SleepPoseNet: Multi-View Multi-Task Learning for Sleep Postural Transition Recognition Using UWB

Maytus Piriya JITAKONKIJ, Patchanon Warin, Payongkit Lakhan, Pitsharponri Leelaarporn, Theerasarn Pianpanit, Nattee Niparnan, Subhas Chandra Mukhopadhyay Fellow, IEEE and Theerawit Wilaiprasitporn Member, IEEE

Abstract—Recognizing the movements during sleep is crucial for monitoring of patients with sleep disorders. However, the utilization of the deep learning approaches ultra-wideband (UWB)-based indoor applications for the classification of human sleeping postures has not been explored widely. This study investigates the performance of the off-the-shelf single antenna UWB in a novel application of sleep postural transition (SPT) recognition. The proposed Multi-View Multi-Task Learning, entitled SleepPoseNet or SPN, with time series data augmentation aims to classify four standard SPTs. SPN exhibits an ability to capture both time and frequency features, including the movement and direction of sleeping positions. Three-cost functions for multi-views, multi-tasks, and classification are proposed to optimize SPN. The data recorded from 26 volunteers displayed that SPN with mean accuracy of 79.7±0.5% significantly outperformed the mean accuracy of 73.5 ± 0.7% obtained from deep convolution neural network (DCNN) in a recent state-of-the-art work on human activity recognition using UWB. Apart from UWB system, SPN with the data augmentation can ultimately be adopted to learn and classify time series data in various applications.

Index Terms—sleep pose recognition, multi-view multi-task deep learning, ultra-wideband (UWB), deep canonical correlation analysis (DCCA), deep convolution neural network (DCNN), time series data augmentation.

I. INTRODUCTION

Sleep is one of the most essential elements to human health. Sleep disorders such as obstructive sleep apnea (OSA), are associated with cardiovascular disease (CVD), stroke, hypertension, and daytime sleepiness [1]–[3], resulting in the lack of work productivity, car accidents, and higher rate of mortality [4], [5]. One important aspect with definite advantages to determine the quality of sleep and the severity of the disorders is to monitor bodily spatial movement, especially different postures during the night [6]–[8]. OSA severity can be identified by using sleep postures as one of the important variables. Furthermore, numerous individuals at risk who are experiencing pressure ulcers or bedsores, which are lesions on the skin resulting from pressure, may face higher mortality rate, especially when they are bedridden with the chronic type [9], [10]. Thus, sleep monitoring can be used to notify the caregivers in order to adjust the sleep postures of the patients, increasing the possibility of pressure ulcers prevention.

Technological advances in Internet of Things (IoT) and deep learning have been used to unleash many in-home sleep monitoring devices through the consumer markets [11]. These devices consist of two types: wearable and non-wearable. Both types have been utilized to recognize physical position and monitor the movement of patients with sleep disorders [12]. Most of the wearable sleep monitoring devices are developed as smartwatches or smartbands for simplicity and equipped with features to track crucial physiological features, such as heart rate (HR) and oxygen saturation, using Photoplethysmography (PPG) [13]. Some devices are also equipped with accelerometer and gyroscope for human activity recognition and sleep stages estimation [14]. Although wearable devices have been proved to provide accurate HR estimation [15], the measurement can only take place on the location where the device is worn such as the upper limb. This may result in high false positive rates and may be inconvenient. Therefore, using non-wearable devices to track body movement are more preferable. Different non-wearable devices have been adopted for sleep monitoring including camera [16], pressure mat [10], [17], [18], and radar-based system [6]. Despite the advantages offered by these devices, some hindrances can be emerged: Cameras are generally susceptible to different light conditions [16] and privacy issue may be concerned when using infrared camera. Pressure mat is needed to place appropriately as the patients movements may shift its placement and require repositioning. In contrast, radar-based sensor does not face the
same dilemmas and has high penetration ability which can detect even through-wall human movements [19] and vital signs [20].

Many studies investigating radar-based systems for sleep monitoring application have been exploring their ability to detect human vital signs, such as HR and breathing rate [21–25]. These systems can also be used to identify breathing disorders [26], body movements and HR in different sleep stages [27], [28]. For instance, Kim et al. applied Continuous Wave (CW) Doppler radar and Deep Convolutional Neural Networks (DCNN), which exploited knowledge in Doppler effect of reflected signals to obtain the velocity of targets, to detect human and classify activity, such as hand gestures [29], with higher accuracy [30]. However, a definite drawback of CW radar is presumably due to the lack of range information in the signals. To solve this issue, a Frequency Modulated Continuous Wave (FMCW) radar, of which the signals contain both range and Doppler information, was used as an alternative [31]. However, multipath interference seems to reduce the performance of FMCW radar especially for indoor application.

