

Developing and comparing machine learning models to detect sleep apnoea using single-lead electrocardiogram (ECG) monitoring

Marcel Hedman^{1,2}, Alex Rojas¹, Anmol Arora³, and David Ola²

¹Harvard University, United States

²Nural Research, United Kingdom

³School of Clinical Medicine, University of Cambridge, United Kingdom

2021/04/18

Manuscript Details

Manuscript Type: Original research

Running head: Single-lead ECG sleep apnoea detection

Word Count: 3441 words

Number of Figures: 1

Number of Tables: 3

Number of References: 40

No financial sources to disclose.

No conflicts of interest.

Corresponding Author: Name: Marcel Hedman, Email: marcelh@nural.cc

Marcel Hedman ORCID: 0000-0003-4316-9618

Alex Rojas ORCID: 0000-0002-5792-6053

Anmol Arora ORCID: 0000-0003-4881-8293

David Ola ORCID: 0000-0002-6836-5298

Abstract

Background

Sleep apnoea has a high disease burden but remains underdiagnosed, in part due to the expensive and resource intensive nature of polysomnography, its definitive investigation. Emerging literature suggests that it may be possible to detect sleep apnoea using single-lead ECG signals, such as those obtained from smartwatches. In this study, we use two forms of recurrent neural networks (RNNs) to detect sleep apnoea events from single-lead ECG signals.

Methods

We use single-lead ECG data from the PhysioNet Apnea-ECG database, which contains data from 70 patients. We train a bidirectional gated recurrent unit (GRU) model and a bidirectional long short-term memory (LSTM) model on labelled ECG signals from 35 patients and test the models on the remaining 35 patients in the dataset.

Results

Both models achieved 97.1% accuracy, sensitivity and specificity to detect whether the ECG recordings belonged to a patient diagnosed with sleep apnoea. This corresponds to 34/35 patients in the dataset. At detecting individual apnoea events, the GRU and LSTM models achieved 90.4% and 91.7% accuracies respectively.

Discussion

The models achieved high levels of accuracy, specificity and sensitivity. Bidirectional RNNs are strengthened by the ability of the models to be informed by both past and future states when analysing sequential data, such as ECGs. The models also require minimal human intervention as they automatically extract features from the data. If single-lead ECGs prove a suitable tool for sleep apnoea detection, this may enhance the diagnosis of sleep apnoea and potentially allow widespread screening for the condition.

Conclusions

We note that using models such as bidirectional RNNs has the potential to augment model performance. However, more research and validation is required in order to test whether these may be applicable to other datasets and in clinical practice.

Key words: machine learning; artificial intelligence; sleep medicine; polysomnography; telemonitoring; remote patient monitoring; telemedicine

Introduction

Sleep Apnoea (SA) is characterised by intermittent interruptions in airflow during sleep [1]. Sleep apnoea can be categorised into three types: 1. central sleep apnoea, which can be triggered by the brain failing to activate breathing muscles, 2. obstructive sleep apnoea, the collapse of the airway during sleep, or 3. a combination of both [2]. The incidence of sleep apnoea has been estimated to be as high as 20% of the population and it has been suggested that approximately 90% of cases may be undiagnosed [3] [4].

While asleep, one apnoea event is defined as a >90% reduction in airflow lasting at least 10 seconds. A hypopnoea is a less severe reduction in airflow. The severity of sleep apnoea is graded by the number of apnoea and hypopnoea events per hour, the Apnoea Hypopnoea Index (AHI), with more than 5 events per hour being diagnostic [5]. The definitive method for diagnosing for sleep apnoea is polysomnography, which is resource intensive and complex to interpret. A polysomnogram consists of several simultaneous measurements: e.g. electroencephalogram (EEG), electrooculogram (EOG), electromyogram (EMG), electrocardiogram (ECG), pulse oximetry and nasal and oral airflow. In recent years, there has been development of home sleep apnoea tests which allow monitoring of patients at home rather than in a controlled clinical setting but these are still resource intensive and involve monitoring several parameters simultaneously. The limited availability of centres and clinicians able to diagnose sleep apnoea leads to long waiting lists for the patients seeking a diagnosis [6].

