# A Deep Learning Approach for Sleep-Wake Detection from HRV and Accelerometer Data

Zhenghua Chen[1], Min Wu[1], Jiyan Wu[1], Jie Ding[1], Zeng Zeng[1], Karl Surmacz[2],and Xiaoli Li[1]
[1] Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way #21-01 Connexis, Singapore 138632
[2] McLaren Applied Technologies, United Kingdom
{chen_zhenghua, wumin, wu_jiyan, ding_jie, zengz, xlli}@i2r.a-star.edu.sg, karl.surmacz@mclaren.com

***Abstract*—Sleep-wake classification is important for measuring the sleep quality. In this paper, we propose a novel deep learning framework for sleep-wake detection by using acceleration and heart rate variability (HRV) data. Firstly, considering the high sampling rate of acceleration data with temporal dependency, we propose a local feature based long short-term memory (LF-LSTM) approach to learn high-level features. Meanwhile, we manually extract representative features from HRV data, as HRV data has a distinct format with acceleration data. Then, a unified framework is developed to combine the features learned by the LF-LSTM from acceleration data and the features extracted from HRV data for sleep-wake detection. We use real data to evaluate the performance of the proposed framework and compare it with some benchmark approaches. The results show that the proposed approach achieves a superior performance over all the benchmark approaches for sleep-wake detection.**

***Index Terms*—Sleep-wake detection, Local features, Deep learning**

## I. Introduction

Sleep is an essential physiological function that affects the performance of various activities in daily life, such as memory, learning, attention and productivity. Moreover, sleep restrictions and disorders have also been found to implicate the physical and mental health conditions of human [1]. Monitoring of sleep by detecting sleep-wake stages can measure both the sleep duration and sleep quality, and it is thus an important step for us to maintain good health condition and improve daily performance.

Traditionally, polysomnography (PSG) is used as a reference tool for sleep stage detection and sleep quality measurement. However, it is labour-intensive, comparatively invasive and expensive. Nowadays, economical wearable sensors have been widely used for sleep tracking. Actigraphy or accelerometer units are common modules to measure activities that depict sleep patterns [2], [3]. Actigraphy, which makes use of movement information to differentiate between sleep and wake, is both cheap and less cumbersome. Actigraphy has found its way into consumer wearables, where products like FitBit, Jawbone and Basis have them embedded within a watch for people to wear throughout the day and night. In addition to actigraphy and accelerometers, wearable sensors, which capture electrocardiogram (ECG), respiratory or heart rate variability (HRV) data, have also been used for sleep-wake detection and sleep monitoring [4], [5].

Given the data collected from the sensors mentioned above, various machine learning algorithms have been proposed for accurate detection of sleep-wake states, such as linear discriminant (LD) classifier [6], [7], decision tree (DT) [2], support vector machine (SVM) [5], artificial neural network (ANN) [2], [4] and random forest (RF) [8]. Recently, deep learning algorithms have also been used for sleep-wake detection. In [9], convolutional neural network (CNN) was introduced to detect sleep-wake patterns from the actigraphy data. In [10], a bidirectional long short-term memory (Bi-LSTM) model was proposed to classify sleep and wake states from multimodal data, i.e., acceleration, skin conductance and skin temperature.

In this paper, we aim to classify sleep-wake states on a dataset with both acceleration and HRV streams. Since the sensory data is typical time series, the LSTM related algorithms are naturally suitable for this problem, due to the strong sequential modeling capacity of LSTM. However, since the acceleration data is with a high sampling rate, each sample will be an extremely long sequence. It is thus infeasible to train the conventional LSTM (or Bi-LSTM [10]) on this extremely long sequence, due to the constrains on memory and computational power. Furthermore, HRV data and acceleration data are with different nature and format. It is also not allowed to directly feed these two heterogeneous sensory data into the LSTM or Bi-LSTM network.

To address the above issues, we propose a unified deep learning framework for sleep-wake detection by using both acceleration and HRV data. Considering the high frequency of acceleration data with temporal dependency, we propose a local feature based long short-term memory (LF-LSTM) approach to learn high-level features. The raw acceleration data is divided into small windows where some statistical features are extracted within each window. These local features represent the properties of raw acceleration data and preserve temporal dependency. Then, the LSTM network is employed to learn high-level features and encode temporal dependency on these local features. In the meantime, some representative features are extracted from HRV data which has a distinct format with the acceleration data. Finally, we present a unified framework to make full use of the features learned by the LF-LSTM from the acceleration data and the extracted features from the HRV data for the better detection of sleep and wake. The results on a real dataset show that our proposed approach achieves a superior performance over all the benchmark methods for sleep-wake detection.

