2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Consumer Smartwatches As a Portable PSG: LSTM Based Neural Networks for a Sleep-Related Physiological Parameters Estimation

Illia Fedorin[1], *Senior Member, IEEE* and Kostyantyn Slyusarenko[2]

***Abstract*— Recently, mobile and wearable devices have become an increasingly integral part of our lives. They provide a possibility of detailed health monitoring using noninvasive and user-friendly techniques. However, lack of continuous monitoring, the need of specific sensors, and the limitations in memory and power consumption are only some of the potential drawbacks of such devices. In the current paper a system based on a deep recurrent neural network is developed for an automatic continuous monitoring of sleep-related physiological parameters by means of a wearable biosignal monitoring systems. Smartwatches based algorithm for non-invasive monitoring of sleep stages, respiratory events (including sleep apnea and hypopnea), snore and blood oxygen saturation is developed. Our experimental results demonstrate that proposed model constitutes a noninvasive and inexpensive screening system for sleep-related physiological parameters and pathological states. The model has shown a 77 % accuracy in sleep stages prediction, more than 80 % accuracy in epoch-by-epoch respiratory events classification, above 60 % accuracy in snore events classification and above 70 % accuracy in blood oxygen saturation (SpO_2) level classification (for a two class problem with a SpO_2 threshold of 95 %).**

I. Introduction

Wearable biomedical sensors and the field of healthcare monitoring systems quickly develop in recent years. Leading high-tech companies increase investments in biomedical research and development every year. Smart belts, rings, watches, and earphones - it is only a part of the examples of consumer electronics available on the market with user health monitoring features. However, there are still several limitations that restrain the development of such devices. These include the limited complexity of sensors, as measuring physiological parameters requires that certain sensors be embedded into the device along with the appropriate methodology be satisfied. Moreover, the accuracy of healthcare analysis of such devices is frequently quite low and sufficiently suffer from particular measurement conditions. Memory limitations and battery life should also be taken into account. Despite the mentioned issues, wearable biomedical systems continue to develop steadily, expanding the possibilities of their use both in terms of a device for screening pathological conditions, and more accurate medical systems [1-3].

It is known, that there are an obvious direct and indirect correlation between different physiological parameters of a body, for example between heart rate and breathing cycle, heart rate variability (HRV) and systolic and diastolic blood pressure values and their trends, respiration rate and HRV and different types of breathing disorders, like sleep apnea-hypopnea syndrome, snore, acute respiratory distress syndrome, etc. Consequently, by using appropriate algorithms and approaches it is possible to estimate (with sufficient accuracy or as a screening) a number of related physiological characteristics for further in-depth analysis or continuous monitoring [4-6].

In terms of mobile and wearable devices, extracted physiological data (for example, using the most popular HRV and activity information which can be simply measured by standard accelerometers, gyroscopes or plethysmography (PPG) sensors available almost in all wearable and mobile devices) can be effectively used to estimate a number of related vital signs and pathological states (such as blood oxygen saturation, sleep apnea and hypopnea, snore, blood pressure, etc.) by applying specialized algorithms [3, 7-9]. Recently, deep-learning based models have shown promising results in the field of biomedical engineering, in particular for the analysis of sensors data, recognition of specific medical patterns, identification of hidden models, and decision-making in the field of healthcare. In particular, a recurrent neural networks (RNN), including gated recurrent units (GRU) and long short term memory (LSTM) neural networks (NN), are appropriate tool for sensors data analysis. Several recent studies have shown the potential of their use for physiological data processing (including ECG and PPG), sleep quality analysis, etc. [10-13].

This paper proposes a method for automatic continuous monitoring of multiple sleep-related physiological parameters. The idea is implemented based on the analysis of the physiological state of a person during sleep by means of wearable devices (smartwatches), including sleep stages and sleep apnea analysis, blood oxygen saturation level and snore episodes' estimation, and respiration pattern reconstruction. The deep learning framework is developed which is based on the LSTM NN with an adaptive output layer for the sleep-related physiological parameters estimation.

II. Methodology

A. General Concept

There are three main blocks in the general flowchart of the proposed system: raw sensor data gathering and processing, physiological data assessment algorithms, and measuring of sleep-related physiological parameters.

