Accurate detection of sleep apnea with long short-term memory network based on RR interval signals

FAUST, Oliver <http://orcid.org/0000-0002-3979-4077>, BARIKA, Ragab, SHENFIELD, Alex <http://orcid.org/0000-0002-2931-8077>, CIACCIO, Edward J. and ACHARYA, U. Rajendra

Published version

FAUST, Oliver, BARIKA, Ragab, SHENFIELD, Alex, CIACCIO, Edward J. and ACHARYA, U. Rajendra (2020). Accurate detection of sleep apnea with long short-term memory network based on RR interval signals. Knowledge-Based Systems, p. 106591.

Accurate detection of sleep apnea with long short-term memory network based on RR interval signals

Oliver Faust[a], Ragab Barika[a], Alex Shenfield[a], Edward J Ciaccio[b], U Rajendra Acharya[c,d,e]

[a]*Sheffield Hallam University, Sheffield, UK*
[b]*Columbia University, Department of Medicine - Cardiology, New York, NY, USA*
[c]*Department of Electronics and Computer Engineering, Ngee Ann Polytechnic, Singapore*
[d]*Department of Bioinformatics and Medical Engineering, Asia University, Taichung, Taiwan*
[e]*School of Management and Enterprise University of Southern Queensland, Springfield, Australia*

Abstract

Sleep apnea is a common condition that is characterized by sleep-disordered breathing. Worldwide the number of apnea cases has increased and there has been a growing number of patients suffering from apnea complications. Unfortunately, many cases remain undetected, because expensive and inconvenient examination methods are formidable barriers with regard to the diagnostics. Furthermore, treatment monitoring depends on the same methods which also underpin the initial diagnosis; hence issues related to the examination methods cause difficulties with managing sleep apnea as well. Computer-Aided Diagnosis (CAD) systems could be a tool to increase the efficiency and efficacy of diagnosis. To investigate this hypothesis, we designed a deep learning model that classifies beat-to-beat interval traces, medically known as RR intervals, into apnea versus non-apnea. The RR intervals were extracted from Electrocardiogram (ECG) signals contained in the Apnea-ECG benchmark Database. Before feeding the RR intervals to the classification algorithm, the signal was band-pass filtered with an Ornstein–Uhlenbeck third-order Gaussian process. 10-fold cross-validation indicated that the Long Short-Term Memory (LSTM) network has 99.80% accuracy, 99.85% sensitivity, and 99.73% specificity. With hold-out validation, the same network achieved 81.30% accuracy, 59.90% sensitivity, and 91.75% specificity. During the design, we learned that the band-pass filter improved classification accuracy by over 20%. The increased performance resulted from the fact that neural

activation functions can process a DC free signal more efficiently. The result is likely transferable to the design of other RR interval based CAD systems, where the filter can help to improve classification performance.

Keywords: Sleep apnea, Deep learning, Heart rate variability, Detrending

1. Introduction

Sleep is a fundamental human activity which is characterized by reduced or suspended consciousness. Hence, the ability to avoid or correct disturbances, such as sleep disordered breathing, is reduced [1]. Sleep apnea is a common cause for sleep-disordered breathing. In the middle-aged workforce about 2% of women and 4% of men were apnea patients in 1993 [2]. In 2003, about 4% of the US population had sleep apnea [3]. The worldwide prevalence was estimated to be 6% in 2008 [4]. It is predicted that this upward trend will continue. Without diagnosis and adequate treatment patients might be exposed to an increased risk of cardiovascular diseases [5], such as stroke and hypertension [6, 7]. Apnea might also disturb recreational activities and by doing so cause mental suffering and in some cases clinical depression [8]. Apnea is also linked to narcolepsy, insomnia, and obesity [9]. Studies show that patients with apnea have a higher chance of being involved in a road traffic accident [10]. The disease is also a risk factor for complications during operations under anesthesia [11]. Finally, patients with untreated apnea have a significantly higher mortality risk when compared to a control group with the same age, sex and Body Mass Index (BMI) [4].

Current diagnostic methods depend on Polysomnography (PSG). The measurements include ECG, Electroencephalogram (EEG), Electrooculogram (EOG), Electromyogram (EMG), respiratory effort, airflow and oxygen saturation (SaO_2) [12, 13, 14, 15]. To capture these signals, the patient must sleep with intrusive measurement equipment in a clinical environment. The process requires supervision by medical specialists. The PSG process makes apnea diagnosis expensive and inconvenient. To improve this situation new methods are required which are less intrusive and more cost effective, but equally accurate. Mobile technology and advanced physiological signal measurement methods might be able to address the intrusiveness and cost issues. One promising measurement technology is single lead ECG for signal acquisition and mobile soft processing for beat-to-beat (RR) interval extraction. As such, that measurement setup has a significantly lower complexity when

compared with PSG. Furthermore, it is notably cheaper to communicate and process the resulting RR interval signals, when compared with the multitude of physiological signals measured during PSG. However, major issues remain with the diagnosis support quality provided by these systems. One critical component to ensure diagnosis support quality are the algorithms which extract the relevant information or provide decision support.

With this study we investigate the diagnosis support quality of deep learning algorithms for sleep apnea. To achieve that, we created a test setup which takes in RR interval signals and returns a decision on whether or not specific signal segments show signs of sleep apnea. The processing structure contains a pre-processing and a classification step. In the pre-processing step, the signal was band-pass filtered with an Ornstein–Uhlenbeck third-order Gaussian process. Subsequently, the filtered signal is partitioned with a sliding window. The resulting signal blocks were passed on to an LSTM network which classifies them into either apnea or non-apnea. The setup was designed with a benchmark dataset from the MIT-BIH Polysomnographic Database. With 10-fold cross-validation, we established an accuracy of 99.80%, a sensitivity of 99.85%, and a specificity of 99.73% for the proposed system. By itself, this result is significant, because it indicates that good diagnostic support is possible even with a less complex data acquisition setup. Apart from these results we also want to report a significant design achievement. We found that low- and high-pass filtering the RR interval signal improved the classification accuracy by over 20%. Filtering, as part of the pre-processing for RR interval signals, might help to improve the detection quality for a wide range of CAD systems, because it allows the deep learning algorithms to focus on the Heart Rate Variability (HRV).

To support these claims, we outline our design of an apnea detection algorithm. The next section introduces the medical background of sleep apnea. Section 3 details the methods used to construct the test setup. Thereafter, we present the results achieved while testing the proposed diagnosis support system. In the Discussion section, we relate our work to other studies done on similar topics. Having this extended scope allows us to show how the RR interval filtering might help to improve the classification accuracy for other detection tasks. The conclusion summarizes the work and puts forward the highlights of the study.

2. Background

During apnea the patient ceases to breath for 10 s or more. Obstructive Sleep Apnea (OSA) and Central Sleep Apnea (CSA) are the two main causes for the pauses in breathing. The pauses usually occur during during rapid eye movement sleep. An OSA event occurs when the airway is blocked completely. The blockage might be due to fatty tissue, musculus geniohyoideus, or musculus genioglossus. In contrast, a CSA event is characterized by a lack of respiratory effort, i.e. there is a problem with respiration control [16]. OSA is diagnosed more often than CSA [17]. There are several therapies for sleep apnea, such as Positive Airway Pressure (PAP) and Palato Pharyngo Plasty (PPP) [18, 12]. In general, these therapies are more effective when sleep apnea is detected early [13, 19].