978-1-7281-0848-3/19/$31.00 ©2019 IEEE

## II. METHODOLOGY

### A. Local Feature Based Long Short-Term Memory

For sleep-wake detection, the three-dimensional acceleration data can be a good indicator. For data preprocessing, a common operation is to use sliding windows for data segmentation. Assume the window size is $a$ seconds and the sampling rate of acceleration is $r$, each sample will have a dimension of $ar \times 3$. Generally, the raw acceleration is noisy and not representative for sleep or wake state. For conventional machine learning based sleep-wake detection, a compulsory operation is to manually extract some representative features from the raw data. Alternatively, deep neural networks can be applied to automatically learn high-level features. Since the acceleration is a typical time series, recurrent neural network (RNN) which is able to encode temporal dependency of data is naturally suitable for this problem. However, the conventional RNN often suffers from the problem of gradient vanishing or exploding. To solve this problem, the long short-term memory (LSTM) network was proposed in [11]. It intends to use some gates, i.e., input gate, output gate and forget gate, to control the information for persevering or discarding, so that it can model long-term dependencies of data.

In experiments, we use a sliding window of $5$ minutes and a sampling frequency of $100$ Hz. Thus, each sample will have a dimension of $30,000 \times 3$. If the conventional LSTM network is utilized for feature learning on this extremely long sequence, it means that $30,000$ LSTM cells need to be connected in cascade for feature learning, which requires extreme large memory to store all the parameters. Therefore, it is not applicable to use the conventional LSTM for feature learning on these data samples. To solve this problem, we propose a LF-LSTM network for efficient learning. It consists of two steps, local feature extraction and high-level feature learning using the LSTM network.

*1) Local features:* Since the sample is too long for learning, we attempt to segment the sample by using sliding windows. Then, for each window, we extract some representative features on the three dimensions to get a feature vector which can be treated as a comprehensive representation of each window. Since the windows are segmented in sequence, they will preserve the temporal dependency of the raw signal. The main objective is to reduce the sample length and preserve the temporal dependency of the sequential data by extracting more efficient representations for the raw data. An illustration on local feature extraction is shown in Fig. 1.

In particular, the local features extracted on each window are mean, absolute mean, maximum, minimum, range, variance, root mean square, interquartile range, and quantile at 25%, 50% and 75%. Totally, eleven features are extracted on each dimension for local features.

*2) High-level feature learning using LSTM:* Due to the strong sequential learning capacity of the LSTM network, it has been widely used for the analysis on time series data, such as natural language processing [12], activity recognition [13] and occupancy estimation [14]. Here, we leverage on the LSTM network to encode temporal dependency and learn high-level features on the extracted sequential local features from raw acceleration data.

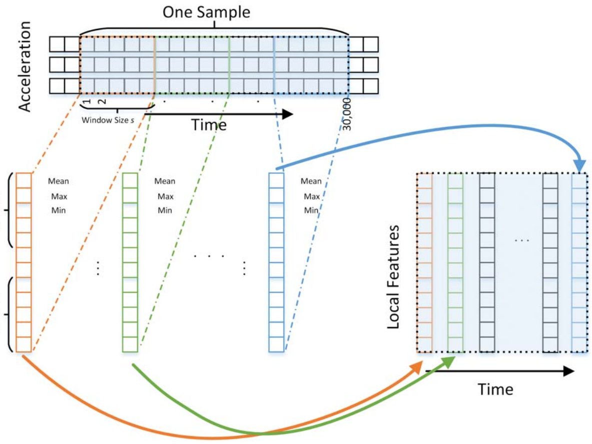


Fig. 1. Illustration on local feature extraction.

Recently, deep architectures have been shown to be effective in representation learning [15]. In this work, we attempt to stack multiple layers of LSTM to learn more representative features from the sequential local features for sleep-wake detection.