Electrocardiogram (ECG) is a test performed to monitor the electrical activity of the heart. ECGs can be performed with varying numbers of leads, typically ranging from one to twelve. In this study, we train machine learning models on data from single-lead ECG investigations to identify sleep apnoea. We use the Physionet Apnea-ECG Database which consists of single-lead ECG signals from 70 patients, 35 allocated for training a model and 35 allocated for its testing [7]. There have been efforts to analyse single-lead ECG signals previously, which have yielded respectable levels of accuracy, specificity and sensitivity. The machine learning models previously tested include feature engineering and feature learning methods. Feature engineering models require processing of raw data to extract features which may be useful in analysing unseen data. Feature learning models are designed to learn the features themselves, reducing the need for pre-processing. Examples of feature learning methods which have been applied to the Physionet database include Singh and Majumder's Pre-trained AlexNet CNN + Decision Fusion model and Wang et al's LeNet-5 CNN model [2] [8]. Examples of feature engineering based methods include: Sharma and Sharma's Feature Engineering + LS-SVM and Song et al's Feature Engineering + HMM-SVM [9] [10].

In this study, we develop novel models using recurrent neural network (RNN) variants such as gated recurrent unit (GRU) and long short-term memory (LSTM). The models we prepared are feature learning models with some pre-processing of data to identify R-R peaks and amplitudes from the ECG signals. We use the PhysioNet Apnea-ECG dataset pre-processing scheme provided by Wang et al [2]. If the analysis of single-lead ECG is proved to be an accurate method of diagnosing sleep apnoea, there is potential to simplify and broaden the screening for sleep apnoea diagnosis given that single-lead ECG data are being made increasingly accessible through 21st century smart watches.

Methods

Sleep apnoea dataset

In this study, we use the PhysioNet Apnea-ECG dataset provided by Philipps University [7]. The dataset consists of 70 records, divided into a learning set and a testing set of 35 records each. The ECG data were sampled at 100Hz with durations ranging between 401 and 587 min. Data are labelled once per minute in each ECG signal, indicating whether or not an apnoea event was in progress at the beginning of that minute. Both apnoea and hypopnoea events were simply labelled as apnoea events. The dataset does include other signals, including chest and abdominal respiratory effort, oronasal airflow and oxygen saturation, however in this study only ECG signals were used for analysis.

Pre-processing of data

We processed our data in line with the method outlined by Wang et al [2]. This involved extracting the R-R intervals and amplitudes from single-lead ECG using the Hamilton algorithm plus the positions and values of calculated R-peaks.

An ECG tracks the electrical activity of the heart, with each beat forming a QRS complex. The QRS complex incorporates the Q, R, and S waves, which represents ventricular depolarisation [11]. R-peaks are crucial for analysing this QRS complex [12], and using the aforementioned pre-processing scheme we can extract RR-intervals and amplitudes. This implementation takes into account the adjacent two segments of each labelled segment because adjacent segment information has been show to be helpful for per-segment sleep apnoea detection [2].

Machine Learning Models

The input of the models is the R-R intervals and amplitudes from the single-lead ECG signal and the output is whether or not the input patient has sleep apnoea. It is extremely important to attain models with high accuracy that are not overparametrized [13]. In conjunction with our goal of deployment, we seek to maximize accuracy while constraining model size during training such that these models can be efficient.

Model 1: LSTM with Convolution

The model architecture involves both convolution and recurrent neural network properties; the input is first passed to a single convolution layer filter size of 5 and stride 2, and the activation map is then passed to max-pooling which subsamples by a factor of 3. Next, we repeat twice bidirectional layers with 64 LSTM nodes, followed by dropout with $p = .8$ and batch normalization. Finally, we pass through fully connected layers with ReLU activation and finally an output layer with softmax activation. This model thus has 760,258 parameters and was trained with an Adam optimizer with a learning rate of $5 \cdot 10^{-3}$, with a learning rate scheduler that decreased the learning rate after 70 epochs by $1 \cdot 10^{-1}$ every 10 epochs. In total, we trained for 100 epochs and built the model with Tensorflow.

