

Received June 25, 2020, accepted July 9, 2020, date of publication July 14, 2020, date of current version July 24, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3009149

Multi-Objective Hyperparameter Optimization of Convolutional Neural Network for Obstructive Sleep Apnea Detection

SHEIKH SHANAWAZ MOSTAFA^{1,2}, (Graduate Student Member, IEEE),
FÁBIO MENDONÇA^{1,2}, **ANTONIO G. RAVELO-GARCIA**^{2,3},
GABRIEL JULIÁ-SERDÁ⁴, AND **FERNANDO MORGADO-DIAS**^{1,2,5}, (Member, IEEE)

¹Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal

²ITI/Larsys, Madeira Interactive Technologies Institute, 9020-105 Funchal, Portugal

³Institute for Technological Development and Innovation in Communications, Universidad de Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain

⁴Pulmonary Medicine Department, Hospital Universitario de Gran Canaria Dr. Negrín, 35010 Las Palmas de Gran Canaria, Spain

⁵Faculty of Exact Sciences and Engineering, Universidade da Madeira, 9000-082 Funchal, Portugal

Corresponding author: Sheikh Shanawaz Mostafa (sheikh.mostafa@tecnico.ulisboa.pt)

This work was supported in part by the Portuguese Foundation for Science and Technology through the Projeto Estratégico under Grant UID/EEA/50009/2019, in part by the Agência Regional para o Desenvolvimento da Investigação, Tecnologia e Inovação (ARDITI) Ph.D. Studentship under Project M1420-09-5369-FSE-000001, and in part by MITIExcell, number M1420-01-0145-FEDER-000002.

ABSTRACT Obstructive sleep apnea (OSA) is a common sleep disorder characterized by interrupted breathing during sleep. Because of the cost, complexity, and accessibility issue related to polysomnography, the gold standard test for apnea detection, automation of the diagnostic test based on a simpler method is desired. Several signals can be used for apnea detection, such as airflow and electrocardiogram. However, the reduction of airflow normally leads to a decrease in the blood oxygen saturation level (SpO₂). This signal is usually measured by a pulse oximeter, a sensor that is cheap, portable, and easy to assemble. Therefore, the SpO₂ was chosen as the reference signal. Feature based classifiers with shallow neural networks have been developed to provide apnea detection using SpO₂. However, two main issues arise, the need for feature creation and the selection of the more relevant features. Deep neural networks can solve these issues by employing featureless methods. Among multiple deep classifiers that have been developed, convolution neural networks (CNN) are gaining popularity. However, the selection of the CNN structure and hyperparameters are typically done by experts, where prior knowledge is essential. With these problems in mind, an algorithm for automatic structure selection and hyper parameterization of a one dimension CNN was developed to detect OSA events using only the SpO₂ signal. Three different input sizes and databases were tested, and the best model achieved an average accuracy, sensitivity, and specificity of 94%, 92%, and 96%, respectively.

INDEX TERMS Biomedical signal processing, CNN, genetic algorithms, machine intelligence, medical expert systems, Pareto optimization, sleep apnea, SpO₂.

I. INTRODUCTION

Sleep is a circadian rhythm that significantly contributes to maintaining a pleasant daily routine. However, more than sixty sleep related disorders have been identified and obstructive sleep apnea (OSA) is one the more prevalent in the population [1]. It is characterized by obstruction or

reduction of the airflow during sleep decreasing the oxygen level in blood. It was also verified that sleep apnea increases cognitive impairment [2] and increases the risk of hypertension [3], coronary artery disease [4], stroke [5] and other diseases. In population based studies it was estimated that around 200 million people suffer from this disorder [6], [7] that is more dominant in adult men (around 4% compared to 2% adult of women) [7]. However, 80% of the patients can be unaware of that they possess the disorder.

The associate editor coordinating the review of this manuscript and approving it for publication was Sabah Mohammed¹.

Polysomnography (PSG) is the gold standard for OSA detection, using signals from multiple sensors to produce accurate diagnostics. Nevertheless, it is a tedious and time-consuming task [8] that requires specialized equipment, with a high economic cost for maintenance, and professionals to perform the exam [9], [10]. So, a simple and less costly system with a lower number of sensors is desirable for analysis.

