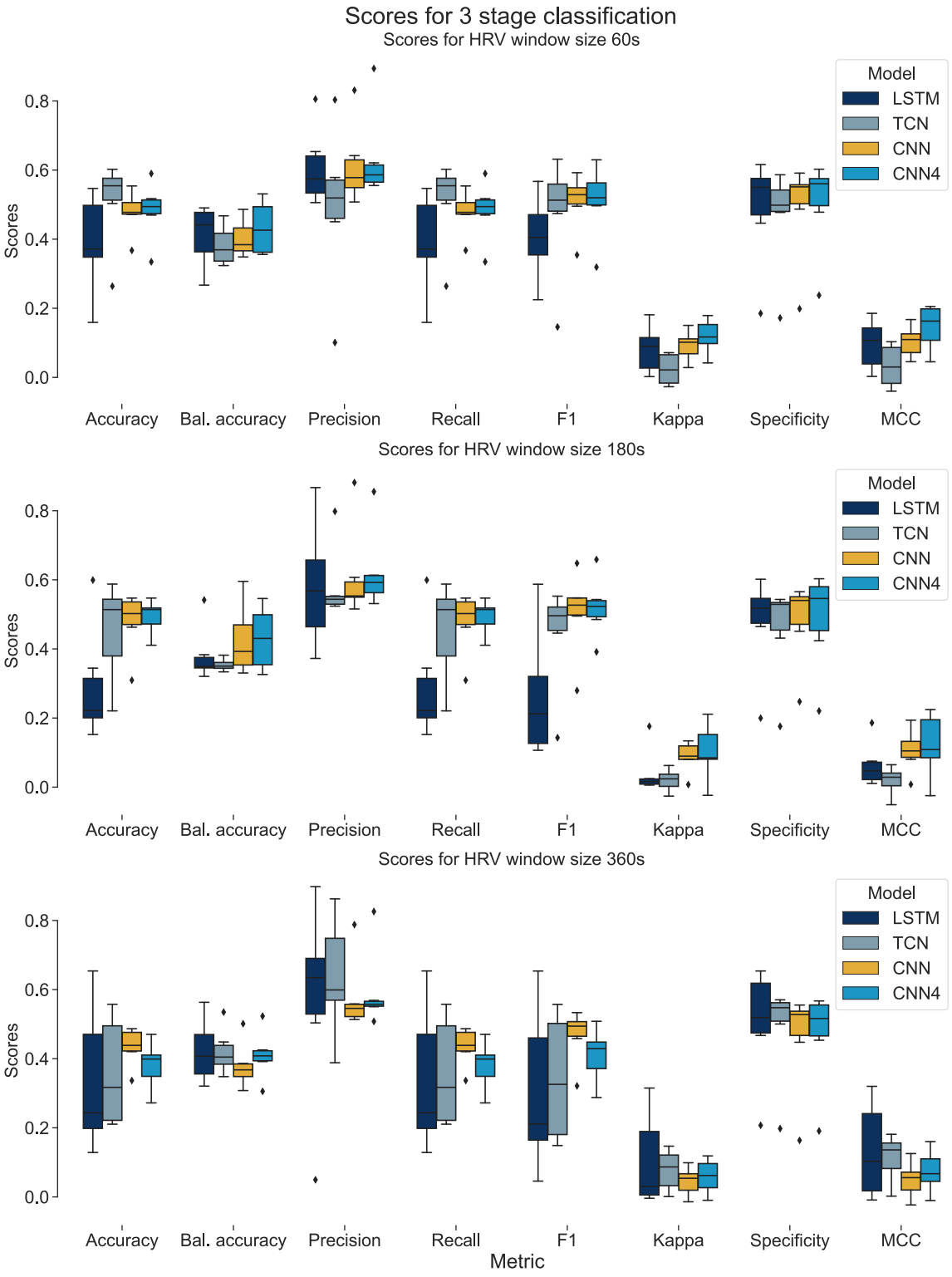

Figure 6.14: Comparing Boxplot for the three stage classification (Wake/NREM/REM) for the different HRV window sizes ($60s$, $180s$, and $360s$). The models (two block 1D-CNN (CNN), four block 1D-CNN (CNN4), LSTM and TCN) are pictured over the metrics (Accuracy, Balanced (Bal.) accuracy, Precision, Recall, F1, Kappa, Specificity and MCC)

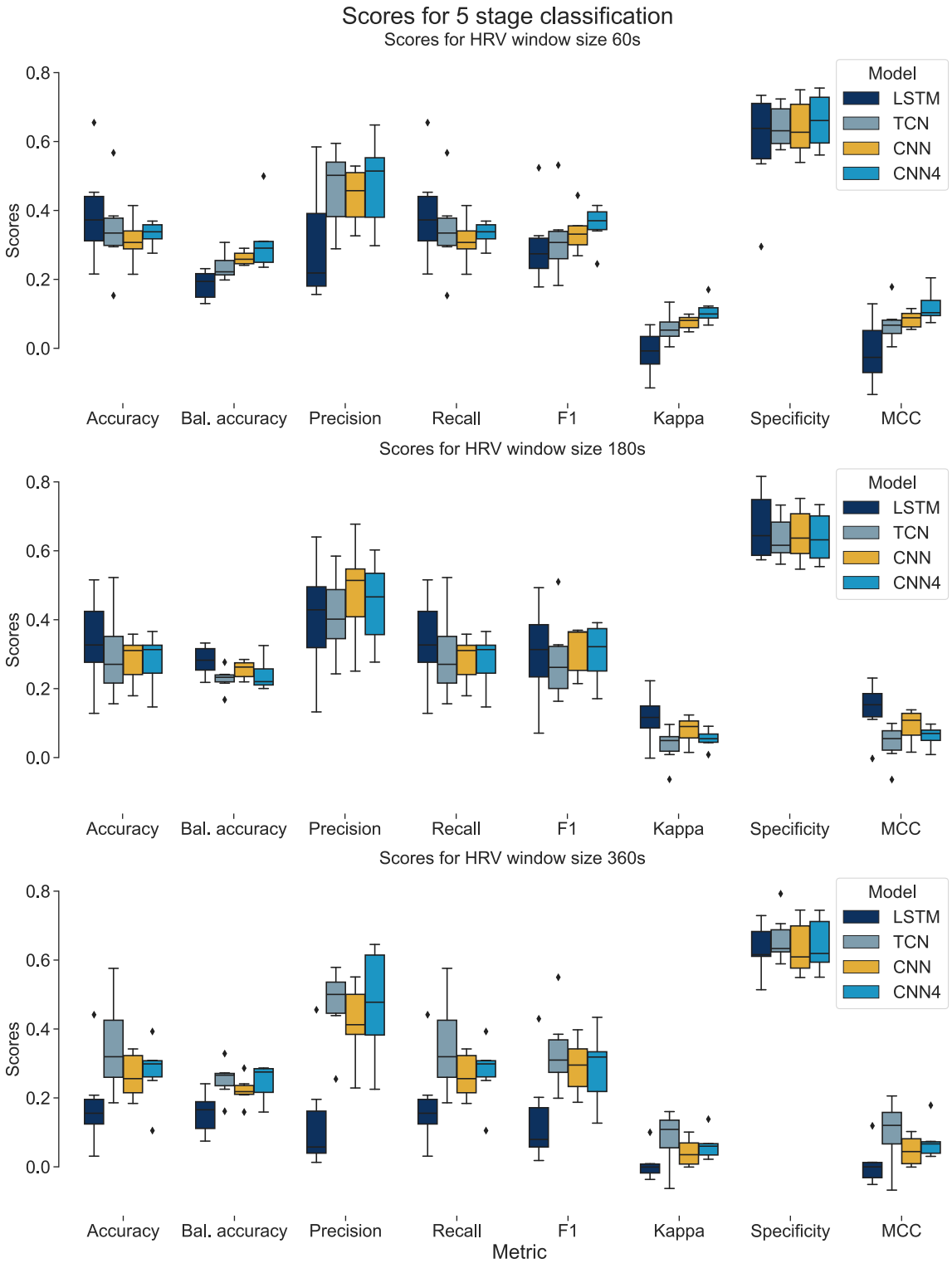

Figure 6.15: Comparing Boxplot for the five stage classification (Wake/N1/N2/N3/REM) for the different HRV window sizes ($60s$, $180s$, and $360s$). The models (two block 1D-CNN (CNN), four block 1D-CNN (CNN4), LSTM and TCN) are pictured over the metrics (Accuracy, Balanced (Bal.) accuracy, Precision, Recall, F1, Kappa, Specificity and MCC)

Chapter 7

Discussion

This chapter discusses the results and limitations of our work, following the general structure of the Methods.

Both CNN models showed the effect of dataset imbalance, evidenced by the high disparity of F1 score and MCC for the two block CNN 0.70 vs. 0.203 for the binary classification of the $60s$ HRV window and a similar different for the four block CNN. We see distribution also across the lager HRV window size and to a lower degree in higher class granularity's. Further supported is this by the high variance during the cross validation and fluctuations in test results across all models. This also indicates a dependency on train split as the test split for all models is the same.