Model 2: GRU with Convolution

This model is almost identical to Model 1, implementing recurrence on top of convolution. Yet, we replaced the LSTM nodes with GRU nodes. This model was trained identically to Model 1, and slightly cuts

down on the number of parameters with 723,906.

Results

We evaluated the performance of our models at both a per minute (segment) apnoea detection level as well as on a per patient (recording) level. This is in line with the traditional method of determining whether a person is diagnosed with sleep apnoea based on the frequency of apnoea events. We use five metrics to evaluate the machine learning model:

- **Accuracy:** Accuracy measures the percentage of predicted labels that matches the ground truth labels as defined by the creators of the dataset.
- **Specificity:** Specificity measures the true negative rate, this is the percentage of true negative predictions over the total number of predictions of the negative class.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Negatives}}$$

- **Sensitivity:** Measures the true positive rate, which is the percentage of all positive predictions which are true positives.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- **Area under receiver operating characteristic (ROC) curve (AUC):** The ROC curve shows the balance of true positive rate to false positive rate as the threshold boundary for classification to a binary class is changed. The area under it reaches a maximum of 1.
- **Correlation:** The correlation measures the strength of the relationship between two variables and was used to compare the predicted AHI index to the actual index. This is in line with previous studies [2] [9].

$$\text{Correlation} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Per Segment

Table 1 contains the results across our approaches and it can be seen that both the LSTM and GRU models outperform a previously constructed modified LeNet5 model which we use as a benchmark [2].

Model	Test Accuracy (%)	Test specificity (%)	Test sensitivity (%)	Test AUC
Bidirectional LSTM	91.7	91.7	86.9	0.959
Bidirectional GRU	90.4	92.5	87.0	0.959
Modified LeNet5 [2]	87.6	90.3	83.1	0.950

Table 1: Per segment results of the bidirectional RNN models

Figure 1 illustrates the receiver operating characteristic (ROC) curve for the high-performing bidirectional GRU model. The AUC is calculated as 0.959, which represents outstanding model performance [14].

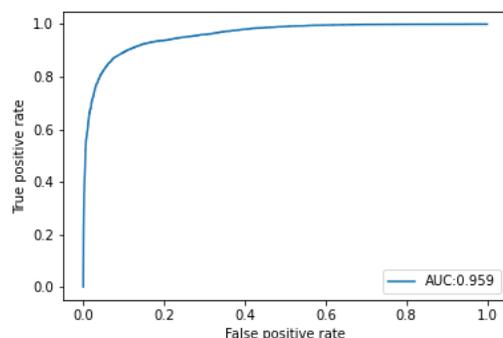


Figure 1: Bidirectional GRU ROC curve

Per Patient

Given the ability to classify individual apnoea events in sleep, a classification then needs to be made on a per patient level. To perform this classification, we used the AHI score. The AHI score gives the number of apnoea and hypoapnoea events per hour and scores less than 5 in an asymptomatic patient are considered normal [15]. Within this dataset, there was an intermediate class for borderline cases and when this dataset was used in a competition, it was advised that these borderline cases need not be used for testing the model [16]. However, in order to ensure that our models correspond with currently accepted clinical practice, we used a simple cut-off of an AHI of 5 when determining the diagnosis. Both models accurately classified 34/35 patients as having sleep apnoea or not, based on the AHI definition. The one patient which the algorithms incorrectly classified was one of the aforementioned borderline cases.

$$AHI = \text{number of apnoea and hypopnoea events} * \frac{60}{T}$$

Table 2 demonstrates that the LSTM and GRU models had high performance in determining whether a recording belonged to someone diagnosed with sleep apnoea and this is to be expected given the high performance on the per segment predictions. The models, therefore, may be potentially highly effective at screening sleep apnoea.

