Design Studies and Intelligence Engineering
L.C. Jain et al. (Eds.)
© 2023 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/FAIA220703

Deep Transfer Learning Approach for Obstructive Sleep Apnea Classification with Photoplethysmography Signal

E SMILY JEYA JOTHI[a,1] J ANITHA[b1], and D JUDE HEMANTH [c]
[a]*Assistant Professor, Department of Biomedical Instrumentation Engineering, Avinashilingam Institute for Home Science and Higher Education for Women*
[b,c]*Professor, Department of Electronics and Communication, Karunya Institute of Technology and Sciences*

Abstract. Human health and quality of life are negatively impacted by apnea, an increasingly prevalent sleep disorder. For monitoring and managing sleep apnea's side effects and consequences, accurate automatic algorithms for detecting sleep apnea are crucial. In this paper, deep transfer learning methods are employed for the detection of OSA events from Electrocardiograph (ECG) and Photoplethysmography (PPG) signals. ResNet34 is a deep learning model based on convolutional neural networks (CNNs). Transfer learning algorithms such as AlexNet, VGG16, VGG19 and ResNet50 are implemented. In order to train the ResNet34 model data augmentation, optimal learning rate finding, and fine-tuning are used. To obtain generalizable models, a training set of data is divided into three sets: a validation set for adjusting hyperparameters and improving generalizability, and a test set for evaluating generalizability on unknown data. Deep transfer learning models have the best accuracy, sensitivity, specificity, precision, and F1 score with 97.86±1.24%, 99.65%, 97.12%, 98.16% and 98.90% respectively. It can assist sleep lab technicians in screening patients for OSA events continuously through PPG and ECG signals.

Keywords.OSA classification, deep transfer learning, CNN, apneic events.

1. Introduction

A sleep disorder in which the upper airway partially or completely collapses is known as obstructive sleep apnea (OSA). Research has demonstrated that OSA impairs nocturnal sleep quality and causes fatigue during the day. OSA has also been linked to cardiovascular and cerebrovascular disease, according to research [1]. Manual visual annotation of PSG data is typically used to score sleep stages, including electrographic measures of brain activity as well as eye and chin movements. An OSA patient can be diagnosed using PSG, which measures apnea-hypopnea index (AHI) [5]. The severity of OSA based on AHI index can be categorized as none (AHI<5), mild (5≤AHI<15), moderate (15≤AHI≤30), and severe (AHI≥30). An automated method of ECG analysis could offer a potential solution to the problem of ECG abnormalities during overnight sleep studies using Artificial Intelligence (AI) and machine learning (ML). Data indicate that AI is capable of classifying sleep apnea more accurately than physicians in categorizing ECG and PPG signals [2].

[1]Dr. D Jude Hemanth

OSA diagnosis costs can be reduced through electrocardiogram (ECG) measurements due to their simplicity. A disadvantage of ECG-based OSA events detection is the complexity of the procedure, which requires qualified practitioners. Detecting subtle changes in signal patterns can be made possible by computer algorithms. Oxygen levels in the blood are reduced because OSA causes insufficient air to enter the lungs. Sleep is frequently interrupted when patients lack oxygen in their brains, which leads to frequent waking. Patients with cardiovascular disease are more likely to suffer from OSA and engage in more cardiovascular risk factors. These include hypertension, coronary heart disease, arrhythmia, heart failure, and stroke [5]. Observed Polysomnography (PSG) is the standard method of diagnosing sleep apnea. Training technicians record several physiological signals and analyze them during overnight hospitalizations in a sleep lab or sleep centre. In addition to being very expensive, time-consuming, and limited in number, this approach involves a lot of paperwork [6].

Saving resources and improving efficiency are two of the main benefits of transfer learning. The use of neural networks to automatically learn features has been proposed in several recent studies. Sparse auto-encoding was performed to encode the unlabelled RR interval from a heartbeat signal. Fine tuning the classifier was accomplished through discrimination in the learners' representations. A back propagation network was optimized using the Newton method. A classifier was then trained to extract the parameter of a Hidden Markov Model. Both the parameters and probability outputs were sent into a Hidden Markov Model to estimate the probability. In the end, different classifiers were combined using the decision fusion method [2]. The 1-min ECG signal was pre-processed with a Butterworth Band-pass filter of fourth order with a frequency of 0.5 Hz to 15 Hz. A 1D CNN was used to identify normal and apneic events from the pre-processed ECG signal. Tensor flow and Keras were used to implement the proposed method. An ECG signal of one minute was used to extract features using 10 identical feature layers. In order to feed the classifier, a two-dimensional feature vector was converted into a one-dimensional feature vector. Based on 1D feature vectors, four identical layers were designed to classify sleep apnea [3]. First, the ECG signals were pre-processed and segmented. Tests were conducted on a variety of deep learning and machine learning methods. There was 92.27% specificity, a sensitivity of 84.26 percent, and an accuracy of 88.13% for the hybrid deep model [4].