In current clinical practice, polysomnograms, which result from PSG sleep studies, are used to evaluate an index score. The score value determines the apnea severity [20, 21]. An important component of these index scores is the airflow signal and blood oxygen content [22, 23]. However, measuring these signals is intrusive and inconvenient for the patient. To reduce the inconvenience, apnea detection methods were developed using respiratory and single-lead ECG signals [24, 19]. In response, PhysioNet held a competition called CinC Challenge 2000 [25, 26], which provided ECG data with minute-by-minute labeling [27, 28]. After the challenge, the training dataset, with 35 recordings, was made publicly available by PhysioNet. Over the years, the dataset was used to design apnea detection algorithms and it is now considered a benchmark that can be used to compare individual method performances.

Digital biomarkers fail to capture all sleep apnea induced morphological changes [29, 30], because transient abnormalities appear randomly, and long-term abnormalities are difficult to quantify [31]. Deep neural networks can refine the information even further and provide medical decision support which can help to diagnose sleep apnea [32, 33, 34, 35, 36, 37]. The research provided precedents of employing Convolutional Neural Network (CNN) to detect disease using ECG signals. In apnea detection tasks, directly feeding original ECG signals to deep neural networks is adopted by some researchers [38, 39, 40], but the high ECG data rate limits the network depth. As such, the RR interval signal is derived from the ECG extracting the beat-to-beat record of RR-intervals and is, as a time series, irregularly sampled. Studies show that there is a physiologic link between the breathing rate and the

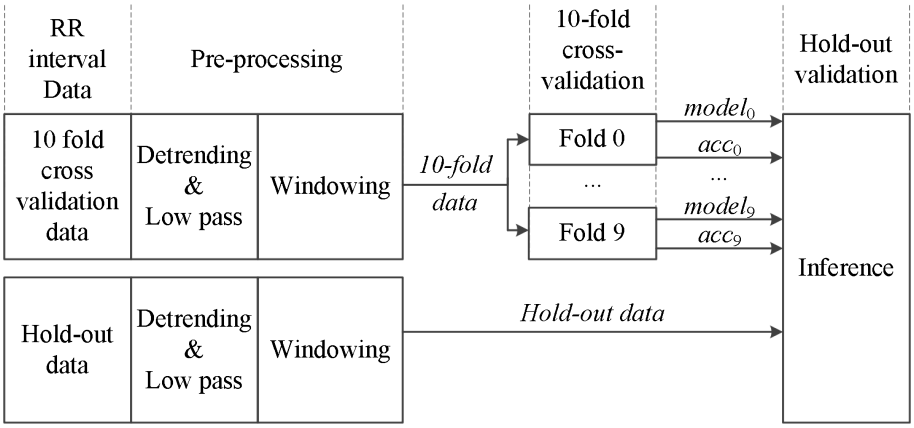

Figure 1: Block diagram for training and validating the deep learning model.

beat-to-beat variations of the human heart [41, 42, 43]. Hence, it is possible to detect sleep disordered breathing based on RR interval signals. The next section describes the methods we have used to detect apnea induced sleep disordered breathing based on RR interval signals.

3. Methods

This section describes the methods used to create the sleep apnea detection system. This is done by describing the data and the methods which process the data to refine and ultimately extract diagnostically relevant information. The block diagram, shown in Figure 1, provides an overview of the system that was used to train and validate the deep learning model. The processing steps are represented by blocks, and the arrows between the blocks represent the data flow. The following sections introduce both processing steps and data in more detail.

3.1. RR interval data

The deep learning model was trained and validated with data from the Apnea-ECG Database [25, 26]. The dataset consisted of 35 records (a01 through a20, b01 through b05, and c01 through c10). The individual recordings vary in length from slightly less than 7 hours to nearly 10 hours. Each record consists of an ECG signal of varying length, and corresponding R beat labels that were generated with automated QRS detection. The shortest signals are just below 7 hours in length and the longest one is almost 10

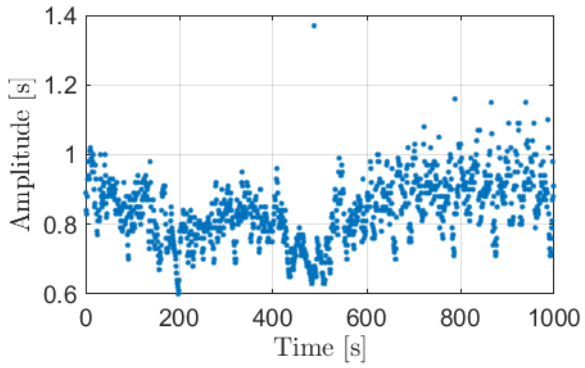

Figure 2: RAW RR interval data from record a01

hours. The subjects of these recordings are men and women between 27 and 63 years of age, with weights between 53 and 135 kg (BMI between 20.3 and 42.1). Crucially for this work, the records also contain apnea annotations established by human experts based on simultaneously recorded signals such as respiration, that were recorded as part of a PSG. Table 1 provides details about the signals for both 10-fold cross- and hold out-validation. We have partitioned that dataset into *Hold-out data* and 10-*fold data* for the two validation methods outlined in Sections 3.3 and 3.4. The *Hold-out data* contains five records (a11, a15, a17, b01, c07). The 10-*fold data* contains the remaining records. Figure 2 shows the RR intervals that occur during the first 1000 seconds of record a01. Note, there is a significant DC bias in the signal. That bias is quantified in the frequency domain as a power level of 192.5 $s^2\,Hz^{-1}$. Figure 3 shows the Power Spectral Density (PSD) of the RAW RR interval data shown in Figure 2.

3.2. Pre-processing

The pre-processing of the RR interval signals for both 10-*fold data* and *Hold-out data* was done with a two-step process. The first step is low and high pass filtering. For RR interval signals, high pass filtering is referred to as detrending. The second pre-processing step is windowing, which partitions the data for the classification algorithm.

3.2.1. Detrending and low-pass filtering

From a time series perspective, RR interval signals are nonuniformly sampled. Therefore, conventional signal conditioning using Infinite Impulse Re-

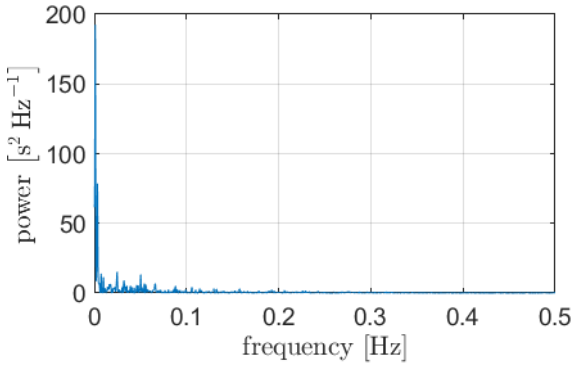

Figure 3: PSD of the RAW RR interval data

Table 1: Number of beats and signal name for 10-fold cross-validation and hold-out-validation data from the Physionet Apnea-ECG Database.

10-fold cross-validation No. beats=935462						Hold-out-validation No. beats=169959	
Name	Beats	Name	Beats	Name	Beats	Name	Beats
a01	29639	a12	33829	b05	26937	a11	32953
a02	34931	a13	39723	c01	27643	a15	33948
a03	33966	a14	28212	c02	32137	a17	36131
a04	30902	a16	34948	c03	23758	b01	35081
a05	28740	a18	29970	c04	28089	c07	31846
a06	27199	a19	38738	c05	27957		
a07	37462	a20	34246	c06	28062		
a08	41102	b02	34877	c08	30360		
a09	31318	b03	28918	c09	31179		
a10	32263	b04	24379	c10	23978		

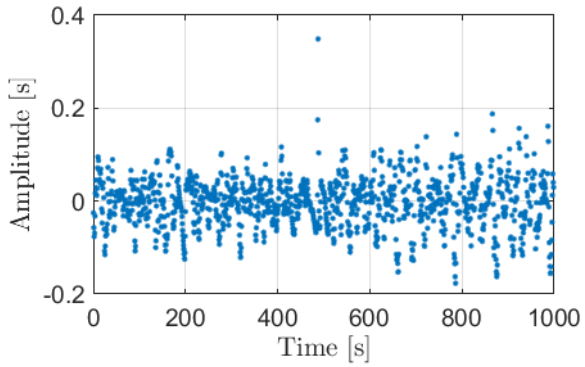

Figure 4: Detrended and low pass filtered RR interval data

sponse (IIR) and Finite Impulse Response (FIR) filters cannot be applied directly. It is necessary to resample the signals such that the resulting samples are at equidistant time intervals, typically at 0.25 s. However, such interpolative resampling introduces noise into the signal, which compromises information quality [44, 45]. Filter methods which act directly on irregularly sampled signals can help to prevent the negative effects of resampling.