Model	Test Accuracy (%)	Test specificity (%)	Test sensitivity (%)	Corr
Bidirectional LSTM	97.1	91.7	100.0	0.966
Bidirectional GRU	97.1	100.0	95.7	0.969
Modified LeNet5 [2]	97.1	91.7	100.0	0.943

Table 2: Per recording results

Discussion

Summary of Key Findings

In this study, we developed and tested two models to detect sleep apnoea using the Physionet Apnea-ECG dataset. We found that both the bidirectional GRU and LSTM models achieved 97.1% accuracy, with both high sensitivity and specificity, in predicting whether a patient would have sleep apnoea.

Comparison to Existing Literature

Even though sleep apnoea is usually a respiratory disease, it is well recognised that it is associated with cardiac rhythm disturbances which may be detected by ECG [17] [18]. The use of ECG alone is not routinely used to diagnose sleep apnoea. However, the use of a two-lead ECG does form part of the polysomnogram which is the current gold-standard method of diagnosing sleep apnoea. The ECG is recognised to include a range of information which is becoming increasingly interpretable with the advent of machine learning analysis [19].

Previous studies attempting to diagnose sleep apnoea with single-lead ECGs or other parameters such as SpO₂ have used several different types of models, including deep vanilla neural networks (DVNNs), artificial neural networks (ANNs), convolutional neural networks (CNNs) and recurrent neural networks (RNNs) [6] [20]. Table 3 compares the results of models generated in this study to models produced in several selected preceding studies. A systematic review in 2019 has provided a more thorough analysis of previous attempts to diagnose sleep apnoea using single-lead ECGs, with the models in this study achieving higher levels of accuracy than most of those identified in the review [6]. However, it should be noted that this high accuracy in a small dataset may not necessarily translate to being the most clinically useful.

Comparison of Machine Learning Models

Artificial Neural Networks

Artificial neural networks (ANN) are computational models designed as a simplified model of the human brain, consisting of layers of nodes and non-linear activation that can model complex functions. Multiple layers form a deep network, and these models train by minimizing a loss function through backpropagation and gradient-descent [21]. ANNs have been demonstrated to predict heartbeat classification tasks with high degrees of accuracy for R-R intervals of ECG recordings [22]. However, these models become overfit when trained on small datasets, failing to generalize well. This overfitting problem is often addressed by a technique known as regularisation which stops any individual weight in the network from growing too large [23]. Within a neural network, this is achieved through dropout which randomly sets nodes to zero in training.

Convolutional Neural Networks

Standard ANNs suffer when applied to certain tasks including image recognition, as they fail to be translation-invariant and contain many parameters to be trained. A convolutional neural network (CNN) addresses these challenges. The CNN was first introduced as the time delay neural network, creating a translation-invariant network trained with backpropagation for the first time [24]. Originally taking inspiration from receptive fields created by neurons in the visual cortex, these networks are most common

for image classification, but can be applied more broadly for feature extraction. CNNs build upon the ANN by introducing convolution, a process by which trainable filters convolve over an input and output an activated feature map, which has the major advantage of reducing the number of trainable parameters for a similar sized network. A common activation function introduced to CNN by AlexNet, a CNN trained on the ImageNet database, is the non-linear ReLU [25]. After the convolutional layers, the activation maps are flattened and passed into a standard ANN. Max-pooling is a method for downsampling the activation maps between certain layers designed to reduce dimensionality by selecting the maximally activated convolved node and discarding the rest.

Recurrent Neural Networks

One weakness of the ANN and CNN is that they can not process data that has inherent sequential meaning in a way that retains memory. The recurrent neural network (RNN) addresses this problem by passing through in each forward pass a hidden state h_t , which taken with the input allows the network to retain memory about the past, often for sequential tasks such as processing language. This is updated by passing the current state and hidden state through an activation function after an affine transformation. However, these networks are subject to the vanishing gradient problem, and to address this, two architectures exist: the Gated Recurrent Unit (GRU) and the Long-Short Term Memory Unit (LSTM) networks.

One earlier attempt to address this is the LSTM, which used additional affine transformations and element-wise selection on the hidden state and inputs to update the states and activated state [26]. The LSTM hidden state is computed where the output gate o_t , an activated affine transformation of the input, previous hidden state, and c_t moderates how much of the activated c_t to pass through. We have that c_t is also computed as an activated affine transformation of the input and previous hidden state with a forget gate and input gate. We can introduce bidirectionality, which allows the network to learn from both the future inputs and past inputs [27]. LSTMs have been shown to perform on ECG data generally and this is to be expected due to its memory ability, it was therefore a natural extension to apply it to sleep apnoea detection using single-lead ECG [28]. A simplification of the LSTM came in the form of the GRU [29].