### B. Feature Extraction on HRV Data

HRV data, which includes the R-R interval values, shows the variation of time intervals between heart beats. Due to the specific format of HRV data, it cannot be used for feature learning via deep learning algorithms. Thus, we manually extract features from the time-domain, frequency-domain, Poincaré plots of R-R values and detrended fluctuation analysis (DFA) for the HRV data [16].

Time-domain features are calculated directly from R-R interval values, including meanRR (mean of the R-R values), StdRR (standard deviation of the R-R values), meanHR (mean of the heart rate), RMSSD (root mean square of the successive differences). Frequency-domain features are calculated from the power spectrum generated by a Fast Fourier Transform (FFT), including the power in different frequency bands, e.g., very low frequency (VLF, 0-0.04 Hz), low frequency (LF, 0.04-0.15 Hz) and high frequency (HF, 0.15-0.4 Hz), and the total power (TP). In addition, the ratio LF/HF is also used as a feature. We derive 3 statistical features derived from the Poincaré plot distribution and 3 slope features from DFA. Table I summarizes the features extracted from HRV data.

### C. Proposed Framework for Sleep-Wake Detection

In this work, we employ two heterogeneous sensor data, i.e., acceleration and HRV, which have distinct formats and properties, for the task of sleep-wake detection. The proposed framework for sleep-wake detection is shown in Fig. 2. Firstly, we design a LF-LSTM network to learn effective features from acceleration data. Meanwhile, we manually extract some representative features from the HRV data which has a special

TABLE I
FEATURES EXTRACTED FROM HRV DATA.

| Categories | Features |
|---|---|
| Time domain | meanRR, meanHR, StdRR, cvRR, RMSSD, SDSD, RR50, pRR50 |
| Frequency domain | VLF, LF, HF, TP, LF/(LF+HF), HF/(LF+HF), LF/HF |
| Distribution features | SD1, SD2, SD1/SD2 |
| DFA features | Three slope coefficients $\alpha$, $\alpha 1$ and $\alpha 2$ |

format. Next, we leverage on a unified framework to combine the features from these two heterogeneous sensors.

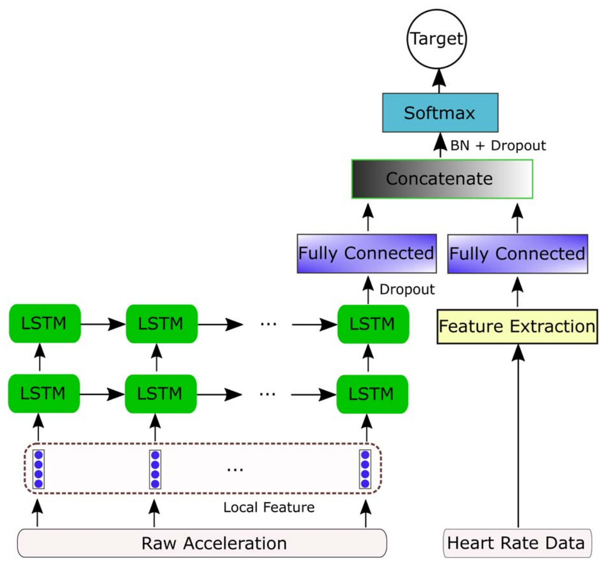


Fig. 2. Proposed framework for sleep-wake detection.

The hyperparamters of the proposed framework are determined by using cross-validation on the training data. Specifically, the hidden nodes of two LSTM layers are 50 and 100, respectively. The dropout rate of the dropout layer after the LSTM layer is set to be 0.5. The both fully connected layers have 100 hidden nodes. After the concatenate layer, we use a batch-normalization (BN) layer and another dropout layer with the same dropout rate. Finally, a softmax layer is utilized for sleep-wake classification.

## III. EVALUATION

### A. Data Collection

We collected data from 11 participants for 28 sleep nights (NUS-IRB Ref Code: B-15-276). All the participants wore 3 types of sensors, including a FAROS device, a CamNtech motionwatch and a Zeo sleep monitor headband sensor as shown in Fig. 3. In particular, FAROS device collected both acceleration and HRV data, CamNtech and Zeo reported the sleep-wake states for the participants. Meanwhile, we also required the participants to record their key events during the sleep nights.