For our study we have used the detrending and low-pass filter proposed by Fisher et al. [46]. The filter combination is based on an Ornstein–Uhlenbeck third-order Gaussian process which acts on the RR interval signal directly. Figure 4 shows the filtered version of the unprocessed signal provided in Figure 2. The DC bias is significantly reduced. This visual observation is confirmed in the PSD plot shown in Figure 5. The effects of the detrending filter can be observed as the absence of low frequency components up to 0.02 Hz of the normalized frequency. In terms of visual interpretation, removing the DC bias helps to focus on the variability of the RR intervals. In the spectrum plot of the RAW signal, the frequency content caused by that variability was overshadowed by the large DC components. Removing that component allowed us to re-scale the y-axis on the PSD plot which essentially means to zoom in on the spectrum component which hold relevant information for apnea classification.

3.2.2. Windowing

To partition the data for the classification algorithm, we have used a sliding window of 100 RR intervals on the data. The window slides with one RR interval at a time. In other words, the windowing method creates one

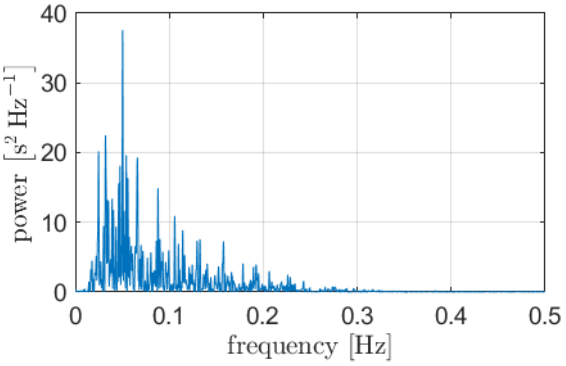

Figure 5: PSD of the detrended and low pass filtered RR interval data

data block of 100 RR intervals for each beat from the database. This creates a good temporal resolution, and it generates sufficient data to train and test the deep learning algorithm. A window was labeled apnea (positive) if at least 25 RR intervals were labeled apnea. All other windows were labeled non-apnea (negative). The labels for the individual RR intervals came from the Apnea-ECG Database.

3.3. 10-fold cross-validation

10-fold cross-validation aims to mitigate the effects of choosing test samples from an available dataset. Kohavi et al. recommend 10-fold cross-validation for model selection [47]. Hence, this performance measure is relevant for comparing classification models; see Table 3 in Section 5. The basic idea is to partition the labelled data into 10 parts. Each of the cross-validation partitions contained mixed data from the cross-validation dataset (as shown in Table 1). This follows common practice within the machine learning and bioinformatics community for tuning models [48, 49, 50, 51]. Once the data is split, the parts are used to generate 10 folds with training and test data. For fold 0, part 0 is used to test and the remaining 9 parts are used to train the network. Similarly, for fold 1, part 1 is used to test and the remaining 9 parts are used to train the network, etc. The left part in the flowchart, shown in Figure 6, depicts the data arrangement for 10-fold cross-validation.

The model fitting process is structured into 40 epochs. Within each epoch the LSTM network is trained and tested. The training step will result in a *model*, i.e. a set of weights. The LSTM network testing step establishes the

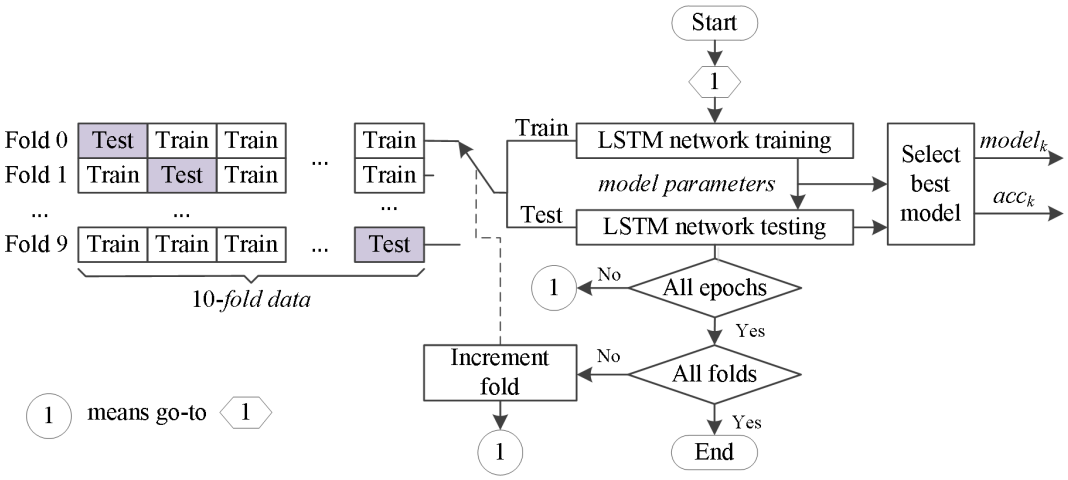

Figure 6: Flow chart for 10-fold cross-validation, where $model_k$ indicates the best LSTM model for fold k, similarly acc_k is the best accuracy for fold k.

prediction quality of the *model*. Based on the prediction quality, the 'Select best model' block decides which model is the best for a particular fold. Once all the epochs are processed, the data from the next fold is loaded. The algorithm returns once all the folds are processed and the K best models, together with their accuracy (acc), are established. The right part in the flow chart depicts the epoch-based fold processing.

3.3.1. Long short-term memory network

Figure 7 shows a functional diagram of the LSTM algorithm. The upper part of the diagram indicates the Recurrent Neural Network (RNN) loop unrolling, which results in individual LSTM cells. The hidden state vector $\vec{h}_t \in \mathbb{R}^h$ and the cell state vector $\vec{c}_t \in \mathbb{R}^h$ are passed from one cell to the next. The cells consume the input vector $\vec{x}_t$ at different time instances t. Each cell A has LSTM functionality, as indicated in the lower part of the figure.

Each cell incorporates the three gates to establish the LSTM functionality [52]. The forget gate regulates the information content stored within the cell and thereby it plays a vital role in modeling the way humans remember and forget [53]. It is implemented as the first multiplier from the left, highlighted in orange. The input gate is implemented as the second multiplier from the left, highlighted in blue. The output gate is implemented as the third multiplier from the left, highlighted in green.