Another attractive spectrum, the Ultra-Wideband (UWB) radar, has been utilized for its range information with high resolution due to its high frequency pulse signals [32], [33]. UWB has been proposed to be more appropriate to an indoor usage due to a higher multipath suppression capability with more affordable price [33]. In 2016, Yin et al. introduced a combination of ECG and UWB radar in order to classify cardiac arrhythmia and found that the accuracy had improved from using only ECG [34]. To evaluate the classification of human activity recognition, many studies used classical machine learning methods such as Support Vector Machine (SVM) [35], [36]. Bryan et al. used Principal Component Analysis (PCA) to reduce dimensions and extract features in time domain combining with the frequency components extracted by Fourier Transform (FT) [37]. Ding et al. created new method for UWB feature extraction called Weighted Range-Time-Frequency Transform (WRTFT), which included range information to a Short Time Fourier Transform (STFT) [38]. WRTFT was later used by Chen et al. with DCNN, obtaining more robust result [39].

With the promising results of the application of the radar-based system on human activity classification tasks, some studies have turned its interest towards analyzing on-bed motions [7], [27], [40], [41]. However, despite the extensive usages of UWB for physical activity recognition, to our best knowledge, none has examined its utilization for detecting and classifying sleep postures. In this study, the development of a deep learning algorithm in order to make an accurate classification of Sleep Postural Transitions (SPT) was investigated. Off-the-shelf UWB radar device was employed in a smart bedroom environment. Due to the ambiguity of the data between classes, time domain features were fused with frequency domain features by Multi-View Learning (MVL) using along with Multi-Task Learning (MTL) to improve classification accuracy. A novel learning method, named SleepPoseNet or SPN, using the combination of MVL and MTL method for the classification problem was introduced. Our method was then compared to the state-of-the-art method to determine the most optimal method for SPT recognition.

Three main contributions are presented in this study:

1) To our best knowledge, this is the first study to propose using Sleep Postural Transitions (SPT) classification models through off-the-shelf Ultra-Wideband (UWB) radar system, which exhibited the potential to support the usage of the current sleep monitoring systems on measuring vital signs and sleep stages.

2) Multiview Learning (MVL) model, which is based on Deep Canonical Correlation Analysis (DCCA), was used to fuse information in time and frequency domains, and was combined with Multitask Learning (MTL), introducing a novel method dubbed SleepPoseNet or SPN.

3) Many data augmentation techniques for time series were investigated and employed to increase classification accuracy.

II. METHODS

A. UWB Principle and Hardware

UWB radar propagates the impulse signals through a transmitter [42]. Once the pulse arrives at the focused object, the pulse is split into two; the reflected pulse returns to the signal receiver, while the transmitted pulse passes through the object. Time of Arrival (TOA) is measured to identify the range of the object. A received signal at an antenna is represented by the scaling value of each reflected signal.

$$y_k(t) = \sum_{i=1}^{N_{path}} a_{ki} x(t - \tau_{ki}) + n(t)$$  \hspace{1cm} (1)$$

where \(y_k(t)\) is the received signal of the \(k\)-th pulse sending from the transmitter and \(k\) is called a slow time index. \(x\) and \(a_{ki}\) represent the original signals sent from transmitter and the scaling value of each reflected signal. \(i\) means \(i\)-th path of travelling pulse between the transmitter and receiver, and \(N_{path}\) is the number of the travelling paths. \(t\) is the fast time index, representing the arrival time of each reflected signal. \(\tau_{ki}\) is the time delay, and \(n(t)\) is noise. The range \(R\) between device and object can be calculated by the TOA \(\Delta t\) and the speed of light \(c\) as in an equation.

$$R = \frac{c \Delta t}{2}$$ \hspace{1cm} (2)$$

The data were collected using the Xethru X4M03 development kit, a state-of-the-art UWB radar device invented by Novelda [43]. Its transceiver operates within the range of 5.9-10.3 GHz. The key features include its flexibility to be customized and suitable for particular application, i.e., proper configurations can be primed to support our applications, such as detection range, number of frames per second, number of pulses per step, and number of sweep iterations. Moreover, the device supports digital down conversion, filtering, and data rate reduction, which are efficient signal processing methods. The radar configurations in our experiments are shown in Table I.
TABLE I: Radar Parameters.

