MorphNet: A 1D-CNN Approach for Obstructive Sleep Apnea Detection using SpO2 Signals

Peter Yang
University of Michigan
Department of Mathematics
yangpe@umich.edu

Abstract

Sleep apnea is a serious sleep disorder in which an individual experiences multiple obstructive apnea events, in which the throat muscles block airflow and prevents breathing of more than eleven seconds. In recent years, Machine Learning methods using biometric signals have proven to be useful in detecting obstructive apnea events, potentially reducing the cost of diagnosis, and reducing the strain on our healthcare system. This paper introduces MorphNet, a 1D-CNN deep learning method for per-minute sleep apnea detection using blood oxygen saturation (SpO2) signal time series data from the Apnea-ECG dataset. The proposed method uses one-dimensional Convolutional Neural Networks in order to extract features from minute intervals of SpO2 data. The model created exhibits an accuracy of 97.47% with a precision of 98.92%, recall of 97.84%, and f1 score of 0.9818. The proposed model is compared against several other state of the art documented machine learning algorithms to classify obstructive sleep apnea.

1. Introduction

Obstructive sleep apnea is a serious sleep disorder that afflicts roughly 936 million people worldwide [1]. It is characterized by episodes of stopped periods of breathing, and has symptoms of increased fatigue, daytime drowsiness, and increased levels of brain fog. In recent years, Machine Learning techniques have been seen as a solution to solve the problem of classifying obstructive sleep apnea. Past approaches use more traditional machine learning algorithms, such as Support Vector Machines and Logistic Regression [7], but in recent years breakthroughs in computing power and network architectures has led to neural network models proving to be very successful in classifying obstructive sleep apnea events. Specifically, breakthroughs in Convolutional Neural Networks and Recurrent Networks like Long-Short-Term-Memory Neural Networks have proven to be effective in capturing the temporal relationships in signals useful to detect sleep apnea, such as electrocardiogram (ECG) signals, or blood oxygen saturation levels (SpO2) [4] [3]. While obstructive sleep apnea has the potential to be a serious condition, most of the definitive symptoms occur during sleep, which makes it hard for individuals to self-diagnose sleep apnea. Definite diagnosis can be done, but usually through an expensive (300-1000$) [9] and laborous polysomnography with a physician. This problem motivates the creation of MorphNet, a lightweight 1D-CNN model used to detect obstructive sleep apnea on a one minute basis using SpO2 signals, designed to be usable in most IoT devices, most notably smartwatches.

In recent years, most smartwatches, including Apple, Samsung, and FitBit smartwatches, allow for the recording and monitoring of SpO2 levels. Recording SpO2 levels in smartwatches is non-invasive and requires minimal effort, which motivates the use of SpO2 as our variable of interest for classifying obstructive sleep apnea. Furthermore, SpO2 levels are useful in detecting reduced airflow through a drop in blood oxygen levels, which is potentially useful for detecting sleep apnea events. Other models use multiple signals in detecting obstructive sleep apnea, especially heart signals through electrocardiogram (ECG) tests [2], and the detection of physical movements in bed [6]. However, many approaches have found success solely using SpO2 to classify sleep apnea [5], which led to our decision to focus exclusively on using SpO2 as our variable of interest.

2. Methods

2.1. Database

The data was collected from the Apnea-ECG Database [8]. It contains 70 overnight biometric recordings of 35 individuals using gold-standard polysomnography techniques. The database primarily focuses on ECG recordings, with ground truth apnea labels available for the test set of 35 overnight biometric readings. It was

Individual	Age(yrs)	Sex	Height (cm)	Weight (kg)
a01	51	M	175	102
a02	38	M	180	120
a03	54	M	168	80
a04	52	M	173	121
b01	44	F	170	63
c01	31	M	179	74
c02	37	M	180	83
c03	39	M	184	65

Table 1. Demographic data of the 8 individuals

collected in 2000, and contributed by Dr. Thomas Penzel of Phillips-University in Marburg, Germany. Because this database was primarily focused on ECG data, of the 35 biometric recordings, 8 of them include SpO2 data. This subset of 8 individuals constitute the dataset used in this report. Demographic details of the individuals are provided in Tab. 1. The recordings are sampled at 100 Hz, and broken up into one minute intervals, which are labelled as apneic or non-apneic, depending on whether or not an apnea has occurred in the interval. An example of the two types of intervals are shown in Fig. 1. In our preprocessing, the data was downsampled to a lower frequency of 10hz in order to reduce the volume of data, and potential noise in the measurements. This was done by splitting the data in chunks of 10 samples and calculating the median to represent the value, so that each datapoint was of length 600, representing a minute of SpO2 levels, with 10 observations per second. Any minute interval containing a value of under 50% SpO2 were assumed to be measurement errors, considered artifacts, and dropped from the dataset. In total, we had 3751 minutes, and therefore datapoints of SpO2 data across the 8 individuals, with 2281 being labels without apnea, and 1470 being labels with apnea.