The superior performance of the four block CNN compared to the two block CNN suggests that a bigger network structure is preferable for classifying sleep stages. This is supported by the high kernel application during hyperparameter selection for the two block CNN. Literature that features CNN also reported the use of higher sized networks with good performances. However, the hyperparameter sets for the four layer CNN only features moderate kernel applications. The CNN size of four blocks therefore for seem promising but needs more testing.

However our findings, showed the need to improve the performance of all models as the MCC is indicates only slightly better performance specially for the LSTM and TCN. The three stage classification confusion matrix for all models but specially the one from the CNN's are highlighting the difficulties to differentiate between sleep stages, specially Wake and REM are hard to distinguish. While the NREM sleep is relatively reliable detected, as the heart rate reaches the lowest point during the NREM sleep. REM and Wake state are both identified with a higher HRV than NREM sleep. Secondly we a shift from in the misclassification of the classes of the higher window sizes. This highlights the the importance of additional features to enhance the difference between the sleep stage and showing that with only the HRV features it is hard to predict the

sleep phases. Adding respiratory waves or body movements is expected to significantly improve classification results, as the explored in the literature. It is common to used approaches with a combination of bio-signals such as the body movements and respiratory waves, reported are results around $80\,\%$ or higher sleep stage accuracy [Lee24; Par24; Hon18].

As expected the more sleep phases the network had to identify, the worse it performs the task. This can be deducted from the overall higher results for binary classification. The CNN results further show that despite the small size of the dataset, for sleep classification a more complex network is preferred.

For the purpose of PD early diagnosis a simple wake sleep classification is not sufficient and requires at lest a three stage classification with Wake,NREM, and REM sleep.

7.1 Limitations

The Section discusses the limitations of the work. In Section 7.1.1 the limitations of the data set are addressed. In Section 7.1.2 the limitations and possible solution concerning the feature quality are discussed. Finally Section 7.1.3 discusses the limitations of the models used in this work.

7.1.1 Dataset

One limitation regarding the dataset is its relatively small size, which is generally bad for generalisation and model performance. A second limitation is the lack of expert labels, this is a important limitation of this work influencing both the training and testing results, as the labels highly impacts the data quality. And bias the model when features and labels are contradicting. This is shown in the high variance and poor MCC score over the models as well the poor detection of the minority classes.

The variation in validation performance and test results depending on data splits further emphasizes the need for expert labels. The automatic sleep labelling by the *SOMNO eco* system is unreliable. Expert labels would not only improve data quality but also increase the dataset size by including currently excluded participants.

7.1.2 Features

The overall quality of the features can be improved in several ways. First, by improving the predicted heartbeats with an advanced combination of the LSTM probability output. As shown in Fig. 5.4, the R-peak predictions of the individual radar nodes are not fully aligned despite being

synchronised with each other. A simple summation leads to wider and lower peak values, which requires a lower threshold for the height and thus increases the chance of detecting false peaks.

Although the presented threshold level of 0.24 for the heartbeats prediction gives a good results prediction of the heartbeats as shown in Section 6.1, it could be beneficial to use a dynamic threshold for height and distance leading to a higher F1 score and precision as the heart rate is variable during the night.

7.1.3 Model

As the training pipelines for the CNN's and provided LSTM and TCN algorithm are containing differences, the results are not completely comparable. The main difference is that for the during the training and validation of the models. While during CNN training a 3 fold cross validation is performed the LSTM and TCN only used a simple 80/20 split of the training data into train and validation set, resulting in different data splits. This further hints at the bad quality of the labels and data as the LSTM and TCN had a bigger training set (18/5), to the CNNs which trained on a (15/8) split on the cross fold validation.

For the CNN the application of a dense CNN might archive better results as the four block CNN revealed to have a better performance.

The missing expert labels are a concern and the results are difficult to interpret and poorly applicable to reveal trends in the underlying data as we may see the sleep stage in the features but when the label is the same as when the feature is not present the models will naturally learn nothing, represented in the poor MCC sores. For the test results it seem not to not help that some splits containing data with good label quality during the training, seen in the high variance during hyperparameter search. This hints that we should perform an sophisticated class balancing than the use of weighted loss functions or improve the class weights, to tackle the class imbalances in the classes data.

Chapter 8

Conclusion and Outlook

In conclusion of this work, this work successfully extracted the heart rate from a sleeping person using a $61\,GHz$ CW radar system. We achieved a high F1 score of 0.803 with for the heart rate extraction. However, when evaluating the HRV feature for different window sizes ($60s$, $180s$ and 360) our results were not comparable to accuracies obtainable from a PSG setup. In this work we tested four DL networks: a two block 1D-CNN, a four block 1D-CNN, a LSTM, and a TCN. The four block 1D CNN emerged as the the best performing network on average, with MCC score of $0,173$ across all the window sizes for a simple Wake/sleep classification. Notably, the highest MCC score of 0.203 ± 0.073 was achieved by the two block CNN model, with an F1 score of 0.732 ± 0.100, and a balanced accuracy of 0.631 ± 0.061 for the binary classification. While the LSTM and TCN both performed worse than the CNN's, they still outperformed random guessing as indicated by their on average MCC scores higher than zero.

These findings demonstrate the general possibility to predict the sleep stages using HRV features. However, they also indicate that the heart rate alone is insufficient to reliable predict the sleep stages. To enhance the model predictions, additional features should be combined to the HRV, ether the respiratory wave or body movements as they can reduce the ambiguity of the features. This need for additional data is further emphasized by the increased difficulty in detecting sleep stages with finer class granularity.

Moreover the high variance during the training and testing reveals a high imbalance in the dataset, which needs to be addressed. Oversampling, is a promising method to mitigate this issue and the small size of our dataset. The overall lower performance of the LSTM and TCN algorithms requires a confirmation trough cross validation.

Further an interesting alternative route to explore would be the direct use of raw radar signals bypassing the feature extraction process with the bi-directional LSTM used during heartbeat

extraction. This however, would greatly impact the explainability of the predicted sleep stages as the understandability of the used features in such a network is not given. Finally, to produce robust results that are usable in real world scenarios, an expert labelling of our dataset is required. This would also affect the current findings, as an inspection revealed the ambiguity of the automated labels produced by the PSG system.

In general, the results can be improved in a few ways. The first way would be to train the models on expert labels, and to address the class imbalance or increasing the dataset. A second way would incorporate the additional biosignals that can be extracted from the radar signals such as the respiratory signals and body movements. Further the HRV feature quality could be improved ether by an improved LSTM algorithm to extract the heartbeats or an advanced summary of the LSTM probability output in order to improve heartbeat detection.

Appendix A

Questionnaires

Fragebogen

1 PSQI

During the past month, what time have you usually gone to bed at night?

Please enter the time using the 24-hour system (e.g. "22.00").

During the past month, how long (in minutes) has it usually taken you to fall asleep each night?

minutes

During the past month, what time have you usually gotten up in the morning?

Please enter the time using the 24-hour system (e.g. 08.00).

During the past month, how many hours of actual sleep did you get at night? (This may be different than the number of hours you spent in bed.)

Please enter the hours of sleep.

hours

During the past month, how often have you had trouble sleeping because you cannot get to sleep within 30 minutes?

During the past month, how often have you had trouble sleeping because you wake up in the middle of the night or early morning?

During the past month, how often have you had trouble sleeping because you have to get up to use the bathroom?

During the past month, how often have you had trouble sleeping because you cannot breathe comfortably?

During the past month, how often have you had trouble sleeping because you cough or snore loudly?

During the past month, how often have you had trouble sleeping because you feel too cold?

During the past month, how often have you had trouble sleeping because you feel too hot?

During the past month, how often have you had trouble sleeping because you had bad dreams?

During the past month, how often have you had trouble sleeping because you have pain?

Other reason(s), please describe:

How often during the past month have you had trouble sleeping because of this?

During the past month, how would you rate your sleep quality overall?

During the past month, how often have you taken medicine to help you sleep (prescribed or "over the counter")?

During the past month, how often have you had trouble staying awake while driving, eating meals, or engaging in social activity?

During the past month, how much of a problem has it been for you to keep up enough enthusiasm to get things done?

Do you have a bed partner or room mate?

If you have a room mate or bed partner, ask him/her how often in the past month you have had loud snoring:

If you have a room mate or bed partner, ask him/her how often in the past month you have had long pauses between breaths while asleep :

If you have a room mate or bed partner, ask him/her how often in the past month you have had legs twitching or jerking while you sleep:

If you have a room mate or bed partner, ask him/her how often in the past month you have had episodes of disorientation or confusion during sleep:

Other restlessness while you sleep; please describe:

How often in the past month have you experienced this other restlessness you described?