| Parameters               | Values                        |
|--------------------------|-------------------------------|
| Tx                       | 7.29 GHz (center frequency)  |
| Pulse Repetition Frequency | 15.18 MHz                   |
| Sampling Frequency       | 23.32 GHz                    |
| Range                    | 0-9 m                        |
| Bins per radar frame     | 180 (baseband)              |
| Frame rate               | 10                            |

B. Experimental Recording

The UWB radar system was attached to the wall with 0.8 m height above the bed. The pitch angle was approximately 45 degrees downward towards the bed. Six STPs were selected, as shown in Figure 1a. There were 26 volunteers, including 19 males and 7 females participated in this study. The height ranged from 1.55 to 1.80 m, the weight ranged from 40 to 90 kg, and the age ranged from 18 to 35 years old. The experiment commenced with the subjects lying down with a supine position in the middle of the bed under the line-of-sight of the radar. The subjects were instructed to perform six motions, in a sample, following an arranged sequence: supine to left lateral, left lateral to supine, supine to right lateral, right lateral to supine, supine to prone, and prone to supine, at their own pace. The subjects were given 10-15 s to rest while inhaling for 4-5 times before performing the next motion. This allowed the radar system to obtain the respiration features before and after each change of posture. Most subjects performed one experimental set, consisting of 30 samples with five samples per class, whereas three subjects performed two sets and one subject performed three sets. Furthermore, left and right lateral positions were treated as side sleep positions. Ultimately, a total of 31 experimental sets, consisting of four classes of SPT, designated as Supine to Side (SUSI), Supine to Prone (SUPR), Side to Supine (SISU), and Prone to Supine (PRSU), were recorded as shown in Figure 2 (a-d). To attain balanced class, some of the samples of the positions SUSI and SISU in every set were undersampled. Therefore, a total of 620 samples for all experiments were finalized.

The radar signals from each sample was labelled manually using the video recordings from 10-15 minutes experiment as the ground truth. Each sample contained 16 s length where the first 5 s began before the first posture change. Informed consents were received from all subjects following the Helsinki Declaration of 1975 (as revised in 2000), which was approved by the internal review board of Rayong Hospital, Thailand (RYH REC No.E010/2562).

C. Data Processing

After sampling, obtained signals are stored in $M \times N$ matrix $R$, where $M = 180$ is the number of fast time indices, also called range bins, and $N = 100$ is the number of slow time indices. $R_{mn}$ is the entry in $m$-th row and $n$-th column of matrix $R$. Subsequently, the following two steps are applied before extracting the meaningful features.

a) Time Average: The data in each fast time index or range bin $n$ may unavoidably contain DC noise which can be eliminated by averaging its value through all the number of slow time indices $N$.

$$\bar{R}_{mn} = \frac{1}{N} \sum_{i=0}^{N-1} R_{mi}$$

The second term is an average by column. Therefore, every $R_{mn}$ is subtracted by its mean along the slow time index, before storing in $\bar{R}_{mn}$.

b) Range Average: The information at an instant slow time composed of static clutter, which are the uninterested objects and obstacles in an environment, and interested target, which referred to an active human. To suppress the static clutter, the data in each slow time index are subtracted with their average along fast time.

$$Y_{mn} = \bar{R}_{mn} - \frac{1}{M} \sum_{i=0}^{M-1} \bar{R}_{in}$$

The second term is an average by row. Therefore, every $\bar{R}_{mn}$ is subtracted by its mean along the fast time index or range bin, before storing in processed $Y_{mn}$.

D. Feature Extraction

In this paper, features in time domain and frequency domain are proposed to be implemented in our classification task.

1) Temporal Difference (TD): This method is based on an assumption that every obstacle and object in a bedroom-like environment are static or not changing its shape over the time. Therefore, the differentiation along the slow time axis $Y^d$ can extract only the information of a moving human, as shown in Figure 1b and Figure 1c. Most non-transitioning parts are suppressed by this discrete differentiation.

$$Y^d_{mn} = Y_{m,n+1} - Y_{mn}$$

Where $Y^d_{mn}$ is the entry in $m$-th row and $n$-th column of matrix $Y^d$. Every $Y^d_{mn}$ is a difference between $Y_{m,n+1}$ and $Y_{mn}$, which represents the pulses from range bin $m$ with different slow time.