Full reviews on commercial devices [11], algorithm performance [12], deep learning [13] and classification methods [14] for OSA detection have been performed. Different source signals, such as oximetry [15], [16], electrocardiogram (ECG) [17], [18], respiration [19], [20], and sound [21], have been previously evaluated. Among the studied sensors, pulse oximetry which estimates blood oxygen saturation (SpO₂), has shown significant potential for OSA detection in a non-invasive way that is easy to use and self-assembly. Therefore, SpO₂ was chosen for this work. The problem of building a reliable system using SpO₂ sensor is mainly two-fold, finding the best features that describe the apnea events and use these features to detect the apnea accurately. Multiple researchers have analyzed different time domain [22] and frequency domain characteristics [23] thus, creating a vast pool of suitable features. The creation of handcrafted features that achieve good performance requires significant domain knowledge. In addition to that, it is becoming significantly harder to find a new set of features that can achieve a higher performance since combining two or more features does not guarantee an improvement [24]. Therefore, a large number of features needed to be sorted according to relevance to increase the accuracy. Various techniques have been employed to address this problem such as minimum Redundancy Maximum Relevance (mRMR), Sequential Forward Search (SFS) [25] and Genetic Algorithms (GA) [26]. However, these techniques are either slow or most of the time it does not guarantee that the best features were chosen. In addition to that, the best features are sometimes dependent upon the classifiers used. Deep learning has the ability to automatically learn features from raw data [27]; thus by using deep network these problems can be solved.

Deep convolution neural network (CNN) is one of the most successful deep networks which are inspired by the vision system. Traditionally, CNN are designed for two-dimensional (2D) images as input with different channels [28] however, it can also be used for one-dimensional (1D) signals with single-channel [29], [30]. In most of the cases, automated detection of obstructive sleep apnea events using a CNN performs better when compared with shallow classifiers [31]. Some authors used nasal airflow [32] or a combination of SpO₂, oronasal airflow, and ribcage and abdomen movements [33] and converted these one dimensional signals to a two dimensional input to employ the two dimensional CNN (CNN2D) directly for apnea detection. One dimensional CNN (CNN1D) is a good alternative requiring far less preprocessing (does not need to convert 1D to 2D) for 1D signals. ECG [34], [30], [35] and nasal pressure or airflow [36], [19] signals have been previously used with a CNN1D

for OSA classification. Haider *et al.* [37] employed three one dimensional signals (nasal airflow, abdominal and thoracic plethysmography) to feed a CNN1D with three channel inputs for OSA detection. Following this research line, the SpO₂ signal which is 1D in nature was selected to be directly fed to CNN1D without any dimensional change stage.

However, implementing CNN presents significant changes. The structure and/or hyperparameters of the network are typically selected through an experimental search. Such methods require a significant amount of time as well as experience and expert knowledge for the creation of handcrafted network structure and hyperparameters [38]. A possible alternative is to use evolutionary algorithms, such as GA, to solve the structural optimization problem. The algorithm starts with a random individual generation and using mutation and crossover over a defined number of generations to achieve the optimized solution by optimizing fitness function. Zhining and Yunming [39] designed a genetic convolutional neural network model based on random sample and found it has a better performance than CNN in MNIST data set. Evolutionary algorithms also achieved significant success in the configuration of topologies [40] and connection for convolution layers [41]. Regarding the selection of the network hyperparameters, an asynchronous evolutionary approach was successfully used on a Titan supercomputer [42]. Also, neuro-evolution was able to construct large, accurate networks from trivial initial conditions while searching a large space without experimenter participation [40]. By combining Dynamic Structured Grammatical Evolution (DSGE) with GA Assunção *et al.* [43] was able to achieve better results without resorting to prior knowledge. Grammatical Evolution (GE) was also used for handwritten digit recognition [44] as well as human activity recognition [45]. EvoDeep, a DNN where GA was used for optimization, was developed by Martín *et al.* [46]. This concept of using GA for choosing the best network was also successfully extended to transfer learning [47]. For sleep apnea detection, Falco *et al.* [48] used Evolutionary Algorithms (EAs) to optimize the hyperparameters of a deep network with Heart Rate Variability (HRV). Therefore, a GA was also employed in this work for hyperparameter optimization of the CNN.