2 STOP BANG

Do you snore loudly?

Louder than talking or loud enough to be heard through closed doors

- ◯ Yes
- ◯ No

Do you often feel tired, fatigued, or sleepy during the daytime?

- ◯ Yes
- ◯ No

Has anyone observed you stop breathing during sleep?

- ◯ Yes
- ◯ No

Do you have (or are you being treated for) high blood pressure?

- ◯ Yes
- ◯ No

BMI

- ◯ <=35
- ◯ >35

Age

- ◯ <=50
- ◯ >50

Gender

- ◯ Female
- ◯ Male
- ◯ Other

3 RBD Screening Questionnaire

I sometimes have very vivid dreams.

- ◯ Yes
- ◯ No

My dreams frequently have an aggressive or action-packed content.

- ◯ Yes
- ◯ No

The dream contents mostly match my nocturnal behaviour.

- ◯ Yes
- ◯ No

It thereby happened that I (almost) hurt my bed partner or myself.

- ◯ Yes
- ◯ No

I have or had the following phenomena during my dreams:
speaking, shouting, swearing, laughing loudly

- ◯ Yes
- ◯ No

I have or had the following phenomena during my dreams:
sudden limb movements, “fights”

- ◯ Yes
- ◯ No

I have or had the following phenomena during my dreams:
gestures, complex movements, that are useless during sleep, e.g., to wave, to salute, to frighten mosquitoes, falls off the bed

- ◯ Yes
- ◯ No

I have or had the following phenomena during my dreams:
things that fell down around the bed, e.g., bedside lamp, book, glasses

- ◯ Yes
- ◯ No

It happens that my movements awake me.

- ◯ Yes
- ◯ No

After awakening, I mostly remember the content of my dreams well.

- ◯ Yes
- ◯ No

My sleep is frequently disturbed.

- ◯ Yes
- ◯ No

I have/had a disease of the nervous system (e.g., stroke, head trauma, parkinsonism, RLS, narcolepsy, depression, epilepsy, inflammatory disease of the brain)

- ◯ Yes
- ◯ No

If yes, which?

4 SF-12 Health Survey

In general, would you say your health is:

- ◯ Excellent
- ◯ Very good
- ◯ Good
- ◯ Fair
- ◯ Poor

The following questions are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

Moderate activities such as moving a table, pushing a vacuum cleaner, bowling, or playing golf.

- ◯ YES, limited a lot
- ◯ YES, limited a little
- ◯ NO, not limited at all

Climbing several flights of stairs.

- ◯ YES, limited a lot
- ◯ YES, limited a little
- ◯ NO, not limited at all

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of your physical health?

Accomplished less than you would like.

- ○ Yes
- ○ No

Were limited in the kind of work or other activities.

- ○ Yes
- ○ No

During the past 4 weeks, have you had any of the following problems with your work or other regular daily activities as a result of any emotional problems (such as feeling depressed or anxious)?

Accomplished less than you would like.

- ○ Yes
- ○ No

Did work or activities less carefully than usual.

- ○ Yes
- ○ No

During the past 4 weeks, how much did pain interfere with your normal work (including work outside the home and housework)?

- ○ Not at all
- ○ A little bit
- ○ Moderately
- ○ Quite a bit
- ○ Extremely

These questions are about how you have been feeling during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling.

How much of the time during the past 4 weeks...

Have you felt calm & peaceful?

- ◯ All of the time
- ◯ Most of the time
- ◯ A good bit of the time
- ◯ Some of the time
- ◯ A little of the time
- ◯ None of the time

Did you have a lot of energy?

- ◯ All of the time
- ◯ Most of the time
- ◯ A good bit of the time
- ◯ Some of the time
- ◯ A little of the time
- ◯ None of the time

Have you felt down-hearted and blue?

- ◯ All of the time
- ◯ Most of the time
- ◯ A good bit of the time
- ◯ Some of the time
- ◯ A little of the time
- ◯ None of the time

During the past 4 weeks, how much of the time has your physical health or emotional problems interfered with your social activities (like visiting friends, relatives, etc.)?

- ◯ All of the time
- ◯ Most of the time
- ◯ A good bit of the time
- ◯ Some of the time
- ◯ A little of the time
- ◯ None of the time

5 Endseite

Appendix B

Detailed HRV Features

B.1 Window size 60s

Table B.1: Feature table win size $60s$

Ind	HRV MeanNN	HRV SDNN	HRV RMSSD	HRV SDSD	HRV CVNN	HRV CVSD
0	989.166345	634.480778	879.417715	887.232551	0.641430	0.889049
1	989.166345	634.480778	879.417715	887.232551	0.64143	0.889049
2	896.930909	482.364586	601.005864	605.631623	0.537795	0.670069
3	1202.654609	1278.319662	1549.321129	1566.701099	1.062915	1.288251
4	983.523019	289.664014	412.317625	416.186297	0.294517	0.419225
	HRV MedianNN	HRV MadNN	HRV MCVNN	HRV IQRNN	HRV SDRMSSD	HRV Prc20NN
0	804.352	165.102336	0.205261	223.104	0.721478	673.3824
1	804.352	165.102336	0.205261	223.104	0.721478	673.3824
2	795.136	130.563686	0.164203	171.008	0.802595	695.8080
3	824.320	162.825062	0.197527	268.800	0.825084	719.3600
4	911.360	46.304563	0.050808	52.736	0.702526	883.7120
	HRV Prc80NN	HRV pNN50	HRV pNN20	HRV MinNN	HRV MaxNN	HRV HTI
0	1026.9696	65.517241	75.862069	363.008	3075.584	19.333333
1	1026.9696	65.517241	75.862069	363.008	3075.584	19.333333
2	939.0080	57.575758	74.242424	301.568	3234.816	13.200000

3	1222.6560	65.217391	84.782609	301.568	8587.776	11.500000
4	967.6800	52.830189	69.811321	617.472	2535.424	8.833333
	HRV TINN	HRV LF	HRV HF	HRV VHF	HRV TP	HRV LFHF
0	351.5625	0.023053	0.047818	0.003592	0.074462	0.482099
1	351.5625	0.023053	0.047818	0.003592	0.074462	0.482099
2	414.0625	0.060723	0.041255	0.002085	0.104063	1.471887
3	460.9375	0.011502	0.001065	0.000026	0.012594	10.798633
4	289.0625	0.050871	0.070591	0.003552	0.125014	0.720647
	HRV LFn	HRV HFn	HRV LnHF	HRV SD1	HRV SD2	HRV SD1SD2
0	0.309591	0.642174	-3.040363	627.368154	594.283492	1.055672
1	0.309591	0.642174	-3.040363	627.368154	594.283492	1.055672
2	0.583519	0.396443	-3.187978	428.246227	534.890503	0.800624
3	0.913333	0.084579	-6.844611	1107.824971	1451.197948	0.763387
4	0.406923	0.564663	-2.650859	294.288153	281.611549	1.045015
	HRV S	HRV CSI	HRV CVI	HRV CSI Modified	HRV PIP	HRV IALS
0	1.171294e+06	0.947264	6.775636	2251.774288	0.637931	0.666667
1	1.171294e+06	0.947264	6.775636	2251.774288	0.637931	0.666667
2	7.196284e+05	1.249026	6.564078	2672.367738	0.545455	0.569231
3	5.050655e+06	1.309952	7.410318	7604.000776	0.478261	0.511111
4	2.603593e+05	0.956925	6.122543	1077.923986	0.566038	0.596154
	HRV PSS	HRV PAS	HRV GI	HRV SI	HRV AI	HRV PI
0	0.894737	0.214286	49.808397	49.301489	50.031079	52.631579
1	0.894737	0.214286	49.808397	49.301489	50.031079	52.631579
2	0.810811	0.200000	51.144476	52.204785	50.105361	56.923077
3	0.739130	0.181818	48.558699	48.300378	53.352664	60.000000
4	0.774194	0.333333	47.215295	46.864692	47.985074	44.230769
	HRV C1d	HRV C1a	HRV SD1d	HRV SD1a	HRV C2d	HRV C2a
0	0.509365	0.490635	447.752450	439.443558	0.472536	0.527464
1	0.509365	0.490635	447.752450	439.443558	0.472536	0.527464
2	20.484050	0.515950	297.972172	307.633787	0.571788	0.428212
3	0.607127	0.392873	863.269338	694.435928	0.359507	0.640493
4	0.501493	0.498507	208.480805	207.859061	0.477108	0.522892