The weights and biases are established during the training phase and they

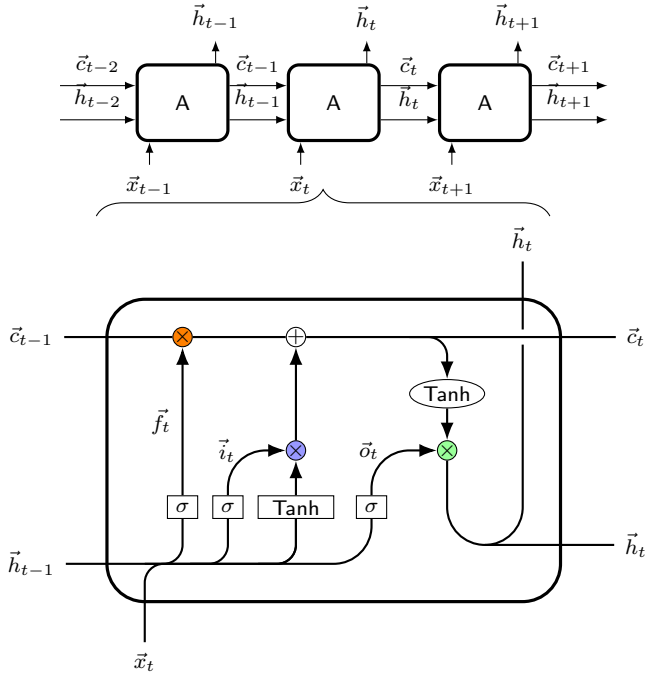

Figure 7: Overview of the deep learning algorithm. Depicted as RNN loop unrolling and LSTM cell. In the LSTM cell, $\sigma(...)$ is the sigmoid activation function and Tanh(...) is the hyperbolic tangent function.

Table 2: Bidirectional LSTM architecture.

Layer	Type	Output shape	Number of parameters
1	Input	100, 1	0
2a	LSTM (forward)	200, 400	161600
2b	LSTM (backward)	200, 400	161600
3	Global 1D max pooling	400	0
4	Fully connected Rectified Linear Unit (ReLU)	50	20050
5	Dropout	50	0
6	Fully connected (Sigmoid)	1	51

constitute the LSTM *model*. During the testing phase, the *model* is used to classify an input sequence $\vec{x}_t$. In our case, the model establishes if there are signs of sleep apnea in a block of 100 RR intervals. The methods used for testing the LSTM model are introduced in the next section.

Table 2 shows the model architecture used in this paper. The model used here is a bidirectional LSTM model [54] - where the RR input sequence is passed simultaneously forward through one LSTM model (i.e. samples $x_0, ..., x_n$) and backward through another LSTM model (i.e. samples $x_n, ..., x_0$). This allows the bidirectional LSTM model to consider time dependencies in both the past and future of a timestep. The outputs of the two LSTMs are then concatenated together and global max pooling (in one dimension) is applied. In these experiments we used both recurrent dropout [55] (with a probability of 0.1) applied to the inputs and hidden states of the LSTM cells and standard dropout [56] (again with a probability of 0.1) applied between the final fully connected layer and the output. These serve to improve the generalization of the model and reduce over-fitting. The model was trained using the Adam optimizer [57] with a learning rate of 1e-3, a batch size of 1024 (providing a good trade-off between available Graphics Processing Unit (GPU) memory and speed of training), and training performance was evaluated using the binary cross-entropy loss function. The same batch size was used in one of our previous models for LSTM based atrial fibrillation detection in RR interval signals [58]. Models were implemented using the Keras and Tensorflow frameworks [59, 60].

3.4. Hold-out testing

The unseen / generalization performance is tested using the held-out dataset (as performed in [51]). During validation we test the best models from each fold with the *Hold-out data*. This is done by accumulating the weighted prediction results. The weight factor reflects the relative prediction accuracy of the specific *model*. It is established by dividing the model accuracy (acc_k) by the sum of all model accuracies ($accAcc$). Equation 1 defines the accumulated accuracy over all folds.

$$accAcc = \sum_{k=0}^{K-1} acc_k \tag{1}$$

where K is the number of all folds. The $inference$ value is established by using the best *model* parameters from the K folds. The prediction result is weight adjusted with the established model accuracy (acc_k) divided by the accumulated accuracies ($accAcc$).

$$inference = \sum_{k=0}^{K-1} \frac{predict(Hold\text{-}out\ data, model_k) \times acc_k}{accAcc} \tag{2}$$

where $predict(data, model)$ used the LSTM algorithm to estimate for a specific *data* based on the *model* parameter.

For hold-out validation testing, the $inference$ results are compared with the data block labels. The comparison results are discussed in the next section.

4. Results

This section provides the hold-out and 10-fold cross-validation results for the proposed sleep apnea detection method. We report a confusion matrix for each of these tests. These matrices detail the number of RR intervals correctly identified as normal (TN), the number of RR intervals falsely identified as apnea (FP), the number of RR intervals falsely identified as normal (FN), and the number of RR intervals correctly identified as apnea (TP). As such, the LSTM network testing algorithm returns a vector with elements in the range of 0 to 1. In order to compare these results with the true labels,

we have used a threshold of 0.5, which was established through Receiver Operating Characteristic (ROC) analysis; see Section 4.1. The confusion matrix has the following form:

$$C = \begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix} \tag{3}$$

With these base results, we calculate the following performance measures:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN}, \\ \text{Sensitivity} &= \frac{TP}{TP + FN}, \\ \text{Specificity} &= \frac{TN}{TN + FP}. \end{aligned} \tag{4}$$

In a final step we evaluate sensitivity and specificity at different threshold levels to establish the true positive rate and false positive rate, respectively. The threshold determines the level below which a result is interpreted as negative, and all other results are interpreted as positive. These results are depicted in a ROC curve which plots the true positive rate over the false positive rate.

4.1. 10-fold cross-validation

Figure 8 shows the confusion matrix for the 10-fold cross-validation, described in Section 3.3. The predicted labels correspond very well with the true labels; this is indicated by the low number of false classifications. The selected operating point maximizes the perpendicular distance between the dashed red line (Luck) the ROC curve. That operating point translates into a threshold of 0.5 which is used to establish the confusion matrix entries. The Area Under Curve (AUC) of 1.00 indicates a perfect result. This outcome indicates that the 1856 misclassifications, reported in the confusion matrix, were not statistically relevant.

Figure 10 shows the accuracy of the models for the test set against the number of epochs. Figure 11 shows the loss of the model against the number of epochs. These plots show the results obtained with the hold-out validation method outlined in Section 3.4. The performance of the LSTM algorithm is similar across the folds, hence the variance is small. Therefore, the shaded area in the graphs, which indicate the variance, is very small, which makes it barely visible.

