공학박사 학위논문

Non-Contact Sleep Monitoring Using Impulse Radio Ultra-Wideband Radar Based on Long Short-Term Memory Network

장단기 메모리 네트워크 기반의 초광대역 레이더를
이용한 비접촉식 수면 모니터링 기술 개발

2021년 8월

서울대학교 대학원
협동과정 바이오엔지니어링 전공
권 현 빈

Non-Contact Sleep Monitoring Using
Impulse Radio Ultra-Wideband Radar Based on
Long Short-Term Memory Network

장단기 메모리 네트워크 기반의 초광대역
레이더를 이용한 비접촉식 수면 모니터링 기술 개발

지도교수 박 광 석

이 논문을 공학박사 학위논문으로 제출함

2021년 8월

서울대학교 대학원
협동과정 바이오엔지니어링 전공
권 현 빈

권현빈의 공학박사 학위논문을 인준함

2021년 6월

위 원 장	최 진 욱
부 위 원 장	박 광 석
위 원	Sungwan Kim
위 원	이 태 수
위 원	박 철 수

Ph.D. Dissertation

Non-Contact Sleep Monitoring Using Impulse Radio Ultra-Wideband Radar Based on Long Short-Term Memory Network

August 2021

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

Hyun Bin Kwon

Non-Contact Sleep Monitoring Using Impulse Radio Ultra-Wideband Radar Based on Long Short-Term Memory Network

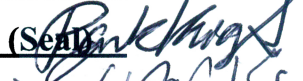*Academic advisor* Kwang Suk Park

Submitting a Ph.D. Dissertation

August 2021

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

Hyun Bin Kwon

Confirming the Ph.D. Dissertation written by

Hyun Bin Kwon

June 2021

Chair	**Choi, Jinwook, M.D., Ph.D.**	(Seal)
Vice Chair	**Park, Kwang Suk, Ph.D.**	(Seal)
Examiner	**Kim, Sungwan, Ph.D.**	(Seal)
Examiner	**Lee, Taesoo, Ph.D.**	(Seal)
Examiner	**Park, Cheol Soo, Ph.D.**	(Seal)

Abstract

Non-Contact Sleep Monitoring Using Impulse Radio Ultra-Wideband Radar Based on Long Short-Term Memory Network

Hyun Bin Kwon

Interdisciplinary Program in Bioengineering

The Graduate School

Seoul National University

Manual scoring of sleep stages and sleep events from polysomnography (PSG) records is essential to understand the sleep architecture and sleep-related breathing disorder. Since the PSG requires a specialized expert, a lab environment, and obtrusive sensors, non-contact methods for sleep staging and sleep disorder monitoring using impulse-radio ultra-wideband (IR-UWB) radar have been investigated over the past years. However, the existing sleep staging models need to be improved interpretability and reliability. Moreover, the existing methods for

monitoring SAHS (sleep apnea hypopnea syndrome), one of the most common sleep disorders, have limitations in that the radar-person distance is fixed, and the detected apnea hypopnea (AH) event cannot be provided in real-time.

In sleep staging study, an attention-based bidirectional long short-term memory (attention-based Bi-LSTM) model for automatic sleep stage scoring using an IR-UWB radar which can remotely detect vital signs is proposed. Sixty-five young (30.0 ± 8.6 yrs.) and healthy volunteers underwent nocturnal PSG and IR-UWB radar measurement simultaneously. However, the 14 subjects were excluded because they had sleep apnea, periodic limb movement, and abnormal sleep. Among 51 recordings, 26 were used for training, 8 for validation, and 17 for testing. Sixteen features including movement-, respiration-, and heart rate variability-related indices were extracted from the raw IR-UWB signals in each 30-s epoch. Sleep stage classification performances of attention-based Bi-LSTM model with optimized hyperparameters were evaluated and compared with those of the conventional LSTM networks for the same test dataset. In the results, an accuracy of 82.6 ± 6.7% and a Cohen's kappa coefficient of 0.73 ± 0.11 for the classification of wake, REM sleep, light (N1+N2) sleep, and deep (N3) sleep which is significantly higher than those using the conventional LSTM networks ($p < 0.01$) were achieved. Moreover, the classification performances were higher than those reported in the comparative studies.

In the SAHS detection study, a novel approach for real-time AH event detection with IR-UWB radar using a deep learning model is proposed. Thirty six PSG recordings and simultaneously measured IR-UWB radar data were used in the experiments. After the clutter was removed, IR-UWB radar images were segmented

by sliding a 20-s window at 1-s shift, and categorized into two classes: AH and N (Normal). A hybrid model combining the convolutional neural networks and LSTM (CNN-LSTM) was trained with the data, which consisted of class-balanced segments. Time sequenced outputs were then fed to an event detector in order to identify valid AH events. Therefore, the proposed method showed a Cohen's kappa coefficient of 0.728, sensitivity of 0.781, specificity of 0.956, and an accuracy of 0.930. According to the apnea-hypopnea index (AHI) estimation analysis, the Pearson correlation coefficient between the estimated AHI and reference AHI was 0.97. In addition, the average accuracy and kappa of SAHS diagnosis was 0.98 and 0.96, respectively, for AHI cutoffs of $\geq$ 5, 15, and 30 events/h. The proposed method achieved the state-of-the-art performance for classifying SAHS severity without any hand-engineered feature regardless of the user's location.

The experimental results demonstrated that the effectiveness of the LSTM networks which coupled with attention mechanism and CNN model for sleep monitoring. Our approaches can be utilized for a cost-effective and reliable sleep monitoring system in a home environment.

Table of Contents

Chapter 3. LSTM model for Real-Time Apnea-Hypopnea Event Detection Based on IR-UWB Radar 57

List of Tables

List of Figures

List of Abbreviations

AASM	American academy of sleep medicine
AH	Apnea and hypopnea
AHI	Apnea hypopnea index
ANS	Autonomic nervous system
AB-LSTM	Attention-based bidirectional LSTM networks
BCG	Ballistocardiography
BMI	Body mass index
CFAR	Constant false alarm rate
CPAP	Continuous positive airway pressure
CNN	Convolutional neural network
CSA	Central sleep apnea
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
FC	Fully connected
FCC	Federal communications commission
FFT	Fast Fourier transform
GDV	General discrimination value
HF	High-frequency
HR	Heart rate

HRV	Heart rate variability
IR-UWB	Impulse-radio ultra-wideband
KAPPA	Cohen's kappa coefficient value
kNN	k-nearest neighbor
LSTM	Long short-term memory
LF	Low-frequency
LFHF	Ratio of the LF power to the HF power
mHR	Mean of HR
NREM	Non-rapid eye movement
OSA	Obstructive sleep apnea
PSG	Polysomnography
PPG	Photoplethysmogram
PLMI	Periodic limb movement index
PVDF	Polyvinylidence fluoride
RF	Radio frequency
RandF	Random forests
R&K	Rechtschaffen & Kales
RNN	Recurrent neural network
REM	Rapid eye movement
SAHS	Sleep apnea hypopnea syndrome
SD	Standard deviation
SOL	Sleep onset latency
SWA	Slow wave activity
SWS	Slow wave sleep

TCN	Temporal convolutional networks
TRT	Total recording time
TST	Total sleep time
t-SNE	t-distributed neighbor embedding

1
Introduction

1.1. Background

Humans sleep almost one-third of their lifetimes. Numerous researches have reported that sleep plays an important role in our physical and mental health and our daily functioning in many ways. Sleep has been proposed as serving an energy-saving function [1], the restoration of our tissues [2], thermoregulation [3], metabolic regulation [4], adaptive immune functions [5], emotion regulation [6], and memory consolidation [7]. In contrast, sleep deprivation and sleep disorders negatively influence mood [8], cognitive performance, and motor function [9, 10] and increase the risk of cardiovascular diseases [11-13], obesity [14] and diabetes [15]. Therefore, continuous monitoring sleep in daily life is very important for a healthy life.

Polysomnography (PSG) is the gold standard for analyzing human sleep. PSG results include sleep duration, sleep stages, sleep-related disorders, sleep fragmentation, and sleep quality. For accurate sleep monitoring, PSG typically requires recordings of multichannel biomedical signals including an electro-

encephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), respiration signals, and blood oxygen saturation. Therefore, PSG can only be conducted by sleep technicians in laboratories equipped to carry out these tests [16]. Moreover, because the sleep technicians score sleep stages and sleep events visually, it is subjective and may have human errors. Several previous studies reported that manual scoring of sleep stages and sleep events exhibit inter-rater variability [17, 18]. In addition, as shown in Fig. 1-1, a large number of electrodes have to be attached to patients that may cause discomfort and affect their sleeping behavior [19]; thus, it is not suitable for long-term continuous sleep monitoring.
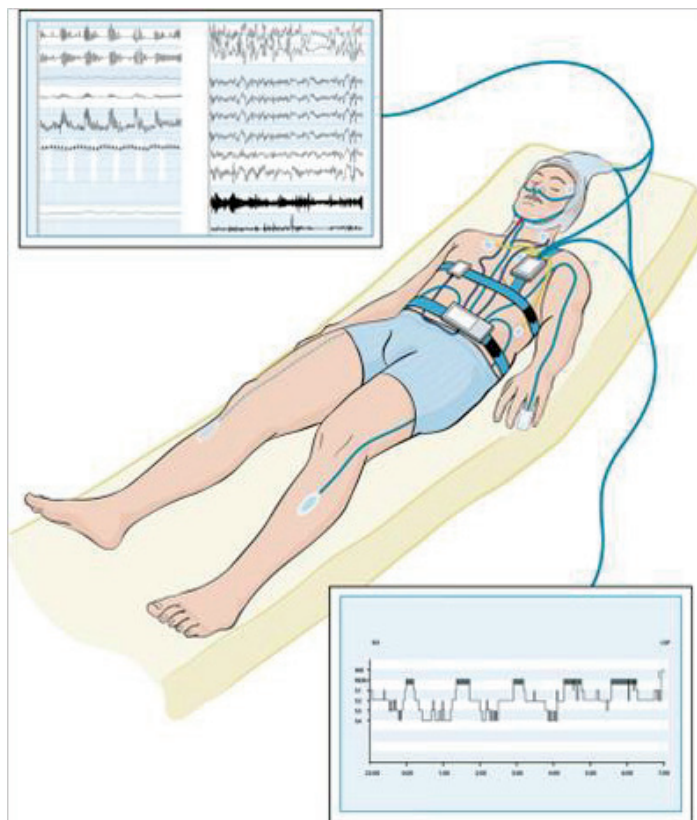
Figure 1-1. Sensors for standard PSG (*Source:* **Aystesis.com**)

1.2. Sleep Architecture

Normal human sleep consists of two core sleep stages: rapid eye movement (REM) sleep and non-rapid eye movement (NREM) sleep, which alternate in a cyclic manner. Sleep stage scoring rules for normal human were first defined by Rechtschaffen & Kales (R&K) in 1968 [20] and adapted by the American Academy of Sleep Medicine (AASM) [21]. Sleep stages are scored by different characteristics of EEG, EOG, and EMG. NREM and REM sleep alternate through the night at about 90-minutes (approximately 4-6 NREM-REM cycles during night sleep). According to the AASM manual, NREM sleep is further divided into N1, N2, and N3 sleep stages as shown in Fig. 1-2 (a). The distribution of each sleep stage across a sleep period is called as sleep architecture and it is visualized in the form of a hypnogram (Fig. 1-2 (b)). N1 sleep stage (2-5% of sleep time) is the lightest sleep stage and is a transition state from wakefulness to sleep. It is defined by attenuated EEG alpha rhythm (8–13 Hz) and the appearance of low-amplitude mixed-frequency EEG activity. N2 sleep stage (45-55% of sleep time) is characterized by the appearance of sleep spindle and K-complex. Sleep spindle is a train of distinct sinusoidal waves with frequency 11–16 Hz lasting at least 0.5 s. K-complex is a negative sharp wave followed by a positive wave lasting more than 0.5 s. N1 and N2 sleep stages are called as light sleep. N3 sleep stage (13-23% of sleep time) is the deepest sleep stage and is called deep sleep or slow wave sleep (SWS). It is defined by high amplitude slow wave activity (SWA, 0.5–4 Hz) accounting for more than 20% of an epoch (30 s). REM sleep stage (20-25% of sleep time) is associated with REM in the EOG, low-amplitude and mixed-frequency EEG, and low chin EMG tone.
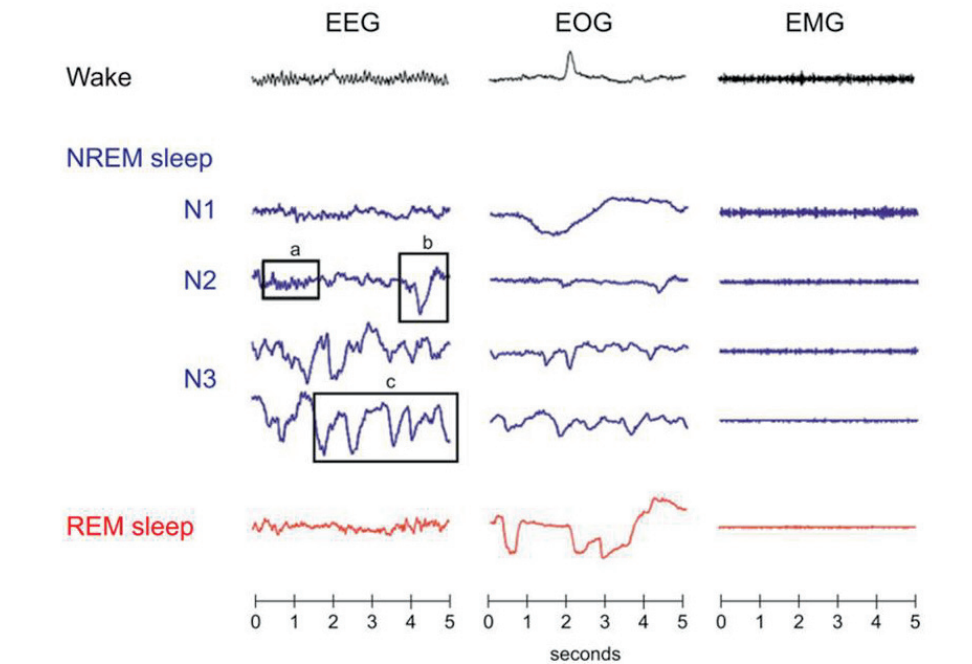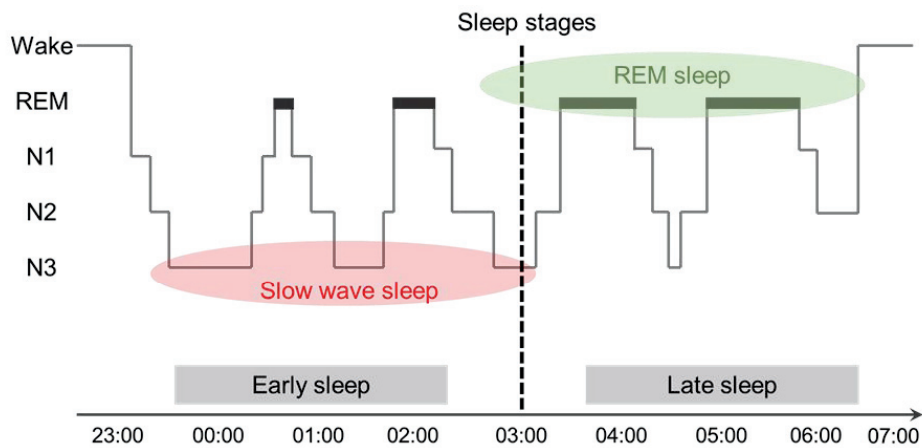
Figure 1-2. (a) EEG, EOG, and EMG signals during wake and each sleep stage. a: Sleep spindle, b: K-complex, c: Slow oscillation, (b) Hypnogram with sleep stages (Wake: wakefulness, REM: REM sleep, N1-N3: NREM sleep stages).

1.3. Sleep Apnea Hypopnea Syndrome

Sleep apnea and hypopnea syndrome (SAHS) is the most common sleep-related breathing disorder in the general population and is caused by partial or complete obstruction of the upper airway [22]. This disorder is characterized by repetitive events in which breathing is shallow or paused during sleep [23]. Apnea event is defined as the absence of tidal volume for more than 10 s and hypopnea is defined as a decrease in the tidal volume of more than 30% for at least 10 s that is accompanied by at least a 3% decrease in oxygen saturation or terminated by an arousal from sleep [21]. Apnea events are classified as 'obstructive' if they are accompanied by inspiratory effort against the occluded pharynx and 'central' if they are not. The apnea-hypopnea index (AHI) is the number of apnea and hypopnea (AH) events per hour of sleep. SAHS severity can be classified according to the AHI; non-SAHS is defined as an AHI of < 5, mild SAHS as an AHI of 5-15, moderate SAHS as an AHI of 15-30, and severe SAHS as an AHI > 30. Fig. 1-3 shows the mechanism of SAHS event.
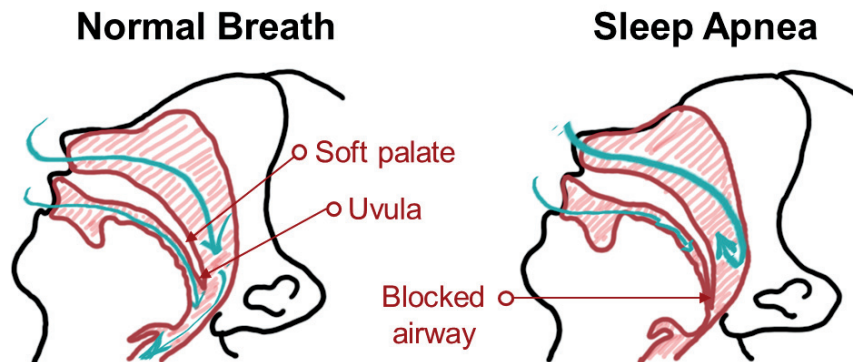
Figure 1-3. The mechanism of sleep apnea event

These events are typically accompanied by blood oxygen desaturation and arousals during sleep, leading to daytime sleepiness, decreased cognitive function and negative mood [24, 25]. Moreover, SAHS is known as a risk factor for several complications including hypertension, type 2 diabetes, cardiovascular disease, stroke, and heart failure in untreated patients [26, 27]. In previous study, Peppard *et al.* [28] reported that approximately 14% of men and 5% of women in population of Wisconsin adults affected by SAHS. Similarly, 27% of men and 17 % of women in the middle-aged Korean population were found to have an AHI of 5 or more [29]. Moreover, Kang *et al.* [30] showed that the prevalence of high-risk group of SAHS was 12.4 % in Korean adults by questionnaire-based study. Hence, monitoring for unrecognized SAHS and appropriate treatment allows preventative measurement to reduce these potential health problems and death.

1.4. Impulse-Radio Ultra-Wideband Radar

Impulse-radio ultra-wideband (IR-UWB) is a short-range wireless communication technique operating in the range of 3.1–10.6 GHz (Fig. 1-4). UWB technology was developed for military purpose, but it was released as a private technology by the federal communications commission (FCC) in 2002.

Due to the wide bandwidth and narrower impulse signals, IR-UWB radars have the following advantages compared to other radio sensors (i.e., Wi-Fi, Bluetooth, and Doppler radar) [31]: (1) high speed communication over short distances; (2) strong material penetration capabilities; (3) robustness to multipath propagation; and (4) high spatial resolution. Therefore, various applications have been demonstrated using IR-UWB radars, such as remotely detecting vital signs, *i.e.,* respiration and heart rate by measuring the distance from the human chest [32-33]. By using the signals from IR-UWB radar, several researchers have proposed a series of algorithm that automatically extracts physiological information during sleep [34-36].
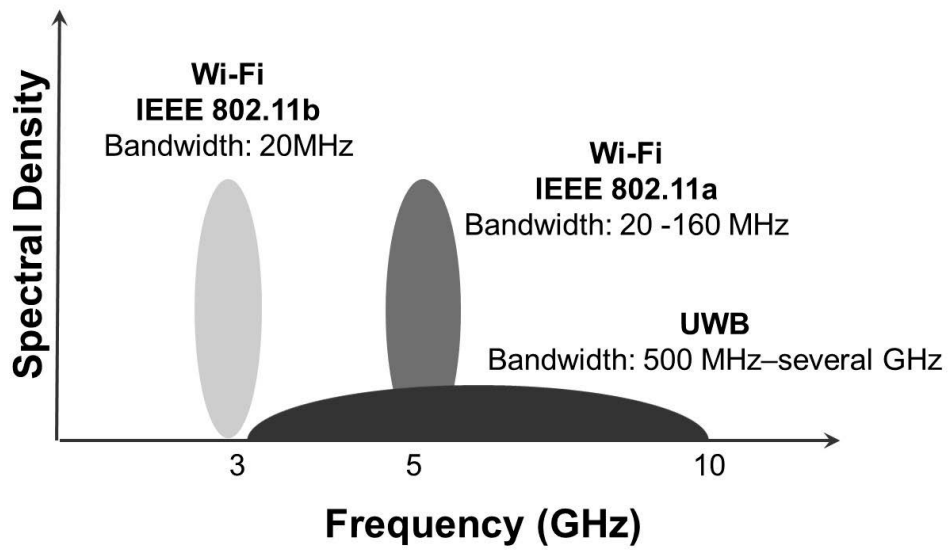
Figure 1-4. Frequency spectrum of UWB

1.5. Related Works

PSG is the gold standard method to determine sleep stages and diagnose SAHS. To overcome the limitations mentioned in Section 1.1, numerous alternative methods have been proposed that classify sleep stages and AH events automatically using the minimum number of signals. In previous studies [38, 39], a single channel EEG measured during PSG was used to score sleep stages. Choi *et al.* detected SAHS events with using a single nasal pressure signal [40]. Pulse oximetry based studies revealed a relatively high correlation coefficient ($r > 0.9$) between the AHI from PSG and the estimated one [41, 42]. Wrist activity was measured by acceleration sensor to develop and evaluate automatic sleep scoring methods in usual home environments [43, 44]. Acceleration sensor was also placed to the diaphragm and suprasternal notch and used to SAHS detection system [45, 46].