2) Range Selection: For our device, the step size between range is 5.14 cm. 40 range bins containing information in range of about 2 m and covering all parts of the human body are chosen. The selection algorithm is done by finding a position of cropping window such that maximized the summation of the slow time difference energy in the window size of $(F - I + 1) \times N$, as shown in Equation 6.

$$\text{maximize} \sum_{m=1}^{F} \sum_{n=0}^{N-1} (Y^d_{mn})^2$$

subject to $F - I + 1 = 40$

Where $I$ and $F$ are the initial range bin and final range bin of the cropping window. A cropped signal is then stored in the $(F - I + 1) \times N$ matrix.
Fig. 1: (a) Experimental bedroom-like environment with UWB radar devices placed above the headboard. The camera was also placed for recording as ground truth. (b) signals and features of SUSI (c) signals and features of SUPR

3) Weighted Range-Time-Frequency Transform (WRTFT): WRTFT was proposed in 2018 [38], to combine spectrograms from all range bins. After the transformation is complete, the output image contain information from range, time, and frequency features of the human motions. WRTFT is performed according to the following steps;

a) Short Time Fourier Transform (STFT): STFT is performed by segmenting time series and employing Fourier transform (FT) on each segment. The obtained result is referred to time-frequency representation (TFR), and each range bin hold its own TFR. The result is stored in three-dimensional array $F$.

$$F_{mkn} = \sum_{p=0}^{N-1} Y_{mp}\omega(n-p)e^{-j2\pi pk/N}, m \in [I, F]$$ (7)

where $k$ is frequency domain index and $\omega(p)$ is a window function.

b) Weighted Average: All spectrograms from all range bins are weighted-averaged by energy signal of each range bin and stored in matrix $W$;

$$W_{kn} = \sum_{m=I}^{F} \sigma_m F_{mkn}$$ (8)

$$\sigma_m = E_m / \sum_{m=I}^{F} E_m$$ (9)

$$E_m = \sum_{n=0}^{N-1} Y_{mn}^2$$ (10)

where $\sigma_m$ is the coefficient weighting a spectrogram and is calculated by the proportion of energy in a range bin.

TD and WRTFT of SUSI and SUPR are shown in Figure 1b and Figure 1c. Both time and frequency features are unlikely to contribute in visually distinguishing between the classes.

E. Data Augmentation

Deep learning normally requires a large training dataset to prevent the occurrence of overfitting. However, for human biosignals, data are commonly small due to the limited human subjects and resources. In order to solve this problem, Time Shift (TS) and Range Shift (RS) were applied. In addition, Time-Warping (TW) and Magnitude-Warping (MW), which were proposed in 2017 for wearable sensor data [45], were also used. The results from data augmentation methods are shown in Figure 3. All possible combinations from these four augmentation methods were applied to the training data.

a) Time Shift (TS): Approximately 2-3 s segments, containing SPT information, are in a long interval of slow time in one sample. The shift on slow time index can possibly assist the model to learn a SPT section in a different position. The parameters for TS include [-10, -5, 5, 10], where positive and negative values indicate right and left shifts of slow time index respectively, and an extension section is padded with zeros.

b) Range Shift (RS): There are different range bins in many samples which represent the respective part of the human body. RS is implied to be capable of aiding in learning the variability of human’s range position. The parameters for RS included [2,
4], where positive values indicate the up shift of range bin and extension section is also padded with zeros.

c) Time-Warping (TW): TW is done by smoothing and randomly distorting the intervals between slow time indices in each sample to change the temporal location of the slow time indices. Cubic spline interpolation is used to impute the missing temporal location values by their neighbors, which are the original value with different temporal location. The variance parameter used in this work is 0.4, representing the variance of interval distortion.

d) Magnitude-Warping (MW): MW create a random smooth curve varying around one and each sample is multiplied with this curve in order to randomly change the amplitude along slow time positions. The variance parameter used in this work is 0.4, representing the variation around one in a random curve.

F. SleepPoseNet

The overview of all processes from the data processing to prediction are illustrated as the flow chart in Figure 4. In order to extract significant features from two views of data and perform two tasks simultaneously, this study proposed a method, entitled SleepPoseNet or SPN, composing of a combination of Deep Convolutional Neural Networks (DCNN), Multi-Task Learning (MTL), and Multi-View Learning (MVL).