In such a system, the physiological data extraction and the sleep-related physiological parameters estimation can be

*Resrach supported by Samsung R&D Institute Ukraine.

[1]Illia Fedorin is with Samsung R&D Institute Ukraine, Kyiv, 01032, Ukraine, i.fedorin@samsung.com

[2]Kostyantyn Slyusarenko is with the Samsung R&D Institute Ukraine, Kyiv, 01032, Ukraine, k.slyusarenk@samsung.com

performed using subject specific and vital sign specific algorithms, which can be, but are not limited to, deep NN, convolutional NN, and RNN. In the current paper, recurrent neural networks were used. In the current paper, RNN were used. In this case, RNN's are the most suitable due to the possibility of time sequence processing and hidden states analysis [10-15].

B. Statement of the problem

The main purpose of the current paper is to develop a portable user friendly system for detailed analysis of physiological state during sleep, including sleep stages, respiratory events (including sleep apnea and hypopnea), snoring and blood oxygen saturation.

It should be noted, that the gold standard for the analysis of physiological state during sleep is a polysomnography (PSG) [1, 14], which is usually carried out in the special medical center under the supervision of a qualified technician and typically includes a dozens of different sensors which are attached to a body and thus extremely inconvenient (ECG, EEG, EMG, etc.).

As a basis of our current system the algorithms developed in our previous papers for sleep stages analysis, respiratory events classification and smart alarm, were used [3, 7-9, 15].

C. Neural Network for sleep-related parameters estimation

NN architecture for the sleep stages (SS) classification, respiration events (RE) pattern reconstruction, blood oxygen saturation estimation and snore events screening represents a combination of fully connected and recurrent layers. Particularly, it consists of a bidirectional LSTM NN (two parallel Bi-LSTM blocks are used) and multiple outputs in the last fully-connected classification layer, which are adjusted to the estimation of each particular sleep-related physiological parameter.

D. Dataset information

We use the same dataset as in our recent studies [3, 7-9, 15]. The full dataset consists of 263 logs from 176 different users and was prepared by Samsung Medical Center. The participant's average age is 39.6 years, BMI index is 23.99 and apnea-hypopnea index (AHI) is 12.69 events/h. The raw data includes full nocturnal PSG along with the green light PPG and 3-axic accelerometer signals from wrist, which were gathered from the Samsung Galaxy Watch at a 20Hz sampling rate. The PSGs data were labeled by a certified technician according to the AASM 2015 guidelines [16]. The dataset was divided into the first part for training and validation of 194 nights of 107 different participants and to the second part for testing of 69 nights of 69 different participants. In addition, in total 7 logs were eliminated at the pre-processing step due to the absence of signals or annotations.

III. Results and Discussion

A. Sleep stages classification

Table I presents classification results for multiclass sleep stages classification problem for each 1-minute interval using LSTM based NN and adjusted FC output layer [15]. The classification is performed into the four stages: deep, light, wake and rapid eye movement. The performance characteristics include weighted precision, recall and F1 score for each sleep stage. The average achieved accuracy is 79 %, while the Cohen's Kappa coefficient is 0.62, which corresponds to a substantial agreement between the true labels and the NN predictions. In general, for all sleep stages all classification metrics shows values greater than 0.5, which indicates a good balance of the classifier predictions. The highest performance is achieved for the Light and REM sleep stages, while the classification of the Deep sleep stages is slightly lower.

TABLE I. Classification Results for the Sleep Stages Subtask

	Deep	Light	Wake	REM
F1-score	0.59	0.85	0.60	0.77
Precision	0.53	0.83	0.67	0.81
Recall	0.56	0.84	0.63	0.79

B. Respiratory pattern

A detailed analysis of the sleep-related RE, and, in particular, sleep-related respiratory pattern reconstruction, is a key aspect of sleep medicine and sleep quality scoring.

The averaged achieved accuracy for the epoch-by-epoch RE classification is 82%, Cohen's Kappa agreement is 0.44 (which corresponds to moderate agreement), and the F1 score is 82 %, which reflects a substantial balance between RE and No RE epochs classification [9].