Previous studies have found that the variation of autonomic nervous system (ANS) is specifically associated with distinct sleep stages [47-51]. In NREM sleep, sympathetic activity decreases and parasympathetic activity increases as the sleep stage deepens from N1 to N3. In contrast, during REM sleep, sympathetic activity is more dominant and becomes unstable like in wakefulness. From these characteristics, heart rate variability (HRV) parameters that provide a quantitative assessment of the ANS activity have been considered as one of the most useful features for sleep staging. The HRV parameters derived from electrocardiogram (ECG) signals were used for automatic sleep scoring [52-55] and photoplethysmogram (PPG)-based sleep stage classification was also studied [56]. HRV parameters were also used to detect AH events in many alternative methods [57-59]. These methods were based on the fact that SAHS affects HRV during sleep.

Although these studies showed reliable performances for automatic sleep stage scoring and SAHS diagnosis using the minimum number of signals, several studies have pointed out that the sensor must be attached to the body, thereby their applicability for long-term monitoring decreases. To overcome the shortcomings, recent studies have proposed alternative methods to classify sleep stages and diagnose SAHS using unconstrained or non-contact sensors such as piezoelectric pressure sensor, depth camera and thermal camera [60-63]. Nevertheless, these methods still require the user to physically contact the sensor, or have privacy issues.

The limitations of the alternative methods have motivated studies on non-contact sleep monitoring methods using radio technology that allow for the reliable acquisition of physiological information. In recent studies, deep learning models such as recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have shown the higher accuracy compared with traditional machine learning methods for the four-stage sleep scoring, i.e., light sleep (N1+N2) stage, deep sleep (N3) stage, wake stage, and REM sleep stage, using IR-UWB radar sensor. Zhao et al. [64] introduced a sleep stage classification model that adapts a convolutional neural network (CNN) to extract features from radio frequency (RF) signals and couples it with RNNs. The mean accuracy for the four sleep stages was 79.8 ± 2.9 %, and the mean Cohen's kappa coefficient was 0.70 [64]. Toften et al. [36] assessed a commercialized product Somnofy®, which applied temporal convolutional networks (TCN) and LSTM networks to the IR-UWB radar for the four sleep stages in 71 PSG data. The results showed an accuracy of 76 ± 7% and Cohen's kappa coefficient of 0.63 ± 0.10. Despite the relatively good performance, there are two issues remain in sleep stage classification using physiological signal

10

with deep learning model. Firstly, these studies have a critical issue that it is difficult to interpret the variables or features that influence the classification because of the black box nature of deep learning. In addition, the commercialized product is not openly available. Secondly, for the measured physiological signals, including cardiorespiratory and movement-related signal, not all timesteps contribute equally to determining the sleep stages. In other words, the features on some timesteps may show more prominent contribution than the others in distinguishing the sleep stages. However, there is not any attempt to consider the temporal dependencies for sleep staging.

Recently, a few studies have developed algorithms for the estimation of apnea hypopnea index (AHI) and diagnosis of SAHS using only IR-UWB radar [65-67]. Javaid et al. [65] implemented an under-the-mattress ultra-wideband (UWB) radar sensor to detect apnea-hypopnea (AH) events using linear discriminant classifier for 4 patients with sleep apnea. Zhou et al. [66] investigated a wireless radar sleep screening device (ZG-S01A) that estimates AHI and total sleep time (TST) based on the IR-UWB radar, and Kang et al. [67] proposed an algorithm for the detection of AH events with the constant false alarm rate (CFAR) algorithm and weight function base on the IR-UWB radar. Both studies showed high correlation coefficient between estimated AHI and PSG AHI and high sensitivity and specificity for diagnostic efficacy in three different AHI cutoffs. Most of the existing radio frequency (RF) sensor-based methods mainly used the adaptive threshold method based on the AASM manual [68]. The adaptive threshold utilizes a procedure of extracting the respiratory signal from the IR-UWB radar and detecting the decrease in amplitude relative to the baseline respiratory signal. However, this procedure requires

information about the location of the human chest to extract the breathing signal from the radar data. Moreover, AH events are usually accompanied by body movements, which act as a factor that prevents radar from not only finding accurate body position but also determining the exact baseline amplitude. In addition, there have been numerous studies on real-time monitoring of SAHS using alternative sensors because it is critical to provide real-time instantaneous feedback for any associated medical treatment, such as continuous positive airway pressure (CPAP) pressure adjustments, when an AH event appears [40, 69, 70]. Nonetheless, none of the existing studies provide real-time AH event detection techniques.

1.6. Motivation and Objectives

To address the issues for sleep staging, I propose the usage of attention mechanism for sleep stage classification in this study. Attention mechanism is one of the most widely used methods for explainable deep learning [71]. Attention-based RNNs selectively encourage the model to focus on and assign heavier weights to certain parts of the physiological information varying over time [72]. Using this attention mechanism, the computational model identifies meaningful and discriminatory features that contribute to the classification. Owing to these characteristics, the attention mechanism has been successfully applied to natural language processing and also biosignal sequence processing in previous studies [72, 73]. In particular, this approach has been previously applied to classifying sleep stages. Recent works have reported that the automatic sleep staging model, which applies RNNs coupled with the attention mechanism to PSG data, exhibits better classification performance compared with conventional machine learning and deep learning methods [74-76]. Therefore, I hypothesized that the attention mechanism is suitable for obtaining an interpretable and a clinically effective deep learning model for sleep staging based on IR-UWB radar. To the best of my knowledge, this is the first study that applies the attention mechanism to a deep learning model for sleep staging with cardiorespiratory signals.

In addition, to effectively address the challenges of SAHS monitoring, I propose a deep learning approach based on a CNN in combination with a LSTM network. CNNs automatically filter out noise and extract the valuable feature from a signal or image without any domain knowledge [77, 78]. Although CNNs are useful in extracting patterns that appear as a local trend or appear the same in different regions

of the time sequence, they are not suitable for capturing temporal dependencies [79]. LSTM network, which is a variant of RNN, contains cyclic feedbacks that are designed to handle the temporal sequence [60]. Thus, LSTM layers can encode relevant information of class-specific characteristics across time [80]. Owing to these characteristics, the models combined with CNN and LSTM, have been successfully applied in detecting SAHS using bio-signal sequences, in recent studies [81-83]. On this premise, I take the temporal characteristics of radar signals into consideration and propose a hybrid model architecture that combines CNNs and LSTM network. To best of my knowledge, none of the previous studies applied a deep learning algorithm to monitor sleep-related breath disorder using RF sensors.

Therefore, this study was conducted to establish non-contact sleep monitoring methods using IR-UWB radar for the continuous and accurate monitoring of sleep stages and SAHS events.

This study has the following two goals:

1) The development of an attention-based bidirectional LSTM networks (attention-based Bi-LSTM) for automatic sleep stage scoring using a single IR-UWB radar

2) The development of a hybrid CNN-LSTM network to accurately detect AH events based on a single IR-UWB radar

1.7. Dissertation Outline

This thesis consists of following chapters.

- Chapter 2 presents the deep neural networks approach for non-contact sleep staging using IR-UWB radar and discusses the potential applicability of the model compared with other methods.

- Chapter 3 describes a hybrid deep neural network model for real-time apnea-hypopnea event detection using IR-UWB radar. In addition, its clinical usability is assessed.

- Chapter 4 summarizes the conclusions of the preceding chapters.

This thesis is based on following scientific articles that have been accepted for publication (chapter 2 and chapter 3):

- Chapter 2

  [84] H. B. Kwon *et al*., "Attention-based LSTM for Non-Contact Sleep Stage Classification using IR-UWB Radar," *IEEE JBHI*, In press.

- Chapter 3

  [85] H. B. Kwon *et al*., "Hybrid CNN-LSTM Network for Real-Time Apnea-Hypopnea Event Detection Based on IR-UWB Radar," *IEEE Access*, In press.

The author of this thesis contributed to the above studies as follows: conception and design of the experiments; data acquisition, analysis, and interpretations; and wrote and reviewed the manuscript.

2

LSTM Model for Non-Contact Sleep Stage Classification Using IR-UWB radar

A non-contact sleep staging method is investigated in this chapter. Based on non-contactly measured IR-UWB signals, attention-based Bi-LSTM network model is proposed for classifying sleep stage automatically. An optimal attention-based Bi-LSTM architecture that produces the best performance is searched. Sleep stage classification performance of the attention-based Bi-LSTM model is compared with those reported in previous works. At last, role of the attention mechanism in sleep staging is analyzed and the application of the model is discussed.

2.1. Methods

2.1.1. Subjects

This study was performed in accordance with the ethical standards in the Declaration of Helsinki, and the Institutional Review Board of Seoul National University Hospital (IRB No. 1906-120-1043) approved this prospective cohort. Subjects were invited to volunteer through posters on the school bulletin board in Seoul National University. Before proceeding with the PSG, subjects were initially screened by questionnaires to ensure that they do not have a history of sleep, psychiatric, neurological, or cardiovascular disorder. Qualifying 65 subjects underwent overnight PSG at the Center for Sleep and Chronobiology of Seoul National University Hospital. As a result of the PSG, six subjects had moderate or severe sleep apnea (AHI > 15), four subjects had periodic limb movement index (PLMI) > 15, and four subjects could not achieve normal sleep due to sensor discomfort (sleep efficiency (SE) < 50%). These 14 PSGs were excluded at the data analysis stage, considering potential factors that affect the quality of IR-UWB radar signals or autonomic controls during sleep [86-88]. A total of 51 people who satisfied the inclusion and exclusion criteria were included in the experiments. All participants were briefed about the objective and procedure of the experiment, and they signed the consent forms.

2.1.2. Polysomnography and IR-UWB Radar

All PSG data were recorded with a NEUVO system (Compumedics Ltd., Victoria, Australia). After PSG recording, the sleep stages were scored by a sleep technologist and verified by two sleep clinicians according to the 2018 AASM manual [68]. The following physiological data were collected: EEG at O2-M1, C4-M1, and F4-M1; submental and tibialis anterior EMG; bilateral EOG; ECG; oronasal airflow, thoracic and abdominal respiratory effort, and nasal pressure using a thermistor, piezoelectric-type belts, and nasal cannula/pressure transducer; body posture from a 3-axis accelerometer; and blood oxygen saturation using a pulse oximeter. Of all these signals, only the blood oxygen saturation was sampled at 200 Hz, and all other signals were measured at 500 Hz. The anthropometric and sleep parameters of the subjects are summarized in Table 2-1. The proportion of each sleep stage was calculated based on total sleep time. Note that the mean percentage of N3 stage is 6.7%, which may appear to be lower than normal healthy population. However, according to [89], the proportion of N3 stage for the normal and mild obstructive sleep apnea group in Korean population tends to be relatively small, with only 8.6% on average.

A commercially available IR-UWB radar system on chip (SoC) X4 (Novelda, Oslo, Norway) was adopted. The IR-UWB radar was fixed on a tripod within a range of approximately 0.5 to 2 m from the human chest and measured simultaneously with the PSG as depicted in Fig. 2-1. The distance was selected as a position where a stable signal can be measured considering that the radar detection range is 3 m and body position on the bed might change during sleep. The transmitter of the radar has a center frequency of 7.29 GHz and a bandwidth of 1.5 GHz. The receiver sampled

the reflected signal at 23.328 GS/s and the radar signals were digitized at a speed of 20 fps. Because the frequency range of the heart rate signal is between 1 - 2 Hz, and the frequency range of the respiration signal is between 0.1 - 0.7 Hz [90], previous studies could obtain reliable heart rate, respiration rate, and HRV using IR-UWB radar with 20 fps or lower sampling rate [32, 86, 91].

Before starting the PSG, the sleep technologist asked the subjects to hold their breath for 15–20 s during calibration time. In this manner, the two devices were synchronized in time by finding the section where the thoracic respiratory effort from PSG and breath signal from IR-UWB appear flat at the same time (Fig. 2-2). Moreover, I could synchronize the devices in time by maximizing the cross-correlation between the movement from the two devices in more detail. For acquiring, processing, and storing data from the IR-UWB radar, I used MATLAB 2019a (MathWorks, New York, MA, USA).

Table 2-1. Summary of anthropometric and sleep parameters

Characteristics	Mean ± S.D.
Sex (male/female)	26 / 25
Age (years)	30.0 ± 8.6
BMI (kg/m²)	22.7 ± 3.9
AHI (events/h)	2.8 ± 3.4
PLMI (events/h)	4.5 ± 9.5
Time in bed (min)	435.4 ± 42.1
Total sleep time (min)	403.4 ± 51.5
Sleep efficiency (%)	92.4 ± 5.1
Sleep Onset Latency (min)	7.4 ± 9.8
Wake after Sleep Onset (min)	24.6 ± 16.9
Stage N1 (%)	10.9 ± 6.2
Stage N2 (%)	61.5 ± 8.3
Stage N3 (%)	6.7 ± 5.1
Stage REM (%)	20.9 ± 4.7

S.D., Standard deviation; BMI, body mass index; AHI, apnea hypopnea index; PLMI, periodic limb movement index.
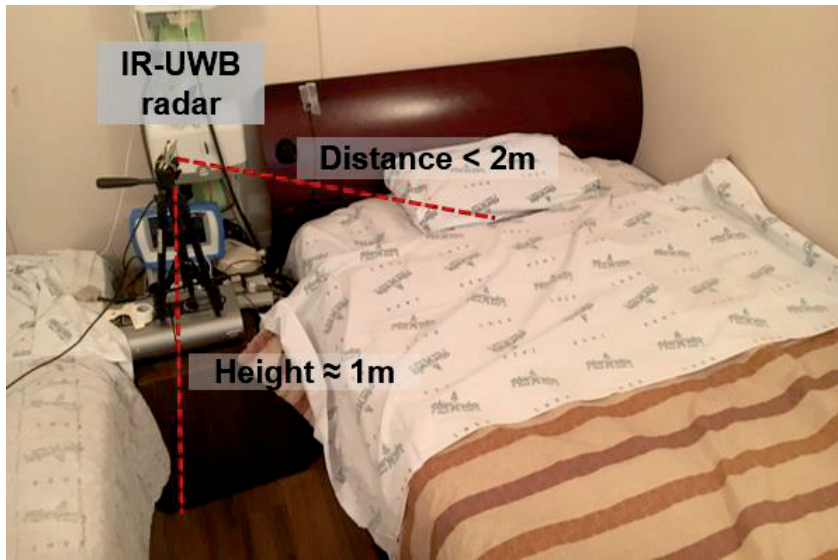
Figure 2-1. Measurement of IR-UWB radar with PSG in the experiment
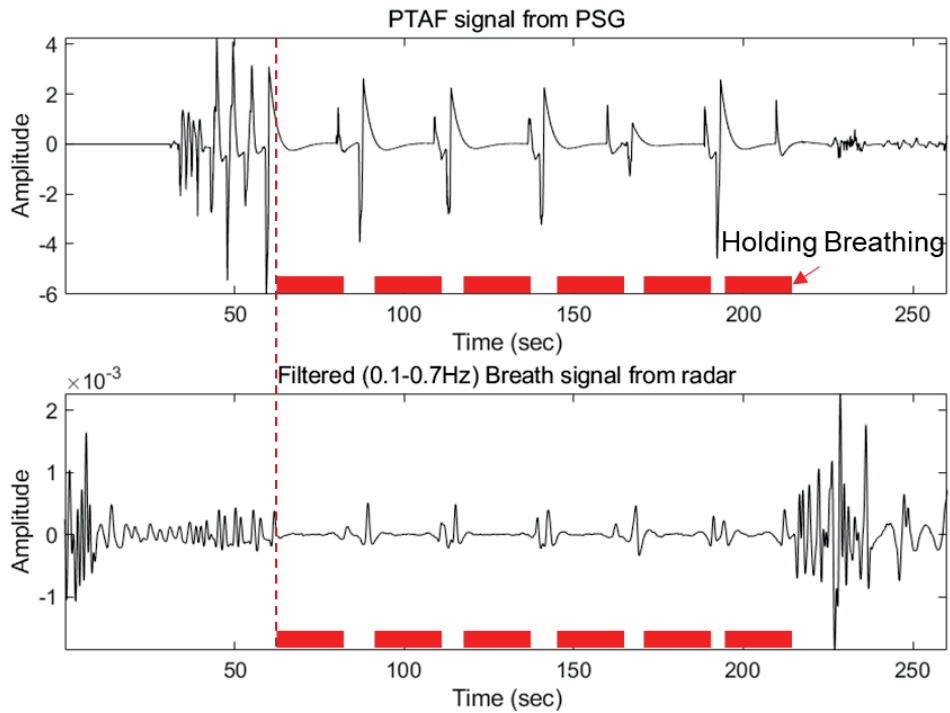
Figure 2-2. Time synchronization between the respiratory signal from PSG and breath signal from IR-UWB radar

2.1.3. Radar Data Processing

After the raw data acquisition with the IR-UWB radar, the features of the training deep learning model were extracted automatically by the software algorithms every 30-s epoch. The data processing is described in Fig. 2-3.

When there are multiple channels, the raw signal received from an IR-UWB radar can be represented as the sum of the responses of several channels, including the subject's respiration and heartbeat [90]:

$$r(t,\tau) = A_d p(\tau - \tau_d(t)) + \sum_i A_i\, p(\tau - \tau_i) \qquad (2\text{-}1)$$

where $r(t,\tau)$ indicates the received signal with a real-time factor t, which is generally called "slow-time," and the observed range factor τ, which is called "fast-time". $p(\tau)$ is the normalized received pulse. A_i and A_d are the amplitudes of each multipath component and the reflected pulse on the body, respectively. τ_i and τ_d represent the corresponding time delay. The time delay $\tau_d(t)$ varies with the movement of the chest with slow time. While the portion of the received pulse associated with respiration and heartbeat is $A_d p(\tau - \tau_d(t))$, the multipath components, called "clutter," should be removed.
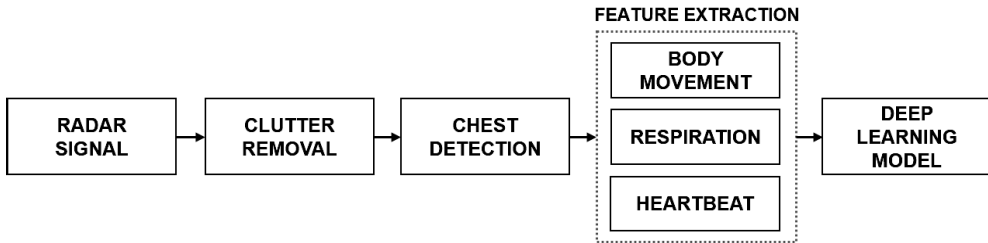
Figure 2-3. Block diagram of data processing for deep learning using IR-UWB radar.