DCNN is a class of neural networks, consisting of convolutional filters, which perform arithmetic operations to find relationship of a data point, and its neighbors, to extract some important features from all data points in a sample. DCNN works especially well in image recognition tasks. Different DCNN models may include 1D convolutional filters, which are used for time series classification and prediction. The data collected in this study are multivariate time series, which can be represented by 2D array similarly to an image. Therefore, DCNN with 2D convolutional filters for the classification tasks is deemed appropriate.

MTL is machine learning model in which multiple tasks are performed concurrently using the same sample. Due to this characteristic, some important features from the sample are strongly shared to solve different tasks. Thus, notably useful features which are important to all tasks can possibly be extracted. MTL can steer the model to achieve better performance and prevent overfitting, especially when solving related tasks. In this study, we essentially aimed to recognize four SPTs, including SUSI, SUPR, SISU, and PRSU. Another auxiliary task is called Sleep Turning Transition (STT). STTs are categorized into two classes: Rotating Up (ROUP), composing of SUSI and PRSU, and Rotating Down (RODO), composing of SISU and SUPR. In other words, the classes in SPT are considered as the sub-classes of STT.

MVL combines and utilizes the information from different views to improve the generalization performance. For instance, video and sound together may contribute to higher classification accuracy than only a single view. In this paper, we combined the information from time and frequency domains of the radar signal to gain higher SPT classification accuracy. Using SPN, we were inspired by Deep Canonical Correlation Analysis (DCCA) and attempted to combine the concept behind DCCA to effectively fuse the information from two views in our task.

DCCA was proposed to deepen the understanding of the representation from two different views of data [46], [47]. The idea was inspired by Canonical Correlation Analysis (CCA) which maximizes the correlation of linear projection between two views. In the same way, DCCA first applies nonlinear transformation on two views, $X_1$ and $X_2$, which are functions of neural networks, and then maximizes the correlation between two extracted features $f_1(X)$ and $f_2(X)$ from transformations.

$$\maximize_{\theta_1, \theta_2} \text{corr}(f_1(X_1; \theta_1), f_2(X_2; \theta_2))$$

(11)

Where $\theta_1$ and $\theta_2$ are the parameters of $f_1$ and $f_2$, both of which are neural networks.

The architecture of SPN is illustrated in Figure 4. Here, the attempt to fuse the information from both time and frequency domains to increase the classification accuracy was executed. Our proposed method was inspired by DCCA, in which the model learns the representation from two different views by maximizing their correlation in feature space. DCCA performs generally well on unsupervised learning for feature representation, and it has been investigated in terms of classification accuracy [47] using SVM. However, it is not end-to-end deep learning. Our model began with receiving TD and WRFT separately as inputs. Two CNNs were then employed to extract the important features from both domains.
The parameters of model are displayed in Figure 4. The receptive field sizes included 2x3 for TD and 2x2 for WRTFT, and max pooling layers are used to reduce the dimension, preventing overfitting, with the same size of receptive field. Two extracted features are maximized for the correlation between each other by CCA loss. The features are then concatenated, before being classified by fully connected layers. Cross entropy loss is also minimized to achieve higher classification accuracy. SPT classification is the main task, following by STT classification using MTL. To summarize, loss function of SPN, which is composed of three optimization objectives to minimize, is shown in Equations 12, 13, 14, and 15.

\[ J_{\text{post}}(\theta) = -\frac{1}{M} \sum_{i=1}^{M} y_p^{(i)} \log \hat{y}_p^{(i)}(\theta) \]  
\[ J_{\text{aux}}(\theta) = -\frac{1}{M} \sum_{i=1}^{M} y_a^{(i)} \log \hat{y}_a^{(i)}(\theta) \]  
\[ J_{\text{cca}}(\theta) = - \text{tr}(T'T)^{1/2} \]  
\[ J_{\text{total}}(\theta) = J_{\text{post}}(\theta) + J_{\text{aux}}(\theta) + J_{\text{cca}}(\theta) \]

Where \( J_{\text{post}}(\theta) \) and \( J_{\text{aux}}(\theta) \) are postural and auxiliary loss function for SPT and STT classification, respectively. \( \theta, y_p, \hat{y}_p, y_a, \) and \( \hat{y}_a \) denote model parameters, SPT predictions, STT true labels, and STT predictions respectively.

For cross-covariance matrix, \( \Sigma_{12} = \frac{1}{M-1} Z_1 Z_2' \) for definition. Therefore, the trace norm \( \text{tr}(T'T)^{1/2} \) intuitively represents the total correlation between the two features.