	HRV SD2d	HRV SD2a	HRV Cd	HRV Ca	HRV SDNNd	HRV SDNNa
0	408.518049	431.608471	0.491947	0.508053	428.584445	435.543634
1	408.518049	431.608471	0.491947	0.508053	428.584445	435.543634
2	404.466359	350.021163	0.537513	0.462487	355.232636	329.509758
3	870.122279	1161.405487	0.450688	0.549312	866.702581	956.844805
4	194.517556	203.636895	0.489842	0.510158	201.620095	205.758808
	HRV DFA alpha1	HRV MFDFA alpha1 Width	HRV MFDFA alpha1 Peak	HRV MFDFA alpha1 Mean	HRV MFDFA alpha1 Max	HRV MFDFA alpha1 Delta
0	0.682497	2.445954	1.381829	1.537270	0.002919	-0.160676
1	0.682497	2.445954	1.381829	1.537270	0.002919	-0.160676
2	0.833684	2.301850	1.626745	1.629737	0.228043	-0.010320
3	0.993351	2.435734	1.664672	1.730758	-0.170763	-0.020711
4	0.775761	1.237567	1.025850	1.003147	-0.222317	-0.349546
	HRV MFDFA alpha1 Asymmetry	HRV MFDFA alpha1 Fluctuation	HRV MFDFA alpha1 Increment	HRV ApEn	HRV SampEn	HRV ShanEn
0	-0.436450	0.011055	0.739575	0.567606	0.864997	5.857981
1	-0.436450	0.011055	0.739575	0.567606	0.864997	5.857981
2	-0.498700	0.012470	0.734613	0.622912	0.668883	6.044394
3	-0.472868	0.007527	0.603164	0.555456	0.513262	5.523562
4	-0.518345	0.001024	0.086600	0.629586	0.771863	5.539241
	HRV FuzzyEn	HRV MSEn	HRV CMSEn	HRV RCMSEn	HRV CD	HRV HFD
0	0.836982	0.630561	0.707253	0.761050	1.101672	1.999975
1	0.836982	0.630561	0.707253	0.761050	1.101672	1.999975
2	0.861208	0.751624	0.921159	0.869954	1.037993	1.909911
3	0.584853	0.214339	0.303849	0.312020	0.658206	2.008628
4	0.680505	0.350746	0.594368	0.563003	0.829114	1.997467
	HRV KFD	HRV LZC				

0	2.937889	1.009997
1	2.937889	1.009997
2	1.993871	0.915817
3	1.639163	1.080697
4	2.394583	0.972666

Appendix C

Score Tables

C.1 Individual scores Heartbeat height threshold

Table C.1: Detailed results of all tested height thresholds,Tt: Total True ECG Beats, Tpb: Total predicted beats, Tp: True positive peaks, Fn: False negative beats, Fp: False positive beats, F1 score: $2Tp/(2Tp + Fp + Fn)$, Precision: $Tp/(Tp + Fp)$, Recall: $Tp/(Tp + Fn)$

Vp, Height threshold	Tt	Tpb	Tp	Fn	Fp	f1	Precision	Recall
02, 0.24	23331	22835	17791	5044	5540	0.770740	0.779111	0.762548
02, 0.26	23331	22064	17346	4718	5985	0.764225	0.786168	0.743474
02, 0.28	23331	21263	16820	4443	6511	0.754362	0.791045	0.720929
02, 0.30	23331	20424	16263	4161	7068	0.743366	0.796269	0.697055
02, 0.32	23331	19507	15606	3901	7725	0.728605	0.800021	0.668895
06, 0.24	27335	23366	19628	3738	7707	0.774265	0.840024	0.718054
06, 0.26	27335	23108	19450	3658	7885	0.771167	0.841700	0.711542
06, 0.28	27335	22840	19250	3590	8085	0.767314	0.842820	0.704225
06, 0.30	27335	22551	19020	3531	8315	0.762539	0.843422	0.695811
06, 0.32	27335	22243	18794	3449	8541	0.758159	0.844940	0.687543
10, 0.24	29119	28082	22638	5444	6481	0.791525	0.806139	0.777431
10, 0.26	29119	27166	22083	5083	7036	0.784685	0.812891	0.758371
10, 0.28	29119	26126	21401	4725	7718	0.774767	0.819146	0.734950
10, 0.30	29119	25109	20690	4419	8429	0.763074	0.824007	0.710533
10, 0.32	29119	23903	19796	4107	9323	0.746709	0.828181	0.679831
28, 0.24	27012	30394	21683	8711	5329	0.755426	0.713397	0.802717
28, 0.26	27012	29697	21337	8360	5675	0.752508	0.718490	0.789908
28, 0.28	27012	28980	20935	8045	6077	0.747785	0.722395	0.775026
28, 0.30	27012	28175	20421	7754	6591	0.740066	0.724791	0.755997
28, 0.32	27012	27262	19788	7474	7224	0.729189	0.725845	0.732563
42, 0.24	39378	37225	34058	3167	5320	0.889208	0.914923	0.864899
42, 0.26	39378	35869	33133	2736	6245	0.880646	0.923722	0.841409
42, 0.28	39378	34240	31906	2334	7472	0.866799	0.931834	0.810249
42, 0.30	39378	32330	30327	2003	9051	0.845847	0.938045	0.770151
42, 0.32	39378	30193	28464	1729	10914	0.818272	0.942735	0.722840

List of Figures

List of Tables

Bibliography

[Alb24] Nils Albrecht, Dominik Langer, Daniel Krauss, Robert Richer, Luca Abel, Bjoern Eskofier, Nicolas Rohleder, and Alexande Koelpin. "EmRad: Ubiquitous Vital Sign Sensing using Compact Continuous-Wave Radars". In: *IEEE Open Journal of Engineering in Medicine and Biology* (2024). DOI: unpublished.

[Alm21] Daniel Gomes de Almeida-Filho, Bruna Del Vechio Koike, Francesca Billwiller, Kelly Soares Farias, Igor Rafael Praxedes de Sales, Pierre-Hervé Luppi, Sidarta Ribeiro, and Claudio Marcos Queiroz. "Hippocampus-retrosplenial cortex interaction is increased during phasic REM and contributes to memory consolidation". en. In: *Scientific Reports* 11.1 (June 2021). Publisher: Nature Publishing Group, p. 13078. ISSN: 2045-2322. DOI: 10.1038/s41598-021-91659-5. URL: https://www.nature.com/articles/s41598-021-91659-5 (visited on 04/29/2024).

[Ang20] Alessandra Angelucci and Andrea Aliverti. "Telemonitoring systems for respiratory patients: technological aspects". In: *Pulmonology* 26.4 (July 2020), pp. 221–232. ISSN: 2531-0437. DOI: 10.1016/j.pulmoe.2019.11.006. URL: https://www.sciencedirect.com/science/article/pii/S2531043719302144 (visited on 04/04/2024).

[Aya23] Zainab Ayaz, Saeeda Naz, Naila Habib Khan, Imran Razzak, and Muhammad Imran. "Automated methods for diagnosis of Parkinson's disease and predicting severity level". en. In: *Neural Computing and Applications* 35.20 (July 2023), pp. 14499–14534. ISSN: 1433-3058. DOI: 10.1007/s00521-021-06626-y. URL: https://doi.org/10.1007/s00521-021-06626-y (visited on 01/30/2024).

[Bae24] Daniel Baena, Balmeet Toor, Nicholas H van den Berg, Laura B Ray, and Stuart M Fogel. "Spindle-slow wave coupling and problem-solving skills: impact of age". In: *Sleep* (Mar. 2024), zsae072. ISSN: 0161-8105. DOI: 10.1093/sleep/zsae072. URL: https://doi.org/10.1093/sleep/zsae072 (visited on 04/29/2024).

[Bai18] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. *An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling*. arXiv:1803.01271 [cs]. Apr. 2018. DOI: 10.48550/arXiv.1803.01271. URL: http://arxiv.org/abs/1803.01271 (visited on 06/01/2024).

[Bar09] Anna L. Bartels and Klaus L. Leenders. "Parkinson's disease: The syndrome, the pathogenesis and pathophysiology". In: *Cortex*. Special Issue on "Parkinson's Disease, Language and Cognition" 45.8 (Sept. 2009), pp. 915–921. ISSN: 0010-9452. DOI: 10.1016/j.cortex.2008.11.010. URL: https://www.sciencedirect.com/science/article/pii/S0010945208002839 (visited on 02/05/2024).

[Bar21] Giuseppe Barbato. "REM Sleep: An Unknown Indicator of Sleep Quality". eng. In: *International Journal of Environmental Research and Public Health* 18.24 (Dec. 2021), p. 12976. ISSN: 1660-4601. DOI: 10.3390/ijerph182412976.

[Bar23] Giuseppe Barbato. "Is REM Density a Measure of Arousal during Sleep?" eng. In: *Brain Sciences* 13.3 (Feb. 2023), p. 378. ISSN: 2076-3425. DOI: 10.3390/brainsci13030378.

[Bei14] Janice M. Beitz. "Parkinson's disease: a review". eng. In: *Frontiers in Bioscience (Scholar Edition)* 6.1 (Jan. 2014), pp. 65–74. ISSN: 1945-0524. DOI: 10.2741/s415.