Because our dataset was relatively balanced, as the number of apnea intervals is 39.19% of our dataset, we did not use any class weights on our loss function or probability threshholding when training and evaluating our model.

2.2. MorphNet

MorphNet is a novel 1D-CNN model which uses convolution filters in order to extract temporal relationships in the SpO2 data to detect drastic fluctuations or drops, which are useful in classifying apneas. The model takes in a minute of SpO2 data as input of shape (1x600), which is first normalized, which helps increase the training speed. After, the normalized inputs are fed into two sequential convolutional blocks, each containing a convolutional layer, a ReLU layer, and a max pooling layer. The convolutional layer outputs are of shape (1, 16) and (16, 32), with kernel sizes of 200 and 100, in each convolutional block respectively. Each

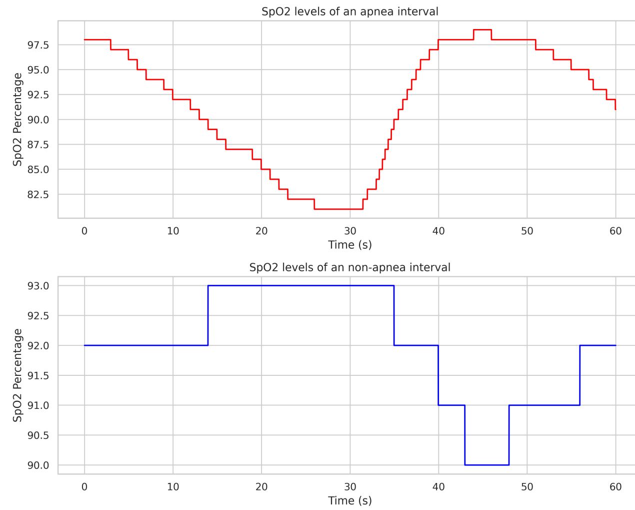

Figure 1. Example SpO2 levels for an apnea and non-apnea interval. Note that the apnea interval is much more erratic and fluctuates significantly more than a non-apnea interval

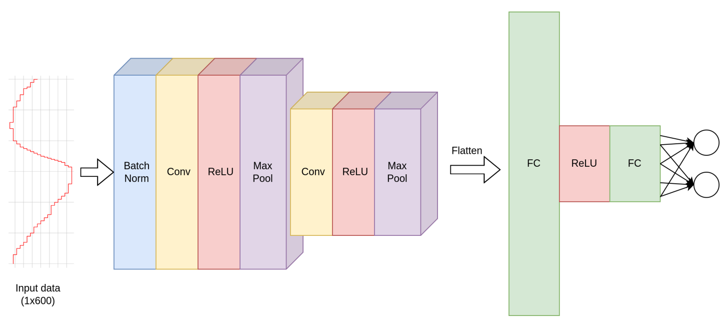

Figure 2. The proposed MorphNet architecture

convolutional layer is of stride 1, with zero padding. The max pooling layers both have a kernel size and stride length of 2. After the two convolutional blocks, the output is flattened, and sent through two fully connected layers of size 4800 and 1200 respectively, before finally being sent to 2 output nodes with a sigmoid activation function to perform binary classification. The full model architecture can be seen in Fig. 2.

2.3. Training

The data was partitioned into a train-test split of 80/20. The model was trained on a Intel(R) Core(TM) i7-10700 CPU, 16 GB of RAM, and a NVIDIA RTX 3070 GPU with 8 GB of memory. The model was trained for 25 epochs, with a batch size of 32, and a learning rate of 5e-6. The loss function used was Binary-Cross Entropy loss. All code for the project was written in Python using Pytorch. Further details about the hyperparameters and model architecture can be seen in Tab. 3 and Tab. 2, respectively.