Using body sounds, an algorithm was developed to extract the heart rate. Using the developed sleep monitor in conjunction with the PSG tests, ten subjects were tested for validation. Apnea detection with the developed instrumentation was 92.8% sensitive and 99.7% specific. The study did not report any significant signal losses [6]. An ECG signal was input to the CNN along with RR interval data and R peak data. A 60-second segment of non-overlapping segments was created for the ECG signal in the pre-processing stage. Noise was reduced using an FIR band pass filter [8]. Using the Hamilton algorithm, the R-peak in the ECG signal was identified, and uninterpretable points were eliminated using a median filter. Per-segment and per-recording classifications were accurate by 87.9% and 97.1%, respectively [9]. OSA events were classified based on ECG scalograms and spectrograms. Continuous Wavelet Transform and Short Time Fourier Transform were used to transform the ECG segments from the time domain to the frequency domain.

Predicting accuracy was compared between three different transfer learning models and the proposed CNN model. ECG scalogram values were correctly predicted with 82.30% accuracy, sensitivity 82.27%, and specificity 82.95%, and ECG spectrogram values were correctly predicted with 80.13% accuracy [15]. Based on 60 second segment OSA classification, the network achieved 86.22% accuracy and 90% sensitivity. Compared to ResNet, Chen et al. proposed an upgraded version of pretrained network ResNet that improved accuracy by nearly 1% [16].

Accordingly, previous published work used pre-apnea segments of the brain for feature extraction, and then used these features as inputs for traditional machine learning algorithms for predicting apnea. We evaluated AlexNet, ResNet50, VGG16 and VGG19 networks in terms of their performance and comparative analysis in identifying OSA segments based on PPG and ECG signals. Further, we demonstrated that a hybrid deep learning architecture outperformed baseline experiments. PPG and ECG recordings are found to be sufficient to predict OSA events, which makes the proposed study novel compared to other prediction studies. This paper continues with a discussion of the methodology in section 2, followed by a discussion and interpretation of the results. The paper is concluded in section 3.

2. Methodology

2.1. Datasets

The deep transfer learning model on OSA detection is trained and evaluated using the PhysioNet Apnea-ECG database from Philipps University, Marburg, Germany. The Apnea-ECG dataset includes 70 single-channel ECG recordings, which range from 401 to 587 minutes in length and are sampled at 100 Hz. There are 35 recordings in total, divided into two groups of 35 recordings each: available recordings (a01 through a20, b01 through b05, and c01 through c10) and withheld recordings (x01 through x35). A one-minute segment of each recording was separated by non-overlapping breaks. There are several files attached to each subject [10].

The Beth Israel Deaconess Medical Centre Patient PPG and Respiration Dataset is publicly available in Physionet [11]. Data were collected from hospitalized critically ill patients undergoing treatment at the Beth Israel Deaconess Medical Centre in Boston, Massachusetts, USA. For each recording, two annotators manually annotated the impedance respiratory signal. Each recording lasts 8 minutes and includes the following information:

- Electrocardiograms, pulse oximetry, and impedance respiration signals are examples of signs that represent bodily functions. Samples at 125 Hz are used.
- There are three physiological parameters: heart rate, respiration rate, and blood oxygen saturation level. One sample is taken every second for each of these parameters
- As well as variable variables like age and gender, fixed variables can be recorded simultaneously.
- Manual annotations can be added to breaths

2.2. Transfer Learning

A CNN model can be built without creating a new one from scratch with transfer learning. By utilizing and optimizing the knowledge of previously acquired data, transfer learning provides training for CNN applications with small training datasets [15]. The study investigates the efficiency of transfer learning on the entire network using AlexNet, ResNet50, VGG16 and VGG19 models. The architecture of these networks is shown in figure 1. In order to optimize weights and ensure appropriate learning, learning rate is an influential hyperparameter. In addition, the mini-batch size determines how many samples should be used during training in each epoch. A single epoch was determined to be 0.0001, 64 epochs were determined, and a maximum number of epochs was determined to be 30 epochs.