[Ben19] Ruth M. Benca and Mihai Teodorescu. "Sleep physiology and disorders in aging and dementia". eng. In: *Handbook of Clinical Neurology* 167 (2019), pp. 477–493. ISSN: 0072-9752. DOI: 10.1016/B978-0-12-804766-8.00026-1.

[Bie20] Lukas Biewald. *Experiment Tracking with Weights and Biases*. en-US. 2020. URL: https://www.wandb.com/ (visited on 07/24/2024).

[Blo97] Konrad E. Bloch. "Polysomnography: a systematic review". en. In: *Technology and Health Care* 5.4 (Jan. 1997). Publisher: IOS Press, pp. 285–305. ISSN: 0928-7329. DOI: 10.3233/THC-1997-5403. URL: https://content.iospress.com/articles/technology-and-health-care/thc075 (visited on 11/05/2023).

[Boo17] Reza Boostani, Foroozan Karimzadeh, and Mohammad Nami. "A comparative review on sleep stage classification methods in patients and healthy individuals". In: *Computer Methods and Programs in Biomedicine* 140 (Mar. 2017), pp. 77–91. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2016.12.004. URL: https://www.sciencedirect.com/science/article/pii/S0169260716308276 (visited on 10/30/2023).

[Bor16] Olga Boric-Lubecke, Victor M. Lubecke, Amy D. Droitcour, Byung-Kwon Park, and Aditya Singh, eds. *Doppler Radar Physiological Sensing*. en. 1st ed. Wiley, Feb. 2016. ISBN: 978-1-118-02402-7 978-1-119-07841-8. DOI: 10.1002/9781119078418. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781119078418 (visited on 11/04/2023).

[Bou19] Mark I Boulos, Trevor Jairam, Tetyana Kendzerska, James Im, Anastasia Mekhael, and Brian J Murray. "Normal polysomnography parameters in healthy adults: a systematic review and meta-analysis". In: *The Lancet Respiratory Medicine* 7.6 (June 2019), pp. 533–543. ISSN: 2213-2600. DOI: 10.1016/S2213-2600(19)30057-8. URL: https://www.sciencedirect.com/science/article/pii/S2213260019300578 (visited on 11/05/2023).

[Bro12] Ritchie E. Brown, Radhika Basheer, James T. McKenna, Robert E. Strecker, and Robert W. McCarley. "Control of Sleep and Wakefulness". In: *Physiological Reviews* 92.3 (2012), pp. 1087–1187. DOI: doi:10.1152/physrev.00032.2011. URL: https://journals.physiology.org/doi/full/10.1152/physrev.00032.2011?rss=1 (visited on 06/02/2024).

[Che00] Kun-Mu Chen, Yong Huang, Jianping Zhang, and A. Norman. "Microwave life-detection systems for searching human subjects under earthquake rubble or behind barrier". In: *IEEE Transactions on Biomedical Engineering* 47.1 (Jan. 2000). Conference Name: IEEE Transactions on Biomedical Engineering, pp. 105–114. ISSN: 1558-2531. DOI: 10.1109/10.817625. URL: https://ieeexplore.ieee.org/abstract/document/817625 (visited on 04/04/2024).

[Con23] Adrien Conessa, Ursula Debarnot, Isabelle Siegler, and Arnaud Boutin. "Sleep-related motor skill consolidation and generalizability after physical practice, motor imagery, and action observation". en. In: *iScience* 26.8 (Aug. 2023), p. 107314. ISSN: 25890042. DOI: 10.1016/j.isci.2023.107314. URL: https://linkinghub.elsevier.com/retrieve/pii/S2589004223013913 (visited on 04/29/2024).

[Coo19] Navin Cooray, Fernando Andreotti, Christine Lo, Mkael Symmonds, Michele T. M. Hu, and Maarten De Vos. "Detection of REM sleep behaviour disorder by automated polysomnography analysis". In: *Clinical Neurophysiology* 130.4 (Apr. 2019), pp. 505–514. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2019.01.011. URL: https://www.sciencedirect.com/science/article/pii/S1388245719300306 (visited on 10/26/2023).

[Erl24] Daniel Erlacher, Daniel Schmid, Stephan Zahno, and Michael Schredl. "Changing Sleep Architecture through Motor Learning: Influences of a Trampoline Session on REM Sleep Parameters". en. In: *Life* 14.2 (Feb. 2024). Number: 2 Publisher: Multidisciplinary Digital Publishing Institute, p. 203. ISSN: 2075-1729. DOI: 10.3390/life14020203. URL: https://www.mdpi.com/2075-1729/14/2/203 (visited on 04/29/2024).

[Fab21] Marco Fabbri, Alessia Beracci, Monica Martoni, Debora Meneo, Lorenzo Tonetti, and Vincenzo Natale. "Measuring Subjective Sleep Quality: A Review". en. In: *International Journal of Environmental Research and Public Health* 18.3 (Jan. 2021). Number: 3 Publisher: Multidisciplinary Digital Publishing Institute, p. 1082. ISSN: 1660-4601. DOI: 10.3390/ijerph18031082. URL: https://www.mdpi.com/1660-4601/18/3/1082 (visited on 10/30/2023).

[Fig23] Michela Figorilli, Mario Meloni, Giuseppe Lanza, Elisa Casaglia, Rosamaria Lecca, Francesca Lea Saibene, Patrizia Congiu, and Monica Puligheddu. "Considering REM Sleep Behavior Disorder in the Management of Parkinson's Disease". In: *Nature and Science of Sleep* 15 (May 2023), pp. 333–352. ISSN: 1179-1608. DOI: 10.2147/NSS.S266071. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10167974/ (visited on 04/28/2024).

[Gro20] Priti Gros and Aleksandar Videnovic. "Overview of Sleep and Circadian Rhythm Disorders in Parkinson Disease". eng. In: *Clinics in Geriatric Medicine* 36.1 (Feb. 2020), pp. 119–130. ISSN: 1879-8853. DOI: 10.1016/j.cger.2019.09.005.

[Gu12] Changzhan Gu, Ruijiang Li, Hualiang Zhang, Albert Y. C. Fung, Carlos Torres, Steve B Jiang, and Changzhi Li. "Accurate Respiration Measurement Using DC-Coupled Continuous-Wave Radar Sensor for Motion-Adaptive Cancer Radiotherapy". In: *IEEE Transactions on Biomedical Engineering* 59.11 (Nov. 2012), pp. 3117–3123. ISSN: 1558-2531. DOI: 10.1109/TBME.2012.2206591.

[Hab23] Ahsan Habib, Mohammod Abdul Motin, Thomas Penzel, Marimuthu Palaniswami, John Yearwood, and Chandan Karmakar. "Performance of a Convolutional Neural Network Derived From PPG Signal in Classifying Sleep Stages". eng. In: *IEEE transactions on bio-medical engineering* 70.6 (June 2023), pp. 1717–1728. ISSN: 1558-2531. DOI: 10.1109/TBME.2022.3219863.

[Han21] Khadija Hanifi and M. Elif Karsligil. "Elderly Fall Detection With Vital Signs Monitoring Using CW Doppler Radar". In: *IEEE Sensors Journal* 21.15 (Aug.

2021). Conference Name: IEEE Sensors Journal, pp. 16969–16978. ISSN: 1558-1748. DOI: 10.1109/JSEN.2021.3079835.

[Hir16] Max Hirshkowitz. "Polysomnography Challenges". en. In: *Sleep Medicine Clinics* 11.4 (Dec. 2016), pp. 403–411. ISSN: 1556407X. DOI: 10.1016/j.jsmc.2016.07.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1556407X16300601 (visited on 10/30/2023).

[Hoc97] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Computation* 9.8 (Nov. 1997). Conference Name: Neural Computation, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. URL: https://ieeexplore.ieee.org/abstract/document/6795963 (visited on 06/01/2024).

[Hon18] Hong Hong, Li Zhang, Chen Gu, Yusheng Li, Guangxin Zhou, and Xiaohua Zhu. "Noncontact Sleep Stage Estimation Using a CW Doppler Radar". In: *IEEE Journal on Emerging and Selected Topics in Circuits and Systems* 8.2 (June 2018). Conference Name: IEEE Journal on Emerging and Selected Topics in Circuits and Systems, pp. 260–270. ISSN: 2156-3365. DOI: 10.1109/JETCAS.2017.2789278.

[Ing21] Thorir Mar Ingolfsson, Xiaying Wang, Michael Hersche, Alessio Burrello, Lukas Cavigelli, and Luca Benini. "ECG-TCN: Wearable Cardiac Arrhythmia Detection with a Temporal Convolutional Network". In: *2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS)*. June 2021, pp. 1–4. DOI: 10.1109/AICAS51828.2021.9458520. URL: https://ieeexplore.ieee.org/document/9458520 (visited on 06/06/2024).

[ITU20] ITU. *Radio Regulations*. en-US. 2020. URL: https://www.itu.int:443/en/publications/ITU-R/Pages/publications.aspx?parent=R-REG-RR-2020&media=electronic (visited on 04/03/2024).