Fig. 1 and Table II show the confusion matrix for the 1-minute epoch's classification as well as the precision, recall and F1 score for each class. The model correctly classifies 84% of epoch's without RE and 70 % of epoch's with RE (sensitivity of the model or a recall). The resulting precision (positive predictive value) of classification of epoch's without RE is 0.94, and for epoch's with RE is 0.45.

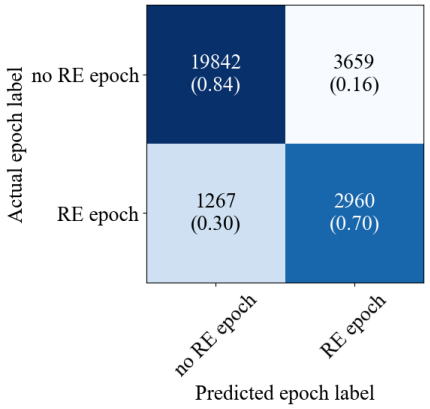

Figure 1. Confusion matrix for epoch-by-epoch respiratory pattern classification.

TABLE II. Performance Metrics of Epoch-by-Epoch Respiratory Pattern Classification

	Precision	Recall	F1 score
No RE	0.94	0.84	0.89
RE	0.45	0.7	0.55

C. AHI prediction

Using the predictions of respiratory events by each epoch of the sleep episode, an estimation of AHI can be carried out.

Unlike the actual AHI value, the predicted AHI value contains a methodological error, since we do not take into account the number of RE during each hour of sleep, as is necessary in the accordance with the classical methodology, but we count the total number of minutes with all types of RE. Therefore, this value should be considered as a screening tool for identifying the patients at risk for whom the predicted AHI value, calculated according to the proposed methodology, should be also greater than a certain threshold value.

Fig. 2 shows the correlation between actual and predicted AHI values, mean absolute error (MAE) and mean squared error (MSE) for predicted AHI for different apnea severity ranges. The correlation coefficients between AHI measured by PSG and AHI measured by proposed model are 0.74 (Spearman correlation coefficient) and 0.91 (Pearson correlation coefficient). As can be seen, 95 % of participants in the test set were either correctly classified or misclassified in the immediate neighbor class (basically, the model makes mistakes only by one severity level) in terms of apnea severity class. The main contribution in the AHI estimation error is caused by the underestimated group of 5 participants with high AHI values measured by PSG.

For a two class problem (as a screening tool for sleep-related respiratory events) with a threshold of 15 events/h the proposed model shows an accuracy of 91%. That is, the model provides high efficiency for the correct classification of participants belonging to the low apnea risk group (including "no apnea" and "mild apnea" classes) and high apnea risk group (including "moderate apnea" and "severe apnea" classes), see Fig. 3.

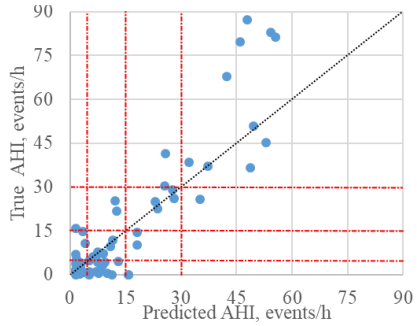

Figure 2. Actual AHI (y-axis) versus predicted AHI (x-axis) (dashed line is the identity line and dash-dotted lines are the apnea severity thresholds).

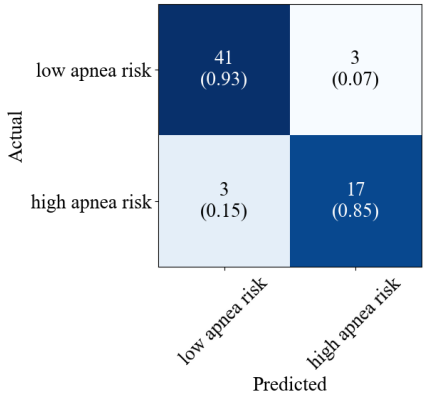

Figure 3. Confusion matrix of the apnea severity prediction: as a screening tool for two severity levels with a threshold of 15 events/h.

D. Blood oxygen saturation level estimation

It would be reasonable to clarify that if our model shows high efficiency in detecting RE epochs during which the level of blood oxygen saturation normally decreases (it is generally known, that during various types of sleep related respiratory events, the level of blood oxygen saturation drops below normal values of 95% and can even achieve critical values below 80%), how well can we evaluate the dynamics of blood oxygen saturation?