Unbalanced data is also a common issue in the sleep apnea detection, having insufficiency of one class (apnea) level and prevalence of another class (normal) level. Thus, a single objective technique (which was applied in the previously mentioned applications) commonly tries to maximize the accuracy, leading to a biased classifier since an increase of the accuracy can sacrifice the sensitivity (apnea events detection) that is related to the less prevalent class. The designed model addresses this issue by simultaneously considering the accuracy (Acc), sensitivity (Sen) and specificity (Sp) in a multi-objective problem. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) [49] was selected for this work due to the large success it has in other areas, such as filter design [50], water distribution system [51]. Therefore, the primary objectives of this work are:

- Design an automatic featureless sleep apnea events detection algorithm using CNN and SpO2 signal;
- Develop an independent algorithm capable of choosing the CNN structure and hyperparameters without any human intervention, using the multi-objective optimization;
- Analyze the effects of input sizes, database dependencies and layer size on the classification.

To achieve the desired objectives, an algorithm was developed for the automatic creation and hyper parameterization of CNN using one database with three input sizes (Section II). The result of the creation and hyper parameterization of CNN is discussed in Section III. In the end, the conclusion, limitation and future works are discussed in Section IV.

II. MATERIALS AND METHODS

The proposed system used CNN with NSGA-II algorithms to solve the multi-objective problem by choosing a suitable structure that can achieve the goal of attaining balanced results. Three different databases were used to create and test the system. A simplified workflow of the developed methodology is shown in Fig. 1. For simplicity, the iterative elements of the workflow are not shown (such as two-fold, leave-one-out, different inputs etc.).

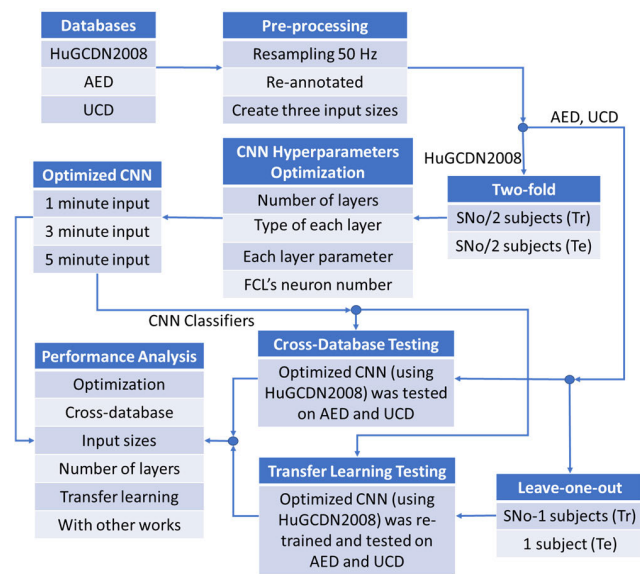


FIGURE 1. Simplified workflow of the optimization and performance analysis methods applied for proposed technique. In the figure Tr is training, Te is testing, FCL is fully connected layer.

First, the databases are re-sampled with a minute based annotation. Then CNN hyperparameters optimization is carried out with the HuGCDN2008 database with a two fold technique using half of the subjects (SNo/2) for training and half of the subjects for testing, where SNo is the total number of the subjects. After achieving the best CNN for three inputs, these classifiers were tested with other two (AED and UCD discussed later) databases using cross-database testing and transfer-learning testing. Finally, the performance of the proposed system was analyzed with different performance

parameters. A brief description of the materials and methods are discussed in the subsequent subsections.

A. DATABASES

Three databases were tested in this work. Two of them were collected from Physionet and are freely available, specifically the Physionet apnea ECG database (AED) and the St. Vincent’s University Hospital / University College Dublin Sleep Apnea Database (UCD). The third one was collected in Gran Canaria University Hospital, named Dr. Negrín University Hospital Database.

The dataset from Sleep Unit of Dr. Negrín was collected in Gran Canaria University Hospital has 70 referred (suspected sleep apnea) patients (51 males and 19 females, from 18 to 82 years old) which will be referred as HuGCDN2008 database. The subjects do not have any arrhythmia and SpO2 signal was sampled at 50 Hz. The annotations were made in 30 s epochs [23].

AED has 70 recordings but only eight have SpO2 signals available. These eight recordings were used and their duration ranges from 7 to 10 hours with minute-by-minute annotation [52], [53]. The sampling frequency of this data is 50Hz.

UCD [54] has 25 referred (suspected sleep apnea) subjects including 21 males and four females. This database is continuously annotated with hypopnea (HYP) and central (C), obstructive (O) and mixed (M) apnea. The sampling rate of the SpO2 signal was 8 Hz. It has a continuous annotation.

