Poster: DeepSleep: A Ballistocardiographic Deep Learning Approach for Classifying Sleep Stages

Shashank Rao
TU Delft
Delft, The Netherlands
shashankpr16@gmail.com

Abdallah El Ali
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
abdallah.el.ali@cwi.nl

Pablo Cesar
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands
p.s.cesar@cwi.nl

ABSTRACT

Current techniques for tracking sleep are either obtrusive (Polysomnography) or low in accuracy (wearables). In this early work, we model a sleep classification system using an unobtrusive Ballistocardiographic (BCG)-based heart sensor signal collected from a commercially available pressure-sensitive sensor sheet. We present *DeepSleep*, a hybrid deep neural network architecture comprising of CNN and LSTM layers. We further employed a 2-phase training strategy to build a pre-trained model and to tackle the limited dataset size. Our model results in a classification accuracy of 74%, 82%, 77% and 63% using Dozee BCG, MIT-BIH's ECG, Dozee's ECG and Fitbit's PPG datasets, respectively. Furthermore, our model shows a positive correlation ($r = 0.43$) with the SATED perceived sleep quality scores. We show that BCG signals are effective for long-term sleep monitoring, but currently not suitable for medical diagnostic purposes.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous computing**.

KEYWORDS

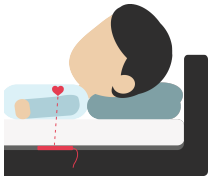
Sleep classification, ballistocardiography, deep learning, dozee

ACM Reference Format:
Shashank Rao, Abdallah El Ali, and Pablo Cesar. 2019. Poster: DeepSleep: A Ballistocardiographic Deep Learning Approach for Classifying Sleep Stages. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2019 International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct), September 9–13, 2019, London, United Kingdom.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3341162.3343758

UbiComp/ISWC '19 Adjunct, September 9–13, 2019, London, United Kingdom

ACM ISBN 978-1-4503-6869-8/19/09.
https://doi.org/10.1145/3341162.3343758

(a) Dozee sleep-monitoring device.



(b) Placement of Dozee sensor-sheet.

Figure 1: Dozee sensor-sheet and its usage.

1 INTRODUCTION

Sleep-related disorders such as sleep deprivation [6] and sleep apnea can be studied and diagnosed, and interventions can occur using ubiquitous sensing technologies [1]. Sleep clinics typically use Polysomnography (PSG), a test conducted to study sleep and to diagnose different forms of sleep disorders [4]. To date, PSG is considered as the most accurate method for diagnosing sleep-related problems and considered the gold standard in clinical sleep medicine. However, it suffers from the fact that it is expensive, complex, time-consuming, and uncomfortable for the users [4]. In this early work, we aim to model a sleep classification system using an unobtrusive Ballistocardiographic (BCG)-based heart sensor signal collected from *Dozee*[1] (Fig. 1), a commercially available non-contact, unobtrusive pressure-sensitive sensor sheet. Here, we ask: *How can a sleep classification system be modelled using BCG sensor data, in order to achieve a performance comparable with PSG?*

2 METHODS

We aim to classify four sleep stages [11]: (1) Wake state (2) Rapid Eye Movement (REM) (3) Light sleep, and (4) Deep sleep. This study is based on four datasets (distribution shown in Table 1): *Dozee*'s *BCG* dataset, *Dozee*'s *ECG* data, the *MIT-BIH Polysomnographic* dataset [7] and the *PPG*-based Fitbit data [10] provided by *Fitabase*[2]. Our model is trained using the *Dozee*'s BCG dataset while the *Dozee ECG*, *MIT-BIH PSG* data and the *Fitbit*'s PPG data are used for transfer learning.

[1]https://www.dozee.io
[2]Fitabase: https://www.fitabase.com/research-library/

Table 1: Overview of the class representation in each of the dataset used in this work. The "Number of Recordings" indicates the total number of data that we have for that particular dataset. The number in parantheses corresponds to the number of unique subjects from which the recordings were obtained.