Layer	Parameters
Input	-
BatchNorm	-
Convolution	in=1, out=16, kernel=200x1
ReLU	-
Max Pool	kernel_size=2, stride=2
Convolution	in=16, out=32, kernel=100x1
ReLU	-
Max Pool	kernel_size=2, stride=2
Flatten	-
Dropout	p=0.2
FC1	in_features=4800, out_features=1200
ReLU	-
Dropout	p=0.2
FC2	in_features=1200, out_features=1
Sigmoid	-
Output	-

Table 2. Detailed description of the MorphNet architecture

Hyperparameter	Value
Train/Test Split	80/20
Optimizer	Adam
Learning Rate	5e-6
Loss function	Binary Cross Entropy
L2 Regularization	1e-4
Epochs	25
Batch Size	32
Dropout Prob in last FC layer	0.2
Activation Function	ReLU

Table 3. Hyperparameters used in training MorphNet

3. Results

Our model is trained to perform binary classification given a minute long interval of SpO2 time series data. MorphNet achieved an accuracy of 97.47% on the test set, with a precision of 98.92%, recall of 97.84%, and F1 score of 0.9818. The resulting confusion matrix and ROC curve made from running our model on the test data can be seen in Fig. 3, and Fig. 4, respectively. A comparison between MorphNet and many state-of-the art models, using a variety of different databases and machine learning methods (LSTM, SVM, 2D-CNN), can be seen in Tab. 4. As seen from Tab. 4, our proposed method performs better compared to many state of the art models. Although direct comparison between models is hard, as most papers use differing datasets, MorphNet notably performs better than the model from Dey *et al.* [2], which uses the same dataset, and a similar method. Furthermore, while other implementations use more complex network architecture such as 2D-CNN and recurrent models like LSTMs, as well as many different variables, and such as ECG alongside SpO2, or uses movement or other biometric data, our model solely uses SpO2 data and a simple 1D-CNN architecture.

Despite this, it is comparable or outperforms more complex models that incorporate many variables, while requiring significantly less data and computational resources. This makes our model lightweight, while still maintaining a good accuracy, making our model ideal for deployment into IoT devices. Furthermore, our choice of variable of SpO2 makes our model exceptionally interpretable, and able to be integrated to be used in supporting professional diagnosis.

Author	Method	Dataset	Accuracy	Recall
Eid *et al.* [3]	LSTM	Charité Apnea	85.10%	90.00%
Ma *et al.* [7]	SVM	UCD Apnea	90.20%	87.60%
Wang *et al.* [11]	CNN	UCD Apnea	71.80%	26.60%
John *et al.* [4]	1D-CNN	UCD Apnea	**99.56%**	96.05%
Dey *et al.* [2]	2D-CNN	Apnea-ECG	94.33%	93.88%
Ours	1D-CNN	Apnea-ECG	97.47%	**97.84%**
Ours (pruned)	1D-CNN	Apnea-ECG	96.14%	94.47%

$$Precision = \frac{TP}{TP + FP} = 98.92\% \quad (1)$$

$$Recall = \frac{TP}{TP + FN} = 97.84\% \quad (2)$$

$$F1\ Score = 2\frac{Precision \times Recall}{Precision + Recall} = 0.9818 \quad (3)$$