In the sleep environment, the background is considered to be stationary, whereas the human body changes its position when rolling over, twitching, and in terms of chest movements caused by breathing and heartbeat. Therefore, I obtained the target signals by subtracting the DC component from the raw radar signals with a moving-average method [90, 91]. The clutter was calculated as the average of each 30-s epoch amplitudes along the fast time range bin. Fig. 2-4 (a) shows an example of the received raw radar signal in this study, and Fig. 2-4 (b) summarizes the result after the clutter component is removed. The row index refers to the received waveforms from different distances, and the column index indicates the sampling times. The waveform after applying the moving-average method is preferable for observing the cardio-respiration activity around 1 m, which is equal to the distance of the radar device from the body.
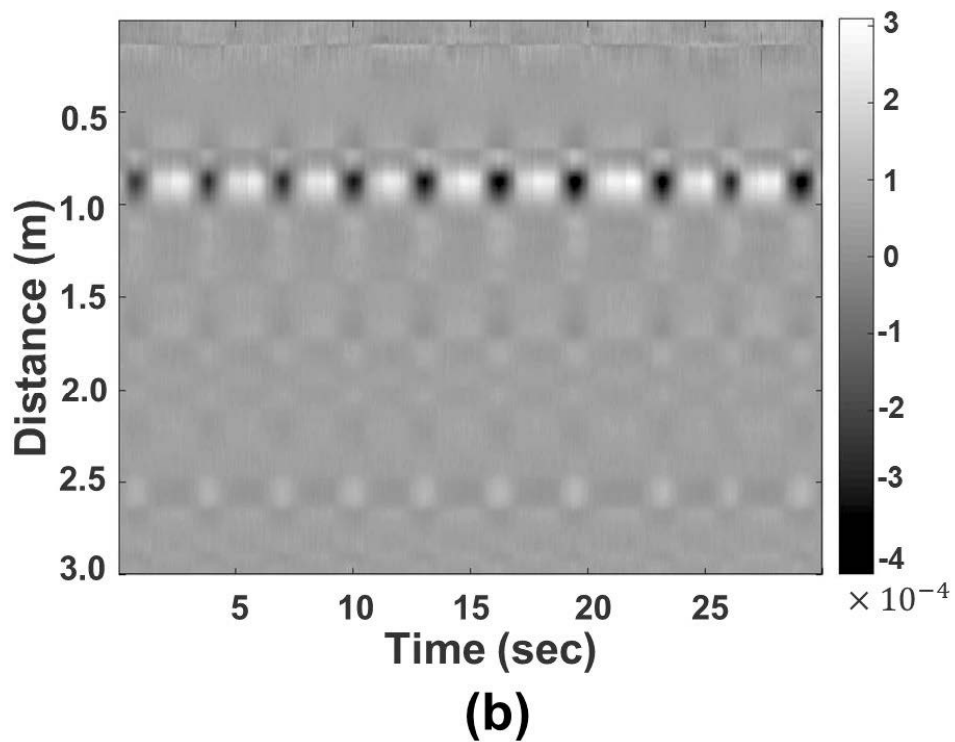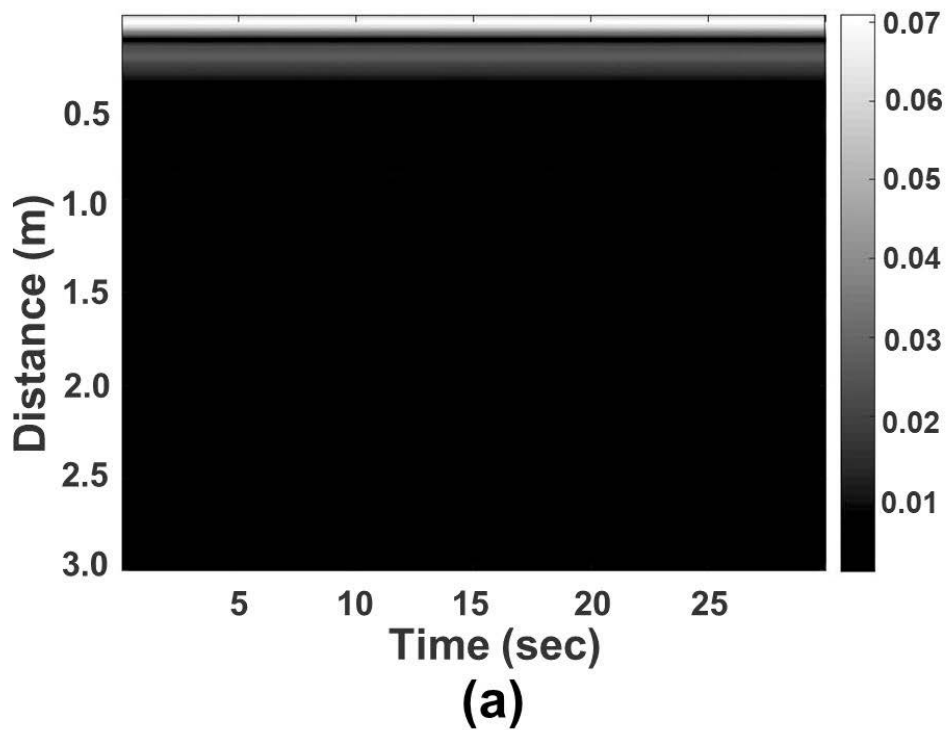
Figure 2-4. Example of clutter removal. (a) Received raw radar signal (b) After application of moving average method

25

After removing the clutter from the raw radar signal, the most prominent location of the human chest was detected for every 30-s epoch. The distance between the radar and the human chest fluctuates periodically around the nominal distance because of respiration and heartbeat motion. However, the amplitude of the breathing signal is considerably higher compared with that of the heartbeat signal [92], and the interference caused by the harmonics of the breathing signal increases the difficulty of measuring the heart rate [33]. Therefore, only the respiration signal was considered in finding the optimal region presumed to be the human chest in this study. By calculating the spectral power in the respiration frequency range of 0.1–0.7 Hz in the fast-time domain, the position of the target with the highest power could be chosen, and the cardiorespiratory signal of the chest could be obtained for every 30-s epoch [90]. Fig 2-5 (a) presents an example of the selected cardiorespiratory signal.

The sleep stages are associated with distinct mechanisms of autonomic nervous control [93]. During non-REM sleep, parasympathetic activity increases and sympathetic activity decreases. In contrast, autonomic balance is shifted toward sympathetic predominance during REM sleep and wakefulness. Moreover, respiratory rhythm during REM sleep and wakefulness become significantly faster and irregular compared with non-REM sleep. Body movement is also an important parameter that distinguishes REM sleep from wakefulness. In normal sleep, human body physically moves during wakefulness however, during REM sleep, more frequent and relatively short-duration twitches are observed compared with other sleep stages [94]. Based on the characterized physiological conditions and previous study, in which an LSTM networks for sleep staging using PVDF film sensor was developed [60], 16 parameters for each epoch were extracted (Table 2-2).
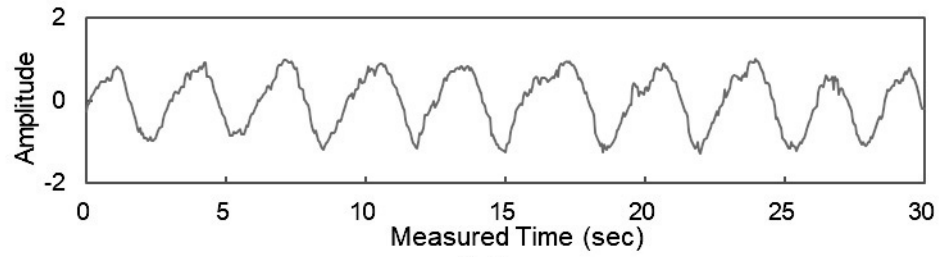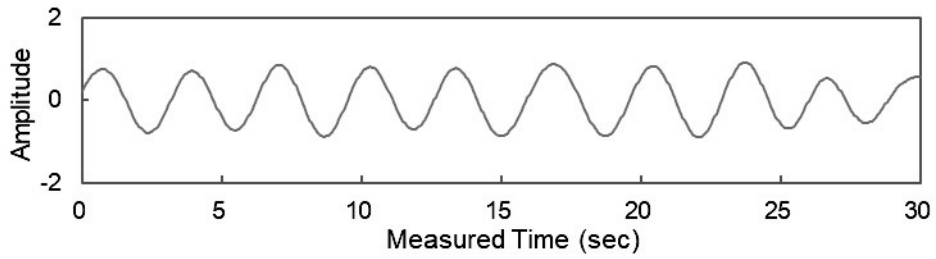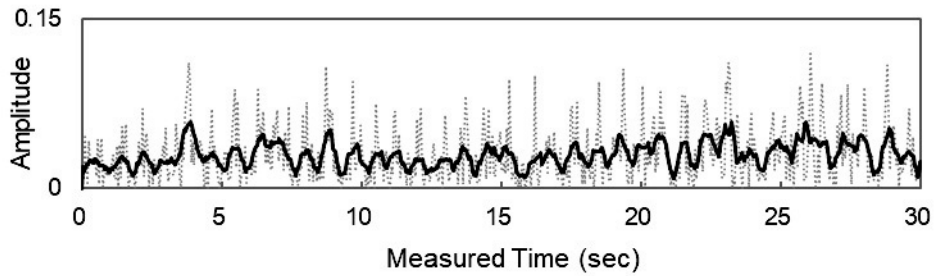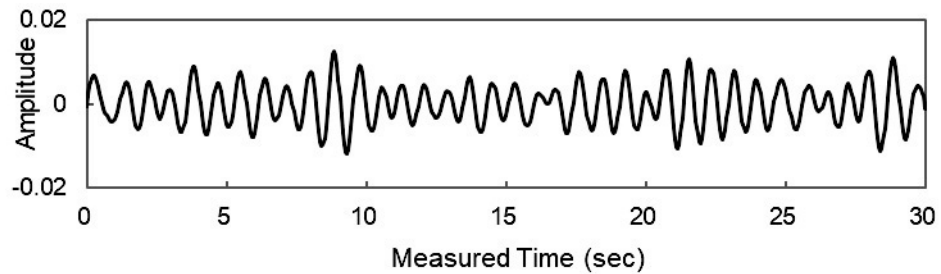
Figure 2-5. Extraction of breathing and heart movements from the raw signal of IR-UWB radar (a) cardio-respiratory signal after clutter removal (b) extracted respiratory signal (c) rectified and smoothed signal and (d) extracted heartbeat signal.

Table 2-2. Summary of features for deep learning model

Association	Parameter	Description
Movement	mov_dens	Movement density from IR-UWB radar signal
Respiration	f_{Resp}	Respiratory frequency derived from filtered IR-UWB radar signal
	tf_{Resp}	Standard deviation of f_{Resp}
	sf_{Resp}	Smoothed value of f_{Resp}
	stf_{Resp}	Smoothed value of tf_{Resp}
	sdf_{Resp}	Smoothed value of absolute difference between f_{Resp} and sf_{Resp}
Heart rate variability	mHR	Mean of heart rate
	LF	Spectral power of inter-beat interval in low-frequency range (0.04–0.15 Hz)
	HF	Spectral power of inter-beat interval in high-frequency range (0.15–0.4 Hz)
	LFHF	LF to HF ratio
	tmHR	Standard deviation of mHR
	smHR	Smoothed value of mHR
	stmHR	Smoothed value of tmHR
	sdmHR	Smoothed value of absolute difference between mHR and smHR
	sHF	Smoothed value of HF
	sLFHF	Smoothed value of LFHF

28

To compute a movement parameter (mov_dens), the IR-UWB radar signal was divided into 60 segments with 0.5-s length, and the variance for each segment was calculated. Then, the percentage of the segments that was higher than the empirically determined threshold was extracted every 30-s epoch. For the respiration-related parameters, the breath signal was extracted from the cardiorespiratory signal by filtering with a band-pass (0.15-0.4 Hz) filter (see Fig. 2-5 (B)), and the respiration frequency (f_{Resp}) was calculated with an autocorrelation method during each 30-s epoch.

HRV is the most widely used tool for quantitatively assessing the ANS activity [26]. By analyzing the HRV in the time and frequency domains, information on cardiac autonomic modulation can be obtained. To estimate the HRV values, the heartbeat was detected as per the method used in my previous study [60]: (1) the IR-UWB radar signal was subtracted by the smoothed radar signal using a moving average filter with a span of 10 samples (signal length of IR-UWB radar during 0.5s); (2) the absolute values of the filtered signal were acquired (see Fig. 2-5 (c) gray line); (3) the 0.8–1.5 Hz band-pass filter was applied to remove baseline drift affected by respiration in order to obtain a clear heartbeat (see Fig. 2-5 (c) black line); and (4) the locations of the local maxima in the extracted heartbeat signal (see Fig. 2-5 (d)) were found, and inter-beat intervals were measured as the peak intervals. The mean heart rate (mHR) could be calculated from the consecutive IBIs for every 30-s epoch. A power spectral analysis of HRV was performed through a fast Fourier transformation (FFT) for 5 min windows; the spectral power in the low-frequency (LF) range between 0.04 to 0.15 Hz representing sympathetic activity, and power in the high-frequency (HF) range between 0.15 to 0.4Hz representing parasympathetic

activity. The LFHF ratio indicates the ratio of sympathetic to parasympathetic activity [15]. The standard deviation of f_{Resp} (tf_{Resp}) and mHR (tmHR) were investigated every 10 epochs with a sliding window of 1 epoch. In addition, the smoothed values of abovementioned parameters, listed in Table 2-2, were obtained by using the Savitzky–Golay filter, which is a finite impulse response smoothing filter of polynomial order of 2 and frame length of 31 epochs. Finally, the z-score normalization was performed on the extracted parameters for entire-night recording, such that the mean value was 0 and the variance was 1.

2.1.4. Proposed Deep Learning Model

Proposed deep learning network, attention-based Bi-LSTM architecture comprises two bidirectional LSTM layers and an attention layer, shown in Fig. 2-6. The LSTM layer comprises 9 sequence length, 16 input dimensions, and 512 units per layer. LSTMs consist of memory blocks that can store long-term dependencies from the input sequence and control the flow of information [95]. Each memory block is comprised of three activation gates: input gate, output gate, and forget gate. The activations of the input and output information flow are controlled by the input and output gates, respectively. The forget gate sets the amount of information forgotten and updated. The hidden state is calculated based on the output gate and the updated memory cell. The sigmoid and tanh functions were set as the recurrent activation function and activation function of the LSTM, respectively. In addition, by combining two forward and backward LSTMs, that is, a bidirectional LSTM, the input data can learn both the previous and following information of the target class in each timestep [96].
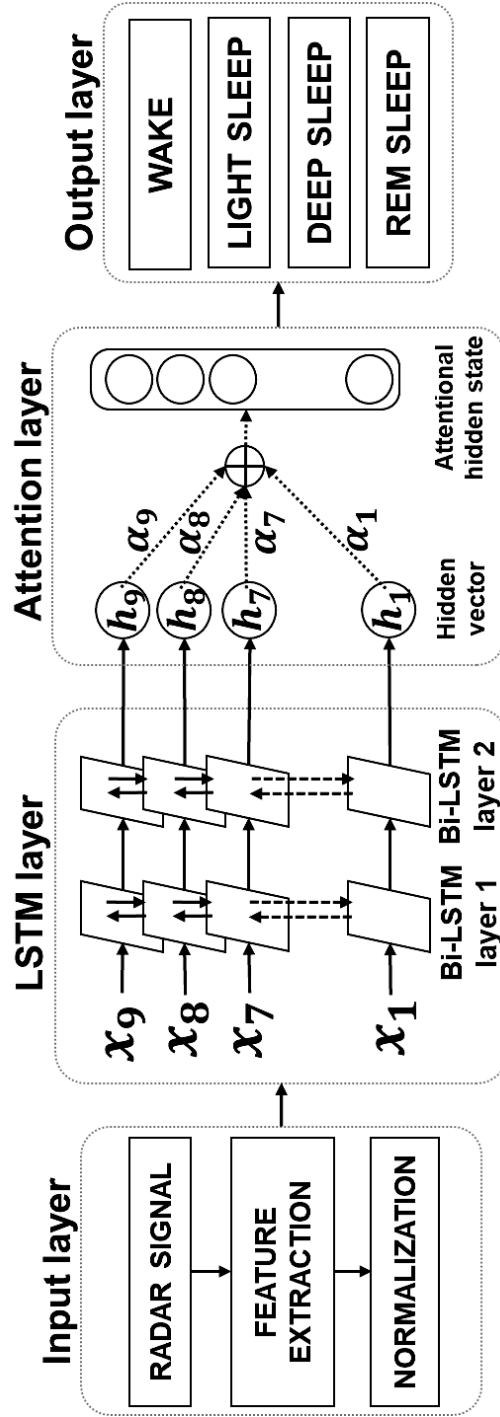
Figure 2-6. Overall structure of the proposed model

Because of the dynamics of physiological interactions in the brain and autonomic controls during each sleep stage [97], I believe the sequential feature vectors from the LSTM layers contribute differently for the classification of the four sleep stages. To improve the performance of the LSTM networks by focusing on certain timesteps with more discriminative sleep stage related features, I added an attention model to automatically learn the importance of feature vectors in each time step. The attention mechanism is applied to the sequence modeling output from the bi-LSTM output hidden state vectors $\boldsymbol{h_i}$ at every timestep $\boldsymbol{i} = 1, 2, 3, ..., $ N. The importance score $\boldsymbol{s_i}$ was calculated with the score function tanh as follows:

$$\boldsymbol{s_i} = tanh(\boldsymbol{W_s}\boldsymbol{h_i} + \boldsymbol{b_s}) \tag{2-2}$$

where $\boldsymbol{W_s}$ and $\boldsymbol{b_s}$ are a trainable weight and bias, respectively. Next, the attention weight α_i was evaluated through the softmax function, and the output vector $\mathbf{v}$ was achieved by multiplying the attention weight vector and with corresponding hidden state vector as follows:

$$\alpha_i = softmax(\boldsymbol{s_i}) = \frac{exp(s_i)}{\sum_i exp(s_i)} \tag{2-3}$$

$$\mathbf{v} = \sum_i \alpha_i \boldsymbol{h_i} \tag{2-4}$$

Finally, the attention layer is followed by a fully connected layer, which classifies the four classes.

To find the optimal value of hyperparameters, I compared the classification performance of the deep learning model on the validation dataset. The candidate of

the hyperparameters were set as follows: the length of input sequence {1-15}; type of LSTM structure {unidirectional, bidirectional}; number of LSTM units {16, 32, 64, 128, 256, 512}; and number of LSTM layers {1, 2, 3}. The target labels were set as the scored sleep stage corresponding to the end of each input sequence for the unidirectional LSTM and the center of each input sequence for bidirectional LSTM as shown in Fig. 2-7.
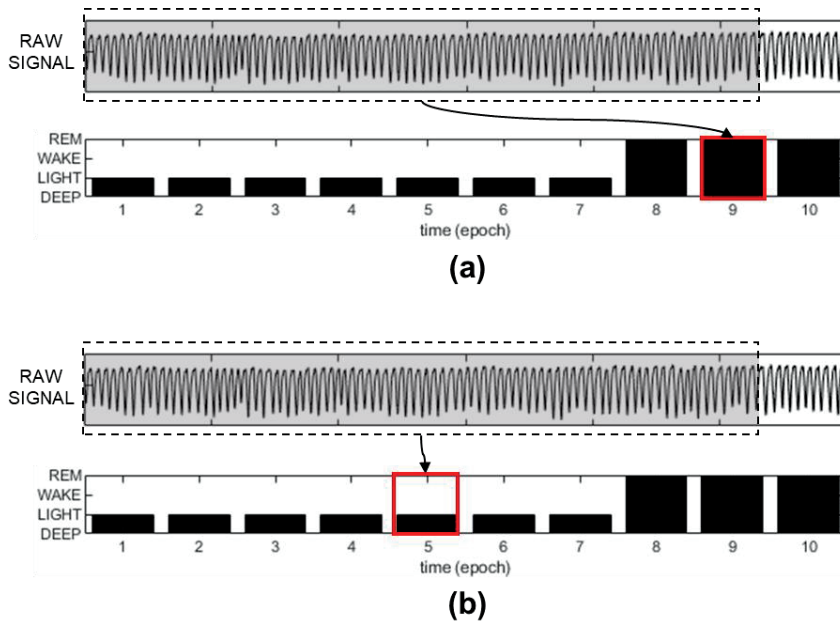
Figure 2-7. Range of input data for the deep learning model when using (a) unidirectional LSTM (b) bidirectional LSTM. The sleep stage indicated by the red line is the target label; the gray and dashed part of the raw signal is the corresponding input sequence.

I trained the model using the Adam optimizer [98] with a learning rate of 0.001, and categorical cross-entropy was used as a loss function. The model was trained for maximum 100 epochs with an early stopping patience of 10 and the batch size of 32. I trained and validated deep learning model with a hold-out method for developing a generalized model. In addition, the number of input segments was sufficient for model training. The numbers of recordings for training/ validation/ testing datasets were 26/ 8/ 17; they were randomly divided without data overlap. In total, the number of segments in the all datasets was 37,732. However, among the segments, the number of wake stage, light sleep, deep sleep, and REM sleep was 2,321, 22,146, 7,623, and 5,642, respectively. Therefore, to prevent the model from overfitting because of class imbalance and to enhance the performance, a dropout layer [99] were applied after every LSTM layer and fully connected layer. The dropout rate was set to 20%. Moreover, I utilized a method that was suggested in previous studies [100]. This method adjusts the class weights inversely proportional to the number of segments in each class as follows:

$$W_{class} = \frac{\sum_{class=\{W,L,D,R\}} N_{class}}{N_{class}} \tag{2-5}$$

where W_{class} is the class weight and N_{class} is the number of segments in the training dataset. In this way, all classes contributed equally to the loss function as if all classes had the same number of segments when training the model. Thus, weighting by the relative frequency of its class label could serve to prevent overfitting for the majority of classes, i.e., light sleep.

To evaluate the performance of the classification of the four sleep stages, the

conventional metric of overall accuracy was calculated by the generating confusion matrix. The accuracy (P_o) was computed from the sum of the diagonal elements divided by the total number of samples (Eq. (2-6)). However, because of the sample imbalance among the sleep stages, the Cohen's kappa coefficient (Kappa) was also considered for evaluating the performance. The Kappa was computed by factoring out the chance agreement as follows:

$$P_o = \sum_{class=\{W,L,D,R\}} \frac{TP_{class}}{Total} \tag{2-6}$$

$$Kappa = \frac{P_o - P_e}{1 - P_e} \tag{2-7}$$

$$P_e = \sum_{class=\{W,L,D,R\}} \frac{TP_{class} + FP_{class}}{Total} \times \frac{TP_{class} + FN_{class}}{Total} \tag{2-8}$$

where P_e, TP, FP, and FN are the hypothetical probability of chance agreement, true positives, false positives, and false negatives, respectively.