G. Experiments

In this section, comparison methods were set in order to evaluate SPN partitioning into three experiments. Experiment I and Experiment II investigated the SPN performance in comparison to other models among different distributions between training, validation, and test data. Experiment III examined the range selection process of the number of range bins, also called Window Size (WS), which affected the performance of SPN. Here, we chose various methods based on WRTFT [38], [39] to compare our multi-view multi-task learning method with a variety of configurations. Model Evaluation: Seven methods were selected for evaluating the performance of SPN. Firstly, SVM was used along with the first 20 principal components of WRTFT, extracted using PCA. Next, DCNN was employed, composing of a convolutional layer with 10 filters and max pooling layer as well as a convolutional layer with 20 filters and max pooling layer, respectively. All filters were primed with the size of \( 2 \times 2 \), ending with 10 hidden units, followed by ReLu activation, and 4 output units, followed by softmax activation. DCNN was then modified into MTL_DCNN with the additional auxiliary head connected to the last max pooling layer, consisting of 10 hidden units, followed by ReLu activation, and 2 output units, followed by softmax activation. Moreover, AUG_MTL_DCNN was formed using MTL_DCNN, in which the training data were processed through our data augmentation methods. Subsequently, NoMTL_SPN, a method similar to SPN but without auxiliary head which was used to predict STT. Lastly, AUG_SPN, a modification of SPN in which the training data were processed through our data augmentation methods, was constructed. All experiments were attained by 10-fold cross validation to shuffle training, validation, and test sets. All schemes were run for five times with different random states. To regularize the
models, spatial dropout and dropout with the probability of 0.3 and 0.5 were employed on every convolutional layer and every hidden layer, which was placed before output layers. All models were trained by ADAM optimization with constant learning rate of 0.0001, and a small batch size of 16 was used along with early stopping. The averaged accuracy of all methods from all folds were then compared and statistically tested using one-way repeated measure ANOVA.

**Experiment I:** We first investigated the performance of all models in predicting data from the unseen subject or person. 26 subjects were divided randomly into 18, 4, and 4 subjects to be used as training, validation, and test sets, respectively.

**Experiment II:** To examine the performance of all models with the prior knowledge of subjects’ data distribution before the test data were predicted, 31 experimental sets were explored. Each set contained 20 samples. 12, 4, and 4 samples from every set were chosen randomly, with balanced class distribution, to be used as training, validation, and test data, respectively.

**Experiment III:** To acquire the most appropriate Window Size (WS) for range bin selection, the method with the most reliable result was examined with WS adjustment. The WS in this experiment were 30, 35, 40, 45, 50, 55, and 60.

### III. RESULTS

All training and validation losses over the epoch are shown in Figure 5. The models, at the minimum validation losses, were selected. The numbers of early stopping epochs of models with data augmentation, which included AUG_MTL_DCNN and AUG_SPN, were reduced from those without augmentation due to the inequality in the number of updating iterations in each epoch between non-augmented and augmented training sets. It was found that every method without data augmentation experienced well-fitting with the close losses between training and validation. All methods were established before becoming overfitting in the later epochs. While AUG_MTL_DCNN and AUG_SPN with data augmentation appeared to show slightly higher training losses than the validation losses, AUG_MTL_DCNN validated with seen subject data, as shown in Figure 5, was an exception due to its training loss close to the validation loss.

**Fig. 5:** Training and validation losses from all methods as described in subsection III-A and subsection III-B. (a)-(f) are for Experiment I: unseen subject, and (g)-(l) are for Experiment II: seen subject.

**Fig. 6:** Confusion matrices of two methods ((a) AUG_MTL_DCNN and (b) AUG_SPN) applied in Experiment I: Unseen Subject.
A. Experiment I: Unseen Subjects

Significant differences ($F(2,884,141.331) = 34.007, p < 0.05$) were found amongst the accuracy means of some models. As shown in Figure 6, SPN with data augmentation (AUG_SPN) performed with higher accuracy than others, revealing highest accuracy of $74.1 \pm 0.7\%$ with the lowest SE. While the accuracy of NoMTL_SPN was not significantly different from AUG_MTL_DCNN, SPN without data augmentation achieved an accuracy of $70.3 \pm 0.9\%$, which was significantly higher than AUG_MTL_DCNN with data augmentation, serving as the best method based on frequency domain features of WRTFT. Confusion matrices of AUG_MTL_DCNN and AUG_SPN were computed from all folds of best WRTFT-based method and our SPN with data augmentation, as shown in Figure 6. All values were normalized with the number of actual samples on each row. As illustrated, PRSU showed high accuracy with sensitivity/recall of $82\%$ for both methods, followed by SISU. Interestingly, SISU appeared to be misclassified as PRSU. SUSI and SUPR were mainly misclassified as one another, while SUSI yielded the lowest accuracy. In general, the accuracy of SUSI, SUPR, and SISU were improved when using AUG_SPN.