Dataset	Sensor type	No. of Recordings	Sample rate (Hz)	Wake	REM	Deep	Light
Dozee BCG	BCG	51 (25)	250	8%	22%	25%	45%
Dozee ECG	ECG	51 (25)	250	8%	22%	25%	45%
MIT-BIH	PSG-ECG	80 (16)	250	15%	25%	30%	30%
Fitabase-Fitbit	PPG	12 (4)	120	20%	20%	15%	45%

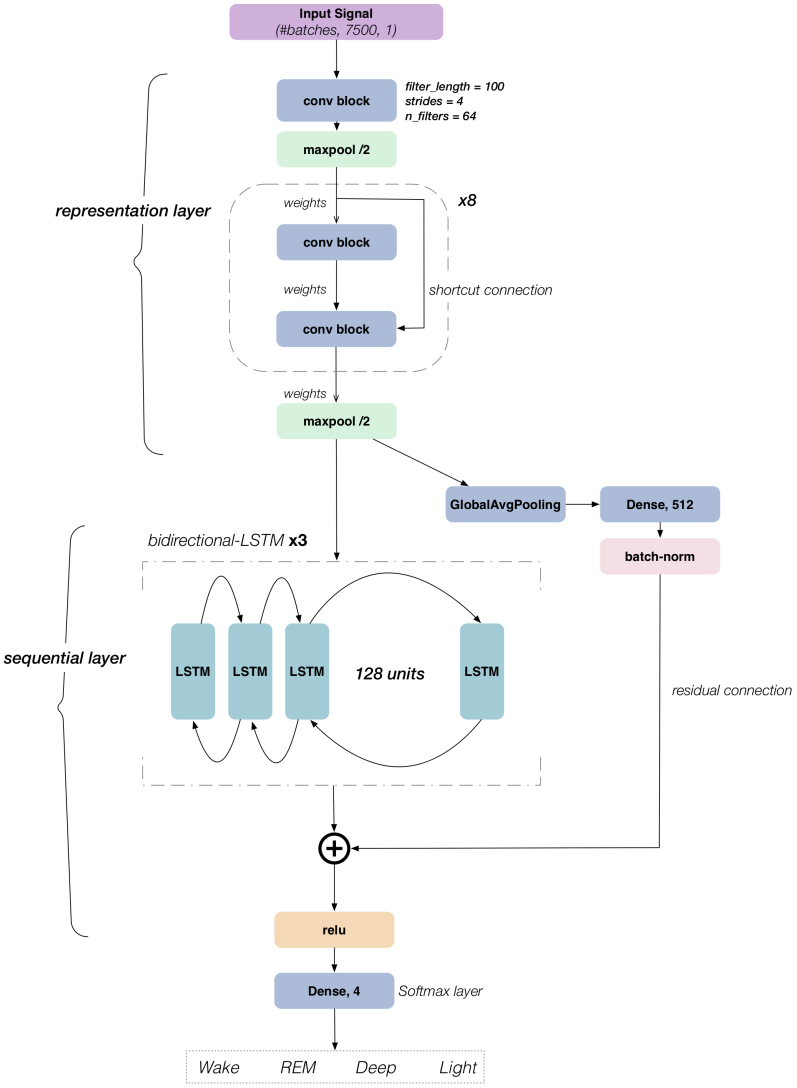
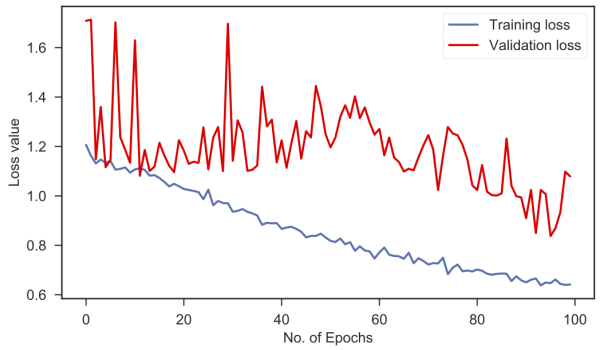
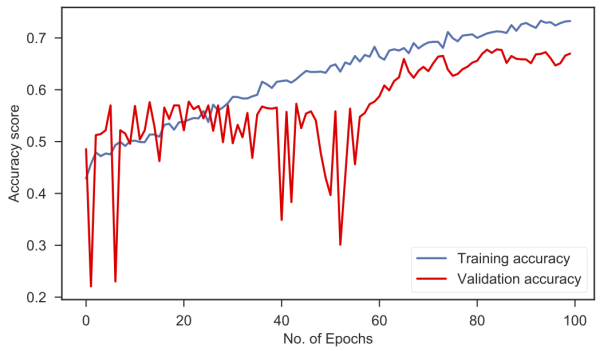