[Jaf10] Behrouz Jafari and Vahid Mohsenin. "Polysomnography". In: *Clinics in Chest Medicine*. Sleep 31.2 (June 2010), pp. 287–297. ISSN: 0272-5231. DOI: 10.1016/j.ccm.2010.02.005. URL: https://www.sciencedirect.com/science/article/pii/S0272523110000286 (visited on 11/05/2023).

[Kag16] Masayuki Kagawa, Kazuki Suzumura, and Takemi Matsui. "Sleep stage classification by non-contact vital signs indices using Doppler radar sensors". In: *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. ISSN: 1558-4615. Aug. 2016, pp. 4913–4916. DOI: 10.1109/EMBC.2016.7591829.

[Kra24] Daniel Krauss, Lukas Engel, Tabea Ott, Johanna Bräunig, Robert Richer, Markus Gambietz, Nils Albrecht, Eva M. Hille, Ingrid Ullmann, Matthias Braun, Peter Dabrock, Alexander Kölpin, Anne D. Koelewijn, Bjoern M. Eskofier, and Martin Vossiek. "A Review and Tutorial on Machine Learning-Enabled Radar-Based Biomedical Monitoring". In: *IEEE Open Journal of Engineering in Medicine and Biology* (2024). Conference Name: IEEE Open Journal of Engineering in Medicine and Biology, pp. 1–22. ISSN: 2644-1276. DOI: 10.1109/OJEMB.2024.3397208. URL: https://ieeexplore.ieee.org/document/10520876 (visited on 05/17/2024).

[Küd23] Arne Küderle, Robert Richer, Raul C. Sîmpetru, and Bjoern M. Eskofier. "tpcp: Tiny Pipelines for Complex Problems - A set offramework independent helpers for algorithms development andevaluation". In: *Journal of Open Source Software* 8.82 (Feb. 2023), p. 4953. ISSN: 2475-9066. DOI: 10.21105/joss.04953. URL: https://joss.theoj.org/papers/10.21105/joss.04953 (visited on 11/13/2023).

[Lah24] Soraya Lahlou, Marta Kaminska, Julien Doyon, Julie Carrier, and Madeleine Sharp. "Sleep spindle density and temporal clustering are associated with sleep-dependent memory consolidation in Parkinson's disease". EN. In: *Journal of Clinical Sleep Medicine* (Mar. 2024). Publisher: American Academy of Sleep Medicine. DOI: 10.5664/jcsm.11080. URL: https://jcsm.aasm.org/doi/10.5664/jcsm.11080 (visited on 04/29/2024).

[Lam22] Richard N. L. Lamptey, Bivek Chaulagain, Riddhi Trivedi, Avinash Gothwal, Buddhadev Layek, and Jagdish Singh. "A Review of the Common Neurodegenerative Disorders: Current Therapeutic Approaches and the Potential Role of Nanotherapeutics". In: *International Journal of Molecular Sciences* 23.3 (Feb. 2022), p. 1851. ISSN: 1422-0067. DOI: 10.3390/ijms23031851. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8837071/ (visited on 01/30/2024).

[Lau20] Timo Lauteslager, Stylianos Kampakis, Adrian J. Williams, Michal Maslik, and Fares Siddiqui. "Performance Evaluation of the Circadia Contactless Breathing Monitor and Sleep Analysis Algorithm for Sleep Stage Classification". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. ISSN: 2694-0604. July 2020, pp. 5150–5153. DOI: 10.1109/EMBC44109.2020.9175419. URL: https://ieeexplore.ieee.org/abstract/document/9175419 (visited on 10/30/2023).

[Le 20] Olivier Le Bon. "Relationships between REM and NREM in the NREM-REM sleep cycle: a review on competing concepts". In: *Sleep Medicine* 70 (June 2020), pp. 6–16. ISSN: 1389-9457. DOI: 10.1016/j.sleep.2020.02.004. URL: https://www.sciencedirect.com/science/article/pii/S1389945720300757 (visited on 11/05/2023).

[Lee24] Ji Hyun Lee, Hyunwoo Nam, Dong Hyun Kim, Dae Lim Koo, Jae Won Choi, Seung-No Hong, Eun-Tae Jeon, Sungmook Lim, Gwang soo Jang, and Baek-hyun Kim. "Developing a deep learning model for sleep stage prediction in obstructive sleep apnea cohort using 60 GHz frequency-modulated continuous-wave radar". en. In: *Journal of Sleep Research* 33.1 (2024). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/jsr.14050, e14050. ISSN: 1365-2869. DOI: 10.1111/jsr.14050. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/jsr.14050 (visited on 02/13/2024).

[Ler21] Itamar Lerner, Shira M. Lupkin, Alan Tsai, Anosha Khawaja, and Mark A. Gluck. "Sleep to remember, sleep to forget: Rapid eye movement sleep can have inverse effects on recall and generalization of fear memories". In: *Neurobiology of Learning and Memory* 180 (Apr. 2021), p. 107413. ISSN: 1074-7427. DOI: 10.1016/j.nlm.2021.107413. URL: https://www.sciencedirect.com/science/article/pii/S1074742721000356 (visited on 04/29/2024).

[Lew21] Laura D. Lewis. "The interconnected causes and consequences of sleep in the brain". eng. In: *Science (New York, N.Y.)* 374.6567 (Oct. 2021), pp. 564–568. ISSN: 1095-9203. DOI: 10.1126/science.abi8375.

[Li18] Meiyu Li and Jenshan Lin. "Wavelet-Transform-Based Data-Length-Variation Technique for Fast Heart Rate Detection Using 5.8-GHz CW Doppler Radar". In: *IEEE Transactions on Microwave Theory and Techniques* 66.1 (Jan. 2018), pp. 568–576. ISSN: 0018-9480, 1557-9670. DOI: 10.1109/TMTT.2017.2730182. URL: http://ieeexplore.ieee.org/document/7997608/ (visited on 12/22/2023).

[Lin21] Yang Lin, Irena Koprinska, and Mashud Rana. "Temporal Convolutional Attention Neural Networks for Time Series Forecasting". In: *2021 International Joint Conference on Neural Networks (IJCNN)*. ISSN: 2161-4407. July 2021, pp. 1–8. DOI: 10.1109/IJCNN52387.2021.9534351. URL: https://ieeexplore.ieee.org/abstract/document/9534351 (visited on 06/01/2024).

[Liu19] Chien-Liang Liu, Wen-Hoar Hsaio, and Yao-Chung Tu. "Time Series Classification With Multivariate Convolutional Neural Network". In: *IEEE Transactions on Industrial Electronics* 66.6 (June 2019). Conference Name: IEEE Transactions on Industrial Electronics, pp. 4788–4797. ISSN: 1557-9948. DOI: 10.1109/TIE.2018.2864702. URL: https://ieeexplore.ieee.org/abstract/document/8437249 (visited on 02/09/2024).

[Mag21] Gianpaolo Maggi, Luigi Trojano, Paolo Barone, and Gabriella Santangelo. "Sleep Disorders and Cognitive Dysfunctions in Parkinson's Disease: A Meta-Analytic Study". eng. In: *Neuropsychology Review* 31.4 (Dec. 2021), pp. 643–682. ISSN: 1573-6660. DOI: 10.1007/s11065-020-09473-1.

[Mak21] Dominique Makowski, Tam Pham, Zen J. Lau, Jan C. Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and S. H. Annabel Chen. "NeuroKit2: A Python toolbox for neurophysiological signal processing". en. In: *Behavior Research Methods* 53.4 (Aug. 2021), pp. 1689–1696. ISSN: 1554-3528. DOI: 10.3758/s13428-020-01516-y. URL: https://doi.org/10.3758/s13428-020-01516-y (visited on 11/19/2023).

[Mal18] John Malik, Yu-Lun Lo, and Hau-tieng Wu. "Sleep-wake classification via quantifying heart rate variability by convolutional neural network". en. In: *Physiological Measurement* 39.8 (Aug. 2018). Publisher: IOP Publishing, p. 085004. ISSN: 0967-3334. DOI: 10.1088/1361-6579/aad5a9. URL: https://dx.doi.org/10.1088/1361-6579/aad5a9 (visited on 02/21/2024).

[Mar20] Joshua M. Martin, Danyal Wainstein Andriano, Natalia B. Mota, Sergio A. Mota-Rolim, John Fontenele Araújo, Mark Solms, and Sidarta Ribeiro. "Structural differences between REM and non-REM dream reports assessed by graph analysis". en. In: *PLOS ONE* 15.7 (July 2020). Publisher: Public Library of Science, e0228903. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0228903. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0228903 (visited on 04/29/2024).

[Mas24] Ihssan S. Masad, Amin Alqudah, and Shoroq Qazan. "Automatic classification of sleep stages using EEG signals and convolutional neural networks". en. In: *PLOS ONE* 19.1 (Jan. 2024). Publisher: Public Library of Science, e0297582. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0297582. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0297582 (visited on 02/13/2024).