B. Experiment II: Seen Subjects

Among some models, significant difference in accuracy means was found ($F(4,803,225.326) = 118.281, P < 0.05$). As shown in Figure 7a, SPN with data augmentation (SPN_AUG) outperformed the others with an accuracy of $79.7 \pm 0.5\%$. However, no significant difference was found among the accuracy of SPN, NoMTL_SPN, and AUG_MTL_DCNN. The overall accuracy of the models with seen subject was revealed to be higher in comparison to the accuracy of the models with unseen subject.

Confusion matrices of AUG_MTL_DCNN and AUG_SPN are shown in Figure 8. PRSU from both models yielded the highest accuracy with equal sensitivity/recall of $87\%$. SUSI was found to be difficult to detect using AUG_MTL_DCNN. SUSI and SUPR were mainly misclassified as one another, while SISU was mainly misclassified as PRSU, by both methods. Interestingly, the overall accuracy of SUSI, SUPR, and SISU increased when using AUG_SPN.

C. Experiment III: Window Size Adjustment

AUG_SPN, which yielded the highest accuracy among the models, was selected for WS adjustment. Figure 7c displays the accuracy of different WS. The accuracy means from some of the configurations were significantly different ($F(6, 294) = 9.608, P < 0.05$). Only the WS of 40 was not statistically significant in comparison to the larger WS. Increasing the WS did not improve the accuracy. Therefore, it was suggested that our first selected WS of 40 was the optimal WS for real-time application, concerning the speed and the optimal amount of data without reducing efficiency.

IV. Discussion

The attractiveness of UWB radar system for non-contact human activity detection has led this study to adapt the radar-based signals for the recognition of sleep postures. Using a novel designed SleepPoseNet (SPN), combining the MVL and MTL deep learning approaches for classification task, we evaluated the performance of different models. The possibility of developing and further implementing the UWB sensors for monitoring patients with sleep disorders was observed. Here in this study, the Xethru UWB radar, an off-the-shelf device, was selected for usage due to its affordable cost, user-friendly attribute, and easy-to-use element suitable as a smart-home device. However, the device has a single Rx/Tx antenna that may induce some limitations in receiving information. It is imperative to improve its efficiency by gathering information from more individuals. This work serves as the instigation for further development of the monitoring aspect of the device. To commence, data from 26 volunteers were first collected and digital signal processing with deep learning approach was performed for the task analysis. Previous studies have been utilizing WRTFT with DCNN for human activity classification by using UWB radar signals [38], [39]. Using their works as foundation, we attempted to further improve the classification performance by incorporating higher complexity learning methods to our model, i.e., MTL and MVL as shown in Figure 4. Figure 7 demonstrated the experimental results, showing that the proposed SPN can achieve 79.7% accuracy, which outperformed the previous method with an improvement of 6.2%. The amount of acquired data generally imposes an issue for the accuracy of the performance. Here, we applied the data augmentation methods for time series, resulting in a higher quality performance. One remark gazed upon the longer period of time which was required to overfit, allowing the models to learn more as observed from the loss of training set and validation set, as shown in Figure 5. Intriguingly, SPN was demonstrated to perform better than the state-of-the-art model. From Experiment I and II, giving some examples from every subject to the model was shown to improve the performance better than leave-subject out approach. Thus, to performance with much higher accuracy, a registration system which requires some information before execution can be beneficial in practicality. Additionally, the result of Experiment III demonstrated that using the WS of 40 increased the accuracy of the models. Nonetheless, increasing the WS does not improve the model performance. With more data collected, the trade-off should be examined and the association of more than 4 SPTs should be considered in future studies for more practical usage in a real-world environment.