Figure 2: *DeepSleep* model architecture with residual connections.



(a) *Categorical loss for pre-training.*



(b) *Categorical accuracy for pre-training.*

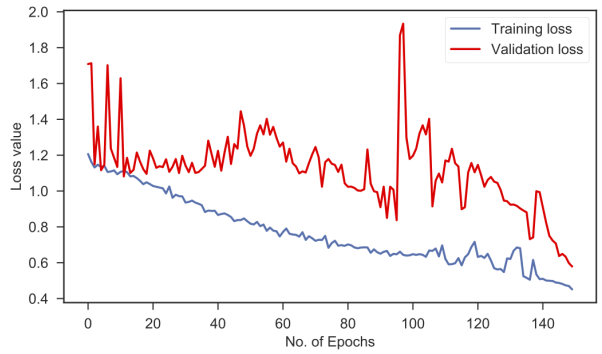
Figure 3: Pre-training phase.

To study the effect of sleep on our cardiac rhythm, *Heart-rate Variability* (HRV) features are extracted [11]. For the Dozee BCG dataset, we had 51 recordings across 25 subjects, where ground truth data was annotated by 2 doctors (Cohen's k = 0.80) from NIMHANS (National Institute of Mental Health and Sciences) in Bangalore, India.

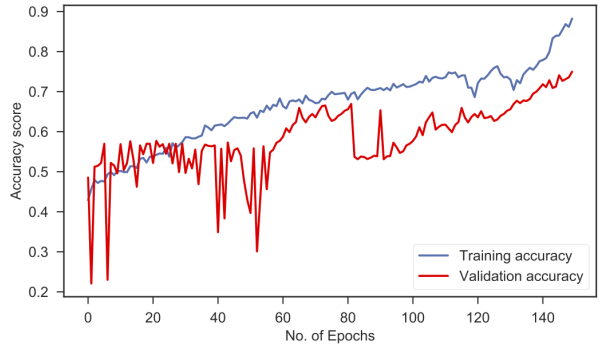
We present *DeepSleep*, a hybrid deep neural network model that can automatically extract heart-related features and learn the time-dependent nature of sleep patterns for classification (Fig. 2). We run pre-training and subsequently fine-tuning to help tackle the limited amount of labelled sensor data. This training strategy enables us to test the pre-trained model's classification ability on ECG and PPG sensor data. A combination of stacks of 1-D convolutional networks

Table 2: Performance comparison between *DeepSleep* model and prior works that perform 4-class classification.

Study	Year	Sensor type	#Features	Classifier	Classes	Accuracy
Längkvist et al. [9]	2012	EEG, EOG, EMG	1	DBN, HMM	W, REM, NREM, L	72%
Samy et al. [12]	2014	BCG	6	KNN, SVM, Naive-Bayes	W, L, REM, Deep (NREM)	72%
Supratak et al. [13]	2017	EEG	1	1D-CNN + LSTM	W, REM, NREM, L	86%
Dong et al. [5]	2018	EEG, EOG	1	LSTM	W, REM, NREM, L	86%
Chambon et al. [3]	2018	EEG, EOG, EMG	1	1D-CNN	W, REN, NREM, L	87%
DeepSleep (proposed)	**2019**	**BCG**	**1**	**1D-CNN + bi-LSTM**	**W, L, REM, Deep (NREM)**	**74%**



(a) *Categorical loss for fine-tuning.*



(b) *Categorical accuracy for fine-tuning.*

Figure 4: Fine-tuning phase.