Fig. 4 represents the blood oxygen saturation level at different values of the predicted probability (by the proposed NN) of belonging to the class "RE epoch" / "no RE epoch". The box and whisker plots show interquartile range, namely the minimum, first quartile, median, third quartile, and maximum. A vertical line is the median. Outliers are omitted. Probabilities exceeding 0.5 refer to the "No RE epoch", and less than 0.5 – "RE epoch". That is, the higher the probability value, the model is more confident that the given epoch does not contain RE. The lower the probability value, the model is more confident that the given epoch contains RE. With a decrease in probability from 1 to zero, the level of blood oxygen saturation also decreases, and with probabilities less than 0.5, the median of the oxygen saturation values becomes below 95%. With probability values in the range from 0 to 0.1, the median of the blood oxygen saturation values decreases to 93%, and the range of values is from 83–100%.

The branch of the proposed model, which is responsible for the blood oxygen saturation level estimation (SpO_2), gives the following results. All 1 minute epochs were considered as epochs with normal SpO_2 (higher or equal to 95 %) and low SpO_2 (lower than 95 %). The accuracy on the test set for the epoch-by-epoch prediction for such model is 79%, the Cohen's Kappa agreement is 0.29 (which corresponds to the fair agreement). The precision, recall and F1 score are summarized in the Table III. The average weighted F1 score is 0.77, reflecting a substantial balance between the "normal" and "low" SpO_2 classes (see Fig. 5 for the details between actual classes distribution).

In general, the model shows good agreement with the SpO_2 values measured by PSG. The trend of SpO_2 changes can be also estimated with sufficient precision.

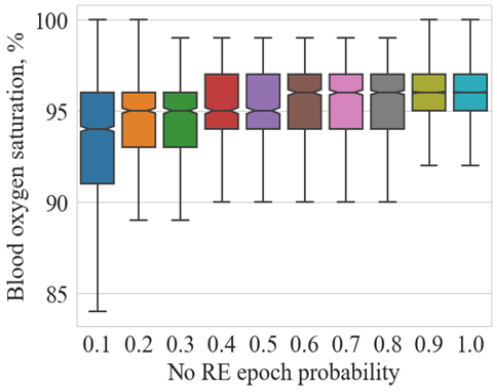

Figure 4. Blood oxygen saturation level versus probability of "RE"/ "no RE" epoch predicted by the proposed model.

TABLE III. PERFORMANCE METRICS OF EPOCH-BY-EPOCH BLOOD OXYGEN SATURATION LEVEL CLASSIFICATION

	Precision	Recall	F1 score
Normal SpO_2	0.82	0.92	0.87
Low SpO_2	0.55	0.33	0.41

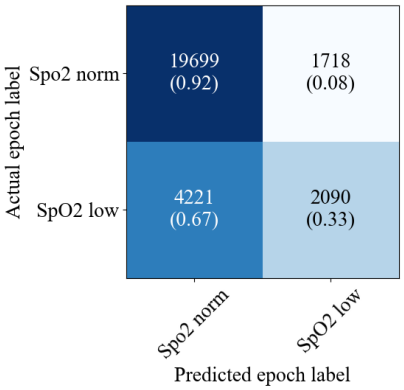

Figure 5. Confusion matrix of the blood oxygen saturation level prediction.

E. Screening of snoring events

Finally, let us analyze the ability of the model under consideration to distinguish the presence of snoring events during sleep. A 1-minute epoch is labeled as containing a snoring event if it contains at least 10 seconds of snoring.

The accuracy for the epoch-by-epoch snore events classification is 75%, Cohen's Kappa agreement is 0.25 (which corresponds to the fair agreement). The precision, recall and F1 score are summarized in the Table IV. The average weighted F1 score is 0.75 which also reflects a substantial agreement in classification of epochs containing snoring events (see Fig. 6 for details of the actual distribution of classes). Thus, the proposed model can be considered as a potential screening tool for snoring events classification.