The authors wished to set this work as an exemplar with an impact on UWB radar applications and escalate the benefits of this technology for future usage. With the increase in development of the off-the-shelf UWB radar systems, the installation procedure of the systems for monitoring purpose in buildings can reduce the difficulty of utilization. This, therefore, eases the usage transitioning to increase the quality of the classification models performance by supplementing other factors such as heart rate or respiratory rate. Moreover,
according to the findings from the previous works, extracting both spatial and time domain features from the human bio-signals is beneficial in deep learning approaches and can be applied in various tasks [48], [49]. Using transfer learning is another interesting addition which may aid the model to continue learning when equipped with more devices [50]. Additionally, from previous studies of our team members [51], [52], we used bio-signal from contact sensors for biometric application and stress monitoring, which can be suitable for non-contact body sensor, i.e., UWB radars. Furthermore, the non-contact equipment could benefit the security issue of assessing the bio-metric signals from wireless and implantable devices in patients [53]. These formulate an excellent adaptation for model improvement in UWB radar applications in the future.

V. CONCLUSION

This study proposed SleepPoseNet (SPN), a novel core architecture featuring Multi-View (MVL) and Multi-Task (MTL) deep learning approaches, to classify different human sleep postures utilizing the signals from single ultra-wideband (UWB) antenna. The classification results toward the four sleep postural transitions (SPTs) demonstrated promising recognition. Moreover, we incorporated time series data augmentation techniques to prevent an overfitting problem during the training session of SPN. This prevention significantly enhanced the classification performance. In summary, SPN will contribute as a pioneer deep learning architecture for various UWB-based applications including human sleep monitoring.

REFERENCES

[1] L. F. Drager, R. D. McEvoy, F. Barbe, G. Lorenzi-Filho, and S. Redline, “Sleep apnea and cardiovascular disease: lessons from recent trials and need for team science,” Circulation, vol. 136, no. 19, pp. 1840–1850, 2017.
[2] M. McDermott and D. L. Brown, “Sleep apnea and stroke,” Current Opinion in Neurology, vol. 33, no. 1, pp. 4–9, 2020.
[3] E. Bixler, F. He, J. Fernandez-Mendoza, L. Dumpring, S. Calhoun, and A. Vgontzas, “Association of childhood sleep apnea with adolescent hypertension: The penn state child cohort,” Circulation, vol. 140, no. Suppl 1, pp. A10 677–A10 677, 2019.
[4] B. Jurado-Gámez, O. Guglielmi, F. Gude, and G. Buela-Casal, “Workplace accidents, absenteeism and productivity in patients with sleep apnea,” Archivos de Bronconeumología (English Edition), vol. 51, no. 5, pp. 213–218, 2015.
[5] S. Garbarino, O. Guglielmi, A. Sanna, G. L. Mancardi, and N. Magnavita, “Risk of occupational accidents in workers with obstructive sleep apnea: systematic review and meta-analysis,” Sleep, vol. 39, no. 6, pp. 1211–1218, 2016.
[6] L. Walsh, S. McLoone, J. Ronda, J. F. Duffy, and C. A. Czeisler, “Non-contact pressure-based sleep/wake discrimination,” IEEE Transactions on Biomedical Engineering, vol. 64, no. 8, pp. 1750–1760, 2016.
[7] F. Lin, Y. Zhuang, C. Song, A. Wang, Y. Li, C. Gu, C. Li, and W. Xu, “Sleepsense: A noncontact and cost-effective sleep monitoring system,” IEEE transactions on biomedical circuits and systems, vol. 11, no. 1, pp. 189–202, 2016.
[8] J. V. Rundo and R. Downey III, “Polysomnography,” in Handbook of clinical neurology. Elsevier, 2019, vol. 289, no. 2, pp. 223–226, 2003.
[9] C. H. Lyder, “Pressure ulcer prevention and management,” Jama, vol. 289, no. 2, pp. 223–226, 2003.
[10] R. Yousefi, S. Ostadabbas, M. Faezipour, M. Farshbaf, M. Nourani, L. Tamil, and M. Pompeo, “Bed posture classification for pressure ulcer prevention,” in 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, 2011, pp. 7175–7178.
[11] S. Fallmann and L. Chen, “Computational sleep behavior analysis: A survey,” IEEE Access, vol. 7, pp. 142 421–142 440, 2019.
[12] Y. Cao, F. Wang, X. Xu, N. Lin, B. Zhang, Z. Liu, and S. Sigg, “Contactless body movement recognition during sleep via wifi signals,” IEEE Internet of Things Journal, vol. 7, no. 3, pp. 2028–2037, 2020.