[Mil22] Dean J. Miller, Charli Sargent, and Gregory D. Roach. "A Validation of Six Wearable Devices for Estimating Sleep, Heart Rate and Heart Rate Variability in Healthy Adults". eng. In: *Sensors (Basel, Switzerland)* 22.16 (Aug. 2022), p. 6317. ISSN: 1424-8220. DOI: 10.3390/s22166317.

[Muñ17] José-María Muñoz-Ferreras, Zhengyu Peng, Roberto Gómez-García, and Changzhi Li. "Review on Advanced Short-Range Multimode Continuous-Wave Radar Architectures for Healthcare Applications". In: *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology* 1.1 (June 2017). Conference Name: IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology, pp. 14–25. ISSN: 2469-7257. DOI: 10.1109/JERM.2017.2735241. URL: https://ieeexplore.ieee.org/abstract/document/8000316 (visited on 10/30/2023).

[Par23] Madan Parajuli, Amy W. Amara, and Mohamed Shaban. "Deep-learning detection of mild cognitive impairment from sleep electroencephalography for patients with Parkinson's disease". en. In: *PLOS ONE* 18.8 (Aug. 2023). Publisher: Public Library of Science, e0286506. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0286506. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0286506 (visited on 10/26/2023).

[Par24] Jonghyun Park, Seungman Yang, Gihsung Chung, Ivo Junior Leal Zanghettin, and Jonghee Han. "Ultra-wideband Radar-based Sleep Stage Classification in Smartphone using an End-to-end Deep Learning". In: *IEEE Access* (2024). Conference Name: IEEE Access, pp. 1–1. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2024.3390391. URL: https://ieeexplore.ieee.org/abstract/document/10504259 (visited on 04/29/2024).

[Pat10] Susheel P. Patil. "What every clinician should know about polysomnography". eng. In: *Respiratory Care* 55.9 (Sept. 2010), pp. 1179–1195. ISSN: 0020-1324.

[Pat23] Giacomo Paterniani, Daria Sgreccia, Alessandro Davoli, Giorgio Guerzoni, Pasquale Di Viesti, Anna Chiara Valenti, Marco Vitolo, Giorgio M. Vitetta, and Giuseppe Boriani. "Radar-Based Monitoring of Vital Signs: A Tutorial Overview". In: *Proceedings of the IEEE* 111.3 (Mar. 2023). Conference Name: Proceedings of the IEEE, pp. 277–317. ISSN: 1558-2256. DOI: 10.1109/JPROC.2023.3244362. URL: https://ieeexplore.ieee.org/abstract/document/10049295 (visited on 02/05/2024).

[Ped11] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent

Dubourg, Jake Vanderplas, Alexandre Passos, and David Cournapeau. “Scikit-learn: Machine Learning in Python”. en. In: *MACHINE LEARNING IN PYTHON* (2011).

[Pha21] Tam Pham, Zen Juen Lau, S. H. Annabel Chen, and Dominique Makowski. “Heart Rate Variability in Psychology: A Review of HRV Indices and an Analysis Tutorial”. en. In: *Sensors* 21.12 (Jan. 2021). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 3998. ISSN: 1424-8220. DOI: 10.3390/s21123998. URL: https://www.mdpi.com/1424-8220/21/12/3998 (visited on 01/09/2024).

[Pos19] Ronald B. Postuma, Alex Iranzo, Michele Hu, Birgit Högl, Bradley F. Boeve, Raffaele Manni, Wolfgang H. Oertel, Isabelle Arnulf, Luigi Ferini-Strambi, Monica Puligheddu, Elena Antelmi, Valerie Cochen De Cock, Dario Arnaldi, Brit Mollenhauer, Aleksandar Videnovic, Karel Sonka, Ki-Young Jung, Dieter Kunz, Yves Dauvilliers, Federica Provini, Simon J. Lewis, Jitka Buskova, Milena Pavlova, Anna Heidbreder, Jacques Y. Montplaisir, Joan Santamaria, Thomas R. Barber, Ambra Stefani, Erik K. St Louis, Michele Terzaghi, Annette Janzen, Smandra Leu-Semenescu, Guiseppe Plazzi, Flavio Nobili, Friederike Sixel-Doering, Petr Dusek, Frederik Bes, Pietro Cortelli, Kaylena Ehgoetz Martens, Jean-Francois Gagnon, Carles Gaig, Marco Zucconi, Claudia Trenkwalder, Ziv Gan-Or, Christine Lo, Michal Rolinski, Philip Mahlknecht, Evi Holzknecht, Angel R. Boeve, Luke N. Teigen, Gianpaolo Toscano, Geert Mayer, Silvia Morbelli, Benjamin Dawson, and Amelie Pelletier. “Risk and predictors of dementia and parkinsonism in idiopathic REM sleep behaviour disorder: a multicentre study”. eng. In: *Brain: A Journal of Neurology* 142.3 (Mar. 2019), pp. 744–759. ISSN: 1460-2156. DOI: 10.1093/brain/awz030.

[Pyt24] Pytorch. *torch.nn — PyTorch 2.3 documentation*. Docs. May 2024. URL: https://pytorch.org/docs/stable/nn.html (visited on 05/24/2024).

[Rah19] Habibur Rahman. *Fundamental Principles of Radar*. en. Google-Books-ID: thhOwQEACAAJ. CRC Press, 2019. ISBN: 978-1-138-38779-9.

[Ric21] Robert Richer, Arne Küderle, Martin Ullrich, Nicolas Rohleder, and Bjoern Eskofier. “BioPsyKit: A Python package for the analysis of biopsychological data”. en. In: *Journal of Open Source Software* 6.66 (Oct. 2021), p. 3702. ISSN: 2475-9066. DOI: 10.21105/joss.03702. URL: https://joss.theoj.org/papers/10.21105/joss.03702 (visited on 11/19/2023).

[Sak23] Jill Sakai. “How sleep shapes what we remember—and forget”. In: *Proceedings of the National Academy of Sciences* 120.2 (Jan. 2023). Publisher: Proceedings of

the National Academy of Sciences, e2220275120. DOI: 10.1073/pnas.2220275120. URL: https://www.pnas.org/doi/abs/10.1073/pnas.2220275120 (visited on 04/29/2024).

[Sar17] Susan J. Sara. "Sleep to Remember". en. In: *Journal of Neuroscience* 37.3 (Jan. 2017). Publisher: Society for Neuroscience Section: Dual Perspectives, pp. 457–463. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0297-16.2017. URL: https://www.jneurosci.org/content/37/3/457 (visited on 03/05/2024).

[Sch17] Anthony H. V. Schapira, K. Ray Chaudhuri, and Peter Jenner. "Non-motor features of Parkinson disease". eng. In: *Nature Reviews. Neuroscience* 18.7 (July 2017), pp. 435–450. ISSN: 1471-0048. DOI: 10.1038/nrn.2017.62.

[Sch20] Daniel Schmid and Daniel Erlacher. "Lucid dream induction by auditory stimulation and reality testing during early-morning sleep". en. In: *International Journal of Dream Research* (Mar. 2020), pp. 99–104. ISSN: 1866-7953. DOI: 10.11588/ijodr.2020.1.71695. URL: https://journals.ub.uni-heidelberg.de/index.php/IJoDR/article/view/71695 (visited on 04/29/2024).

[sci24] scikit. *3.4. Metrics and scoring: quantifying the quality of predictions*. en. Docs. 2024. URL: https://scikit-learn/stable/modules/model_evaluation.html (visited on 07/04/2024).

[Sco21] Alexander J. Scott, Thomas L. Webb, Marrissa Martyn-St James, Georgina Rowse, and Scott Weich. "Improving sleep quality leads to better mental health: A meta-analysis of randomised controlled trials". In: *Sleep Medicine Reviews* 60 (Dec. 2021), p. 101556. ISSN: 1087-0792. DOI: 10.1016/j.smrv.2021.101556. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8651630/ (visited on 10/26/2023).

[Scu15] Michael K. Scullin and Donald L. Bliwise. "Sleep, Cognition, and Normal Aging: Integrating a Half Century of Multidisciplinary Research". en. In: *Perspectives on Psychological Science* 10.1 (Jan. 2015). Publisher: SAGE Publications Inc, pp. 97–137. ISSN: 1745-6916. DOI: 10.1177/1745691614556680. URL: https://doi.org/10.1177/1745691614556680 (visited on 06/02/2024).

[Sej22] Monika Sejbuk, Iwona Mirończuk-Chodakowska, and Anna Maria Witkowska. "Sleep Quality: A Narrative Review on Nutrition, Stimulants, and Physical Activity as Important Factors". eng. In: *Nutrients* 14.9 (May 2022), p. 1912. ISSN: 2072-6643. DOI: 10.3390/nu14091912.