To analyze the sleep data, we divide it into 5-min segments. We derive 3 sleep-wake labels for each segment from CamNtech, Zeo and participant's event log and keep all the segments whose 3 labels are consistent. As such, we have 1,658 sleep segments and 200 wake segments in this study. Note that this data is naturally imbalanced and we have many more sleep segments than wake segments.


Fig. 3. The devices for data collection.

### B. Experimental Setup

To evaluate the performance of the proposed approach, we have compared it with some well-known benchmark approaches, including conventional machine learning approaches of DT [2], LD [6], [7], SVM [5], ANN [4] and random forest (RF) [8], and deep learning approach of CNN [9]. Here, the conventional machine learning approaches apply the HRV features described in Table I. They also utilize the same local features for the acceleration data, while these local features are calculated over the entire 5-min sample. The CNN in [9] takes the acceleration data only as input and cannot include the HRV data due to its special format. Since each sample has 30,000 steps, the Bi-LSTM in [10] cannot be trained due to the constrains on memory and computational power.

Since sleep-wake detection is a typical imbalance data problem, the classification accuracy may overlook the minority class (i.e., wake) in the evaluation of model performance. Hence, we also utilize the metric of G-mean which is widely used for performance evaluation on imbalanced data [17]. According to the definition of confusion matrix in Table II, the G-mean can be defined as

$$\begin{aligned} \text{sensitivity} &= \text{TP}/(\text{TP}+\text{FN}) \\ \text{specificity} &= \text{TN}/(\text{TN}+\text{FP}) \\ \text{G-mean} &= \sqrt{\text{sensitivity} * \text{specificity}} \end{aligned} \tag{1}$$

TABLE II
CONFUSION MATRIX

| | Predicted positives | Predicted negatives |
|---|---|---|
| True positives | TP | FN |
| True negatives | FP | TN |

In evaluation, we randomly select 70% of data for model training, and the remaining for testing. In addition, to address the data imbalance issue, we use the oversampling technique of SMOTE (Synthetic Minority Over-sampling Technique) [18] for data augmentation on training data.

### C. Experimental Results and Discussion

*1) Results:* The experimental results of all the approaches are shown in Table III. The ensemble learning of RF performs better than the other conventional machine learning algorithms, i.e., DT, LD, SVM and ANN. Due to the extreme long sequence of samples, the CNN cannot capture the long-term dependencies, resulting a limited performance. Besides, it cannot include the HRV data which is also helpful in separating the sleep and wake states. Lastly, our proposed approach outperforms all the benchmark approaches in terms of both accuracy and G-mean, owing to the proposed LF-LSTM for feature learning on acceleration data and the unified framework to include the features from the HRV data.

TABLE III
EXPERIMENTAL RESULTS OF VARIOUS APPROACHES WITH SMOTE.

| Models | Accuracy (%) | G-mean |
|---|---|---|
| DT [2] | 89.6 | 0.718 |
| LD [6], [7] | 86.9 | 0.802 |
| SVM [5] | 82.4 | 0.816 |
| ANN [4] | 91.4 | 0.817 |
| RF [8] | 92.3 | 0.854 |
| CNN [9] | 76.0 | 0.755 |
| Proposed | **95.1** | **0.884** |

*2) Impact of SMOTE and HRV:* To evaluate the impact of the SMOTE and the HRV data, we conduct additional experiments with different combinations. The results are shown in Table IV. Due to the imbalance property of data, the accuracy of the model without SMOTE is higher than that with SMOTE. But the model has opposite results in terms of G-mean which is widely used for the evaluation of imbalance data. This means the criterion of accuracy is not stable and proper for the evaluation of model performance in this problem.

According to Table IV, it is clear that the SMOTE will dramatically improve the performance of sleep-wake detection in terms of G-mean due to the high imbalance of the data. Moreover, the HRV data further enhances the performance of the proposed approach with and without SMOTE for data augmentation. This indicates the effectiveness of SMOTE for imbalance data and the usefulness of the HRV data for sleep-wake detection.