2.2.1. AlexNet

A total of eight layers are present in the AlexNet, each with a learnable parameter. In each of the five layers except the output layer, they use ReLU activation in combination with max pooling followed by three fully connected layers [4]. The layers included are shown in table 1.

Table 1. Layer specification of AlexNet

Layer	Filter Size	Activation function
Input	-	-
Conv 1	11 x 11	ReLU
Max Pooling 1	3 x 3	-
Conv 2	5 x 5	ReLU
Max Pooling 2	3 x 3	-
Conv 3	3 x 3	ReLU
Conv 4	3 x 3	ReLU
Conv 5	3 x 3	ReLU
Max pooling 3	3 x 3	-
Dropout	-	-
Fully Connected	2	

2.2.2. VGG16

With a dimension of 3 x 1, the VGG16 is composed of two convolutional layers. In the next stage, there is a max pooling layer of 2 x 1, followed by two convolutional layers with 128 kernels of size 3 x 1 stacked together. Finally, a convolutional layer with 512 kernels of size 3 x 1 is included. Lastly, a pooling layer with 2 x 1 is included in the fourth layer. Two layers consisting of 512 and 2 nodes flatten the data before it is fed into two fully connected layers [4].

2.2.3. ResNet50

As part of its Artificial Intelligence research efforts, Google introduced the ResNet50, a convolutional neural network with 50 layers and 7 million parameters. It is trained with the help of the Physionet database. ResNet50 differs from other architectures by having nine layers of inception. An organized convolutional layer consists of convolutional layers of 1x1, and 3x3 filters activated by ReLU functions, each surrounded by a pooling layer. The output of every inception module concatenates the acquired feature maps. In addition to reducing the number of parameters in the network, ResNet50 also dramatically reduces its complexity. A global average pooling is applied before the output layer, instead of a fully connected layer in traditional CNNs [15, 17].

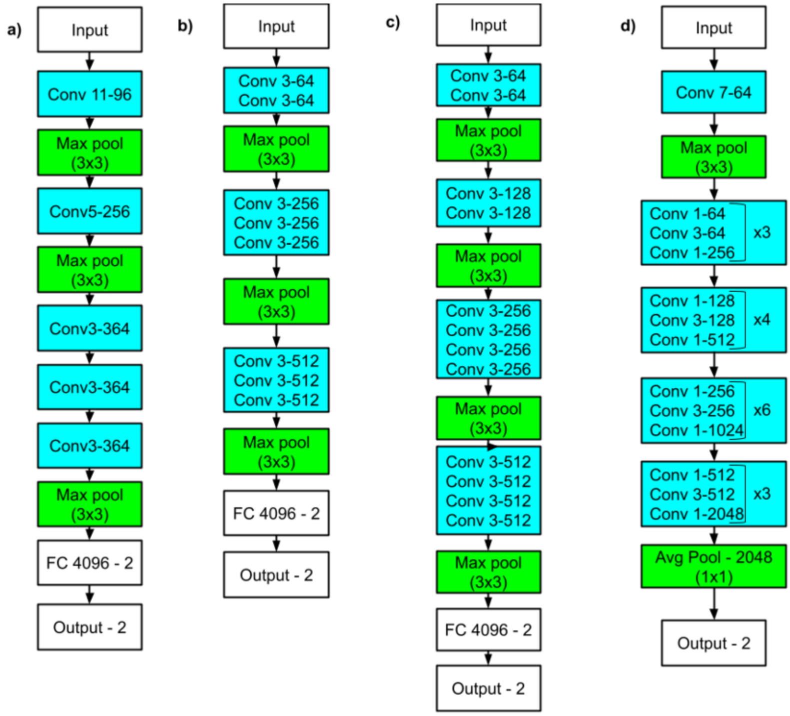

Figure 1. Architecture of a) AlexNet, b) VGG16, c) VGG19, and d) ResNet50

2.2.4. VGG19

Table 2. Layer description of VGG19

Layers	Kernel	Size
2 Convolutional	64	3 x 1
Max-Pooling	-	2 x 1
2 Convolutional	128	3 x 1
Max-Pooling	-	2 x 1
4 Convolutional	256	3 x 1
Max-Pooling	-	2 x 1
4 Convolutional	512	3 x 1
Max-Pooling	-	2 x 1
4 Convolutional	512	3 x 1
Fully Connected	2	

Table 2 shows the layer description of VGG19 [4].