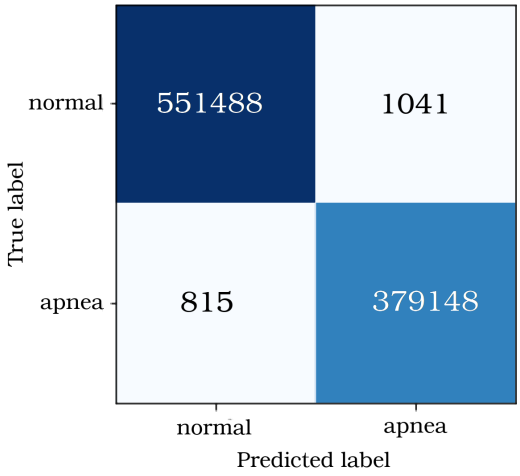

Figure 8: Confusion matrix for 10-fold cross-validation

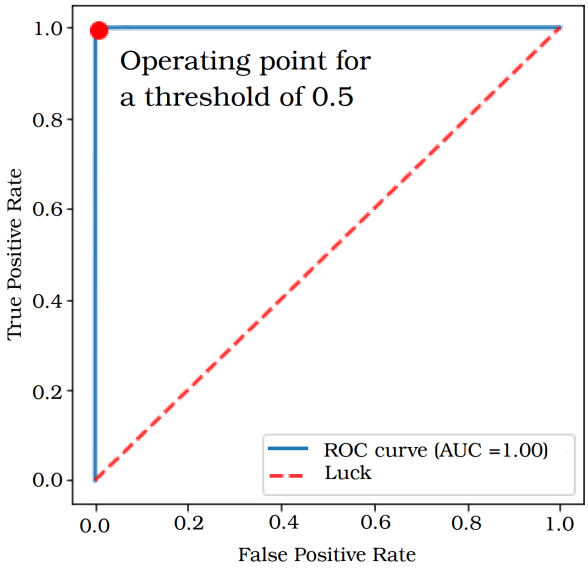

Figure 9: ROC for the 10-fold cross-validation test

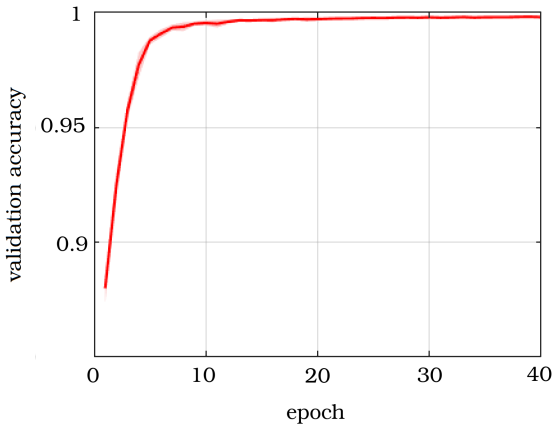

Figure 10: Validation accuracy over 40 training epochs. The solid red line represents the mean valuation accuracy of the 10 folds and the shaded area indicates the variance.

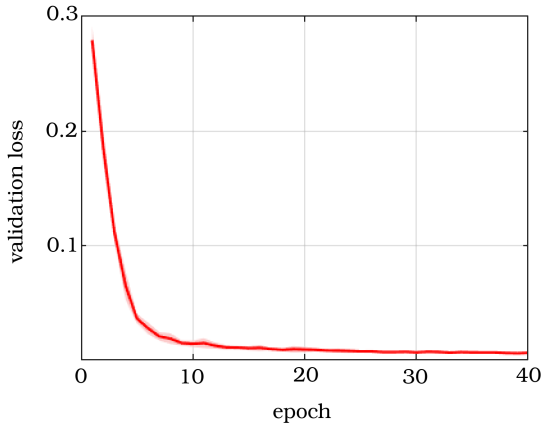

Figure 11: Validation loss function over 40 epochs. The solid red line represents the mean and the shaded area indicates the variance.

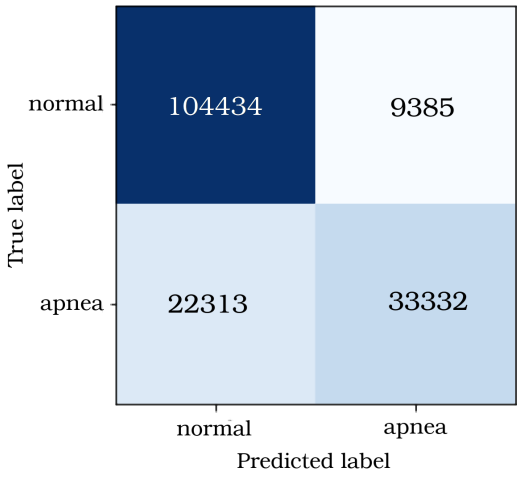

Figure 12: Hold-out confusion matrix

4.2. Hold-out validation

Once the 10 best LSTM models were established during 10-fold cross-validation, we were in a position to conduct the hold-out validation, as described in Section 3.1. The confusion matrix for the hold-out validation is shown in Figure 12. Based on these measures, the classification performance was established. The last row in Table 3 provides the hold-out performance values. Figure 13 shows the corresponding ROC curve.

5. Discussion

In this study we show that it is possible to detect sleep apnea through RR interval analysis. The following list details the advantages of the proposed method:

- Low measurement complexity – this translates into low energy requirements, which is beneficial for wireless sensor applications. Furthermore, the measurement can be done in the patient environment, potentially even by the patient.

- Low data rate – It makes RR interval signals energy efficient to communicate, store, and process. In many cases, this energy efficiency translates into cost efficiency.

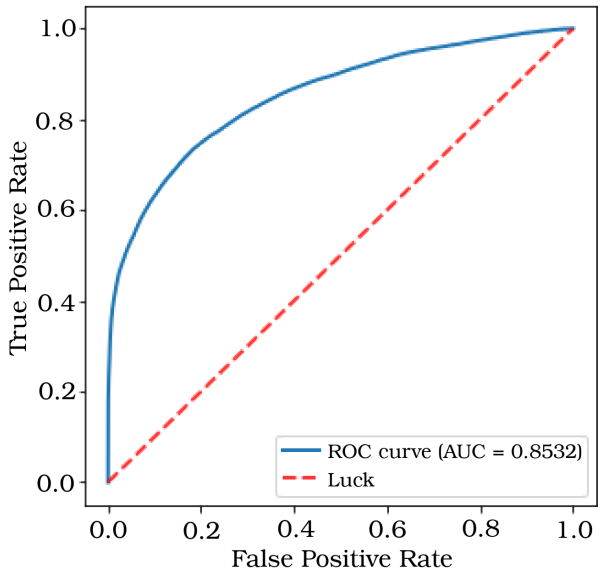

Figure 13: ROC for Hold-Out validation

- Low complexity of the algorithm chain – to classify the RR interval section we use only a two-step process. There is no feature engineering which complicates and in some cases even dilutes the information extraction.

- Real-time processing – RR intervals can be measured, communicated, and processed such that the results are available for efficient diagnostic support, and treatment monitoring can be guaranteed.

This work is based on the assumption that variations in the beat-to-beat interval of the human heart holds information that can help to detect sleep apnea. As a corollary, we assume that all components of the RR signal which do not hold information about the beat variations are irrelevant. With these ground rules in place, we set about investigating appropriate pre-processing methods. Initially, we focused our efforts on detecting and correcting outliers in RR interval data and adjusting the method used for labeling data RR interval blocks. However, with these pre-processing methods, the classification accuracy remained below 80%. Furthermore, the graph which documents the training progress showed a split between training- and valuation-accuracy, which indicates that the network could not extract decision relevant information from the RR interval signal. Only after the band-pass filter, described

in Section 3.2.1, initial model fitting tests showed that the valuation accuracy jumped to over 99% and there was no split between the training and valuation performance of the network. As such, detrending the RR interval signals removes a narrow frequency band around DC from the signal. This band does not carry information about the beat-to-beat variability. Hence, the irrelevance reduction does not impact on the beat-to-beat variability as it turns out the opposite effect was observed: detrending improved the classification accuracy significantly. We have selected LSTM as classification algorithm, because previous studies showed that LSTM performed well on time series data. Several researchers have compared the performance of Gated Recurrent Units (GRU) and LSTM model architectures on a range of natural language processing and sequence modelling tasks with no overall winner emerging [61, 62, 63]. Generally, GRU models seem to perform better when data sets are small, with LSTM models exhibiting greater expressive power in capturing long term dependencies in larger data sets.

Our study was based on data from the well known PhysioNet Apnea-ECG Database. That enabled us to compare our results with the classification results that are available from other research projects. Table 3 summarizes the outcome of these research projects. Some classical studies were focused on the design of digital biomarkers, which extract in specific properties from the available signals. For example, Varon et al. used orthogonal subspace projections to extracted 7 digital biomarkers from an ECG-Derived Respiration (EDR) signal [64]. Mendez et al. combined an autoregressive model with a K-Nearest Neighbor (K-NN) classifier to achieve a classification accuracy of above 85% [65]. An extreme learning machine was used by Tripathy to classify digital biomarkers, extracted with intrinsic band functions, from both EDR and HRV signals [66]. Song et al. extracted 11 digital biomarkers hidden in the ECG [48]. The resulting values were fed into a Markov model to refine the information further. Janbakhshi and Shamsollahi extracted digital biomarkers from ECG to derive EDR [67]. Other studies used adaptive boosting (AdaBoost) [49] and even threshold methods [68] for apnea detection. Apart from focusing on detection algorithms, researchers also investigated the practicality of such systems by using data from wearable sensors [50] and by analyzing the real-time properties of the information extraction algorithms [69]. Both studies used Support Vector Machine (SVM) for classification.