To validate the effect of attention mechanism, I evaluated the classification performance of the proposed attention-based Bi-LSTM networks and compared it with the performance of the conventional bidirectional LSTM networks for the same dataset. Next, I visualized the attention weight for each sleep stage and the response of the attention-based Bi-LSTM to analyze the role of the attention layer. Furthermore, I measured the class separability using a measure called the general discrimination value (GDV), which is defined as the difference between the mean inter-cluster separation and the mean intra-cluster variability [100]. The GDV Δ was calculated considering that N points $x_{n=1,\dots,N} = (x_{n,1}, \dots, x_{n,D})$ are distributed

within D-dimensional space and label l_n assigns each point to one of L distinct classes $C_{l=1,\dots,L}$ as follows:

$$\Delta = \frac{1}{\sqrt{D}} \left[\frac{1}{L} \sum_{l=1}^{L} \bar{d}(C_l) - \frac{2}{L(L-1)} \sum_{l=1}^{L-1} \sum_{m=l+1}^{L} \bar{d}(C_l, C_m) \right] \qquad (2\text{-}9)$$

Here, the mean intra-class distances $\bar{d}(C_l)$ and the mean inter-class distances $\bar{d}(C_l, C_m)$ were calculated as follows:

$$\bar{d}(C_l) = \frac{2}{N_l(N_l-1)} \sum_{i=1}^{N_l-1} \sum_{j=i+1}^{N_l} d\left(s_i^{(l)}, s_j^{(l)}\right) \qquad (2\text{-}10)$$

$$\bar{d}(C_l, C_m) = \frac{1}{N_l N_m} \sum_{i=1}^{N_l} \sum_{j=1}^{N_m} d\left(s_i^{(l)}, s_j^{(m)}\right) \qquad (2\text{-}11)$$

N_k is the number of points in class k, and $s_i^{(k)}$ is the i-th point of class k. The quantity $d(a, b)$ is the distance between a and b in the Euclidean distance.

Models were implemented in Python 3.7 and the Keras framework [101] with TensorFlow backend [102]. The training was done using a GTX1080 8GB GPU and a 3.4 GHz Intel i7-6700 CPU.

2.2. Results

2.2.1. Hyperparameter tuning

I selected the hyperparameter combination that achieved the highest Kappa for the validation set (Table 2-3). The model with the longer input sequence length or a higher number of units showed better classification performance, regardless of the type of LSTM or the number of LSTM layers. Therefore, the results with the number of units fixed at 512 are summarized in Fig. 2-8 (a). The highest Kappa was obtained for the model with two bidirectional LSTM layers (named as 'L2_bi', red line) with an input sequence length of 9 and unit number of 512; thus, the best-performance model was selected. Fig. 2-8 (b) shows the details of the performance in terms of the length of input sequence and the number of units.

Table 2-3. Hyperparameters in the proposed model

Parameters	Values
Input (# of features × sequence length)	16×9
sequence length	9
Directionality of LSTM	Bidirectional
Learning rate	0.001
# of LSTM layers	2
Hidden units (the first LSTM)	512
Hidden units (the second LSTM)	512
Output	4
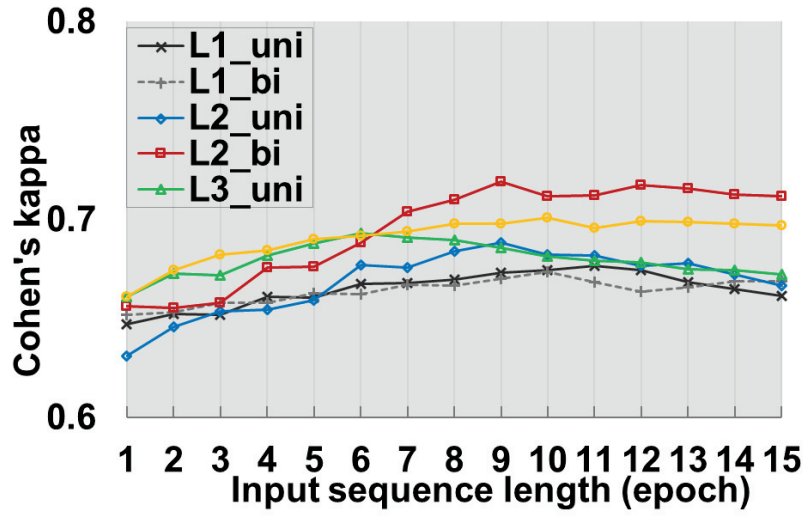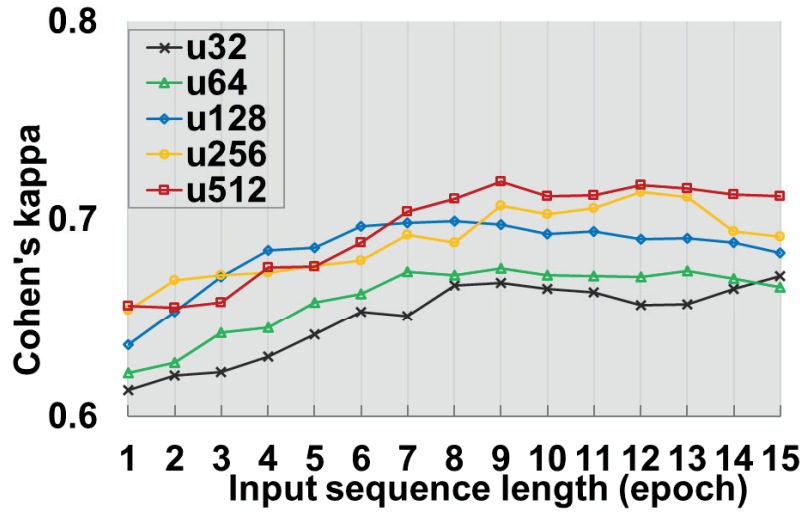
Figure 2-8. Performance of attention-based LSTM model with (a) input sequence, uni-/bi-directional LSTM and number of layers, (b) input sequence length and units with two bi-LSTM layer. LSTM, long-short-term memory; L1~3, the number of LSTM layers; uni-, unidirectional; bi-, bidirectional.

2.2.2. Model Performance on Test Dataset

The accuracy curve of the training and validation phase to investigate the over or underfitting of the designed deep learning model was presented in Fig. 2-9 (a), and Fig 2-9 (b) shows the loss curve of the training and validation dataset. There is no overfitting in the proposed model.

Table 2-4 lists the average accuracies and Kappa values along with their standard deviations for the proposed attention-based Bi-LSTM model and the conventional LSTM model for the same dataset. In addition, a paired t-test was conducted between the results to compare the performance. The proposed model has an accuracy of 82.6% and a Kappa of 0.73, which is significantly higher than the results of the conventional LSTM model ($p < 0.01$). The detailed values for the test subjects are illustrated in Fig. 2-10. Fig. 2-11 show the confusion matrix of the conventional LSTM model and the attention-based Bi-LSTM model. As shown in Fig. 2-11, the attention-based Bi-LSTM model outperforms the conventional LSTM model for all sleep stages.

Table 2-4. Accuracy and Cohen's kappa coefficient on each dataset

Dataset	Network	Accuracy (%)	Kappa
Training	attention-based Bi-LSTM	86.5 ± 7.2	0.75 ± 0.20
Validation	attention-based Bi-LSTM	80.1 ± 8.6	0.71 ± 0.15
Test	attention-based Bi-LSTM	82.6 ± 6.7	0.73 ± 0.11
	LSTM	79.2 ± 7.6	0.68 ± 0.09
	p-value[*]	0.006	0.0006

[*] Wilcoxon signed-rank test for accuracy and Kappa between Attention Bi-LSTM and LSTM on test dataset
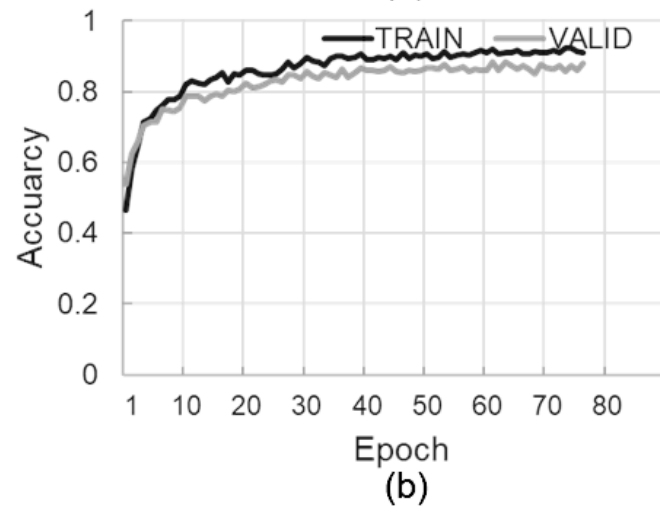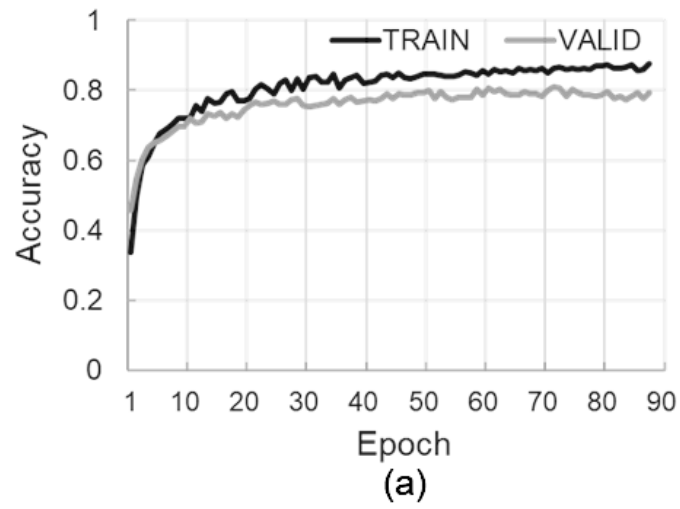
Figure 2-9. Performance graph of (a) the accuracy and (b) loss function from training and validation dataset
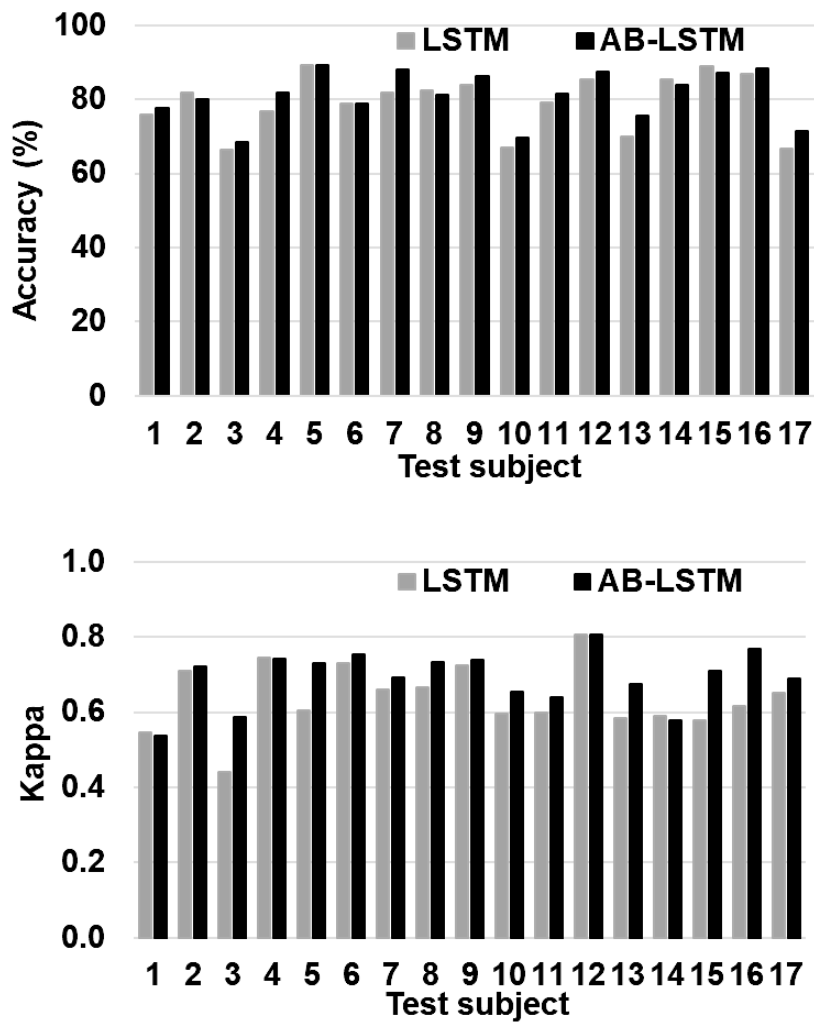
Figure 2-10. Comparison of accuracy and Cohen's kappa coefficient between LSTM model and attention-based Bi-LSTM model (AB-LSTM) on test subjects.
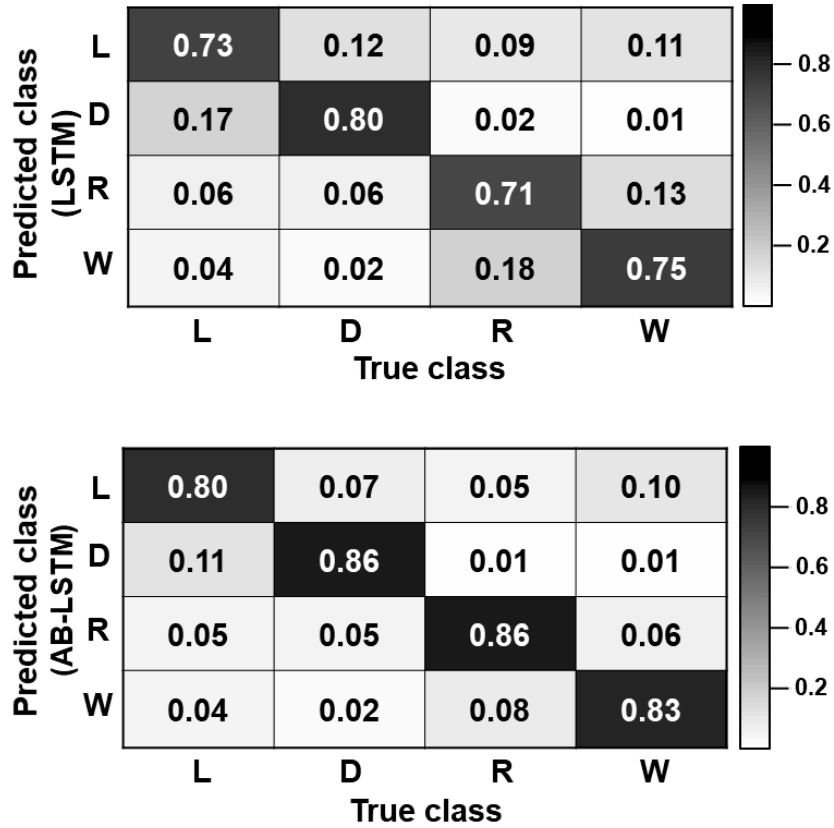
Figure 2-11. Confusion matrix of LSTM model and attention-based Bi-LSTM model (AB-LSTM). D, deep sleep; L, light sleep; W, wake; R, REM sleep.

2.2.3. Attention Weight

To verify that my model can pick up the stage-specific importance information, I plotted the distribution of the attention weights assigned to each timestep from 0 to 1. A high weight in certain regions indicates that the corresponding timestep contains a relatively important characteristic. Fig. 2-12 shows the distribution for each sleep stage for the training and test datasets. According to Fig. 2-12, the attention layer assigned the highest weight on the current epoch (timestep = 0) during deep sleep and wake. However, during light sleep and REM sleep, the highest weight in both training and test datasets was assigned to one epoch before (timestep = -1) and two epochs before (timestep = -2) the current epoch, respectively.
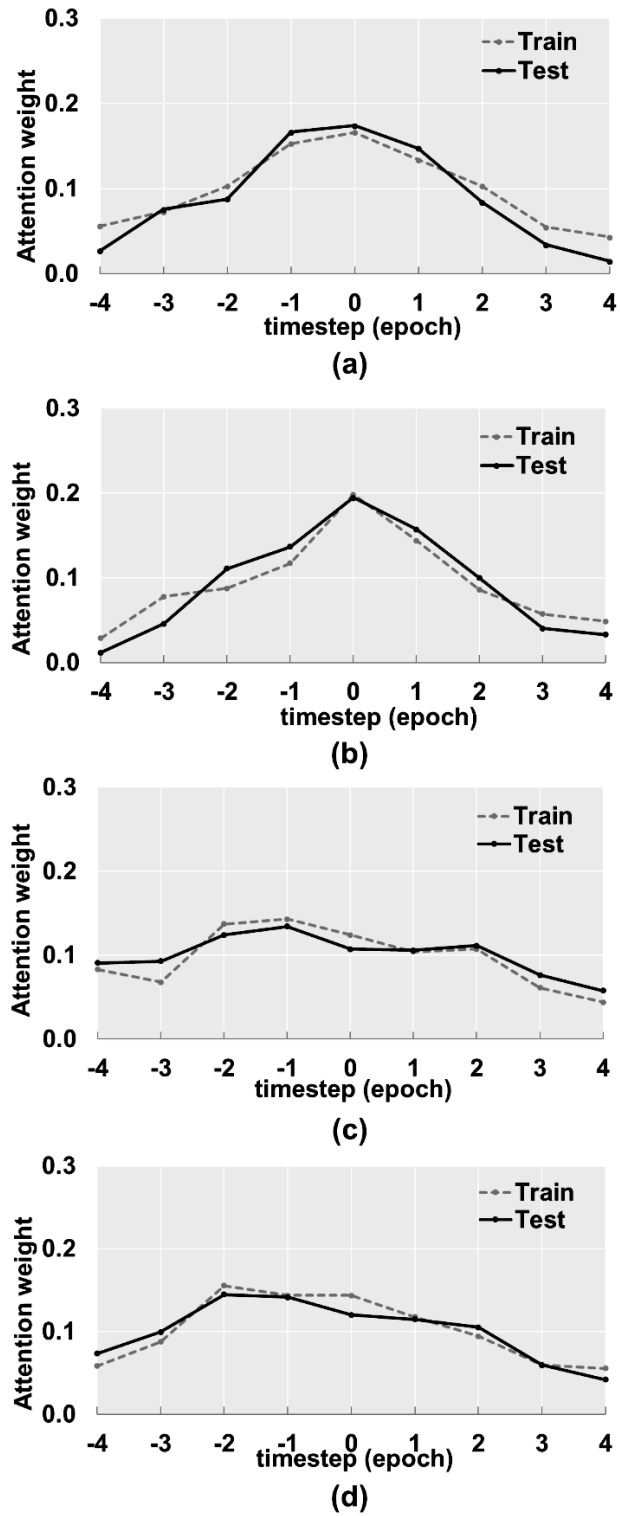
Figure 2-12. Mean attention weight across train and test dataset for each timestep in 4 sleep stages (a) deep sleep (b) wakefulness (c) light sleep, and (d) REM sleep

2.2.4. Visualization of Class Separability

I visualized the sleep stage embeddings by dimensionality reduction into two dimensions by using the t-distributed neighbor embedding (t-SNE) method [103]. The perplexity for t-SNE was set to 30.0, the metric was "Euclidean," which is interpreted as the squared Euclidean distance, and the method was "Barnes_Hut." Moreover, I applied a GDV algorithm to quantify the class separability. A GDV of 0 indicates randomly shuffled labels, and a GDV of -1 is the case of perfect class separability [100]. Fig. 2-13 shows the response of each layer organizing attention-based Bi-LSTM model to data points which randomly sub-sampled for better viewing. It can be observed that the separability for the four sleep stages improved with each deeper network layer. In addition, the GDV value decreased for deeper layers.
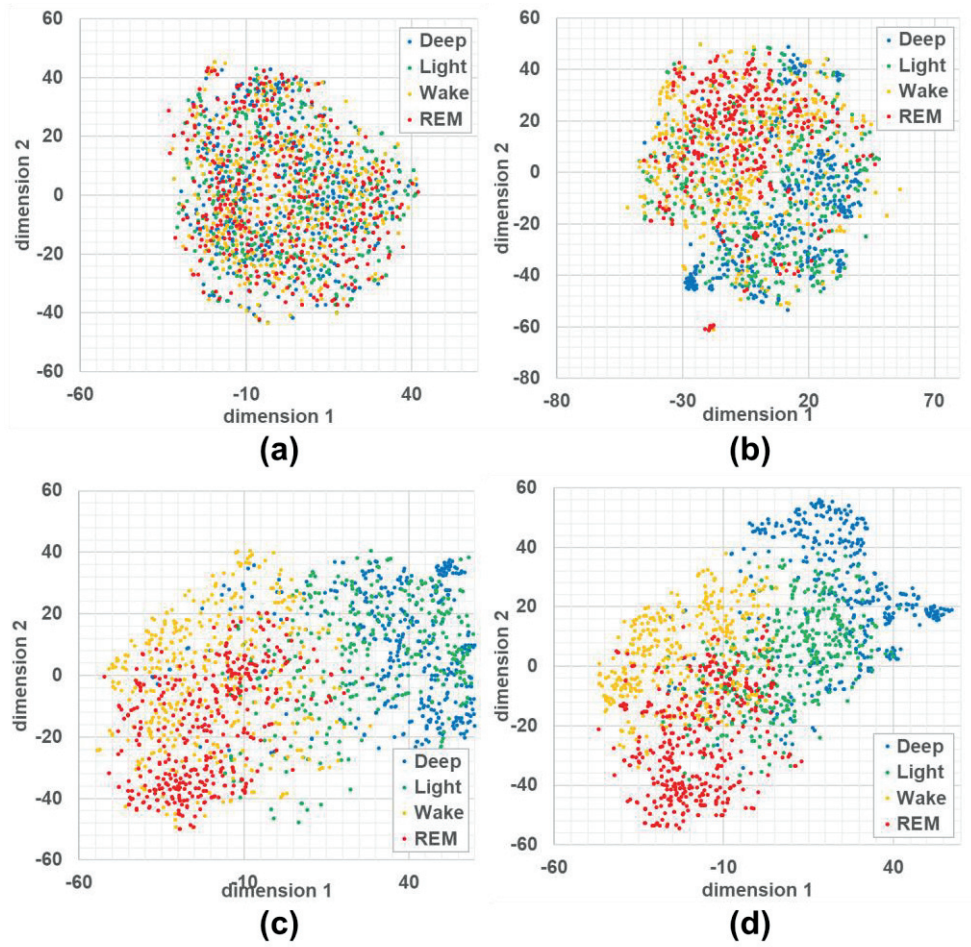
Figure 2-13. Visualization of the separability of outputs using t-SNE from (a) input layer, GDV = –0.02 (b) first LSTM layer, GDV = –0.13 (c) second LSTM layer, GDV = –0.28 and (d) attention layer, GDV = –0.42.