The two bidirectional models we produced have the advantage that they can learn from both past (backwards) and future (forwards) states, which is a useful characteristic when analysing data with long-term dependencies between time steps in sequential data [30]. RNNs are ideally suited to time-series data such as ECGs, as applied in this study. The models used in this study are classified as 'feature learning', such that they require no manual feature engineering. In our case, human processing the data was limited to pre-processing of the data to extract R-R peaks and amplitudes.

Author	Dataset	Model	Accuracy (%)	Specificity (%)	Sensitivity (%)	Correlation
Our study	Apnea-ECG	Bidirectional LSTM	97.1	91.7	100.0	0.966
Our study	Apnea-ECG	Bidirectional GRU	97.1	100.0	95.7	0.969
Wang et al (2019) [2]	Apnea-ECG	Modified LeNet5	97.1	91.7	100.0	0.943
Li et al (2018) [31]	Apnea-ECG	Decision Fusion	100.0	100.0	100.0	-
Sharma and Sharma (2016) [9]	Apnea-ECG	LS-SVM	97.1	100.0	95.8	0.841
Song et al (2016) [10]	Apnea-ECG	HMM-SVM	97.1	100.0	95.8	0.860
Chang et al (2020) [32]	Apnea-ECG	1D deep CNN	97.1	100.0	95.7	0.865

Table 3: Per recording performance comparison of our models with selected models using the same Physionet Apnea-ECG dataset

Limitations

In this dataset, hypopnoea events were labelled as apnoea events which means that our models are unable to distinguish between the two [33]. Furthermore, the dataset only contains apnoeas which are obstructive or mixed, with no episodes of purely central apnoea. Future iterations of these models should involve larger datasets which are labelled for both apnoea and hypopnoea events, although both are used when calculating AHI. The small sample size used in this study dictates that further research is required to validate these findings. Only one dataset was used, in which the same ECG techniques were applied. In order for these findings to have implications for future practice, the machine learning system must be interoperable. Without testing a more diverse dataset, we are unable to comment on whether our model would be effective if applied to novel datasets or data taken from other ECG monitoring devices. The purpose of this study was to highlight potential models which offer most potential for future avenues of research. Such future avenues should involve external validation and randomised trials conforming to SPIRIT-AI and CONSORT-AI guidelines in order to truly determine whether the algorithms may be suitable for clinical practice [34] [35].

It should be noted that in clinical practice, sleep apnoea diagnosis does not solely rely on AHI but also on the clinical presentation and symptoms experienced by the patient [16]. Sleep studies are to be used in combination with other clinical tools, for example the Epworth Sleepiness Scale [36].

In recent years, the development of machine learning algorithms has led to the issue of 'black-box' algorithms, whereby it is difficult to understand how the outputs are derived. In this case, the machine learning models were able to detect whether a single-lead ECG signal indicated sleep apnoea, which is a difficult task for even the most skilled ECG readers. In current clinical practice, diagnoses of sleep apnoea are derived from polysomnograms, which is both an expensive and resource intensive method. Conversely, it is also difficult to monitor sleep apnoea at home, given that the resources required are typically unavailable outside a secondary care setting.

Implications of Findings

In recent years, there has been an exponential increase in the adoption of wearable devices, including Apple watches, Oura rings and other smartwatches. This has led to the use of such devices to assist in remote patient monitoring. Certain novel watches are capable of providing ECG monitoring as well as measuring breathing rate in some cases. The use of ECG watches to predict, diagnose and monitor sleep apnoea seems to be within the realms of possibility in the context of a growing body of literature suggesting that it might be possible to diagnose using only a single-lead ECG. There remain a number of uncertainties regarding how such practice would be implemented, including whether the analysis of ECG leads would occur on the device, such as in edge computing, or centrally. The issue of accountability has been raised in the literature as a key barrier to adoption of AI systems, since there is a plausible risk that the systems would have imperfect accuracy.