2.3. Deep CNN ResNet34

CNN ResNet34 architecture [12] is used in this study to classify OSA events via transfer learning. An existing trained model is taken and learned to solve a similar classification problem instead of creating one from scratch. Based on the Physionet database, the ResNet34 architecture is trained. Deep learning tasks are commonly performed using this architecture. Convolutional neural networks are state-of-the-art neural networks in computer vision that construct new models. Compared to other pre-trained models such as Reception and VGG, the ResNet34 architecture converges faster, and there is only one parameter to adjust. Comparing the architecture to other pre-trained models like Inception and VGG, it is very simple to apply to different datasets. The gradient approach cannot be used to update the weights of the first layers as the number of layers increases. It transmits input data in a way that maintains gradients and prevents information loss. When insufficient data is available to train a model, transfer learning can provide significant benefits.

Convolutional neural networks consist of convolutional, pooling, and fully connected layers. There are alternated convolutional and pooling layers in the ResNet34 architecture. Convolutional base is a term used to describe this part of the model. Rather than the ResNet34 model's dense layer, a new dense layer produces a binary vector representing normal and abnormal classes. On top of the convolutional base, the newly added dense layer is trained. In addition, a SoftMax activation function is used in a final layer that returns logs of predictions instead of probabilities. There are 33 layers in the CNN convolutional base. These layers have a filter size of 3*3 and a stride of 2 in the complete model. The hidden layers are activated using the rectified linear unit (ReLU). Table 3 shows the number of parameters for each layer.

Table 3. Number of parameters for each layer

Layer	Input size	Output size	Number of parameters
ResNet-34	3*128*128	64*64*64	9408
Dense	512	2	1026
SoftMax	2	2	0

2.3.1. Data Augmentation

It is generally the case that training a model with only a few data points leads to over fitting during training. Using the validation set, the model generalizes but does not preserve the details of the training set. This over fitting problem is mitigated by employing the data augmentation technique during training. An approach to augmenting existing training data is known as data augmentation. As a result, a novel signal set for the training is created using data augmentation methods. Virtualizing data through data augmentation techniques such as zooming, fusing, rotating, etc., is an easy way to have more data.

2.3.2. Optimal Learning Rate Finder

Model parameters are updated according to the gradient based on a hyper parameter called learning rate. It takes a long time and produces tiny changes in the model weights if the learning rate is set too low. The optimizer will overshoot the minimum if the learning rate is too high, and it may even diverge if the learning rate is too low. A model's performance is greatly affected by it [13].

2.3.3. Stochastic Gradient descent with restarts (SGDR)

The learning rate during training is moderately reduced by stochastic gradient descent with restarts. With this approach, closer desired parameters can be identified, so optimizations can be made with fewer changes to model parameters [14].

2.3.4. Fine-Tuning

Pre-trained models are fine-tuned by lightly adjusting their weights. ResNet34 is retrained with a new dense layer, as well as fine-tuned some parameters in the convolutional base. A block diagram showing the proposed method for OSA classification is shown in Figure 2. Raw data from PPG and ECG recordings is retrieved from the Physionet database. As part of the data augmentation process, the sample size is increased and the data is fed into the classifier. Having convolutional base layers, ResNet34 is proposed as a deep transfer technique. Classifiers are tuned to achieve the optimal learning rate. A classifier outputs two classes, OSA and Non-OSA, according to the input signal.

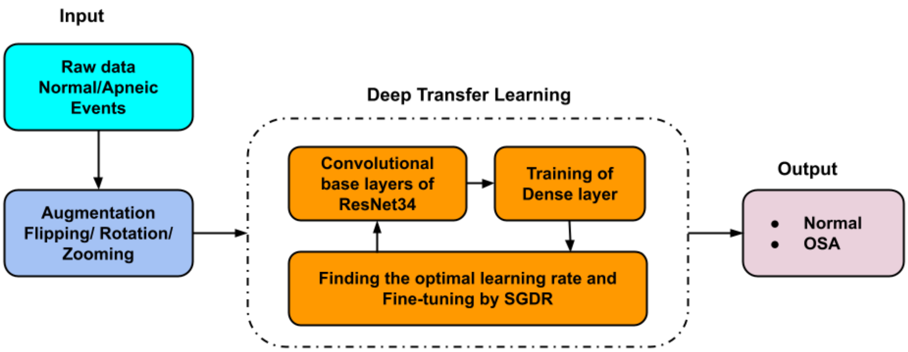

Figure 2. Overview of the Proposed System

3. Discussions

The comparison between the proposed method and existing models can be found in table 4. The pre-processing and the transfer learning techniques are implemented using MATLAB R2020b environment. In order to optimize the parameters, a stochastic gradient descent algorithm was used with a learning rate of 0.0001 in order to minimize the cross-entropy loss function. After multiplying the learning rate and the bias of the last fully connected layer, a constant is calculated. A set of measures, including accuracy, sensitivity, specificity, precision, and F1 score, were calculated to evaluate the accuracy, sensitivity, specificity, and precision of all models. The mathematical expression can be formulated as in equation [1]-[5]