2.2.5. Additional Analysis

Although 17 PSG recordings were used as the test dataset in this experiment, there is a possibility that the model shows high performance only in the test dataset. Therefore, I additionally conducted 3-fold cross validation without hyperparameter tuning. 51 subjects were randomly divided into 3 equal-size subsets. In each 3-fold loop, 2 subsets (N = 34) were used as the training dataset, and 1 subset (N = 17) was used for test dataset. Hyperparameters were selected as the combination chosen in the hold-out method. The results are shown in Table 2-5. The model achieved a Kappa of 0.72 ± 0.14 and accuracy of 82.0 ± 7.3% across the entire dataset. In addition, the result also showed significant difference with the performance of LSTM model. Considering that there is no significant difference in the performance for the test datasets in all folds compared to the results presented in Table 2-4, it can be argued that my model performs well for all subjects.

Table 2-5. Accuracy and Cohen's kappa coefficient on 3-fold cross validation

Dataset		N	Network	Accuracy (%)	Kappa
Fold 1	Training	34	AB-LSTM	85.8 ± 6.6	0.76 ± 0.16
	Test	17	AB-LSTM	82.7 ± 7.4	0.73 ± 0.12
Fold 2	Training	34	AB-LSTM	86.0 ± 7.4	0.76 ± 0.20
	Test	17	AB-LSTM	81.5 ± 7.6	0.70 ± 0.18
Fold 3	Training	34	AB-LSTM	85.5 ± 7.1	0.74 ± 0.16
	Test	17	AB-LSTM	81.8 ± 7.1	0.71 ± 0.11
Overall	Test	51	AB-LSTM	82.0 ± 7.3	0.72 ± 0.14
		51	LSTM	75.6 ± 7.8	0.66 ± 0.12
			p-value[*]	0.001	0.024

AB-LSTM, attention-based Bi-LSTM, * Wilcoxon signed-rank test

To interpret the better performance of the proposed model, I need to understand how the different features contribute to the classification of sleep stages. Therfore, I set the value of parameters one by one to zero, and then trained the model and evaluated the Cohen's kappa value for sleep stage classification. The reduction of the classification performance according to the exclusion of each parameter indicates how useful the parameter is for classifying sleep stage. Fig. 2-14 shows the degradation in sleep stage classification performance corresponding to the parameters with the surrogate data. A higher degradation indicates the higher importance of the feature in classifying sleep stages.
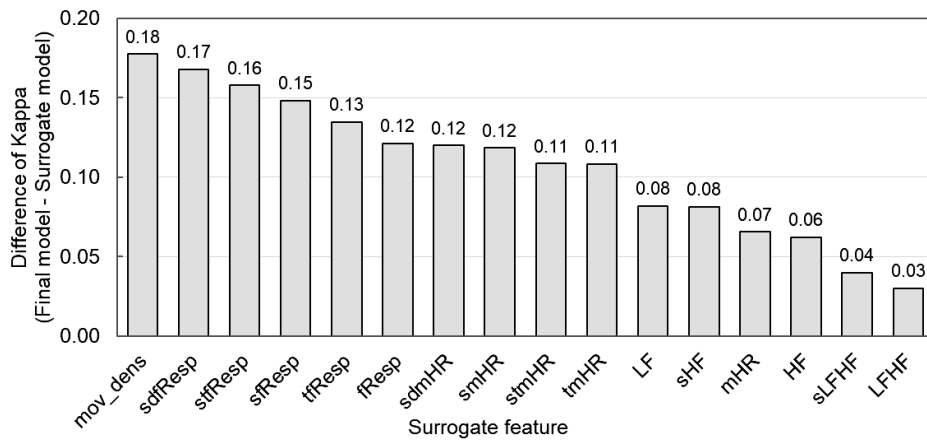
Figure 2-14. Difference of Cohen's kappa coefficient according to the surrogate feature.

In order to verify how reliable the extracted heartbeat and respiration signals were reliable, I conducted two additional comparative experiments. First, I compared the heart rate (mHR) and respiration rate (f_{Resp}) calculated using the heartbeat and respiration signals extracted from the radar to ECG and respiratory effort of the PSG. To rule out the effect of any movements, any epoch with mov_dens > 0 was excluded and interpolated for analysis. As a result, the correlation coefficient of heart rate was 0.87 ± 0.08 and respiratory rate was 0.83 ± 0.10. Moreover, the correlation coefficient did not differ by sex (BR: Male = 0.86 ± 0.07, Female = 0.88 ± 0.08, p-value = 0.382; HR: Male = 0.85 ± 0.10, Female = 0.81 ± 0.10, p-value = 0.182). Second, I investigated the influence of the radar signal according to the posture during sleep. For reference on posture during sleep (supine, left, and right), body position recorded at 500Hz on PSG was used. As a result, both accuracy and Kappa were almost the same in all posture (supine position: accuracy = 82.5 %, Kappa = 0.73, N = 11151 epochs; left position: accuracy = 80.7 %, Kappa = 0.70, N = 1188 epochs; right position: accuracy = 82.8 %, Kappa = 0.72 ± 0.15, N = 1675). Based on these results, it can be argued that the heartbeat and respiration signals obtained by my algorithm were sufficiently reliable and that the input parameters of deep learning networks could be accurately calculated.

2.3. Discussion

2.3.1. Comparison with Automatic Scoring Model of Previous Studies

Table 2-6 lists the recent studies on sleep stage classification that use cardiorespiratory and movement-related signal based on RF sensors. The sizes of dataset in this study was sufficient for training the deep learning model. The attention-based Bi-LSTM model outperformed the models of previous works.

Table 2-6. Comparison of performance with previous studies

Study	RF Sensors	N	Feature	Classifier	Performance (Mean ± S.D.)
Class = 2 Stages					
Pallesen et al. [35]	IR-UWB	14	MOV, RR	Static threshold	Acc = 93.1 ± 3.8 %, kappa = 0.67 ± 0.14
Class = 3 Stages					
Chung et al. [104]	Doppler	38	MOV, RR, HR	DNN + RanF	Acc = 67.9 %, kappa = 0.37
Zhang et al. [105]	Wifi CSI	52	MOV, RR, HR	Threshold, SVM	Acc = 88.4 %
Class = 4 Stages					
Tataraidze et al. [107]	FMCW	32	MOV, RR	GBM + LDA	Acc = 63.5 ± 0.08 %, kappa = 0.49 ± 0.12
Zaffaroni et al. [108]	S+, ResMed	40	-	-	Acc = 70 %, kappa = 0.53
Zhao et al. [64]	RF	100	-	CNN + LSTM	Acc = 79.8 ± 2.9 %, kappa = 0.70
Hong et al. [109]	Doppler	13	MOV, RR, HR	k-NN	Acc = 81 %
Toften et al. [36]	IR-UWB	71	MOV, RR	TCN, LSTM	Acc = 76 ± 7 %, kappa = 0.63 ± 0.10
This study	**IR-UWB**	**51**	**MOV, RR, HR**	**AB-LSTM**	**Acc = 82.6 ± 6.7 %, kappa = 0.73 ± 0.11**

RF, Radio frequency; S.D., Standard deviation; FMCW, Frequency-modulated continuous-wave radar; RF, Radio Frequency; IR-UWB, Impulse-radio ultra-wideband; Wifi CSI, Wireless fidelity channel state information; MOV, Movement; RR, Respiration rate; HR, Heart rate; GBM, Gradient boosted machine; LDA, Linear discriminant analysis; CNN, Convolutional neural network; LSTM, Long short-term memory; DNN, Deep neural network; RanF, Random forest; k-NN, k-nearest neighbor; SVM, Support vector machine; TCN, Temporal convolutional network; AB-LSTM, attention-based Bi-LSTM; Acc, Accuracy; kappa, Cohen's kappa coefficient.

2.3.2. Role of Attention Mechanism in Sleep Staging

According to Table 2-4, the attention mechanism enhances the sleeps stage classification capability of LSTMs based on IR-UWB radar signals. In particular, as shown in Fig. 2-11, the increase in TP for light sleep was mainly because of the decrease in FN in deep sleep and vice versa. Similarly, the attention-based Bi-LSTM differentiates between REM sleep and wake more precisely. This observation agrees with the visualization of the outputs from the LSTM layers and attention layer (Fig. 2-13). In Fig. 2-13 (c), although the LSTMs showed successful separability for wake and REM sleep from non-REM sleep, they failed to learn coherent information between light sleep and deep sleep, as well as REM sleep and the wake stage. In contrast, the outputs from the attention layer showed clearly distinct clusters. Thus, it can be interpreted that the attention mechanism plays a key role in separating light sleep and deep sleep, and REM sleep and wake in the attention-based Bi-LSTM.

Fig. 2-12 shows that the model utilizes the attention module to give different weight distributions based on the importance of each timestep according to the sleep stage. As can be seen from the results, the current epoch contains particularly prominent characteristics in deep sleep and wake, whereas less attention was given to the current epoch during light sleep and REM sleep. The variance in the attention weights over timesteps is relatively higher in the deep sleep and wake stage. Considering that the probability of each epoch remaining the same stage of deep sleep and wake is lower than that for light and REM sleep in both training and test datasets (probability: REM sleep = 0.962, light sleep = 0.933, deep sleep = 0.793, wake = 0.624), it might be supposed that the meaningful information for identifying deep sleep and wake is mainly contained in the current epoch. Moreover, it is well

known that body movement, which is the most important phenomenon characterizing wake stage, mostly occurs in short periods during the wake epoch [110].

In Fig. 2-12, significant timesteps to determine the sleep stage during light sleep and REM sleep are 1 and 2 epochs before the current epoch, respectively. ANS is closely related along anatomical and physiological lines to the sleep regulation system, and also plays a role in the interaction between central nervous system (CNS) and cardiorespiratory system. The interaction is associated with sleep control and modulated according to CNS-defined sleep stages [93]. Notably, the previous studies have shown that the distinct physiological states, i.e., sleep stages, have an influence on the characteristic time lags underlying the interaction [97]. Considering these physiological relationships, my result is in line with previous studies reporting that cardiorespiratory dynamics precedes modulation in EEG activity [111, 112]. Therefore, I can speculate that attention mechanism can capture such physiologically stage-specific characteristics, which effectively distinguish the sleep stages.

2.3.3. Importance of Features

According to Fig. 2-14, movement was the most important for determining the sleep stage, followed by respiration-related features and time-domain HRV indices. Finally, the frequency-domain HRV indices were the least important.

In my previous work, I conducted a similar analysis to evaluate the importance of features for sleep staging using PVDF [60]. Movement was the most important feature consistently in that study as well. However, the frequency-domain HRV parameters were the second-most important feature group, which is contrary to the results of the present study. Considering this, it can be suggested that the significance of features for sleep staging depends on the device.

2.3.4. Limitations

The proposed method has a few limitations. First, this study investigated the recordings of only healthy subjects. Considering the prevalence of sleep disorders and its comorbidities, further studies are required to validate my model to generalized populations. Moreover, the IR-UWB sensor has potentials for application in monitoring and diagnosing patients with sleep disorders such as sleep-related breathing disorders, movement disorder, insomnia, and narcolepsy. Second, my study was conducted in a controlled laboratory environment. Given the real-world environment, more complex noise or non-static clutter might occur during nocturnal sleep. Therefore, to confirm the feasibility, my model ought to be tested in home environments in future studies. Finally, for an interpretable deep learning model with good performance, I used the calculated features and RNNs in this study. However, further studies on reducing the model complexity should be investigated.

3

LSTM Model for Real-Time Apnea-Hypopnea Event Detection Based on IR-UWB Radar

A non-contact sleep staging model using attention-based Bi-LSTM network model was investigated in the previous chapter, in which it was concluded that the proposed model achieved the best classification performance compared with previous works. To enhance the usability of the sleep monitoring, a technology that extends the range of users to patients with sleep disorders as well as normal people is required. For this purpose, a method for monitoring sleep breathing disorder is investigated in this chapter, particularly, sleep apnea hypopnea event detection is studied. A hybrid CNN-LSTM network model is proposed for classifying apnea-hypopnea event in real-time. Based on the classified outputs, SAHS diagnostic efficacy of the proposed approaches is compared with that of previous works. Finally, the clinical effectiveness of the model is discussed.

3.1. Methods

3.1.1. Subjects and Measurements

This study was performed in accordance with the ethical standards in the Declaration of Helsinki, and the Institutional Review Board of Seoul National University Hospital (IRB-SNUH No. 1807-190-964) approved this prospective cohort. All participants were briefed about the objective and procedure of the experiment, and they signed the consent forms. Subjects with suspected SAHS were recruited from clinic populations and the online clinical trials center of SNUH.

Subjects were initially screened by study coordinators to ensure that they met inclusion criteria and exclusion criteria. The inclusion criteria were adults whose ages were $\geq$ 18 yrs. and who were judged as a part of the high-risk group in both the and STOP-BANG questionnaire [115] and Berlin questionnaire [116] (Fig. 3-1 (a) and (b)). Exclusion criteria were people who had any history of sleep disorders other than SAHS, psychiatric, neurological, or cardiovascular disorders. Qualifying 40 subjects underwent overnight PSG at the Center for Sleep and Chronobiology of SNUH. As a result of the PSG, 4 PSG recordings for which incomplete data were collected due to a defective cable connecting the IR-UWB radar and PC were excluded, leaving 36 PSG recordings for analysis. SAHS was diagnosed with an AHI > 5 events/h, and they were classified into three groups: mild SAHS (5 $\leq$ AHI < 15 events/h), moderate SAHS (15 $\leq$ AHI < 30 events/h), and severe SAHS (AHI $\geq$ 30 events/h). The IR-UWB radar was measured simultaneously with PSG and time-synchronized with PSG system in time. The anthropometric and sleep parameters of the subjects are summarized in Table 3-1.

(a) STOP-BANG Questionnaire	
S (snore): Do you snore loudly?	**B** (BMI): Is your body mass index more than 35 kg/m^2?
T(tired): Do you often feel tired, fatigued, or sleepy during the daytime?	**A** (age): Are you older than 50 years?
O (observed to have stopped breathing): Has anyone observed that you stop breathing, or choke or gasp during your sleep?	**N** (neck size): Is your neck circumference greater than 43 cm (for male)/ 41 cm (for female)
P (high blood pressure): Do you have or are you being treated for high blood pressure?	**G** (gender): Are you male?
STOP-BANG score: OSA (obstructive sleep apnea) low risk: (0-2) OSA (obstructive sleep apnea) intermediate risk: (3-4) OSA (obstructive sleep apnea) high risk: (5-8) or STOP score $\geq$ 2 and one of the following conditions is satisfied: B, N, G	

Figure 3-1. (a) STOP-BANG questionnaire

(b) Berlin Questionnaire	
Category 1	**Category 2**
1. Do you snore?	**6. How often do you feel tired or fatigued after your sleep?**
a. Yes	a. Nearly every day
b. No	b. 3-4 times a week
c. Don't know	c. 1-2 times a week
	d. 1-2 times a month
2. If yes, your snoring is?	e. Never or nearly never
a. slightly louder than breathing	
b. As loud as talking	**7. During your waking time, do you feel tired, fatigued or not up to par?**
c. Louder than talking	a. Nearly every day
d. Very loud – can be heard in adjacent room	b. 3-4 times a week
	c. 1-2 times a week
3. How often do you snore?	d. 1-2 times a month
a. Nearly every day	e. Never or nearly never
b. 3-4 times a week	
c. 1-2 times a week	**8. Have you ever nodded off or fallen asleep while driving a vehicle?**
d. 1-2 times a month	a. Yes
e. Never or nearly never	b. No
4. Has your snoring ever bothered other people?	**If yes, how often does it occur?**
a. Yes	a. Nearly every day
b. No	b. 3-4 times a week
c. Don't know	c. 1-2 times a week
	d. 1-2 times a month
5. Has anyone noticed that you quit breathing during your sleep?	e. Never or nearly never
a. Nearly every day	
b. 3-4 times a week	**Category 3**
c. 1-2 times a week	**9. Do you have high blood pressure?**
d. 1-2 times a month	a. Yes
e. Never or nearly never	b. No

Scoring Question: Any answer with gray highlight is a positive response
Scoring Categories: Category 1 is positive with 2 or more positive response to questions 1-5
   Category 2 is positive with 2 or more positive response to questions 6-8
   Category 3 is positive with positive response to questions 9

High risk: if two to three categories are scored as "positive"
Low risk: if zero to one category is scored as "positive"

Figure 3-1. (b) Berlin questionnaire

Table 3-1. Subject demographics and sleep-related variables

Variable /Group	Non-SAHS	Mild SAHS	Moderate SAHS	Severe SAHS
N (M/F)	4/2	6/4	5/4	7/4
Age (yrs.)	30.4 (8.9)	37.4 (16.9)	43.2 (13.6)	40.7 (8.3)
BMI (kg/m^2)	22.1 (4.9)	24.2 (4.1)	28.4 (4.8)	30.3 (7.4)
TRT (min)	443.2 (52.9)	427.2 (39.5)	394.9 (72.1)	435.6 (30.9)
SE (%)	89.9 (6.8)	92.5 (4.6)	91.7 (4.8)	88.4 (9.1)
SOL (min)	8.5 (7.8)	8.1 (13.6)	3.9 (5.2)	4.5 (5.8)
PLMI (evts/h)	1.3 (1.6)	3.8 (5.1)	2.4 (4.0)	0.6 (1.2)
AHI (evts/h)	3.3 (0.7)	8.6 (2.8)	22.8 (3.2)	92.0 (80.4)
# A evts	2.0 (2.1)	10.8 (16.1)	20.8 (29.3)	172.5 (161.3)
# H evts	19.7 (4.9)	45.1 (17.3)	117.7 (42.7)	197.8 (172.7)

All results presented as mean (standard deviation). N-number of subjects; M-male; F-female; BMI-body mass index; TRT-total recording time; SE-sleep efficiency; SOL-sleep onset latency; PLMI-periodic limb movement index; AHI-apnea–hypopnea index; evts-events; A-apnea; H-hypopnea; SAHS-sleep apnea and hypopnea syndrome.

3.1.2. Signal Preprocessing and Segmentation

Because the multiple baseband signals obtained by the IR-UWB radar sensor can be expressed as a 2-D virtual image, the raw signals were preprocessed and segmented to train and test the CNN-LSTM model. The window size was set to 20 s with a shift of 1 s to sufficiently reflect the normal breathing section. The image segment included all ranges from 0 to 2 m. In the sleep environment, the background is considered to be stationary, while the human body changes its position in terms of chest movements caused by breathing. Therefore, I obtained the target signals by subtracting the DC component from the raw radar signals with a moving-average method [90]. The clutter was calculated as the average of each 20-s epoch amplitudes along the fast time range bin. Then, the image was downsized from 300×400 pixels to 80×300 pixels with area interpolation method to increase learning speed.

Fig. 3-2 summarizes the examples of synchronized IR-UWB radar image, nasal airflow, and thoracic respiratory effort during no apnea event, central sleep apnea (CSA), obstructive sleep apnea (OSA), and hypopnea for 1 min. During no apnea event (Fig. 3-2 (a)), respiration activity that occurs at a distance of approximately 1 m in the radar image is clearly visible. In contrast, in central apnea (Fig. 3-2 (b)), it can be seen that the airflow and thoracic waveform disappear, and at the same time, the breathing pattern in the radar image also disappears. In the obstructive apnea (Fig. 3-2 (c)) and hypopnea (Fig. 3-2 (d)), a significant decrease appears in the thoracic waveform, and the contrast due to breathing in the radar image is weakened.