The potential use of AI in clinical practice is accompanied by discussion of practical and ethical issues that may surround its implementation, including the need for interpretability of algorithms [37]. Although the models were able to detect sleep apnoea with a high degree of accuracy, it is often unclear how this is achieved. Other ethical arguments relate to the interaction between the AI system and the clinicians providing care to the patient since there is a need to supervise the models, which is difficult if the operator has a limited understanding of how the models are operating [38].

There are a number of obstacles before such technologies may be used in practice. Yoo et al highlight three characteristics of a transformative innovation which are relevant in this instance: interoperability, homogeneity of data and self-referential nature [39] [40]. In order for algorithms to be useful in clinical practice, it is important to ensure that they are interoperable, in such a way that they are generalisable to different sources of input data e.g. different devices collecting ECG signals in different individuals. At the very least, there needs to be an understanding of any limitations of the generalisability of the model. There should also be effort spent trying to homogenise data between different devices, or data sources, in order to make achieving interoperability a realistic prospect. Lastly, the utility of an algorithm would increase substantially if it could be used to improve the quality of or generate other algorithms. For example, a system developed to interpret ECG signals in such a way that it can detect sleep apnoea may further be applied to detect other conditions, such as through the implementation of techniques such as transfer learning.

Unfortunately, there is a notable shortage of single-lead ECG datasets labelled for sleep apnoea events. In order to develop the findings of this study and to improve the models to a standard which would be acceptable for routine care, there is a need to develop larger datasets with more diverse populations. A number of studies assessing the utility of machine learning algorithms to diagnose sleep apnoea have used the same datasets, limiting the interoperability and generalisability of findings. This study as well as previous studies are important in proving that there is potential for machine learning models to detect sleep apnoea using only a single-lead ECG signal, however further research is required using larger datasets. The accretion of datasets for developing models is complicated by the fact that it is difficult for human clinicians to label single-lead ECG signals with sleep apnoea diagnoses without clinical context. Developing large scale datasets, which includes patients known to have sleep apnoea remains a priority for advancing research in the field.

Conclusions

This study adds to the growing body of literature suggesting that, in the future, single-lead ECGs may offer potential to diagnose sleep apnoea using only a smartwatch. We propose two novel machine learning models applied to ECG analysis to detect sleep apnoea. Both the bidirectional GRU and the bidirectional LSTM models outperformed most models in existing literature. They highlight the potentially important role of recurrent neural networks when developing future models. Further research is required to confirm these findings using larger datasets and data obtained using smartwatches.

Data

All data are freely accesible on Physionet [7]: <https://physionet.org/content/apnea-ecg/1.0.0/>