Wang et al. [51] used five records (a11, a15, a17, b01, c07) as *Hold-out data*. These are the same five records we used for hold-out validation. Thus, the

Table 3: Summary of studies on algorithmic sleep apnea detection based RR interval signals from records in the Apnea-ECG Database.

Author	Classifier	Validation method	No. features	Acc. in %	Sen. in %	Spe. in %
Mendez et al. [65]	K-NN	Leave-One-Out	52	85.7	81.4	88.4
Surrel et al. [50]	SVM	10-fold	88	88.4	73.3	87.6
Bsoul et al. [69]	SVM	Variable-folds	111	88.49	96.77	83.62
Song et al. [48]	SVM+LR	10-fold	32	86.2	80.0	89.9
Hassan [49]	adaboost	10-fold	18	87.33	81.99	90.72
Janbakhshi et al. [67]	assemble	cross-validation	85	90.90	89.60	91.80
Chazal et al. [70]	LD/QD	Many-fold	52	92.5	91.4	93.1
Dong et al. [68]	threshold	single-fold	6	90.10	88.29	90.50
Wang et al. [51]	residual network	10-fold Hold-out	0	94.39 80.60	93.04 -	94.95 -
Proposed method	**LSTM**	**10-fold** **Hold-out**	**0**	**99.80** **81.30**	**99.85** **59.90**	**99.73** **91.75**

results achieved are strictly comparable. Table 3 shows the hold-out performance measures for both studies. The hold-out performance of our study is 0.7% better than the results from Wang et al. However, the main point is that both studies could not confirm the 10-fold cross-validation results with equally good hold-out results. This and other limitations will be discussed in the next section.

5.1. Limitations

The main limitation of this work comes about from the low hold validation accuracy of 81.30%. We suspect that the number of training cases was insufficient to extract knowledge concerning sleep apnea changes in the RR interval signal. Therefore, more varied data is needed to improve the knowledge extracted during training and establish robust hold-out testing. Concerning the data used for this study, there is also a shortcoming in terms of instrumentation. The RR intervals were extracted from ECG signals via automated QRS detection. Changing the instrumentation setup might alter the QRS detection algorithm as well. These different QRS detection algorithms can show variations in the RR interval signal produced from the same ECG signal.

Our study is also limited by the rectangular window we use to create data blocks with 100 RR intervals. The window function alters the PSD of the RR interval sequence. The blocks of 100 RR interval blocks might not contain sufficient data to capture all relevant information present in the nonlinear signal characteristics. Hence, the LSTM algorithm might not receive all of the available information. However, the 10-fold cross-validation and the training progress, indicated by the graphs shown in Figures 10 and 11, indicate the 100 beats were sufficient to answer the apnea non-apnea question with a high degree of accuracy.

5.2. Future work

The 10-fold cross-validation results show that the proposed deep learning model is robust for the datasets it was trained on. However, the hold-out performance needs to be improved in the future. This should be done by training and testing the model with more varied data. Apart from improving the model, there is also scope to extend the role of the deep learning system from detection to prediction. Recent work by Hu et al. indicates that RR interval based sleep apnea detection might be possible [71].

Hypopnea is defined as abnormally slow or shallow breathing [72]. The airways are partially blocked, in contrast for apnea in which the airways are fully blocked. Hence, hypopnea can be considered a milder form of breathing disorder, which makes it harder to detect. However, hypopnea might lead to apnea, and therefore hypopnea detection can help to initiate treatment which prevents patients from developing sleep apnea [1]. Therefore, in the future we plan to train and test our deep learning model with hypopnea data in order to detect this breathing disorder as well in RR interval signals.

6. Conclusion

In this paper we proposed a processing architecture for sleep apnea detection in RR interval signals. In a pre-processing step we filtered the RR interval signal and partitioned it with a sliding window. The resulting RR interval blocks were fed into an LSTM network for classification. Filtering the signal helped the deep learning system to focus on the information contained in the HRV. As a consequence, the LSTM algorithm could extract relevant knowledge from the signal to achieve a 10-fold cross-validation accuracy of 99.80%. The variance between the folds was low. The hold-out accuracy was 81.30%.

Having accurate and robust processing methods for RR interval based sleep apnea detection is prerequisite for cost-effective CAD systems. These systems could be used for the initial diagnosis and during treatment monitoring. In such a CAD setting, the deep learning results constitute an independent second opinion on the data. In the clinical workflow, a human expert should validate the machine decision through an independent review of the evidence, i.e. the measured signal, information from the patient record, and personal interaction with the patient. Having these two independent opinions during diagnosis and treatment monitoring can help to improve safety, reliability, and quality of the decisions. Safety comes from the human interpretation of the algorithm results. The human expert has to decide whether or not the machine results make sense and act accordingly. This allows machine algorithms and human experts to work symbiotically on the sleep apnea detection problem. The machine algorithms provide real-time monitoring of patient data without risk of inter- and intra-observer variability. Furthermore, computer-based systems do not suffer from fatigue, and the results are reproducible. The decision model can be updated, which will improve the decision support over time. The human expert then becomes involved

only if apnea is detected. That will improve reliability and efficiency of the clinical process, because both machine algorithms and human experts will work according to their strength. Diligent machine work is then supervised with human creativity and intuition. Hence, accurate detection of sleep apnea with an LSTM network based on RR interval signals has the potential to become a key component for delivering appropriate diagnostic support and convenient uninterrupted treatment monitoring.

Acronyms

AUC	Area Under Curve
BMI	Body Mass Index
CAD	Computer-Aided Diagnosis
CNN	Convolutional Neural Network
CSA	Central Sleep Apnea
ECG	Electrocardiogram
EDR	ECG-Derived Respiration
EEG	Electroencephalogram
EMG	Electromyogram
EOG	Electrooculogram
FIR	Finite Impulse Response
GPU	Graphics Processing Unit
GRU	Gated Recurrent Units
HRV	Heart Rate Variability
IIR	Infinite Impulse Response
K-NN	K-Nearest Neighbor
LSTM	Long Short-Term Memory
OSA	Obstructive Sleep Apnea
PAP	Positive Airway Pressure
PPP	Palato Pharyngo Plasty
PSD	Power Spectral Density
PSG	Polysomnography
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine

References

[1] Oliver Faust, U Rajendra Acharya, EYK Ng, and Hamido Fujita. A review of ecg-based diagnosis support systems for obstructive sleep apnea. *Journal of Mechanics in Medicine and Biology*, 16(01):1640004, 2016.

[2] Terry Young, Mari Palta, Jerome Dempsey, James Skatrud, Steven Weber, and Safwan Badr. The occurrence of sleep-disordered breathing among middle-aged adults. *New England Journal of Medicine*, 328(17):1230–1235, 1993.

[3] Vishesh Kapur, Kingman P Strohl, Susan Redline, Conrad Iber, George O'Connor, and Javier Nieto. Underdiagnosis of sleep apnea syndrome in us communities. *Sleep and Breathing*, 6(02):049–054, 2002.

[4] Terry Young, Laurel Finn, Paul E Peppard, Mariana Szklo-Coxe, Diane Austin, F Javier Nieto, Robin Stubbs, and K Mae Hla. Sleep disordered breathing and mortality: eighteen-year follow-up of the wisconsin sleep cohort. *Sleep*, 31(8):1071–1078, 2008.