The preprocessed images were categorized into two classes: AH and N. If at least 10 s of a segment occurred within an apnea and hypopnea event period, it was labeled as class AH. Other cases were labeled as class N. As a result, 138,067

segments were labeled as class AH, and 778,471 segments were labeled as class N. Next, to prevent the model from overfitting to the majority number of the class, I made the training set consist of the same number of samples for each class. Because class N had more segments than class AH, for each subject, class N segments were randomly subsampled by the imbalance ratio. The original imbalance ratio of the training set, which is the number of class N segments divided by the number of class AH, was 5.64.
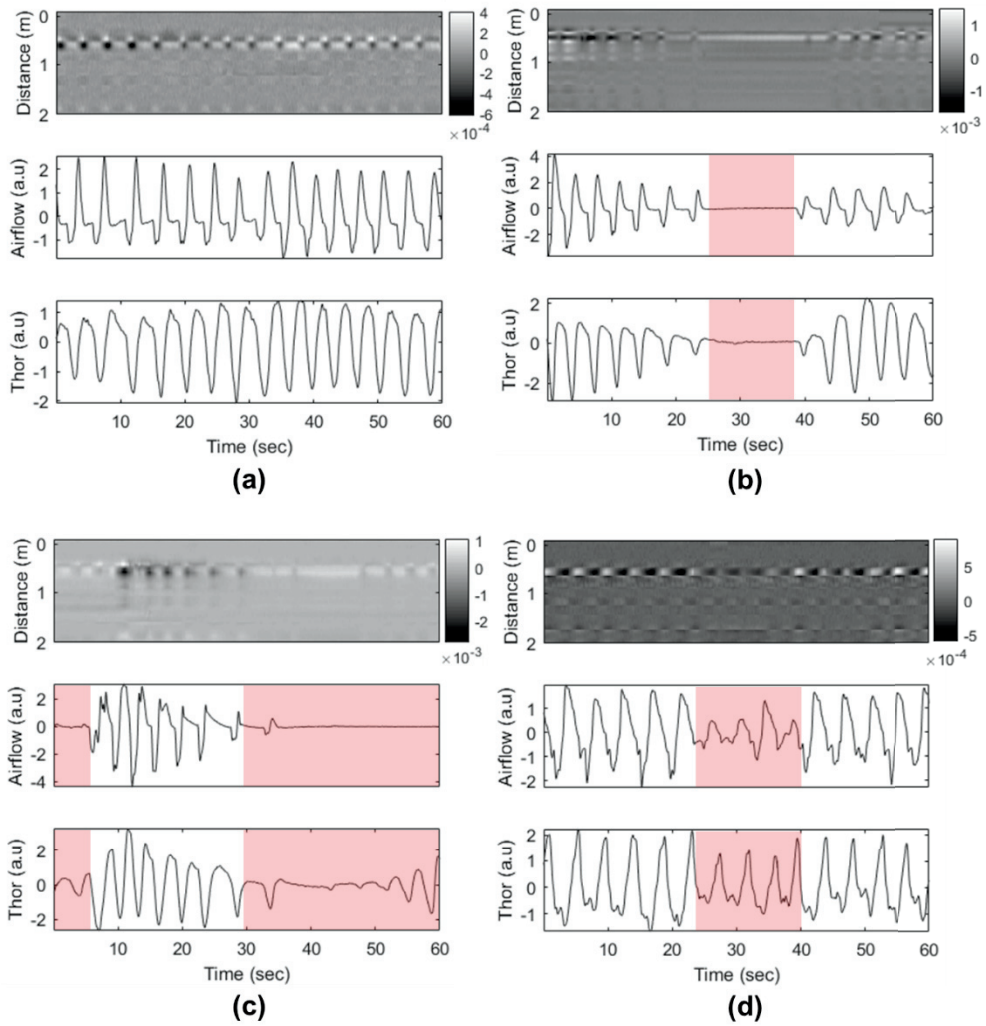
Figure 3-2. Example of IR-UWB raw data and different normalized PSG signals with normal breath and three types of respiratory events (a) no apnea event (b) central sleep apnea (c) obstructive sleep apnea, and (d) hypopnea. Red areas indicate where the respiratory event occurred. The PSG signals were recorded at 500 Hz, and the IR-UWB signals were sampled with 20 frame per second.

3.1.3. CNN-LSTM Architecture

As shown in Fig. 3-3, the proposed deep learning network CNN-LSTM architecture comprises three convolutional layers, two max-pooling layers, one bidirectional LSTM layer, and one fully connected layer. Note that the input image is treated as a 1-D signal based on time axis, and 1-D CNNs were applied. This is to better preserve the temporal characteristics of the breathing pattern and feed the time-dependent feature vectors into the LSTM units. Moreover, 1-D CNN has less computational complexity than 2-D CNN [117].

To find the optimal hyperparameters and evaluate the model performance, I used a nested 6-fold cross-validation. To do this, subjects were randomly divided into 6 equal-size subsets. In the outer 6-fold cross-validation loop, 5 subsets were used as the training dataset, and 1 subset was used for test dataset. At each inner fold, the training dataset was further divided into 6 equal-size subsets. Then, 5 subsets were used as the training dataset, and 1 subset was used for the validation dataset. From the results of the 6×6 inner folds, the best hyperparameters were selected by maximizing Cohen's kappa coefficient for the validation dataset. The candidate hyperparameters were set as follows: number of CNN layers {1-5}; number of convolution filters {32, 64, 128, 256, 512}; kernel size for convolution {3, 5, 7, 9, 11}; type of LSTM structure {unidirectional, bidirectional}; and number of LSTM units {32, 64, 128, 256, 512}. Table 3-2 shows the detailed configuration of various layers of the proposed model. The total number of parameters required for learning in all layers was 1,793,986.

In each CNN layer, the layer input and the kernel were convolved with the stride of 2 and same padding. I trained the model using the Adam optimizer [98] and He

normal initializer [118]. Learning rate was set to 0.001, and binary cross-entropy was used as a loss function. The model was trained for maximum 100 epochs with an early stopping patience of 10 and the batch size of 128. Models were implemented in Python 3.7 and the Keras framework [101] with TensorFlow backend [102]. The training and test process were done using a GTX 1080 8GB GPU and a 3.4 GHz Intel i7-6700 CPU.
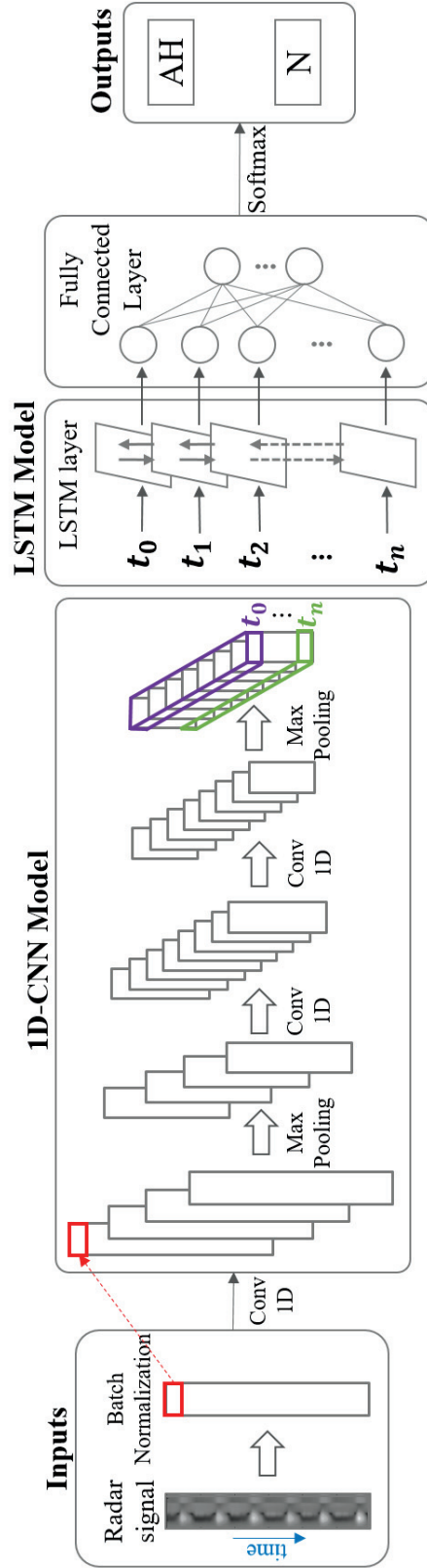
Figure 3-3. Overall structure of the proposed CNN-LSTM model

Table 3-2. Configuration of various layers of the proposed model

Layer Type	Filter Size@ Kernel Size	Activation function	Output Shape	Number of parameters
Input			300×80	
BatchNorm_1			300×80	320
Conv 1D_1	64@1× 9	ReLu	150×64	46,144
Max pool_1	2 x 1		75×64	
Conv 1D_2	128@1× 5	ReLu	38×128	41,088
Conv 1D_3	256@1× 3	ReLu	19×256	98,560
Max pool_2	2 x 1		9×256	
LSTM_1	512	Sigmoid	512	1,574,912
FC_1		ReLu	64	32,832
Output		Softmax	2	130

BatchNorm-batch normalization; Conv-convolutional neural network; Max pool-max pooling; LSTM-long short-term memory network; FC-fully connected.

3.1.4. Performance Evaluation

After finding the optimal hyperparameters, I evaluated the performance of the sleep apnea event detection for the test dataset. Test segments were applied to the CNN-LSTM model, and classification outputs, representing class AH or class N, were received. Finally, these time sequenced classified labels were then fed to the event detector to identify valid AH events. The event detector judges valid AH events if at least six consecutive segments are classified as class AH.

To evaluate the performance of the sleep apnea event detection, I performed three analyses. First, the conventional metric of accuracy (ACC), sensitivity (SENS), specificity (SPEC), and Cohen's kappa coefficient (Kappa) were calculated by the generating a confusion matrix between the estimated results and reference PSG results, according to a segment-by-segment analysis. Second, I estimated AHI based on the number of valid AH events. Then, the Pearson's correlation analysis and Bland-Altman analysis between the estimated AHI and reference PSG AHI were conducted. Lastly, the SAHS diagnostic performance for AHI cutoff $\geqq$ 5, 15, and 30 events/h was validated with ACC, SENS, SPEC, positive predictive value (PPV), and Kappa.

3.2. Results

3.2.1. Segment-by-segment analysis

I calculated the time taken to classify all the test segments to verify the real-time application of the trained model. As a result, because it took 211.9 seconds to classify the 916,538 test segments, it took approximately 0.00023 seconds to classify one segment on average. Therefore, this model is sufficiently capable of real-time event detection.

Fig. 3-4 (a) presents the performance graph of the accuracy from the training and validation dataset, and Fig 3-4 (b) illustrates the performance graphs of the loss function from the training and validation dataset.

Table 3-3 shows the performance of the proposed model for the test dataset based on segment-by-segment analysis. AH events classified by the CNN-LSTM model were compared with the scored AH events from the reference PSG. When computing the performance for apnea event detection in each of the severity groups, SENS and Kappa value tended to gradually increase from the non-SAHS group (AHI < 5 events/h) to the severe SAHS group (AHI $\geq$ 30 events/h). For the overall test segments, I obtained a ACC of 0.930, SENS of 0.781, SPEC of 0.956, and Kappa of 0.728.

Table 3-3. Event detection performance in each SAHS severity group

		Reference PSG		Group	ACC	SENS	SPEC	Kappa
		AH	**N**					
IR-UWB								
	AH	107835	34309	Non-SAHS	0.975	0.689	0.980	0.504
				Mild SAHS	0.959	0.698	0.973	0.618
				Moderate SAHS	0.919	0.748	0.946	0.668
	N	30232	744162	Severe SAHS	0.856	0.806	0.923	0.737
				Overall	0.930	0.781	0.956	0.728

71
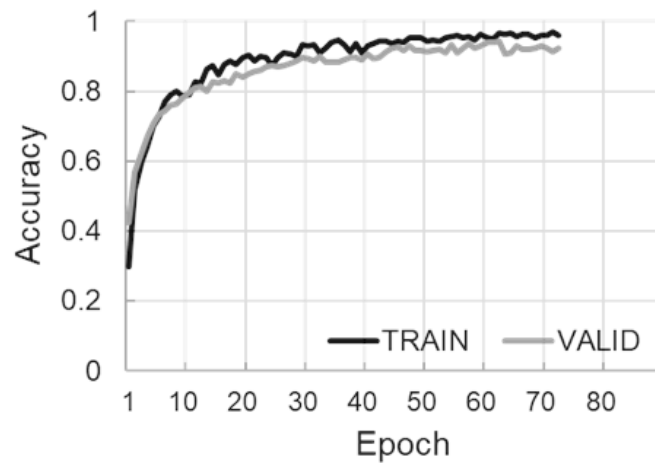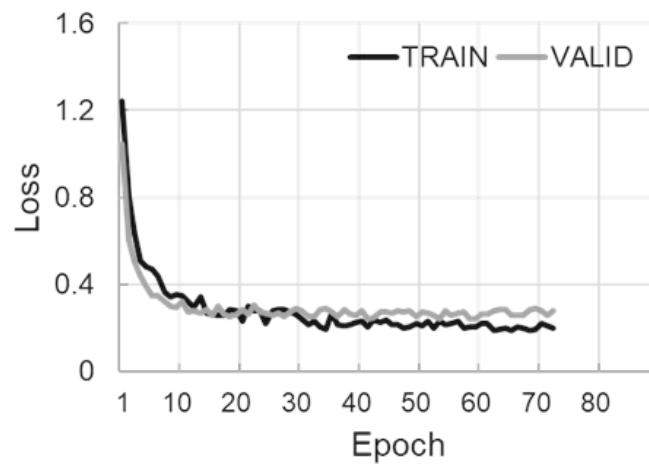
Figure 3-4. Performance graph of (a) the accuracy and (b) loss function from training and validation dataset

3.2.2. AHI estimation analysis

Intra-class correlation coefficients and Bland-Altman plots were used to assess the agreement between the estimated AHI (AHI_{EST}) from IR-UWB radar and PSG-derived AHI (AHI_{PSG}). Fig. 3-5 (a) and (b) show the scatter plots of the AHI_{EST} versus AHI_{PSG} and Bland-Altman plots for the entire night sleep. The Pearson correlation coefficient (N = 36) was 0.970 with $p < 0.001$. Bland-Altman shows low mean biases (-1.983), and the limits of agreement were -14.655 to 10.689. To verify the performance in detail, the AHI of each hour was additionally calculated from both the IR-UWB radar and PSG for each subject. Fig. 3-5 (c) shows that the Pearson correlation coefficient for overall samples (N = 258) was 0.955 with $p < 0.001$. Moreover, Bland-Altman also shows low mean biases (-1.567), and the limit of agreement was -17.710 to 15.375 (Fig 3-4 (d)).

Table 3-4 and Table 3-5 summarize the SAHS severity classification and diagnostic performance for all test subjects. The diagnostic performance was calculated for AHI cutoffs of 5, 15, and 30 events/h. Therefore, the average values for ACC, SENS, SPEC, PPV, and Kappa were 0.98, 0.97, 1.00, 1.00, and 0.96.
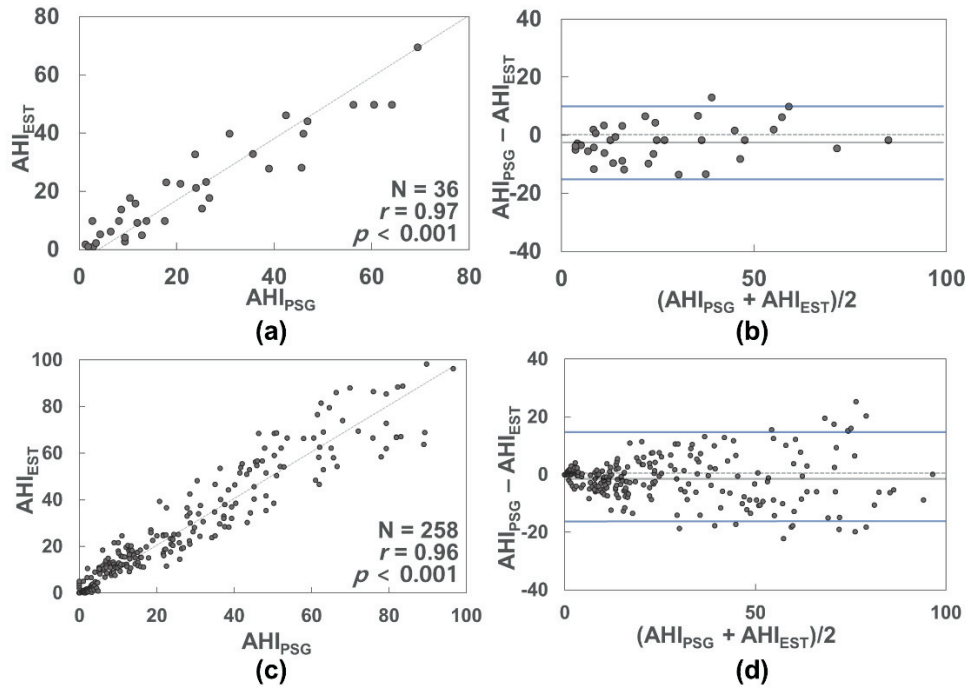
Figure 3-5. Scatter plots of estimated AHI using the proposed method (AHI_{EST}) versus reference AHI obtained from polysomnography (AHI_{PSG}) for (a) total sleep time and (c) each hour from all subjects. Bland-Altman plots for visualization of the agreement between AHI_{EST} and AHI_{PSG} for (b) total sleep time and (d) each hour from all subjects. Gray line indicates an identity line in (a, c). Gray bold line and blue lines in (b, d) indicate the average difference (Bias) and the average $\pm 1.96*$standard deviation, respectively.

<div align="center">

Table 3-4. SAHS severity classification

</div>

		Reference SAHS severity			
		Non-	**Mild**	**Moderate**	**Severe**
IR-UWB	**Non-**	6	0	0	0
	Mild	0	9	0	0
	Moderate	0	1	8	0
	Severe	0	0	1	11

<div align="center">

Table 3-5. SAHS diagnostic performance

</div>

AHI cutoff (events/h)	**ACC**	**SENS**	**SPEC**	**PPV**	**Kappa**
≥ 5	1.00	1.00	1.00	1.00	1.00
≥ 15	0.97	1.00	0.94	0.95	0.94
≥ 30	0.97	1.00	0.96	0.92	0.94
Average	0.98	1.00	0.97	0.96	0.96

3.3. Discussion

The purpose of this study was to develop a new diagnostic algorithm for real-time SAHS monitoring by using a deep learning method based on a non-contact sensor. A single IR-UWB radar was used as a breathing monitoring device, and a hybrid model combining CNNs and LSTM network was used as a classifier for the AH event in this study. To demonstrate the ability to detect an AH event in real-time, the radar image with overlapping window was input into the CNN-LSTM model.

An important feature of the proposed method is that it detects individual AH events in real-time and reports the results as per-segment classification performance. Most previous studies do not report event detection performance and only present AHI for the entire sleep. Javaid *et al.* [65], which detected AH events using a machine learning technique with an under-the-mattress IR-UWB radar sensor, only showed an accuracy, sensitivity, and specificity of 0.73, 0.71, and 0.71, respectively. However, the number of participants in this study was very small, and the classification performance also could be improved.

3.3.1. Agreement between Proposed Method and Reference PSG

In the segment-by-segment analysis, the Kappa value of the proposed model for all dataset segments was 0.72, which is a substantial agreement. Table 3-3 reports a confusion matrix for overall segment-by-segment comparison between the proposed CNN-LSTM model and reference PSG. The total number of false positives was 34309. Notably, of the 34309 false positives, about 68% contain events in which the nasal airflow signal had an amplitude reduction of more than 30% compared to the baseline. But the segments were recorded as "no event" because of the absence of the desaturation or arousal of SpO_2. Because additional information for classifying hypopneas, such as EEG or SpO_2, was not considered in this study, I speculate that these segments would have led the model to misclassify it into class AH and cause overestimated AH events. In Table 3-3, the Kappa value for all SAHS severity groups exceeded 0.5, which reveals the proposed model detects the AH events with substantial agreement. Note that the performance indices tend to gradually increase from the non-SAHS group to severe SAHS group. When the ratio of hypopnea event and apnea event was calculated for each severity group, it was found to be 9.85, 7.22, 3.87, and 1.15 in each group. In other words, the aforementioned overestimation problem for hypopnea event deteriorated the model performance in the non-SAHS and mild SAHS groups, where the ratio of hypopnea is relatively high. In addition, according to Fig. 3-5 (b), the average bias value is -1.983, which means that the model overestimated AH events as well.

However, despite the problem of overestimation, the sensitivity in the non- and mild SAHS groups is lower than that of the moderate and the severe SAHS groups (Table 3-3). Considering that the proportion of hypopnea events is relatively large in

the non- and mild SAHS groups, it is reasonable to think that the factor that induced misclassification into class N is a hypopnea event rather than an apnea event. In fact, about 92% of false negatives in non- and mild SAHS groups contained hypopnea events. These misclassifications can be understood from Fig. 3-2 in which the hypopnea event shows breathing weakens and the amplitude decreases, but has a similar the waveform image as no event. In addition, the average duration of hypopnea events in the normal and mild groups was 26.3 s, whereas the average duration of hypopnea corresponding to false negatives was only 13.5 s. In other words, it can be assumed that the characteristic is not well reflected in the radar image, so that it is misclassified as class N when a hypopnea event has a short duration in non- and mild SAHS groups.

3.3.2. Comparison with SAHS Diagnostic Models in Previous Studies

As shown in Table 3-5, despite the bias of IR-UWB radar compared to the AHI of the reference PSG, the SAHS diagnostic performance showed a high value. The outstanding diagnostic performance and the statistically high AHI correlation of the model supports the fact that it can be used to clinically screen or continuously monitor SAHS severity. When the results are compared with the previous studies based on IR-UWB-radar technologies [66, 67], my results achieved better or equal level of performances among the used metrics (Average diagnostic performance in [66]: ACC = 0.99, SENS = 0.99, SPEC = 0.99, PPV = 0.99, Kappa = 0.98, correlation coefficient of AHI = 0.97; Average diagnostic performance in [67]: ACC = 0.95, SENS = 0.88, SPEC = 0.98, PPV = 0.98, Kappa = 0.88, correlation coefficient of AHI = 0.93). This study is the first to investigate the effectiveness of deep learning for sleep-related breathing disorders based on RF sensor. Fig. 3-2 shows that when an AH event occurs, the amplitude of the sinusoidal form of thoracic movement due to breathing decreases, which in turn weakens the contrast by oscillation in the radar image. CNN modules can learn these prominent characteristics and extract robust features [119]. Therefore, contrary to the existing rule-based algorithms based on hand-engineered features that might miss important sleep apnea markers, my method does not include the process of extracting and selecting hand-engineered features [120]. Moreover, LSTM layer plays a key role in identifying the sequence pattern information and short-term and long-term dependencies [121]. Thus, the proposed model was able to determine whether an AH event occurred in the segment with high performance in real-time with only 20-s of radar image.