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (1)$$

$$Sensitivity = \frac{TP}{TP+FN} * 100 \quad (2)$$

$$Specificity = \frac{TN}{TN+FP} * 100 \quad (3)$$

$$Precision = \frac{TP}{TP+FP} * 100 \quad (4)$$

$$F1\ Score = \frac{TP}{TP+\frac{1}{2}(FP+FN)} * 100 \quad (5)$$

A proposal for a prediction model for OSA detection and prediction is shown in Table 4. This model achieves the highest accuracy of 97.86%, as well as sensitivity of 99.65%, specificity of 97.12%, precision of 98.16% and F1 score of 98.90%. Other pre-trained networks and existing methods are also summarized in the table. Rather than improving OSA detection success, we sought to validate prediction performance by comparing the proposed model with detection results on AlexNet, ResNet50, VGG16, and VGG19. All models achieved the highest detection performance. For prediction experiments, detection results were considered a reliable indicator. Figure 3 represents the graphical form of the performance of transfer learning and deep transfer learning model.

Table 4. Comparison with the existing models

Reference	Classifier	Signal	Acc (%)	Sens (%)	Spec (%)	Precision (%)	F1 Score (%)
Li, K et al., [2]	HMM, DNN, Decision Fusion	Single lead ECG	85	88.9	-	-	-
Bahrami, M., &Forouzanfar, M. [3]	Deep CNN	ECG	87.9	81.1	92.0	-	-
Bahrami, M., &Forouzanfar, M. [4]	Hybrid deep models	ECG	88.13	84.26	92.27	-	-
Gao, Q et al., [9]	1D CNN	ECG	87.9	83.8	91.4	-	-
Kalkbrenner, C et al., [6]	-	PPG	-	92.8	99.7	-	-
Radha, M et al., [7]	LSTM	PPG and ECG	76.36±7.57	-	-	-	-
Nasifoglu, H., &Eroğul, O [15]	CNN	ECG scalograms	82.30	83.22	82.27	82.95	

		ECG Spectrogram	80.13	81.99	77.87	-	-
Proposed model	**CNN ResNet34**	PPG and ECG	**97.86±0.24**	**99.65**	**97.12**	**98.16**	**98.90**
	AlexNet		88.09±0.89	82.64	89.87	88.12	83.10
	VGG16		87.12±0.36	82.06	90.50	90.78	83.13
	VGG19		86.75±0.38	81.44	90.07	87.15	82.52
	ResNet50		88.13±0.40	81.49	92.27	91.65	84.04

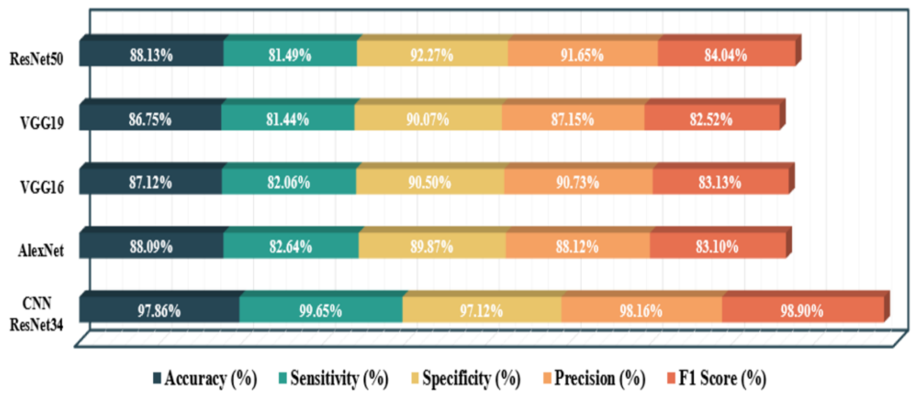

Figure 3. Performance of transfer learning and deep transfer learning model

There are some limitations to this study. The study excludes hypopnea, central sleep apnea, and mixed sleep apnea as other types of sleep apnea. By having a larger set of training and testing subjects, it would have been possible to evaluate models more accurately. Because deep neural networks can automatically extract the most important features, deep neural networks are superior to conventional machine learning. In summary, the proposed approach has the following advantages:

- According to our knowledge, this is the first study to use deep transfer learning to identify OSA events.
- A deep model that has proven its performance can be rapidly adapted through deep transfer learning instead of creating a new model. As a result, long validation processes are avoided as well as complicated layer parameters.
- This study examines as many signals as possible.