TABLE IV
THE RESULTS OF THE PROPOSED APPROACH UNDER DIFFERENT SETTINGS.

| SMOTE | Sensor | Accuracy | G-mean |
|---|---|---|---|
| No | Acceleration | 94.8 | 0.794 |
| No | Acceleration + HRV | **95.2** | 0.818 |
| Yes | Acceleration | 93.6 | 0.876 |
| Yes | Acceleration + HRV | 95.1 | **0.884** |

## IV. CONCLUSION

In this paper, we proposed a deep learning framework for sleep-wake detection using two different types of sensor data, i.e., acceleration and heart rate variability (HRV). Based on the evaluation on real data, the proposed method outperformed the state-of-the-art methods in the literature. Since our data is quite imbalanced, we employed an oversampling technique SMOTE for data imbalance correction. We also tested the impact of SMOTE and HRV on the performance of sleep-wake detection. The results showed that both SMOTE and HRV data are important for classifying sleep and wake states.

## REFERENCES


[1] O. M. Buxton, S. W. Cain, S. P. Oconnor, J. H. Porter, J. F. Duffy, W. Wang, C. A. Czeisler, and S. A. Shea, “Adverse metabolic consequences in humans of prolonged sleep restriction combined with circadian disruption,” *Science Translational Medicine*, vol. 4, no. 129, p. 129ra43, 2012.

[2] J. Tilmanne, J. Urbain, M. V. Kothare, A. V. Wouwer, and S. V. Kothare, “Algorithms for sleep–wake identification using actigraphy: a comparative study and new results,” *Journal of sleep research*, vol. 18, no. 1, pp. 85–98, 2009.

[3] V. T. van Hees, S. Sabia, K. N. Anderson, S. J. Denton, J. Oliver, M. Catt, J. G. Abell, M. Kivimäki, M. I. Trenell, and A. Singh-Manoux, “A novel, open access method to assess sleep duration using a wrist-worn accelerometer,” *PloS one*, vol. 10, no. 11, p. e0142533, 2015.

[4] W. Karlen, C. Mattiussi, and D. Floreano, “Sleep and wake classification with ecg and respiratory effort signals,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 3, no. 2, pp. 71–78, 2009.

[5] M. Adnane, Z. Jiang, and Z. Yan, “Sleep–wake stages classification and sleep efficiency estimation using single-lead electrocardiogram,” *Expert Systems with Applications*, vol. 39, no. 1, pp. 1401–1413, 2012.

[6] S. Devot, R. Dratwa, and E. Naujokat, “Sleep/wake detection based on cardiorespiratory signals and actigraphy,” in *32th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2010, pp. 5089–5092.

[7] 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 journal of Biomedical and Health Informatics*, vol. 18, no. 4, pp. 1272–1284, 2014.

[8] M. B. Pouyan, M. Nourani, and M. Pompeo, “Sleep state classification using pressure sensor mats,” in *37th Annual International Conference of Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 1207–1210.

[9] L. Granovsky, G. Shalev, N. Yacovzada, Y. Frank, and S. Fine, “Actigraphy-based sleep/wake pattern detection using convolutional neural networks,” *arXiv preprint arXiv:1802.07945*, 2018.

[10] W. Chen, A. Sano, D. L. Martinez, S. Taylor, A. W. McHill, A. J. Phillips, L. Barger, E. B. Klerman, and R. W. Picard, “Multimodal ambulatory sleep detection,” in *IEEE-EMBS International Conference on Biomedical and Health Informatics*, vol. 2017. NIH Public Access, 2017, p. 465.

[11] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[12] S. Wang and J. Jiang, “Learning natural language inference with lstm,” *arXiv preprint arXiv:1512.08849*, 2015.

[13] Z. Chen, L. Zhang, C. Jiang, Z. Cao, and W. Cui, “Wifi csi based passive human activity recognition using attention based blstm,” *IEEE Transactions on Mobile Computing*, 2018.

[14] Z. Chen, R. Zhao, Q. Zhu, M. K. Masood, Y. C. Soh, and K. Mao, “Building occupancy estimation with environmental sensors via cd-blstm,” *IEEE Transactions on Industrial Electronics*, vol. 64, no. 12, pp. 9549–9559, 2017.

[15] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[16] M. Wu, H. Cao, H.-L. Nguyen, K. Surmacz, and C. Hargrove, “Modeling perceived stress via hrv and accelerometer sensor streams,” in *37th Annual International Conference of IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 1625–1628.

[17] Y. Tang, Y.-Q. Zhang, N. V. Chawla, and S. Krasser, “Svms modeling for highly imbalanced classification,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 39, no. 1, pp. 281–288, 2009.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “Smote: synthetic minority over-sampling technique,” *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.