3.3.3. Robustness of Classification algorithm for Human Location

In addition, the high-performance detection ability was maintained in various positions where the breath occurs. To confirm that my experiment showed high AH event detection performance evenly, regardless of the location of the human chest, I additionally calculated the distribution of the distances from the radar device to the subject's chest, and analyzed the correlation with the detection performance. The distance of the chest with the highest power could be chosen by calculating the spectral power of the radar signal in the respiration frequency range (0.1–0.7 Hz) in the fast-time domain. The calculation was performed every 30-s epoch, but the epochs recorded as the stage "Wake" or containing AH events were excluded from the analysis. Therefore, the mean distance during the whole night's sleep for each subject was 0.98 ± 0.31 m with a range of $0.84 - 1.32$ m, and it did not show any significant correlation with all AH event detection performance indices.

3.3.4. Limitations

The proposed method has a few limitations. First, although the information about actual sleep time is necessary for accurate AH event classification and AHI calculation, the method could not take the sleep stage into account because the EEG is not included. Moreover, the occurrence of AH events is greatly influenced by sleep posture as well as sleep stage [122], but accelerometer was also not considered in this study. However, recently, studies have been conducted to classify the sleep stages and sleep posture based on IR-UWB radar signals [64, 84, 123]. Combining these approaches with my method will allow the model to classify AH event and calculate AHI accurately from a single IR-UWB radar. Second, this study could not compare the agreement between respiratory disturbance index (RDI) between IR-UWB radar and reference PSG. RDI is also a criterion for classifying the severity of SAHS and is very similar to AHI, but it includes the number of respiratory effort-related arousals (RERAs) as well as apnea and hypopnea [68]. However, as mentioned above, since EEG information could not be included in this study, arousal-associated respiratory events could not be classified. Future studies including additional EEG information will enhance the usefulness. Third, this study did not differentiate detailed types of AH events, i.e., obstructive sleep apnea, central sleep apnea, mixed sleep apnea, and hypopnea. As show in Fig. 3-2, the respiration signals from PSG show different patterns in which the waveform disappears or the amplitude decreases depending on the type of AH event. However, there is a limitation in recognizing the difference in the pattern only with the radar signal reflected from the chest movement. Moreover, it is difficult to train the model classifying the detailed type of AH events owing to an imbalance in the samples for

each type. Finally, the model was tested in a controlled laboratory environment. To confirm usability, the model ought to be validated at home environments in future studies.

4

Conclusion

PSG has been considered to be the gold standard tool for the assessment of sleep; however, it has a few limitations. Various alternative sleep monitoring methods have been developed to overcome the limitations of PSG. Nevertheless, most of these methods still have limitations such as the need for physical contact with the sensor, privacy issues, and inaccuracies. Even though non-contact sleep monitoring methods using RF sensors have also been proposed, they need to be improved interpretability and reliability and ensure real-time capability to increase clinical applicability.

In this thesis, an optimal deep learning model for classifying the four sleep stages using a non-contact sensor was developed. The proposed model adopted bidirectional LSTM networks based on the attention mechanism to improve interpretability and achieve better performance. A single IR-UWB radar was used, and features related to cardiorespiratory information and physical movements were extracted from the raw signal. The attention mechanism played an essential role in effectively distinguishing light and deep sleep, and REM sleep and wake stage.

According to previous studies [124, 125], inter-scorer agreement and Kappa for the five sleep-stage classification (W, N1, N2, N3, and REM) based on the AASM manual was approximately 82% and 0.76, respectively. Although the performance of the proposed method is slightly lower than the clinical level, it is higher on average compared to a state-of-the-art performance for four sleep staging using ECG signal (accuracy = $77.0 \pm 8.9\%$, Kappa = 0.61 ± 0.15) [126]. Therefore, it could be said that my method is potentially practical for long-term and in-home sleep monitoring. Moreover, these findings show that the attention mechanism might be useful in future researches about sleep stage classification based on autonomic nervous activity.

In addition, I developed a deep learning model combining CNN and LSTM network, and detected AH events based on overlapping images of IR-UWB radar, allowing it to identify actual events in real-time. Despite not using any hand-engineered features as input, the proposed method achieved the state-of-the-art performance for classifying SAHS severity regardless of the user's location. Moreover, in this model, the hybrid architecture that exploits the benefits of both deep learning techniques did not require a feature extraction and selection process. Because of the advantages of that users do not need to attach any sensor to their body, the IR-UWB radar is drawing attention as a sleep monitoring device that has the potential as an alternative to PSG [16]. The proposed method based on the IR-UWB radar and LSTM network can be utilized for a cost-effective and reliable SAHS monitoring in both hospital and home environments.

5

References

[1]   R. J. Berger and N. H. Phillips, "Energy conservation and sleep," *Behav. Brain Res.*, vol. 69, no. 1, pp. 65–73, 1995.

[2]   I. Oswald, "Sleep as restorative process: human clues," *Prog. Brain Res.,* vol. 53, pp. 279–88, 1980.

[3]   A. Rechtschaffen and B. M. Bergmann, "Sleep deprivation in the rat by the disk-over-water method," *Behav. Brain Res.*, vol. 69, no. 2, pp. 55–63, 1995.

[4]   K. L. Knutson, K. Spiegel, P. Penev, and E. V. Cauter, "The metabolic consequences of sleep deprivation," *Sleep Med. Rev.*, vol. 11, no. 3, pp. 163–178, 2007.

[5]   T.   Lange, S. Dimitrov, and J. Born, "Effects of sleep and circadian rhythm on the human immune system," *Ann. N. Y. Acad. Sci.*, vol. 1193, pp. 48–59, 2010.

[6]   M. Vandekerckhove and Y. Wang, "Emotion, emotion regulation and sleep: An intimate relationship," *AIMS Neurosci.*, vol. 5, no. 1, pp. 1–17, 2018.

[7]   B. Rasch and J. Born*,* "About Sleep's Role in Memory", *Phyisol. Rev*, vol. 93, no. 2, pp. 681–766, 2013.

[8]   J. D. F. Dinges *et al.,* "Cumulative Sleepiness, Mood Disturbance, and Psychomotor Vigilance Performance Decrements During a Week of Sleep

Restricted to 4–5 Hours per Night," *Sleep*, vol. 20, no. 4, pp. 267–277, 1997.

[9]   G. Curcio, M. Ferrara, and L. De Gennaro, "Sleep loss, learning capacity and academic performance," *Sleep Med. Rev.*, vol. 10, no. 5, pp. 323–337, 2006.

[10]   N. Goel, H. Rao, J. S. Durmer, and D. F. Dinges, "Neurocognitive consequences of sleep deprivation," *Semin. Neurol.,* vol. 29, no. 4, pp. 320, 2009.

[11]   W. M. A. Van Leeuwen *et al.*, "Sleep restriction increases the risk of developing cardiovascular diseases by augmenting proinflammatory responses through IL-17 and CRP," *PLoS One*, vol. 4, no. 2, 2009.

[12]   B. Faraut, K. Z. Boudjeltia, L. Vanhamme, and M. Kerkhofs, "Immune, inflammatory and cardiovascular consequences of sleep restriction and recovery," *Sleep Med. Rev.*, vol. 16, no. 2, pp. 137–149, 2012.

[13]   J. M. Mullington, M. Haack, M. Toth, J. M. Serrador, H. K. Meier-Ewert, "Cardiovascular, Inflammatory, and Metabolic Consequences of Sleep Deprivation," *Prog. Cardiovasc. Dis.*, vol. 51, no. 4, pp. 294–302, 2009.

[14]   M. A. Cappuccio *et al.,* "Meta-Analysis of Short Sleep Duration and Obesity in Children and Adults," *Sleep*, vol. 31, no. 5, pp. 619–626, 2008.

[15]   C. Touma and S. Pannain "Does lack of sleep cause diabetes", *Clevel. Clin. J. Med.*, vol. 78, no. 8, pp. 549–558, 2011.

[16]   K.S. Park and S.H. Choi, "Smart technologies toward sleep monitoring at home," *Biomed. Eng. Lett.,* vol. 9, no. 1, pp. 73–85, 2019.

[17]   M. Younes *et al.,* "Reliability of the American Academy of Sleep Medicine Rules for Assessing Sleep Depth in Clinical Practice," *J. Clin. Sleep Med.*, vol. 14, no. 2, pp. 205–213, 2018.

[18]   R. S. Rosenberg and S. Van Hout, "The American Academy of Sleep Medicine inter-scorer reliability program: Respiratory events," *J. Clin. Sleep Med.*, vol. 10, no. 4, pp. 447–454, 2014.

[19]   S.H. Choi *et al.,* "Real-time apnea-hypopnea event detection during sleep by

convolutional neural network," *Comput. Biol., Med,*. vol. 100, pp. 123–131, 2018.

[20]   A. Rechtschaffen, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *Natl. Institutes Heal.*, 1968.

[21]   R. B. Berry *et al.,* "The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, Version 2.0," *Am. Acad. Sleep Med.*, 2012.

[22]   I. Sadek, E. Seet, J. Biswas, B. Abdulrazak, and M. Mokhtari, "Nonintrusive Vital Signs Monitoring for Sleep Apnea Patients: A Preliminary Study," *IEEE Access*, vol. 6, pp. 2506–2514, 2017.

[23]   H. Azimi, P. Xi, M. Bouchard, R. Goubran, and F. Knoefel, "Machine Learning-Based Automatic Detection of Central Sleep Apnea Events From a Pressure Sensitive Mat," *IEEE Access*, vol. 8, pp. 173428–173439, 2020.

[24]   R. D. Chervin, "Sleepiness, Fatigue, Tired, and Lack of Energy in Obstructive Sleep Apnea," *Chest*, vol. 118, pp. 372–379, 2000.

[25]   H. Yoon *et al.,* "Sleep-Dependent Directional Coupling of Cardiorespiratory System in Patients with Obstructive Sleep Apnea," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 12, pp. 2847–2854, 2018.

[26]   Y. Peker, H. Kraiczi, J. Hedner, S. Loth, A. Johansson, and M. Bende, "An Independent Association Between Obstructive Sleep Apnoea And Coronary Artery Disease," *Eur. Respir. J.,* vol. 14, no. 1, pp. 179–184, 1999.

[27]   H. K. Yaggi *et al.,* "Obstructive Sleep Apnea as a Risk Factor for Stroke and Death," *N. Engl. J. Med.*, vol. 353, no. 19, pp. 2034–2041, 2005.

[28]   P. E. Peppard *et al.,* "Increased prevalence of sleep-disordered breathing in adults," *Am. J. Epidemiol.*, vo. 177, pp. 1006–1014, 2013.

[29]   J. Kim *et al.,* "Prevalence of sleep-disordered breathing in middle-aged Korean men and women," *Am. J. Respir. Crit. Care. Med.,* vol. 170, no. 10, pp. 1108–1113, 2004.

[30]   K. Kang, J. G. Seo, S. H. Seo, K. S. Park, and H. W. Lee, "Prevalence and Related Factors for High-Risk of Obstructive Sleep Apnea in a Large Korean Population: Results of a Questionnaire-Based Study," *J. Clin. Neurol.*, vol. 10, no. 1, pp. 42–49, 2014.

[31]   J.D. Kim *et al.,* "Non-contact respiration monitoring using impulse radio ultrawideband radar in neonates," *R. Soc. Open Sci.*, vol. 6, no. 6, 2019.

[32]   X. Liang, J. Deng, H. Zhang, and T. A. Gulliver, "Ultra-wideband impulse radar through-wall detection of vital signs," *Sci. Rep.,* vol. 8, no. 1, p. 13367, 2018.

[33]   Y. Wang, Q. Liu, and A. E. Fathy, "Simultaneous localization and respiration detection of multiple people using low cost UWB biometric pulse Doppler radar sensor," *in IEEE MTT-S Int. Microw. Symp. (IMS) Dig*, Montreal, QC, Canada, Jun. pp. 1–3, 2012.

[34]   J. M. Lee, J. W. Choi, and S. H. Cho, "Movement analysis during sleep using an IR-UWB radar sensor," *in Proc. IEEE Int. Conf. Netw. Infrastruct. Digit. Content (IC-NIDC)*, Sep. pp. 486–490, 2016.

[35]   S. Pallesen *et al.,* "A pilot study of impulse radio ultra wideband radar technology as a new tool for sleep assessment," *JCSM,* vol. 14, no. 7, pp. 1249–1254, 2018.

[36]   S. Toften, S. Pallesen, M. Hrozanova, F. Moen, and J. Grønli, "Validation of sleep stage classification using non-contact radar technology", *Sleep Med.*, 2020.

[37]   C. C. Yang and Y. L. Hsu, "A review of accelerometry-based wearable motion detectors for physical activity monitoring," *Sensors*, vol. 10, no. 8, pp. 7772–7788, 2010

[38]   L. Fraiwan, K. Lweesy, N. Khasawneh, H. Wenz, and H. Dickhaus, "Automated sleep stage identification system based on time–frequency analysis of a single EEG channel and random forest classifier," *Comput. Meth. Prog. Bio.*, vol. 108, no. 1, pp. 10–19, 2012.

[39]   L. Fraiwan, N. Khasawneh, and K. Lweesy, "Automatic sleep stage scoring with

wavelet packets based on single EEG recording," *World Acad. Sci. Eng. Technol.*, vol. 54, pp. 485–488, 2009.

[40]   S. H. Choi *et al.,* "Real-time Apnea-Hypopnea Event Detection during Sleep by Convolutional Neural Networks," *Comput. Biol. Med.*, vol. 100, pp. 123–131, 2018.

[41]   N. T. Ayas, S. Pittman, M. MacDonald, and D. P. White, "Assessment of wrist-worn device in the detection of obstructive sleep apnea," *Sleep Med.*, vol. 4, pp. 435–442, 2003.

[42]   D. P. Whie, "Assessment of accuarcy and analysis time of a novel device to monitor sleep and breathing in the home," *Sleep*, vol. 18, pp. 115–126, 1995.

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

[44]   M. Marino *et al.,* "Measuring Sleep: Accuracy, Sensitivity, and Specificity of Wrist Actigraphy Compared to Polysomnography," *Sleep*, vol. 36, no. 11, pp. 1747–1755, 2013.

[45]   D. S. Morillo, J. L. R. Ojeda, L. F. C. Foix, and A. L. Jiménez, "An accelerometer-based device for sleep apnea screening,*" IEEE T. Inf. Technol. B.*, vol. 14, no. 2, pp. 491–499., 2009.

[46]   A. H. Yüzer, H. Sümbül, M. Nour, and K. Polat, " A different sleep apnea classification system with neural network based on the acceleration signals," *Appl. Acoust.*, vol. 163, pp. 107225, 2020.

[47]   J. Trinder *et al.,* "Autonomic activity during human sleep as a function of time and sleep stage," *J. sleep*, vol. 10, no. 4, pp. 253–264, 2001.

[48]   Z. Shinar, S. Akselrod, Y. Dagan, and A. Baharav, "Autonomic changes during wake-sleep transition: A heart rate variability based approach," *Auton. Neurosci.*, vol. 130, no. 1–2. pp. 17–27, 2006.

[49]   I. M. Voronin and E. V. Biryukova, "Heart rate variability in healthy humans

during night sleep," *Hum. Physiol.*, vol. 32, no. 3, pp. 258–263, 2006.

[50] A. Baharav, S. Kotagal, V. Gibbons, B. K. Rubin, G. Pratt, J. Karin, and S. Akselrod, "Fluctuations in autonomic nervous activity during sleep displayed by power spectrum analysis of heart rate variability," *Neurology*, vol. 45, no. 6, pp. 1183–1187, 1995.

[51] V. K. Somers, M. E. Dyken, A. L. Mark, and F. M. Abboud, "Sympathetic-Nerve Activity during Sleep in Normal Subjects," *N. Engl. J. Med.*, 1993.

[52] H. Yoon *et al.,* "Slow-Wave Sleep Estimation for Healthy Subjects and OSA Patients Using R-R Intervals," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 1, pp. 119–128, 2018.

[53] M. Xiao, H. Yan, J. Song, Y. Yang, and X. Yang, "Sleep stages classification based on heart rate variability and random forest," *Biomed. Signal Process. Control*, vol. 8, no. 6, pp. 624–633, 2013.

[54] B. Yilmaz, M. H. Asyalı, E. Arıkan, S. Yetkin, and F. Özgen, "Sleep stage and obstructive apneaic epoch classification using single-lead ECG," *Biomed. Eng. Online*, vol. 9, pp. 1–14, 2010.

[55] H. Yoon *et al.,* "REM sleep estimation based on autonomic dynamics using R-R intervals.," *Physiol. Meas.*, vol. 38, no. 4, pp. 631–651, 2017.

[56] P. Fonseca *et al.,* "Validation of photoplethysmography-based sleep staging compared with polysomnography in healthy middle-aged adults," *Sleep*, vol. 40, no. 7, 2017.

[57] F. Roche *et al.,* "Screening of obstructive sleep apnea syndrome by heart rate variability analysis," *Circulation*, vol. 100, no. 13, pp. 1411–1415, 1999.

[58] T. Penzel, J.W. Kantelhardt, L. Grote, J.H. Peter, and A. Bunde, "Comparison of detrended fluctuation analysis and spectral analysis for heart rate variability in sleep and sleep apnea," *IEEE Trans. Biomed. Eng.*, vol. 50, no. 10, pp. 1143–1151, 2003.

[59] E. Gil *et al.,* "Discrimination of sleep-apnea-related decreases in the amplitude

fluctuations of PPG signal in children by HRV analysis," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1005–1014, 2008.

[60]   S.H. Choi *et al.,* "Long short-term memory networks for unconstrained sleep stage classification using polyvinylidene fluoride film sensor," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 12, pp. 3606–3615, 2020.

[61]   S. H. Hwang *et al.,* "Unconstrained sleep apnea monitoring using polyvinylidene fluoride film-based sensor," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 7, pp. 2125–2134, 2014.

[62]   J. Fei, I. Pavlidis, and J. Murthy, "Thermal vision for sleep apnea monitoring," In *International Conference on Med. Image Comput. Comput.-Assist. Interv.*. pp. 1084–1091, 2009.

[63]   C. Yang, G. Cheung, V. Stankovic, K. Chan, and N. Ono, " Sleep apnea detection via depth video and audio feature learning," *IEEE Trans. Multimedia*, vol. 19, no. 4, pp. 822–835, 2016.

[64]   M. Zhao, S. Yue, D. Katabi, T. S. Jaakkola, and M. T. Bianchi, "Learning sleep stages from radio signals; A conditional adversarial architecture," *PMLR*, vol. 70, pp. 4100–4109, 2017.

[65]   A. Q. Javaid, C. M. Noble, R. Rosenberg, and M A. Weitnauer, "Towards sleep apnea screening with an under-the-mattress IR-UWB radar using machine learning," *in Proc 14th Int. Conf. IEEE Mach. Learn. Appl. ICMLA*, Dec. 2015.

[66]   Y. Zhou *et al.,* "Validation of novel automatic ultra-wideband radar for sleep apnea detection,*" J. Thorac. Dis.,* vol. 12, no. 4, pp. 1286–1295, 2020.

[67]   S. Kang, D. K. Kim, Y. Lee, Y. H. Lim, H. K. Park, S. H. Choi, and S. H. Cho, "Non-contact diagnosis of obstructive sleep apnea using impulse-radio ultra-wideband radar," *Sci. Rep.*, vol. 10, no. 1, pp. 6–12, 2020.

[68]   R. B. Berry *et al.,* "The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications, Version 2.5," *Am. Acad.*

Sleep Med., 2018.

[69]   M. Bsoul, H. Minn, and L. Tamil, "Apnea MedAssist: Real-time Sleep Apnea Monitor Using Single-Lead ECG," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 3, pp. 416–427, 2011.

[70]   B. L. Koley, and D. Dey, "Real-time Adaptive Apnea and Hypopnea Event Detection Methodology for Portable Sleep Apnea Monitoring Devices," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 12, pp. 3354–3363, 2013

[71]   A. Vaswani *et al.,* "Attention is all you need," *arXiv Preprint,* arXiv:1706.03762, 2017.

[72]   G. Zhang, V. Davoodnia, A. Sepas-Moghaddam, Y. Zhang, and A. Etemad, "Classification of hand movements from EEG using a deep attention-based LSTM network," *IEEE Sens. J.*, vol. 20, no. 6, pp. 3113–3122, 2020.

[73]   H. S. Eom *et al.,* "End-To-End Deep Learning Architecture for Continuous Blood Pressure Estimation Using Attention Mechanism," *Sensors*, vol. 20, no. 8, pp. 23–38, 2020.

[74]   H. Phan, F. Andreotti, N. Cooray, O. Y. Chen, and M. De Vos, "SeqSleepNet: End-to-End Hierarchical Recurrent Neural Network for Sequence-to-Sequence Automatic Sleep Staging," *IEEE Trans on Neural Syst. Rehabilitation Eng.*, vol. 27, no. 3, pp. 400–410, 2019.

[75]   Y. Yuan and K. Jia, "FusionAtt: Deep Fusional Attention Networks for Multi-Channel Biomedical Signals," *Sensors*, vol. 19, no. 11, pp. 24–29, 2019.

[76]   T. Zhu, W. Luo, and F. Yu, "Convolution- and attention-based neural network for automated sleep stage classification," *Int. J. Environ. Res. Public Health*, vol. 17, no. 11, pp.41–52, 2020.

[77]   W. Yin, X. Yang, L. Zhang, and E. Oki, "ECG Monitoring System Integrated with IR-UWB Radar Based on CNN," *IEEE Access*, vol. 4, pp. 6344–6351, 2016.

[78]   S. S. Mostafa, F. Mendonça, A. G. Ravelo-Garcia, G. Gabriel Juliá-Serdá and F.