[Ser15] Ioffe Sergey and Christian Szegedy. “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”. In: *Proceedings of the 32nd International Conference on Machine Learning*. Vol. 37. Proceedings of Machine Learning Research. Lille, France: PMLR, 2015, pp. 448–456. URL: https://arxiv.org/pdf/1502.03167 (visited on 05/03/2024).

[Spr13] Fabienne Sprenger and Werner Poewe. “Management of Motor and Non-Motor Symptoms in Parkinson’s Disease”. en. In: *CNS Drugs* 27.4 (Apr. 2013), pp. 259–272. ISSN: 1179-1934. DOI: 10.1007/s40263-013-0053-2. URL: https://doi.org/10.1007/s40263-013-0053-2 (visited on 02/08/2024).

[Sti05] Robert Stickgold. “Sleep-dependent memory consolidation”. en. In: *Nature* 437.7063 (Oct. 2005). Publisher: Nature Publishing Group, pp. 1272–1278. ISSN: 1476-4687. DOI: 10.1038/nature04286. URL: https://www.nature.com/articles/nature04286 (visited on 04/30/2024).

[Sve16] Sigurlaug Sveinbjornsdottir. “The clinical symptoms of Parkinson’s disease”. eng. In: *Journal of Neurochemistry* 139 Suppl 1 (Oct. 2016), pp. 318–324. ISSN: 1471-4159. DOI: 10.1111/jnc.13691.

[Tof20] Ståle Toften, Ståle Pallesen, Maria Hrozanova, Frode Moen, and Janne Grønli. “Validation of sleep stage classification using non-contact radar technology and machine learning (Somnofy®)”. eng. In: *Sleep Medicine* 75 (Nov. 2020), pp. 54–61. ISSN: 1878-5506. DOI: 10.1016/j.sleep.2020.02.022.

[Tol21] Eduardo Tolosa, Alicia Garrido, Sonja W. Scholz, and Werner Poewe. “Challenges in the diagnosis of Parkinson’s disease”. eng. In: *The Lancet. Neurology* 20.5 (May 2021), pp. 385–397. ISSN: 1474-4465. DOI: 10.1016/S1474-4422(21)00030-2.

[Vir20] Pauli Virtanen et al. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. en. In: *Nature Methods* 17.3 (Mar. 2020), pp. 261–272. ISSN: 1548-7091, 1548-7105. DOI: 10.1038/s41592-019-0686-2. URL: https://www.nature.com/articles/s41592-019-0686-2 (visited on 07/17/2024).

[Wan12] Jianqing Wang and Qiong Wang. *Body Area Communications: Channel Modeling, Communication Systems, and EMC*. en. 1st ed. Wiley, Nov. 2012. ISBN: 978-1-118-18848-4 978-1-118-18849-1. DOI: 10.1002/9781118188491. URL: https://onlinelibrary.wiley.com/doi/book/10.1002/9781118188491 (visited on 04/04/2024).

[Wan14] Guochao Wang, José-María Muñoz-Ferreras, Changzhan Gu, Changzhi Li, and Roberto Gómez-García. "Application of Linear-Frequency-Modulated Continuous-Wave (LFMCW) Radars for Tracking of Vital Signs". In: *IEEE Transactions on Microwave Theory and Techniques* 62.6 (June 2014). Conference Name: IEEE Transactions on Microwave Theory and Techniques, pp. 1387–1399. ISSN: 1557-9670. DOI: 10.1109/TMTT.2014.2320464. URL: https://ieeexplore.ieee.org/abstract/document/6810197 (visited on 10/30/2023).

[Wan22] Zixia Wang, Shuai Zha, Baoxian Yu, Pengbin Chen, Zhiqiang Pang, and Han Zhang. "Sleep Staging Using Noncontact-Measured Vital Signs". en. In: *Journal of Healthcare Engineering* 2022 (July 2022). Publisher: Hindawi, e2016598. ISSN: 2040-2295. DOI: 10.1155/2022/2016598. URL: https://www.hindawi.com/journals/jhe/2022/2016598/ (visited on 02/02/2024).

[Wat10] G. Stennis Watson and James B. Leverenz. "Profile of Cognitive Impairment in Parkinson's Disease". en. In: *Brain Pathology* 20.3 (2010). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1750-3639.2010.00373.x, pp. 640–645. ISSN: 1750-3639. DOI: 10.1111/j.1750-3639.2010.00373.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1750-3639.2010.00373.x (visited on 03/05/2024).

[Wei22] Daniel Weintraub, Dag Aarsland, Kallol Ray Chaudhuri, Roseanne D. Dobkin, Albert Fg Leentjens, Mayela Rodriguez-Violante, and Anette Schrag. "The neuropsychiatry of Parkinson's disease: advances and challenges". eng. In: *The Lancet. Neurology* 21.1 (Jan. 2022), pp. 89–102. ISSN: 1474-4465. DOI: 10.1016/S1474-4422(21)00330-6.

[Wen23] Marvin Wenzel, Nils C. Albrecht, Dominik Langer, Markus Heyder, and Alexander Koelpin. "Catch Your Breath! Vital Sign Sensing With Radar". In: *IEEE Microwave Magazine* 24.3 (Mar. 2023). Conference Name: IEEE Microwave Magazine, pp. 75–82. ISSN: 1557-9581. DOI: 10.1109/MMM.2022.3226546. URL: https://ieeexplore.ieee.org/abstract/document/10035798 (visited on 04/04/2024).

[WHO24a] WHO. *Parkinson disease*. en. 2024. URL: https://www.who.int/news-room/fact-sheets/detail/parkinson-disease (visited on 04/28/2024).

[WHO24b] WHO. *World Health Statistics 2023 - A visual summary*. en. 2024. URL: https://www.who.int/data/stories/world-health-statistics-2023-a-visual-summary/ (visited on 04/28/2024).

[Yan18] Zhicheng Yang, Maurizio Bocca, Vivek Jain, and Prasant Mohapatra. "Contactless Breathing Rate Monitoring in Vehicle Using UWB Radar". In: *Proceedings of the 7th International Workshop on Real-World Embedded Wireless Systems and Networks*. RealWSN'18. New York, NY, USA: Association for Computing Machinery, Nov. 2018, pp. 13–18. ISBN: 978-1-4503-6048-7. DOI: 10.1145/3277883.3277884. URL: https://doi.org/10.1145/3277883.3277884 (visited on 03/21/2023).

[Yav16] Ehsan Yavari, Olga Boric-Lubecke, and Shuhei Yamada. "Radar Principles". en. In: (Jan. 2016). Book Title: Doppler Radar Physiological Sensing ISBN: 9781119078418 Publisher: John Wiley & Sons, Ltd, pp. 21–38. DOI: 10.1002/9781119078418.ch2. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781119078418.ch2 (visited on 11/04/2023).

[Zha17] Bendong Zhao, Huanzhang Lu, Shangfeng Chen, Junliang Liu, and Dongya Wu. "Convolutional neural networks for time series classification". In: *Journal of Systems Engineering and Electronics* 28.1 (Feb. 2017). Conference Name: Journal of Systems Engineering and Electronics, pp. 162–169. ISSN: 1004-4132. DOI: 10.21629/JSEE.2017.01.18. URL: https://ieeexplore.ieee.org/abstract/document/7870510 (visited on 01/28/2024).

[Zhu20] Tianqi Zhu, Wei Luo, and Feng Yu. "Convolution- and Attention-Based Neural Network for Automated Sleep Stage Classification". en. In: *International Journal of Environmental Research and Public Health* 17.11 (Jan. 2020). Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, p. 4152. ISSN: 1660-4601. DOI: 10.3390/ijerph17114152. URL: https://www.mdpi.com/1660-4601/17/11/4152 (visited on 02/13/2024).

[Zuz20] José Rafael P. Zuzuárregui and Emmanuel H. During. "Sleep Issues in Parkinson's Disease and Their Management". eng. In: *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics* 17.4 (Oct. 2020), pp. 1480–1494. ISSN: 1878-7479. DOI: 10.1007/s13311-020-00938-y.

Appendix D

Acronyms

DNN deep neural network

DL deep learning

ML machine learning

WHO World Health Organization

REM rapid eye movement

NREM non rapid eye movement

EEG electroencephalography

PSG polysomnography

EMG electromyography

EOG electrooculography

ECG electrocardiography

SW slow wave

PD Parkinson's disease

CNN convolutional neural network

HRV heart rate variability

MLP multilayer perceptron,

SVM support vector machine,

TCN temporal convolutional network

LSTM long short-term memory.

CW continuous wave

ITU International Telecommunication Union

ISM industrial, scientific and medical

RBD REM sleep behaviour disorder

LBD Lewy body dementia

MSA multiple system atrophy

iRBD isolated RBD

RF random forest

BN batch norm

FC fully connected

TCNN time convolutional neural network

ReLU rectified linear unit function

HRV heart rate variability

IMU inertia measurement unit

Std standard deviation