[5] Takatoshi Kasai, John S Floras, and T Douglas Bradley. Sleep apnea and cardiovascular disease: a bidirectional relationship. *Circulation*, 126(12):1495–1510, 2012.

[6] Paul E Peppard, Terry Young, Mari Palta, and James Skatrud. Prospective study of the association between sleep-disordered breathing and hypertension. *New England Journal of Medicine*, 342(19):1378–1384, 2000.

[7] H Klar Yaggi, John Concato, Walter N Kernan, Judith H Lichtman, Lawrence M Brass, and Vahid Mohsenin. Obstructive sleep apnea as a risk factor for stroke and death. *New England Journal of Medicine*, 353(19):2034–2041, 2005.

[8] Carmen M Schröder and Ruth O'Hara. Depression and obstructive sleep apnea (osa). *Annals of general psychiatry*, 4(1):13, 2005.

[9] Stijn Verhulst. Sleep-disordered breathing and sleep duration in childhood obesity. In *Pediatric obesity*, pages 241–252. Springer, 2010.

[10] Laiali Almazaydeh, Khaled Elleithy, and Miad Faezipour. Detection of obstructive sleep apnea through ecg signal features. In *2012 IEEE International Conference on Electro/Information Technology*, pages 1–6. IEEE, 2012.

[11] Cindy Den Herder, Joachim Schmeck, Dick JK Appelboom, and Nico de Vries. Risks of general anaesthesia in people with obstructive sleep apnoea. *Bmj*, 329(7472):955–959, 2004.

[12] R Doug McEvoy, Nick A Antic, Emma Heeley, Yuanming Luo, Qiong Ou, Xilong Zhang, Olga Mediano, Rui Chen, Luciano F Drager, and Zhihong Liu. Cpap for prevention of cardiovascular events in obstructive sleep apnea. *New England Journal of Medicine*, 375(10):919–931, 2016.

[13] Barbara Schmidt, Robin S Roberts, Peter J Anderson, Elizabeth V Asztalos, Lorrie Costantini, Peter G Davis, Deborah Dewey, Judy D'Ilario, Lex W Doyle, and Ruth E Grunau. Academic performance, motor function, and behavior 11 years after neonatal caffeine citrate therapy for apnea of prematurity: an 11-year follow-up of the cap randomized clinical trial. *JAMA pediatrics*, 171(6):564–572, 2017.

[14] Claudia E Korcarz, Ruth Benca, Jodi Barnet, and James Stein. Treatment of sleep apnea rapidly reduces arterial tone, improves endothelial function and left ventricular diastolic function in normotensive adults. *Journal of the American College of Cardiology*, 67(13 Supplement):1941, 2016.

[15] Ahnaf Rashik Hassan and Abdulhamit Subasi. A decision support system for automated identification of sleep stages from single-channel eeg signals. *Knowledge-Based Systems*, 128:115–124, 2017.

[16] Virend K Somers, David P White, Raouf Amin, William T Abraham, Fernando Costa, Antonio Culebras, Stephen Daniels, John S Floras, Carl E Hunt, and Lyle J Olson. Sleep apnea and cardiovascular disease: An american heart association/american college of cardiology foundation scientific statement from the american heart association council for high blood pressure research professional education committee, council on clinical cardiology, stroke council, and council on cardiovascular nursing in collaboration with the national heart, lung, and blood institute national center on sleep disorders research (national institutes of health). *Journal of the American College of Cardiology*, 52(8):686–717, 2008.

[17] David P White. Sleep apnea. *Proceedings of the American Thoracic Society*, 3(1):124–128, 2006.

[18] Rahul K Kakkar and Richard B Berry. Positive airway pressure treatment for obstructive sleep apnea. *Chest*, 132(3):1057–1072, 2007.

[19] Christian Guilleminault, Robert Riley, and Nelson Powell. Obstructive sleep apnea and abnormal cephalometric measurements: implications for treatment. *Chest*, 86(5):793–794, 1984.

[20] Thomas Penzel, J McNames, P De Chazal, B Raymond, A Murray, and G Moody. Systematic comparison of different algorithms for apnoea detection based on electrocardiogram recordings. *Medical and Biological Engineering and Computing*, 40(4):402–407, 2002.

[21] Kemal Polat, Şebnem Yosunkaya, and Salih Güneş. Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome. *Journal of Medical Systems*, 32(3):243–250, 2008.

[22] Sheikh Shanawaz Mostafa, Fábio Mendonça, Fernando Morgado-Dias, and Antonio Ravelo-García. Spo2 based sleep apnea detection using deep learning. In *2017 IEEE 21st international conference on intelligent engineering systems (INES)*, pages 000091–000096. IEEE, 2017.

[23] Regine Ragette, Yi Wang, Gerhard Weinreich, and Helmut Teschler. Diagnostic performance of single airflow channel recording (apnealink) in home diagnosis of sleep apnea. *Sleep and Breathing*, 14(2):109–114, 2010.

[24] Tom Van Steenkiste, Willemijn Groenendaal, Dirk Deschrijver, and Tom Dhaene. Automated sleep apnea detection in raw respiratory signals using long short-term memory neural networks. *IEEE journal of biomedical and health informatics*, 23(6):2354–2364, 2018.

[25] Thomas Penzel, George B Moody, Roger G Mark, Ary L Goldberger, and J Hermann Peter. The apnea-ecg database. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pages 255–258. IEEE, 2000.

[26] Ary L Goldberger, Luis AN 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. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *circulation*, 101(23):e215–e220, 2000.

[27] Joseph E Mietus, Chung-Kang Peng, P Ch Ivanov, and Ary Louis Goldberger. Detection of obstructive sleep apnea from cardiac interbeat interval time series. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pages 753–756. IEEE, 2000.

[28] Zvi Shinar, A Baharav, and S Akselrod. Obstructive sleep apnea detection based on electrocardiogram analysis. In *Computers in Cardiology 2000. Vol. 27 (Cat. 00CH37163)*, pages 757–760. IEEE, 2000.

[29] U Rajendra Acharya, Dhanjoo N Ghista, Zhu KuanYi, Lim Choo Min, EYK Ng, S Vinitha Sree, Oliver Faust, Liu Weidong, and APC Alvin. Integrated index for cardiac arrythmias diagnosis using entropies as features of heart rate variability signal. In *2011 1st Middle East Conference on Biomedical Engineering*, pages 371–374. IEEE, 2011.

[30] U Rajendra Acharya, Oliver Faust, Vinitha Sree, G Swapna, Roshan Joy Martis, Nahrizul Adib Kadri, and Jasjit S Suri. Linear and nonlinear analysis of normal and cad-affected heart rate signals. *Computer methods and programs in biomedicine*, 113(1):55–68, 2014.

[31] Philip De Chazal, Maria O'Dwyer, and Richard B Reilly. Automatic classification of heartbeats using ecg morphology and heartbeat interval features. *IEEE transactions on biomedical engineering*, 51(7):1196–1206, 2004.

[32] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, Oh Shu Lih, and U Rajendra Acharya. Deep learning for healthcare applications based on physiological signals: A review. *Computer methods and programs in biomedicine*, 161:1–13, 2018.

[33] U Rajendra Acharya, Hamido Fujita, Oh Shu Lih, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. Automated detection of arrhythmias using different intervals of tachycardia ecg segments with convolutional neural network. *Information sciences*, 405:81–90, 2017.

[34] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, Yuki Hagiwara, Jen Hong Tan, and Muhammad Adam. Application of deep convolutional neural network for automated detection of myocardial infarction using ecg signals. *Information Sciences*, 415:190–198, 2017.