Morgado-Dias, "Multi-Objective Hyperparameter Optimization of Convolutional Neural Network for Obstructive Sleep Apnea Detection," *IEEE Access*, vol. 8, 2020.

[79] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting," *IEEE Access,* vol. 8, pp. 180544–180557, 2020.

[80] D. Lee *et al.,* "Beat-to-Beat Continuous Blood Pressure Estimation Using Bidirectional Long Short-Term Memory Network," *Sensors*, vol. 21, no. 1, 2020.

[81] M. Hafezi *et al.,* "Sleep Apnea Severity Estimation From Tracheal Movements Using a Deep Learning Model," *IEEE Access*, vol. 8, 2020.

[82] X. Liang, X. Qiao, and Y. Li, "Obstructive Sleep Apnea Detection Using Combination of CNN and LSTM Techniques," *IEEE ITAIC*, pp. 1733–1736, 2019.

[83] U. Erdenebayar *et al.,* "Deep Learning Approaches for Automatic Detection of Sleep Apnea Events from an Electrocardiogram," *Comput. Method Programs Biomed..*, vol. 180, pp. 105001, 2019.

[84] H. B. Kwon *et al.,* "Attention-based LSTM for Non-Contact Sleep Stage Classification using IR-UWB Radar," *IEEE JBHI*, In press.

[85] H. B. Kwon *et al.,* "Hybrid CNN-LSTM Network for Real-Time Apnea-Hypopnea Event Detection Based on IR-UWB Radar," *IEEE Access*, In press.

[86] S. Kang, Y. Lee, Y. H. Lim, H. K. Park, S. H. Cho, and S. H. Cho, "Validation of noncontact cardiorespiratory monitoring using impulse-radio ultra-wideband radar against nocturnal polysomnography," *Sleep Breath*, 2019.

[87] J. A. Palma, M. Alegre, M. Valencia, J. Artieda, J. Iriarte, and E. Urrestarazu, "Basal cardiac autonomic tone is normal in patients with periodic leg movements during sleep," *J. Neural Transm*, vol. 121, no. 4, pp. 385–390, 2014.

[88] H. B. Kwon *et al.,* "Heart rate variability changes in major depressive disorder during sleep: fractal index correlates with BDI score during REM sleep,"

Psychiatry Res. Vol. 271, pp. 291–298, 2019.

[89]   J. W. Kim, T. B. Won, C. S. Rhee, Y. M. Park, I. Y. Yoon, and S. W. Cho, "Polysomnographic phenotyping of obstructive sleep apnea and its implications in mortality in Korea," *Sci. Rep.*, vol. 10, no. 1, pp. 1–12, 2020.

[90]   A. Lazaro, D. Girbau, and R. Villarino, "Techniques for clutter suppression in the presence of body movements during the detection of respiratory activity through UWB radars," *Sensors*, vol. 14, no. 2, pp. 2595–2618, 2014.

[91]   X. Hu and T. Jin, "Preliminary results of noncontact respiration and heartbeat detection using IR-UWB radar," in *Proc. 1st IEEE Int. Conf. Comput. Commun. Internet (ICCCI)*, Oct. 2016, pp. 320–323.

[92]   T. Kondo, T. Uhlig, P. Pemberton, and P. D. Sly, "Laser monitoring of chest wall displacement," *Eur. Resp. J.,* vol. 10, no. 8, pp. 1865–1869, 1997.

[93]   H. Yoon *et al.,* "Sleep-dependent directional coupling of cardiorespiratory system in patients with obstructive sleep apnea," *IEEE Trans. Biomed. Eng.,* vol. 65, no. 12, 2847–2854, 2018.

[94]   A. M. Reiter, G. D. Roach, C. Sargent, and L. Lack, "Finger twitches are more frequent in REM sleep than in non-REM sleep," *Nat. Sci. Sleep*, vol. 12, pp. 49–56, 2020.

[95]   S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[96]   H. Sak, A. W. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," Google, USA, 2014.

[97]   M. Zambotti, J. Trinder, A. Silvani, I. M. Colrain, and F. C. Baker, "Dynamic coupling between the central and autonomic nervous systems during sleep: a review," *Neurosci. Biobehav. Rev.*, vol. 90, pp. 84–103, 2018.

[98]   D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Prepr*. arXiv1412.6980, pp. 1–15, 2014.

[99]   N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[100]   A. Schilling, C. Metzner, J. Rietsch, R. Gerum, H. Schulze, and P. Krauss, "How deep is deep enough? Quantifying class separability in the hidden layers of deep neural networks," *arXiv Prepr.* arXiv1811.01753, pp. 1–39, 2019.

[101]   F. Chollet, "Keras (2015)," URL http//keras. io, 2017.

[102]   M. K. Martín Abadi "TensorFlow: A System for Large-Scale Machine Learning," *Proc. 12th USENIX Conf. Oper. Syst. Des. Implement.*, pp. 44, 2005.

[103]   L. van der Maaten, and G. Hinton, "Visualizing data using t-sne," *JMLR,* vol. 9, pp. 2579–2605, 2008.

[104]   K. Chung, K. Song, S. H. Cho, and J. H. Chang, "Noncontact sleep study based on an ensemble of DNN and random forests," *IEEE Sens. J.*, vol. 18, no. 17, pp. 7315–7324, 2018.

[105]   P. Krauss *et al.,* "Analysis and visualization of sleep stages based on deep neural networks," *Neurobiol. Sleep. Circadian Rhythms,* vol. 10, pp. 100064, 2021.

[106]   F. Zhang *et al.,* "SMARS sleep monitoring via ambient radio signals," *IEEE Trans. Mobile Comput.*, 2019.

[107]   A. Tataraodze, L. Korostovtseva, L. Anishchenko, M. Bochkarev, Y. Sviryaev, and S. Ivashov, "Bioradiolocation-based sleep stage classification," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2016, pp. 2839–2842.

[108]   A. Zaffaroni *et al.,* "Non-contact estimation of sleep staging," *EMBC & NBC 2017. EMBEC 2017, NBC 2017. IFMBE Proceedings*, vol. 65, pp. 77–80, 2017.

[109]   H. Hong, L. Zhang, C. Gu, Y. Li, G. Zhou, and X. Zhu, "Noncontact sleep stage estimation using a CW doppler radar," *IEEE J. Em. Sel. Top. C.*, vol. 8, no. 2, pp. 260–270, 2018.

[110] X. Long, P. Fonseca, J. Foussier, R. Haakma, and R. M. Aarts, "Sleep and wake classification with actigraphy and respiratory effort using dynamic warping," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 4, pp. 1272–1284, 2013.

[111] F. Jurysta *et al.,* "A study of the dynamic interactions between sleep EEG and heart rate variability in healthy young men," *Clin. Neurophysiol.,* vol. 114, no. 11, pp. 2146–2155, 2003.

[112] L. Faes, G. Nollo, F. Jurysta, and D. Marinazzo, "Information dynamics of brain–heart physiological networks during sleep," *New J. Phys.*, vol 16, no. 10, pp. 105005, 2014

[113] R..P. Bartsch, K. K. L. Liu, A. Bashan, and P. C. Ivanov, "Network physiology: how organ systems dynamically interact," *PloS onem*, vol. 10, no. 11, e0142143, 2015.

[114] P. C. Ivanov, K. K. L. Liu, and R. P. Bartsch "Focus on the emerging new fields of network physiology and network medicine,*" New J. Phys.*, vol. 18, no. 10, 100201, 2016.

[115] F. Chung, H. R. Abdullah, and P. Liao, "STOP-Bang Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea," *Chest*, vol. 149, pp. 631–638, 2016.

[116] N. C. Netzer. R. A. Stoohs, C. M. Netzer, K. Clark, and K. P. Strohl, "Using the Berlin Questionnaire to Identify Patients at Risk for the Sleep Apnea Syndrome," *Ann. Intern. Med.*, vol. 131, pp. 485–491, 1999.

[117] J. Zhu, H. Chen, and W. Ye, "A Hybrid CNN–LSTM Network for the Classification of Human Activities Based on Micro-Doppler Radar," *IEEE Access*, vol. 8, pp. 24713–24720, 2020

[118] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: surpassing human-level performance on imagenet classification," *Proc. IEEE Int. Conf. Comput. Vis,* pp. 1024–1034, 2015.

[119] G. Jayatilaka, H. Weligampola, S., Sritharan, P. Pathmanathan, R. Ragel, and I. Nawinee, "Non-contact infant sleep apnea detection," *arXiv,* pp. 18–20, 2019.

[120] T. Van Steenkiste, W. Groenendaal, D. Deschrijver, and T. Dhaene, "Automated Sleep Apnea Detection in Raw Respiratory Signals using Long Short-Term Memory Neural Networks," *IEEE J. Biomed. Health Inform.,* vol. 23, no. 6, pp. 2354–2364, 2018.

[121] V. Y. Senyurek, M. H. Imtiaz, P. Belsare, S. Tiffany, and E. Sazonov, "A CNN-LSTM neural network for recognition of puffing in smoking episodes using wearable sensors," *Biomed. Eng. Lett.,* vol. 10, pp. 195–203, 2020.

[122] N. A. Eiseman, M. B.Westover, J. M. Ellenbogen, and M. Bianchi, "The impact of body posture and sleep stages on sleep apnea severity in adults," *J. Clin. Sleep Med.,* vol. 8, no. 6, pp. 655–666, 2012.

[123] M. Piriyajitakonkij *et al.,* "SleepPoseNet: Multi-View Learning for Sleep Postural Transition Recognition Using UWB," *IEEE J. Biomed. Health Inform.,* vol. 25, no. 4, pp. 1305–1314, 2020.

[124] R. S. Rosenberg and S. Van Hout, "The American Academy of Sleep Medicine inter-scorer reliability program: Respiratory events," *J. Clin. Sleep Med.,* vol. 10, no. 4, pp. 447–454, 2014.

[125] H. Danker-Hopfe *et al.,* "Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the new AASM standard," *J. Sleep Res.,* vol. 18, no. 1, pp. 74–84, 2009.

[126] M. Radha *et al.,* "Sleep stage classification from heart-rate variability using long short-term memory neural networks," *Sci. Rep.* vol. 9, p. 14149, 2019.

국문초록

장단기 메모리 네트워크 기반의 초광대역 레이더를 이용한 비접촉식 수면 모니터링 기술 개발

수면다원검사를 통한 수면 단계와 수면 이벤트의 수동 판독은 수면 구조와 수면 관련 호흡 장애를 분석하기 위해 필수적이다. 그러나 수면다원검사는 전문 인력, 연구실 환경, 그리고 불편한 센서 부착을 요구하기 때문에 지난 몇 년간 임펄스 무선 초광대역 레이더를 이용한 비접촉식 수면 단계 분류 및 수면 무호흡 저호흡 증후군 관찰 방법이 제안되어 왔다. 기존의 연구들의 수면 단계 분류 모델은 설명력과 신뢰도를 높일 필요가 있으며, 수면 무호흡 저호흡 증후군 관찰 모델 또한 레이더-사람 간 거리가 고정되어 있고, 수면 무호흡 저호흡 이벤트가 실시간으로 검출되지 않는다는 제한점이 있다.

수면 단계 분류 연구에서는, 초광대역 레이더로 측정한 생체 정보를 기반으로 수면 단계를 분류하는 양방향성 장단기 기억 네트워크와 주의 집중 메커니즘을 결합한 모델을 제안한다. 65명 (30.0 ± 8.6 세)의 젊고 건강한 참여자들이 야간 수면다원검사를 수행하였고, 동시에 초광대역 레이더를 측정하였다. 51건의 기록 중, 26건은 학습, 8건은 검증, 17건은 테스트 데이터로 사용하였다. 움직임, 호흡, 심박 변이율 관련 16개의 특징점들이 초광대역 레이더 신호에서 매 30초 에폭마다 추출되었다.

최적의 하이퍼파라미터로 구성된 주의 집중 기반 양방향 장단기 기억 모델의 수면 단계 분류 성능을 평가하고, 이를 기존의 장단기 기억 모델과 같은 데이터 세트로 비교하였다. 그 결과, 수면 4단계 (깸, 렘 수면, 얕은 수면, 깊은 수면)으로 분류하는 정확도 가 82.6 ± 6.7%, Cohen의 카파값이 0.73 ± 0.11으로 나타났고 이 성능은 기존 장단기 기억 네트워크에 비해 유의미하게 높은 값이 였다 ($p < 0.01$). 그리고, 수면 단계 분류 성능은 기존 연구들에서 보고된 결과값보다 더 높았다.

　　수면 무호흡 저호흡 검출 연구에서는, 초광대역 레이더를 이용한 실시간 수면 무호흡 저호흡 검출을 위한 딥러닝 모델을 제안한다. 36건의 수면다원검사 데이터, 그리고 동시에 측정한 초광대역 레이더 데이터가 연구에 사용되었다. 클러터가 제거된 후, 초광대역 레이더의 이미지는 1초씩 이동하는 20초 윈도우에 의해 분할되었고, 수면무호흡저호흡 또는 정상으로 분류되었다. 합성곱 신경망과 결합된 장단기 기억 네트워크는 클래스 간 샘플 수의 균형이 맞춰진 데이터를 기반으로 학습되었다. 분류된 시계열 결과는 이벤트 검출기에 입력되어 유효한 수면무호흡저호흡 이벤트가 확인되었다. 그 결과, 본 모델은 평균 Cohen's 카파값 0.728, 민감도 0.781, 특이도 0.956, 그리고 정확도 0.930으로 나타났다. 수면 무호흡 저호흡 지수 분석에서는 예측된 지수값과 레퍼런스 지수값 간의 피어슨 상관 계수가 0.97로 나타났다. 또한, 수면 무호흡 저호흡 지수의 컷오프 기준을 5, 15, 30 이벤트/시간을 기준으로 설정하였을 때, 수면 무호흡 저호흡 증후군의 중증도 분류 결과로 평균 정확도 0.98, Cohen's 카파값 0.96이

나타났다. 본 방법은 사용자의 위치에 상관없이 가공된 특징점 없이도 수면 무호흡 저호흡 증후군의 중증도를 분류하는데 가장 최신 기술의 상태를 달성하였다.

본 실험 결과들은 수면 단계 분류를 위해 주의 집중 메커니즘과 결합한 장단기 기억 네트워크와 수면 무호흡 저호흡 증후군 진단을 위해 합성곱 신경망과 결합 장단기 기억 네트워크 모델의 효과를 보여준다. 이러한 접근법은 가정 환경에서 경제적이고 신뢰할 수 있는 수면 모니터링 시스템을 위한 새로운 전략으로 활용될 것으로 기대된다.

감사의 글

중학교 졸업을 앞둔 해, 막내 손주였던 저를 누구보다 예뻐해주셨던 할아버지께서 당시 원인을 알 수도, 예방을 할 수도 없던 알츠하이머병를 앓으시다 돌아가셨고, 이는 제게 어떠한 병도 찾아내는 장치를 만들겠다는 막연한 포부를 품은 계기가 되었습니다. 대학과 대학원에서 의공학을 전공하는 약 10 여년간의 학위기간은 치열한 배움과 인내의 연속이였지만 수많은 분들의 도움으로 그 포부와 목적을 잃지 않고 이 학위논문을 마무리할 수 있었습니다. 모두 진심으로 감사드립니다.

바쁘신 와중에도 흔쾌히 심사위원을 맡아주신 서울대학교 최진욱 교수님, 서울대학교 김성완 교수님, 충북대병원 이태수 교수님, 광운대학교 박철수 교수님께 진심으로 감사드리고 큰 영광이었습니다. 교수님들의 꼼꼼하고 면밀한 조언과 격려 덕분에 이 졸업논문이 완성될 수 있었습니다.

2014 년 겨울, 처음 지도교수님이신 박광석 교수님을 뵙기 위해 연건동 서울대병원의 언덕을 올랐었습니다. 교수님께서 허락해주신 6 년반이라는 시간 동안의 연구실 생활은 미숙한 저의 인생의 많은 부분을 바꿔놓았고, 일상에서는 경험하지 못했을 성찰과 지식을 얻을 수 있었습니다. 그간 교수님께 연구자로서, 스승으로서의 존경심을 느끼고 동기부여를 받을 수 있었던 것을 큰 축복이라고 생각하고 있습니다. 비록 아직 많이 부족하다고 느끼지만, 사회의 일원이 되어서도 은혜를 기억하며 교수님과 같은 공학자이자 연구자로 살도록 노력하겠습니다.

학위 기간동안 주 연구분야가 수면 공학이였던만큼 서울대병원 정신건강의학과 교수님들의 도움으로 많은 연구 논문들의 결실을 맺을 수 있었습니다. 특히 이유진 교수님께서는 대부분의 논문들을 공동으로 지도해주시며 많은 영향을 주셨습니다. 대학원 초년생 때의 연구 프로포절을 돌이켜보면 미흡하고 좁은 식견에 낯부끄러운 마음이 그지없지만 교수님께서 늘 따뜻한 격려와 함께 공학자로서는 깨닫기 힘들었던 임상의의 관점으로 아낌없는 조언을 해주셨습니다. 또 다른 지도교수님과 다름없었던 교수님께 진심으로 감사드립니다. 더불어 서울대병원 정도언 교수님, 이미현 펠로우 선생님, 을지대병원 최재원 교수님, 동국대병원 오성민 교수님, 유익한 조언과 도움으로 연구의 견고함을 높일 수 있도록 해주신 점 감사드립니다.

이 논문은 수면다원검사를 대체 가능한 UWB 레이더 기반의 AI 모델 개발에 관한 연구이며, 2018 년부터 2020 년까지 장기간의 실험을 토대로 획득한 100 여건의 수면 데이터를 바탕으로 하고 있습니다. 이와 관련하여 산학 협력 프로젝트를 함께 진행해주신 삼성리서치 스마트머신랩의 한종희 박사님과 정기성 박사님의 배려와 지원에 감사드립니다. 또한 데이터 획득에 주말없이 수고해주신 가천대학교 마효진 수면기사님과 서울대학교병원 수면의학센터 수면 기사님들의 도움에 감사를 표합니다.

지난 약 7 년간의 생체신호 및 정보 연구실에서 함께 생활했던 동료이자, 선후배분들께도 진심으로 감사드립니다. 특히 가족과도 같은 수면팀의 윤희남 교수님, 최상호 교수님, 같은 길을 걷고 있는 사람이 있다는 것이 제게 얼마나 큰 힘이 되었는지 모릅니다. 새벽까지 함께 논문을 쓰고, 퇴근하는 길에 술 한잔 하며 수면 연구에 대한 사색을 나누고, 직접 실험해보자며 주말없이 매트리스를 들고 병원을 뛰어다녔던 시간들. 당시에는 매일이 바쁜 나날들이었지만 함께 고민하고 상상하며 성취해가는 과정을 나눌 수 있었기에 즐거웠던 추억이자 연구 경험으로 남아있습니다. 앞으로도 두 교수님의 연구자로서의 길을 늘 진심으로 응원합니다.

연구실에서 동고동락하며 많은 추억을 함께한 김지훈 박사님, 이동석 박사님, 주광민 석사님, 함께 맨몸으로 뛰어다녔던 제주도의 새벽 해변, 겁을 내며 뭉쳐다녔던 오사카의 어둑한 밤거리, 온몸이 짜릿하게 스카이다이빙을 뛰었던 하와이의 하늘, 대낮부터 쉴새없이 퍼마셨던 베를린의 맥주, 기나긴 연구실 생활에 잊지못할 즐거운 시간을 함께 보내며 늘 막내동생으로 배려해주고 챙겨주셔서 감사합니다. 처음 309 호 연구실에 들어와 낯선 환경에서 따뜻하게 대해주셨던 김한별 박사님, 심수영 박사님, 김현석 박사님, 연구실에서의 첫 산학 연구들을 이끌어 주셨던 이홍지 박사님, 이원규 박사님, 홍승혁 교수님, 807 호에서 함께 생활하며 많은 조언을 해주신 황수환 박사님과 허정 박사님께도 감사드립니다. 텍사스에 방문했을 당시에 먼 거리를 와주셨던 한정민 박사님, 유일한 동갑내기 친구였던 성동석 석사님, 그리고 함께 입학하여 서로 많은 고민을 나누었던 고명준 석사님. 외국에서도 건승하시길 바라겠습니다. 그리고 선배로서 큰 도움을 주지는 못해 미안한 마음뿐인 진형원 학생, 손동연 학생, 정종혁 학생, 많은 시간을 함께 하지 못해 아쉽지만, 좋은 연구를 잘 마무리하고 무사히 졸업하시길 바랍니다.

마지막으로, 지난 5 년간 옆에서 큰 힘이 되어준 노정원 변호사님, 변함없는 마음으로 같은 자리를 지켜주어서 제가 마음 편히 학위를 끝마칠 수 있었습니다. 기나긴 대학원 생활동안 지적 호기심과 책임감은 연구를 수행하는 원동력이었지만, 정작 삶을 이끄는 어떠한 가치는 찾지 못할 때가 많았습니다. 지치는 일상에서도 매일 웃을 수 있게 해주고 자존감을 잃지 않도록 한없이 배려해주어서 고맙습니다. 지금처럼 서로를 존경하고 서로에게 기대며 살아가면 좋겠습니다. 앞으로도 잘 부탁합니다.

　지면으로 언급하지 못했지만 저를 아끼고 격려해주셨던 모든 분들께 다시 한번 진심으로 감사합니다. 더욱 정진하며 견실한 모습으로 한층 성장해 헬스케어 분야에서 꼭 필요한 존재가 되도록 노력하겠습니다.

<div align="right">2021 년 6 월 30 일 권현빈 올림</div>