4. Conclusion

An ECG and PPG signal classification method using deep transfer learning is presented in this study. CNN-ResNet34 architecture is proposed for the automatic detection of OSA events. It is comparable or better than existing deep learning methods as well as traditional machine learning methods. In order to evaluate the model, 70 ECG records and 53 PPG recordings are taken from the Physionet dataset. As an alternative to manual scoring, this method can provide sleep technicians with an alternative way to annotate SA events, and can promote clinical and household applications of SA detection methods based on ECG and PPG signals.

References

[1] Gahungu, N., Shariar, A., Playford, D., Judkins, C., &Gabbay, E. Transfer learning artificial intelligence for automated detection of atrial fibrillation in patients undergoing evaluation for suspected obstructive sleep apnoea: a feasibility study. *Sleep medicine, 85*, 166-171. 2021

[2] 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, 294, 2018*

[3] Bahrami, M., &Forouzanfar, M.. Detection of Sleep Apnea from Single-Lead ECG: Comparison of Deep Learning Algorithms. *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 1-5. 2021

[4] Bahrami, M., &Forouzanfar, M. Sleep Apnea Detection From Single-Lead ECG: A Comprehensive Analysis of Machine Learning and Deep Learning Algorithms. *IEEE Transactions on Instrumentation and Measurement, 71*, 1-11. 2022

[5] Chang, H., Yeh, C., Lee, C., & Lin, C. A Sleep Apnea Detection System Based on a One-Dimensional Deep Convolution Neural Network Model Using Single-Lead Electrocardiogram. *Sensors (Basel, Switzerland), 20*. 2020

[6] Kalkbrenner, C., Eichenlaub, M.M., Rüdiger, S., Kropf-Sanchen, C., Rottbauer, W., &Brucher, R. Apnea and heart rate detection from tracheal body sounds for the diagnosis of sleep-related breathing disorders. Medical & Biological Engineering & Computing, 56, 671-681. 2017

[7] Radha, M., Fonseca, P., Moreau, A. *et al.* A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. *npj Digit. Med.* **4,** 135 2021.

[8] Ramachandran, A., Pant, A., &Karuppiah, A Performance Analysis of Machine Learning Algorithms for Sleep Apnea Detection Using ECG. *Lecture Notes in Networks and Systems*. 2021

[9] Gao, Q., Shang, L., & Zhang, Y. Automated Detection of Sleep Apnea Using Convolutional Neural Network from a single-channel ECG signal. Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence. 2020

[10] Penzel, T., Moody, G. B., Mark, R. G., Goldberge, A. L., Peter, J. H. 2000. The apnea-ECG data-base. Computers in Cardiology 27, 255-258 (2000).

[11] Goldberger, A., Amaral, L., Glass, L., Hausdorff, J., Ivanov, P. C., Mark, R., J. E. Mietus, G. B. Moody, C. K. Peng, and & Stanley, H. E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation [Online]. 101 (23), pp. e215–e220. 2000

[12] He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778). 2016

[13] Smith, L. N. Cyclical learning rates for training neural networks. In Applications of Computer Vision (WACV), 2017 IEEE Winter Conference on (pp. 464-472). IEEE, 2017

[14] Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J. E., & Weinberger, K. Q. Snapshot ensembles: Train 1, get M for free. arXiv preprint arXiv:1704.00109, 2017

[15] Nasifoglu, H., &Eroğul, O. Convolutional Neural Networks based OSA Event Prediction from ECG Scalograms and Spectrograms. 2021

[16] Chen Q, Yue H, Pang X, Lei W, Zhao G, Liao E, Wang Y Mr-ResNeXt: A multi-resolution network architecture for detection of obstructive sleep apnea. International Conference on Neural Computing for Advanced Applications 420-432. 2020

[17] Rao, A.S., Nguyen, T., Palaniswami, M.S., & Ngo, T.D. Vision-based automated crack detection using convolutional neural networks for condition assessment of infrastructure. Structural Health Monitoring, 20, 2124 - 2142. 2020