Table 4. Comparison of MorphNet to other state-of-the-art models

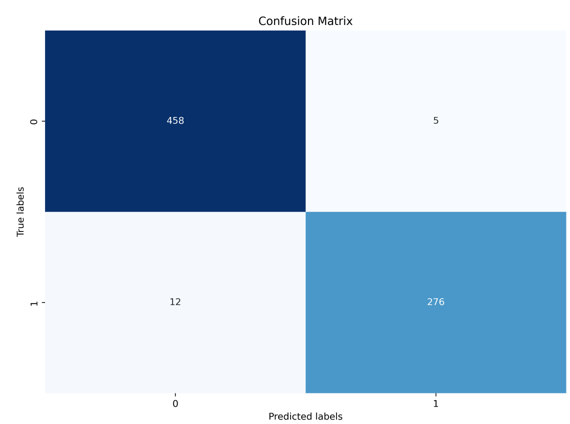

Figure 3. Confusion matrix of MorphNet on the Testing set

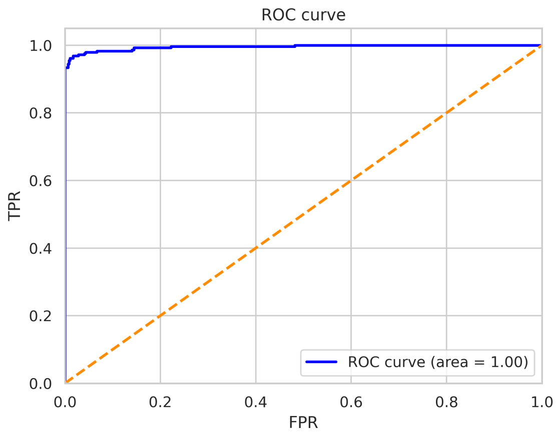

Figure 4. ROC curve of MorphNet on the Testing set

3.1. Model Pruning and Quantization

Our original model can be stored using 24.5 MB. Storage in IoT devices is limited, so it is crucial that we optimize the size of our model to make it fit within the confined storage of IoT devices. We do this with quantization and model pruning, both techniques are used to reduce the size of the model by removing or rounding model weights. Quantization is the process of converting the weights in our model to lower resolution storages. For example, converting the weight datatype from 64-bit floats into 8-bit integers. Model Pruning, on the other hand, is the process of removing a certain percentage of weights from certain layers. In order to reduce the size of our model, we incorporate both of these techniques. Using pruning, we remove 30% of the weights in the last two fully connected layers of model, and then convert the weights in the last two fully connected layers to 8-bit integers. The results of the pruned model on the test data is shown in Tab. 4. Doing this pruning reduces the size of our model substantially, reducing our model size to 6.4 MB, making it much more suitable for IoT devices with storage restraints.

4. Conclusion

In this paper, we introduce a method of per-minute detection of obstructive sleep apnea events using SpO2 levels, compatible for IoT devices. Our proposed method exhibits an accuracy of 97.47%, with a precision of 98.92%, Recall of 97.84%, and F1 Score of 0.9818, with the sparse version of the model performing similarly. Given the low storage space of the pruned model, deployment of this model to IoT sensors like smartwatches is feasible. While databases that include obstructive apnea data that includes SpO2 levels are limited, work in the future could be done on exploring the efficacy of MorphNet on different polysomnography databases, such as the University College Dublin Sleep Apnea Database. Future work could also be done in exploring attention based neural network architectures, such as transformers [10]. More work could also be done implementing the real-time monitoring system outlined in this paper into an IoT device capable of extracting real-time SpO2 data, such as an Apple, Galaxy, or Fitbit smartwatch.

References

[1] AV Benjafield, NT Ayas, PR Eastwood, R Heinzer, MSM Ip, MJ Morrell, CM Nunez, SR Patel, T Penzel, JL Pépin, PE Peppard, S Sinha, S Tufik, K Valentine, and A Malhotra. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis, 2019. Epub 2019 Jul 9. PMID: 31300334; PMCID: PMC7007763. 1

[2] D. Dey, S. Chaudhuri, and S. Munshi. Obstructive sleep apnoea detection using convolutional neural network based deep learning framework. *Biomedical Engineering Letters*, 8(1):95–100, 2017. 1, 3

[3] H. ElMoaqet, M. Eid, M. Glos, M. Ryalat, and T. Penzel. Deep recurrent neural networks for automatic detection of sleep apnea from single channel respiration signals. *Sensors*, 20(18):5037, 2020. 1, 3

[4] Arlene John, Barry Cardiff, and Deepu John. A 1d-cnn based deep learning technique for sleep apnea detection in iot sensors. *CoRR*, abs/2105.00528, 2021. 1, 3

[5] A. John, K.K. Nundy, B. Cardiff, and D. John. Somnnet: An spo2 based deep learning network for sleep apnea detection in smartwatches. pages 1961–1964, Nov 2021. 1

[6] J.L. Kelly, R. Ben Messaoud, M. Joyeux-Faure, R. Terrail, R. Tamisier, J.B. Martinot, N.N. Le-Dong, M.J. Morrell, and J.L. Pépin. Diagnosis of sleep apnoea using a mandibular monitor and machine learning analysis: One-night agreement compared to in-home polysomnography. *Frontiers in Neuroscience*, 16:726880, 2022. 1

[7] B. Ma, Z. Wu, S. Li, et al. Development of a support vector machine learning and smart phone internet of things-based architecture for real-time sleep apnea diagnosis. *BMC Medical Informatics and Decision Making*, 20(Suppl 14):298, 2020. 1, 3

[8] Thomas Penzel, G.B. Moody, R.G. Mark, Ary Goldberger, and J.H. Peter. The apnea-ecg database. *Computers in Cardiology*, 27:255 – 258, 02 2000. 1

[9] S. A. Stewart, E. Penz, M. Fenton, and R. Skomro. Investigating cost implications of incorporating level iii at-home testing into a polysomnography based sleep medicine program using administrative data. *Canadian Respiratory Journal*, 2017:8939461, 2017. 1

[10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. 4

[11] T. Wang, C. Lu, G. Shen, and F. Hong. Sleep apnea detection from a single-lead ecg signal with automatic feature-extraction through a modified lenet-5 convolutional neural network. *PeerJ*, 7:e7731, 2019. 3