References

- [1] Dempsey JA, Veasey SC, Morgan BJ, O'Donnell CP. Pathophysiology of sleep apnea [published correction appears in *Physiol Rev*.2010 Apr;90(2):797-8]. *Physiol Rev*. 2010;90(1):47-112. doi:10.1152/physrev.00043.2008
- [2] Wang, T, Lu, C, Shen, G, Hong, F. Sleep apnea detection from a single-lead ECG signal with automatic feature-extraction through a modified LeNet-5 convolutional neural network. *PeerJ*. 2019;7, e7731. <https://doi.org/10.7717/peerj.7731>
- [3] Kevin J. Finkel, Adam C. Searleman, Heidi Tymkew, Christopher Y. Tanaka, Leif Saager, Erika Safer-Zadeh, Michael Bottros, Jacqueline A. Selvidge, Eric Jacobsohn, Debra Pulley, Stephen Duntley, Colleen Becker, Michael S. Avidan, Prevalence of undiagnosed obstructive sleep apnea among adult surgical patients in an academic medical center. *Sleep Med*. 2009;10(7):753-8. doi: 10.1016/j.sleep.2008.08.007. Epub 2009 Jan 30.
- [4] Young T, Evans L, Finn L, Palta M. Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. *Sleep*. 1997 Sep;20(9):705-6. doi: 10.1093/sleep/20.9.705. PMID: 9406321.
- [5] David T.R. Berry, Wilse B. Webb, A.Jay Block, Sleep Apnea Syndrome: A Critical Review of the Apnea Index as a Diagnostic Criterion. *Chest*. 1984 Oct;86(4):529-31. doi: 10.1378/chest.86.4.529.
- [6] Mostafa SS, Mendonça F, Ravelo-García AG, Morgado-Dias F. A Systematic Review of Detecting Sleep Apnea Using Deep Learning. *Sensors (Basel)*. 2019;19(22):4934. Published 2019 Nov 12. doi:10.3390/s19224934
- [7] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., ... Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation [Online]*. 101 (23), pp. e215–e220., <https://physionet.org/content/apnea-ecg/1.0.0/>
- [8] Singh SA, Majumder S. A novel approach osa detection using single-lead ECG scalogram based on deep neural network. *Journal of Mechanics in Medicine and Biology*. 2019 Jun 13;19(04):1950026. <https://doi.org/10.1142/S021951941950026X>
- [9] Hemant Sharma, K.K. Sharma, An algorithm for sleep apnea detection from single-lead ECG using Hermite basis functions. *Comput Biol Med*. 2016 Oct 1;77:116-24. doi: 10.1016/j.combiomed.2016.08.012. Epub 2016 Aug 13.
- [10] C. Song, K. Liu, X. Zhang, L. Chen and X. Xian, "An Obstructive Sleep Apnea Detection Approach Using a Discriminative Hidden Markov Model From ECG Signals," in *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1532-1542, July 2016, doi: 10.1109/TBME.2015.2498199
- [11] Sattar Y, Chhabra L. Electrocardiogram. [Updated 2020 Nov 26]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2021 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK549803/>

- [12] Kaur, A., Agarwal, A., Agarwal, R. et al. A Novel Approach to ECG R-Peak Detection. Arab J Sci Eng 2019;44, 6679–6691. <https://doi.org/10.1007/s13369-018-3557-8>
- [13] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International Conference on Machine Learning. PMLR, 2019.
- [14] Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. J Thorac Oncol. 2010 Sep;5(9):1315-6. doi: 10.1097/JTO.0b013e3181ec173d.
- [15] Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, Harrod CG. Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea: An American Academy of Sleep Medicine Clinical Practice Guideline. J Clin Sleep Med. 2017 Mar 15;13(3):479-504. doi: 10.5664/jcsm.6506.
- [16] Detecting and Quantifying Apnea Based on the ECG - the PhysioNet Computing in Cardiology Challenge 2000, Accessed 12th April 2021, <https://physionet.org/content/challenge-2000/1.0.0/>
- [17] Effects of obstructive sleep apnoea on heart rhythm Valentina A. Rossi, John R. Stradling, Malcolm Kohler European Respiratory Journal 2013;41: 1439-1451; DOI: 10.1183/09031936.00128412
- [18] Alonso-Fernández A, García-Río F, Racionero MA, et al. Cardiac rhythm disturbances and ST-segment depression episodes in patients with obstructive sleep apnea-hypopnea syndrome and its mechanisms. Chest. 2005;127(1):15-22. doi:10.1378/chest.127.1.15
- [19] Topol EJ. What's lurking in your electrocardiogram? Lancet. 2021;397(10276):785. doi: 10.1016/S0140-6736(21)00452-9.
- [20] Alvarez, D., Cerezo-Hernandez, A., Lopez-Muniz, G., Alvaro-De Castro, T., Ruiz-Albi, T., Hornero, R., del Campo, F. (2017). Usefulness of Artificial Neural Networks in the Diagnosis and Treatment of Sleep Apnea-Hypopnea Syndrome. In Sleep Apnea - Recent Updates. InTech. <https://doi.org/10.5772/66570>
- [21] Keiron O'Shea, Ryan Nash, "An Introduction to Convolutional Neural Networks". arXiv:1511.08458. <https://arxiv.org/abs/1511.08458>
- [22] Jiang W, Kong SG. Block-based neural networks for personalized ECG signal classification. IEEE Trans Neural Netw. 2007;18(6):1750-1761. doi:10.1109/TNN.2007.900239
- [23] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov.. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 2014;15, 1, 1929–1958. <https://dl.acm.org/doi/10.5555/2627435.2670313>
- [24] Kevin J. Lang, Alex H. Waibel, Geoffrey E. Hinton, A time-delay neural network architecture for isolated word recognition, Neural Networks. Neural Netw. 1990; 3, 23–43. DOI: 10.1016/0893-6080(90)90044-L
- [25] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E, "ImageNet Classification with Deep Convolutional Neural Networks, <https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf>.2012.