(1D-CNNs) and Bidirectional Long-Short Term Memory (bi-LSTM) [8] layers are incorporated in the model design to enable unsupervised feature learning and sequential learning, respectively. We run our experiments using two NVIDIA GTX 1080Ti GPU clusters, random weight initialization, random oversampling to balance data, learning rate of 1e-3, use Adam optimizer, and perform an 80-20% training-test split, with pre-train and fine-tune steps (shown in Fig. 3 and Fig. 4, respectively) of 150 epochs each.

Lastly, we test how well the sleep quality scores from our DeepSleep model correlate with the perceived quality score

Table 3: SATED framework questionnaire [2] used for collecting perceived sleep quality scores. Total for all items ranges from 0 (poor sleep health) to 10 (good sleep health).

		Rarely/ Never (0)	Some-times (1)	Usually/ Always (2)
Satisfaction	Are you satisfied with your sleep?			
Alertness	Do you stay awake all day without dozing?			
Timing	Are you asleep (or trying to sleep) between 2:00 a.m. and 4:00 a.m.?			
Efficiency	Do you spend less than 30 minutes awake at night? (This includes the time it takes to fall asleep and awakenings from sleep.)			
Duration	Do you sleep between 6 and 8 hours per day?			

(1hr and 24 hr after PSG recording) as reported by 16 different users (N=16), using the 5-item self-rated sleep quality questionnaire called *SATED* [2] (Satisfaction, Alertness, Timing, Efficiency, Duration). The SATED questionnaire is shown in Table 3. Since the SATED dimensions include questions about alertness and satisfiability, we instructed the subjects to fill in the same questionnaire again after a gap of one day from their study. This way we attempted to collect better alertness scores which may not be perceived immediately after waking up. We used the mean SATED scores to test its correlation with the PSG's and *DeepSleep*'s sleep score.

Table 4: Classification performance (W, L, REM, Deep (NREM)) of DeepSleep model on different datasets.

Dataset	Sensor type	#Recordings	Accuracy
Dozee BCG	BCG	51	74%
Dozee ECG	ECG	51	77%
MIT-BIH ECG	ECG	80	82%
Fitbit-PPG	PPG	12	63%

Table 5: Precision, Recall and F1-score of *DeepSleep* model.

	Precision	Recall	F1-score	Samples
Deep	0.74	0.76	0.75	236
Light	0.79	0.84	0.82	497
REM	0.77	0.64	0.71	193
Wake	0.59	0.70	0.64	98
avg / total	0.73	0.74	0.73	1024

3 RESULTS & FUTURE WORK

Our DeepSleep model has a mean accuracy of 74% using BCG signals only (Table 2), with confusion matrix shown in Fig. 5. Precision, recall and F1-scores are shown in Table 5. We further employed a 2-phase training strategy to build a pre-trained model and to tackle the limited dataset size. With a classification accuracy (Table 4) of 82%, 77% and 63% using MIT-BIH's ECG, Dozee's ECG and Fitbit's PPG datasets, we find lowest performance on Fitbit-PPG (likely due to lower number of recordings). In our transfer learning setting on ECG data, we reach an accuracy of 82%, likely due to shape similarities of BCG and ECG signals.

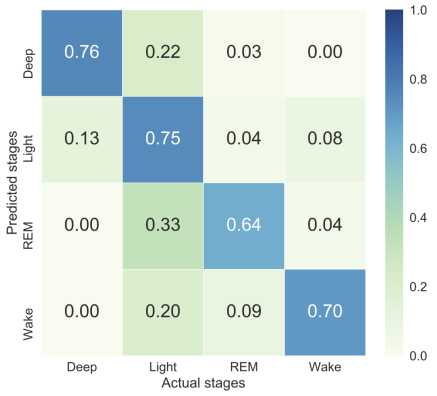


Figure 5: Classification accuracy confusion matrix for our DeepSleep model.