[35] Jen Hong Tan, Yuki Hagiwara, Winnie Pang, Ivy Lim, Shu Lih Oh, Muhammad Adam, Ru San Tan, Ming Chen, and U Rajendra Acharya. Application of stacked convolutional and long short-term memory network for accurate identification of cad ecg signals. *Computers in biology and medicine*, 94:19–26, 2018.

[36] U Rajendra Acharya, Hamido Fujita, Shu Lih Oh, U Raghavendra, Jen Hong Tan, Muhammad Adam, Arkadiusz Gertych, and Yuki Hagiwara. Automated identification of shockable and non-shockable life-threatening ventricular arrhythmias using convolutional neural network. *Future Generation Computer Systems*, 79:952–959, 2018.

[37] Pranav Rajpurkar, Awni Y Hannun, Masoumeh Haghpanahi, Codie Bourn, and Andrew Y Ng. Cardiologist-level arrhythmia detection with convolutional neural networks. *arXiv preprint arXiv:1707.01836*, 2017.

[38] Debangshu Dey, Sayanti Chaudhuri, and Sugata Munshi. Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. *Biomedical engineering letters*, 8(1):95–100, 2018.

[39] Mashail Alsalamah, Saad Amin, and Vasile Palade. Detection of obstructive sleep apnea using deep neural network. In *Applications of Big Data Analytics*, pages 97–120. Springer, 2018.

[40] Erdenebayar Urtnasan, Jong-Uk Park, and Kyoung-Joung Lee. Multiclass classification of obstructive sleep apnea/hypopnea based on a convolutional neural network from a single-lead electrocardiogram. *Physiological measurement*, 39(6):065003, 2018.

[41] M Baumert, J Smith, P Catcheside, DR McEvoy, D Abbott, and E Nalivaiko. Changes in rr and qt intervals after spontaneous and respiratory arousal in patients with obstructive sleep apnea. In *2007 Computers in Cardiology*, pages 677–680. IEEE, 2007.

[42] K Dingli, T Assimakopoulos, PK Wraith, I Fietze, C Witt, and NJ Douglas. Spectral oscillations of rr intervals in sleep apnoea/hypopnoea syndrome patients. *European Respiratory Journal*, 22(6):943–950, 2003.

[43] Janet H Smith, Mathias Baumert, Eugene Nalivaiko, RONALD DOUGLAS McEVOY, and Peter G Catcheside. Arousal in obstructive sleep apnoea patients is associated with ecg rr and qt interval shortening and pr interval lengthening. *Journal of sleep research*, 18(2):188–195, 2009.

[44] Gari D Clifford, Francisco Azuaje, and Patrick Mcsharry. Ecg statistics, noise, artifacts, and missing data. *Advanced methods and tools for ECG data analysis*, 6:18, 2006.

[45] Pablo Laguna, George B Moody, and Roger G Mark. Power spectral density of unevenly sampled data by least-square analysis: performance and application to heart rate signals. *IEEE Transactions on Biomedical Engineering*, 45(6):698–715, 1998.

[46] Anthony C Fisher, Antonio Eleuteri, David Groves, and Christpher J Dewhurst. The ornstein–uhlenbeck third-order gaussian process (ougp) applied directly to the un-resampled heart rate variability (hrv) tachogram for detrending and low-pass filtering. *Medical & biological engineering & computing*, 50(7):737–742, 2012.

[47] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14(2), pages 1137–1145. Montreal, Canada, 1995.

[48] Changyue Song, Kaibo Liu, Xi Zhang, Lili Chen, and Xiaochen Xian. An obstructive sleep apnea detection approach using a discriminative hidden markov model from ecg signals. *IEEE Transactions on Biomedical Engineering*, 63(7):1532–1542, 2015.

[49] Ahnaf Rashik Hassan. Computer-aided obstructive sleep apnea detection using normal inverse gaussian parameters and adaptive boosting. *Biomedical Signal Processing and Control*, 29:22–30, 2016.

[50] Grégoire Surrel, Amir Aminifar, Francisco Rincón, Srinivasan Murali, and David Atienza. Online obstructive sleep apnea detection on medical wearable sensors. *IEEE transactions on biomedical circuits and systems*, 12(4):762–773, 2018.

[51] Lei Wang, Youfang Lin, and Jing Wang. A rr interval based automated apnea detection approach using residual network. *Computer methods and programs in biomedicine*, 176:93–104, 2019.

[52] Felix A Gers, Nicol N Schraudolph, and Jürgen Schmidhuber. Learning precise timing with lstm recurrent networks. *Journal of machine learning research*, 3(Aug):115–143, 2002.

[53] Felix A Gers, Jürgen A Schmidhuber, and Fred A Cummins. Learning to forget: Continual prediction with lstm. *Neural Computation*, 12(10):2451–2471, 2000.

[54] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.

[55] Stanislau Semeniuta, Aliaksei Severyn, and Erhardt Barth. Recurrent dropout without memory loss. *arXiv preprint arXiv:1603.05118*, 2016.

[56] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[57] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[58] Oliver Faust, Alex Shenfield, Murtadha Kareem, Tan Ru San, Hamido Fujita, and U Rajendra Acharya. Automated detection of atrial fibrillation using long short-term memory network with rr interval signals. *Computers in biology and medicine*, 102:327–335, 2018.

[59] François Chollet. Keras, 2015.

[60] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, and Michael Isard. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*, pages 265–283, 2016.

[61] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*, 2014.

[62] Rafal Jozefowicz, Wojciech Zaremba, and Ilya Sutskever. An empirical exploration of recurrent network architectures. In *International conference on machine learning*, pages 2342–2350, 2015.

[63] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.

[64] Carolina Varon, Alexander Caicedo, Dries Testelmans, Bertien Buyse, and Sabine Van Huffel. A novel algorithm for the automatic detection of sleep apnea from single-lead ecg. *IEEE Transactions on Biomedical Engineering*, 62(9):2269–2278, 2015.

[65] Martin O Mendez, Davide D Ruini, Omar P Villantieri, Matteo Matteucci, Thomas Penzel, Sergio Cerutti, and Anna M Bianchi. Detection of sleep apnea from surface ecg based on features extracted by an autoregressive model. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6105–6108. IEEE, 2007.

[66] RK Tripathy. Application of intrinsic band function technique for automated detection of sleep apnea using hrv and edr signals. *Biocybernetics and Biomedical Engineering*, 38(1):136–144, 2018.

[67] P Janbakhshi and MB Shamsollahi. Sleep apnea detection from single-lead ecg using features based on ecg-derived respiration (edr) signals. *Irbm*, 39(3):206–218, 2018.

[68] Zhao Dong, Xiang Li, and Wei Chen. Frequency network analysis of heart rate variability for obstructive apnea patient detection. *IEEE journal of biomedical and health informatics*, 22(6):1895–1905, 2017.

[69] Majdi Bsoul, Hlaing Minn, and Lakshman Tamil. Apnea medassist: real-time sleep apnea monitor using single-lead ecg. *IEEE Transactions on Information Technology in Biomedicine*, 15(3):416–427, 2010.

[70] Philip De Chazal, Conor Heneghan, Elaine Sheridan, Richard Reilly, Philip Nolan, and Mark O'Malley. Automated processing of the single-lead electrocardiogram for the detection of obstructive sleep apnoea. *IEEE Transactions on Biomedical Engineering*, 50(6):686–696, 2003.

[71] Jun Hu and Wendong Zheng. Multistage attention network for multivariate time series prediction. *Neurocomputing*, 383:122–137, 2020.

[72] U Rajendra Acharya, Eric Chern-Pin Chua, Oliver Faust, Teik-Cheng Lim, and Liang Feng Benjamin Lim. Automated detection of sleep apnea from electrocardiogram signals using nonlinear parameters. *Physiological measurement*, 32(3):287, 2011.