- [26] Hochreiter, Sepp, and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*. 1997;9(8): 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [27] Alex Graves, Jürgen Schmidhuber, Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Networks*. 2005;18(5–6): 602–610. <https://doi.org/10.1016/j.neunet.2005.06.042>.
- [28] Vykintas Maknickas, Algirdas Maknickas, Atrial Fibrillation Classification Using QRS Complex Features and LSTM. *Computing in Cardiology*. 2017; VOL 44. DOI:10.22489/CinC.2017.350-114.
- [29] Cho, Kyunghyun, et al. "Learning phrase representations using RNN encoder-decoder for statistical machine translation." arXiv preprint. 2014. arXiv:1406.1078.
- [30] ElMoquet H, Eid M, Glos M, Ryalat M, Penzel T. Deep Recurrent Neural Networks for Automatic Detection of Sleep Apnea from Single Channel Respiration Signals. *Sensors (Basel)*. 2020;20(18):5037. doi:10.3390/s20185037
- [31] Li et al. (2018) Li K, Pan W, Li Y, Jiang Q, Liu G. A method to detect sleep apnea based on deep neural network and hidden markov model using single-lead ECG signal. *Neurocomputing*. 2018;294:94–101. doi: 10.1016/j.neucom.2018.03.011.
- [32] Chang HY, Yeh CY, Lee CT, Lin CC. A Sleep Apnea Detection System Based on a One-Dimensional Deep Convolution Neural Network Model Using Single-Lead Electrocardiogram. *Sensors (Basel)*. 2020;20(15):4157. Published 2020 Jul 26. doi:10.3390/s20154157
- [33] George Moody, Roger Mark. Detecting and Quantifying Apnea Based on the ECG - the PhysioNet Computing in Cardiology Challenge 2000. Accessed: 12th April 2021. Published: Feb. 10, 2000. Version: 1.0.0. Available from: <https://physionet.org/content/challenge-2000/1.0.0/>
- [34] Liu, X, Cruz Rivera, S, Moher, D et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med* 2020;26, 1364–1374. <https://doi.org/10.1038/s41591-020-1034-x>
- [35] Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers. *Korean J Radiol*. 2019;20(3):405-410. doi:10.3348/kjr.2019.0025
- [36] NICE. Obstructive sleep apnoea syndrome. Accessed 12th April 2021, Last Revised: March 2021, <https://cks.nice.org.uk/topics/obstructive-sleep-apnoea-syndrome/diagnosis/assessment/>
- [37] Liu, Xiaoxuan et al., "A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis", *The Lancet Digital Health*. 2019;1(6): e271 - e297, DOI:[https://doi.org/10.1016/S2589-7500\(19\)30123-2](https://doi.org/10.1016/S2589-7500(19)30123-2)
- [38] Arora A. Shooting from the hip into our own foot? A perspective on how artificial intelligence may disrupt medical training. *Future healthcare journal*. 2020;7(2), e7–e8. <https://doi.org/10.7861/fhj.Let-7-2-4>

- [39] Yoo, Youngjin, et al. “Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research.” *Information Systems Research*, vol. 21, no. 4, 2010, pp. 724–735. JSTOR, www.jstor.org/stable/23015640. Accessed 31 Mar. 2021.
- [40] Arora A. Conceptualising Artificial Intelligence as a Digital Healthcare Innovation: An Introductory Review. *Med Devices (Auckl)*. 2020;13:223-230 <https://doi.org/10.2147/MDER.S262590>