Finally, with a correlation coefficient of $r = 0.43$, our model shows a positive correlation with the SATED questionnaire perceived sleep quality scores, by contrast to a coefficient of $r = 0.48$ with PSG, and $r = 0.54$ between SATED and PSG. Although our current proposed model's performance is not yet comparable to PSG, we show that heart rate signals alone are an effective means for long-term sleep monitoring, but currently not suitable for medical diagnostic purposes. Our next steps are to validate our approach using leave-one-subject out cross-validation (despite testing on different sets of subjects), and to test our model against non-Artificial Neural Network approaches. Finally, we aim at experimenting with better oversampling techniques, such as *Seq2Seq autoencoders* [14] to encode a signal length of 30 seconds and create a new synthetic sequence of the same signal length. The generative property of the *Seq2Seq* network could retain a high amount of correlation and temporal order of the original sequence when generating a new sequence.

REFERENCES

[1] Saeed Abdullah, Mark Matthews, Elizabeth L. Murnane, Geri Gay, and Tanzeem Choudhury. 2014. Towards Circadian Computing: "Early to Bed and Early to Rise" Makes Some of Us Unhealthy and Sleep Deprived. In *Proc. UbiComp '14*. ACM, New York, NY, USA, 673–684. https://doi.org/10.1145/2632048.2632100

[2] Daniel J Buysse. 2014. Sleep health: can we define it? Does it matter? *Sleep* 37, 1 (2014), 9–17.

[3] Stanislas Chambon, Mathieu N Galtier, Pierrick J Arnal, Gilles Wainrib, and Alexandre Gramfort. 2018. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* (2018).

[4] Alexandru Corlateanu, Serghei Covantev, Victor Botnaru, Victoria Sircu, and Raffaella Nenna. 2017. To sleep, or not to sleep – that is the question, for polysomnography. *Breathe* 13, 2 (jun 2017), 137 LP – 140. http://breathe.ersjournals.com/content/13/2/137.abstract

[5] Hao Dong, Akara Supratak, Wei Pan, Chao Wu, Paul M Matthews, and Yike Guo. 2018. Mixed neural network approach for temporal sleep stage classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 26, 2 (2018), 324–333.

[6] Jeffrey S Durmer and David F Dinges. 2005. Neurocognitive consequences of sleep deprivation. In *Seminars in neurology*, Vol. 25. Copyright© 2005 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA., 117–129.

[7] Ary L Goldberger, Luis A Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101, 23 (2000), E215–20.

[8] Alex Graves, Navdeep Jaitly, and Abdel-rahman Mohamed. 2013. Hybrid speech recognition with deep bidirectional LSTM. In *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*. IEEE, 273–278.

[9] Martin Längkvist, Lars Karlsson, and Amy Loutfi. 2012. Sleep stage classification using unsupervised feature learning. *Advances in Artificial Neural Systems* 2012 (2012).

[10] Hawley E Montgomery-Downs, Salvatore P Insana, and Jonathan A Bond. 2012. Movement toward a novel activity monitoring device. *Sleep and Breathing* 16, 3 (2012), 913–917.

[11] National Institutes of Health et al. 2014. Brain basics: understanding sleep. *NIH Publication* 06-3440 (2014).

[12] Lauren Samy, Ming-Chun Huang, Jason J Liu, Wenyao Xu, and Majid Sarrafzadeh. 2014. Unobtrusive sleep stage identification using a pressure-sensitive bed sheet. *IEEE Sensors Journal* 14, 7 (2014), 2092–2101.

[13] Akara Supratak, Hao Dong, Chao Wu, and Yike Guo. 2017. DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 11 (2017), 1998–2008.

[14] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*. 3104–3112.