TABLE IV. PERFORMANCE METRICS OF EPOCH-BY-EPOCH SNORE EVENTS CLASSIFICATION

	Precision	Recall	F1 score
No Snore	0.84	0.84	0.84
Snore	0.41	0.4	0.41

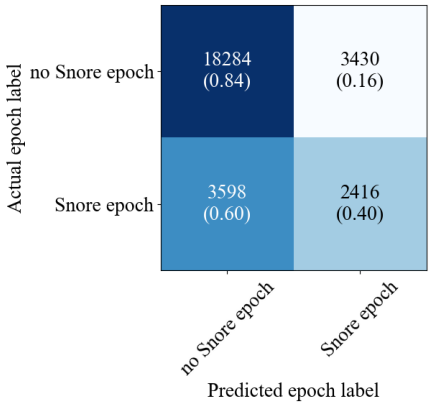

Figure 6. Confusion matrix of the snoring events classification.

IV. CONCLUSION

This research provides the deep neural network based algorithm for estimation, screening and/or continuous monitoring of multiple sleep-related physiological parameters. The hidden states, trends and correlations between the measured sleep-related physiological parameters are evaluated by means of time sequence sensors data processing using RNN. The proposed novel approach provides substantial agreement with the measured physiological parameters using PSG. It shows comparable to the human expert's results, as well as improved performance compared to the state-of-the-art solutions for sleep stages and respiratory events classification. At the same time, the assessment of AHI and SpO_2 level, as well as snoring pattern reconstruction can also be done with substantial precision, which is suitable for screening pathological conditions.

REFERENCES

[1] M. Matsuo et al., "Comparisons of portable sleep monitors of different modalities: potential as naturalistic sleep recorders," *Frontiers in Neurology*, vol. 7, pp. 110, July 2016.

[2] L. Xie et al., "Sleep drives metabolite clearance from the adult brain," *Science*, vol. 342, pp. 373-377, 2013.

[3] I. Fedorin et al., "Sleep stages classification in a healthy people based on optical plethysmography and accelerometer signals via wearable devices," in *Proc. 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON 2019)*, Lviv, Ukraine, 2019, pp. 1201-1204.

[4] A. Roebuck et al., "A review of signals used in sleep analysis," *Physiological measurement*, vol. 35, pp. R1-57, 2014.

[5] C. Orphanidou, "A review of big data applications of physiological signal data," *Biophys Rev*, vol. 11, pp. 83-87, 2019.

[6] A. Muzet et al., "Assessing sleep architecture and continuity measures through the analysis of heart rate and wrist movement recordings in healthy subjects: comparison with results based on polysomnography," *Sleep Medicine*, vol. 21, pp. 47-56, 2016.

[7] A. Havriushenko et al., "Smartwatch based respiratory rate estimation during sleep using CNN/LSTM neural network," in *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO)*, Kyiv, Ukraine, 2020, pp. 584-587.

[8] K. Slyusarenko et al., "Smart alarm based on sleep stages prediction, in *Proceedings of the 42st IEEE International Engineering in Medicine and Biology Conference (EMBC 2020)*, Montreal, QC, Canada, 2020, pp. 4286-4289.

[9] I. Fedorin et al., "Respiratory events screening using consumer smartwatches," in *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct)*, 2020, pp. 25-28.

[10] P. Warrick and M. N. Homsi, "Cardiac arrhythmia detection from ECG combining convolutional and long short-term memory networks," *2017 Computing in Cardiology (CinC)*, vol. 44, pp.1-3, 2017.

[11] A. Zeeshan et al., "Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine," *Database*, vol. 2020, pp. 1-35, 2020.

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

[13] X. Zhang et al., "Sleep stage classification based on multi-level feature learning and recurrent neural networks via wearable device," *Computers in Biology and Medicine*, vol. 103, pp. 71-81, 2018.

[14] T. Penzel et al., "Dynamics of heart rate and sleep stages in normals and patients with sleep apnea," *Neuropsychopharmacology: official publication of the American College of Neuropsychopharmacology*, vol. 28 Suppl 1, pp. S48-53, 2003.

[15] K. Slyusarenko et al., "Sleep stages classifier with eliminated apnea impact," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers (UbiComp/ISWC '19 Adjunct)*, London, United Kingdom, 2019, pp. 210-213.

[16] R. Berry et al., *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications: Version 2* (American Academy of Sleep Medicine), Darien, Illinois, 2015.