B. PREPROCESSING OF THE DATABASES

Due to the non-uniformity between datasets, such as sampling frequency and annotation methods, a normalization preprocess was performed.

To have all of the databases at the same sample rate UCD [54] was resampled at 50Hz. In this work, apnea events were detected with one minute epochs (as employed by AED [53]). Therefore, the annotations for HuGCDN2008 were produced by labeling the minute as apnea minute if any or both of its 30 second windows were annotated as apnea by the physician. For UCD [54] if 10 seconds or more, in a minute, were annotated as apnea by the physician, then the one minute epoch was labeled as apnea.

The input sizes of 1 minute, 3 minute (with 2 overlapping minutes) and 5 minute (with four overlapping minutes) were created considering the central minute as the one that defines the label. Therefore, taking into consideration the selected sampling frequency the 1 minute (60 seconds), 3 minute (180 seconds), and 5 minute (300 seconds) windows had 3000, 9000 and 15000 sampling points, respectively.

C. CONVOLUTION NEURAL NETWORK

CNN commonly comprise different types of layers. In this work, the CNN has an input layer, convolution layers, nonlinear layers, fully connected layers, batch normalization layers, softmax layer and a class output layer.

The input layer is the first layer of the network and receives the raw data. Thus, the size of the input layer is the same as the input data so three input layers were tested.

Different convolution kernels of the convolution layer, or filters, slide over the input providing a diversity of features which capture different local information [55]. The learners choose the values of the filters during the training process [56]. The feature map size depends on the number of filters (sometimes referred as depth), filter size (for 1D only width is considered), stride (which is the number of sample points the filters slides in each step) and padding (adding zeros at the end of the data input so that filters can be run on the bordering elements).

If the whole convolution layer is considered, the feature maps can be seen as a $n + 1$ dimension map where n is the dimension of the input [57]. The equation for the feature map of the convolution layer is given by

$$C_d = \varphi(k_d \otimes f + b_d) \quad (1)$$

where $1 \leq d < nk_d$, nk_d is the number of convolution kernels in a layer, C is the feature map of the entire convolution layer ($C \in \mathbb{R}^{i \times j \times nk_d}$), \otimes is the $n(= 1)$ dimensional convolution operation, k is the kernel, f is the input matrix, b is the bias and d is the number of kernels.

In this work, Nonlinear Layer uses the Rectified Linear Unit (ReLU) for a non-linear operation that introduces the non-linearity in the network by replacing all negative values by zero. This non-linearity can improve the classification performance [58] therefore, ReLU was chosen as the nonlinear function. It is defined as

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2)$$

where x is the input of the function.

The Pooling layer uses a subsampling operation for reducing the dimensionality of the data that passes through the network. It has the same parameters as a convolution filter and can perform operations such as average and maximum (max). Max pooling commonly provided the best results thus it was used [59], defined as

$$y_p = \max(x) \quad (3)$$

where x is the input of the maxpooling layer with the same number of input samples as the pooling size of maxpooling.

Batch normalization layer normalizes its input over a batch as [60]

$$y_b = \gamma \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} + \beta \quad (4)$$

where μ_B is the mean and σ_B the variance over the batch size. ϵ is a constant which stabilizes the system if the batch size is too small. This layer is added to increase the training speed and decrease the network initialization sensitivity.

A fully connected layer infers that every neuron in the previous layer is connected to every neuron on the next layer

and helps to learn the nonlinear parameter. In this layer, a neuron has the same function as a common perceptron with an activation function defined by [61]

$$y = \varphi \left(\sum_{j=1}^n x_j \times w_j \right) + b \quad (5)$$

where x is the input, w is the weight, b is the bias, n is the number of inputs and φ is the nonlinear function. This layer is followed by a SoftMax layer, also known as normalized exponential function. The softmax function $\varphi_{softmax}$ allows to represent a categorical distribution which is a probability distribution over k different possible outcomes. In this implementation $k = 2$ (binary case), thus

$$\varphi_{softmax}(x) = \frac{e^x}{\sum_{j=1}^k e^{x_j}} \quad (6)$$

Afterwards, a class out layer generates the output of the desired class level of the input.

D. OPTIMIZATION OF CNN HYPERPARAMETERS USING MULTI-OBJECTIVE HYPERPARAMETER OPTIMIZATION

CNN's hyperparameters were optimized using a multi objective genetic algorithm named NSGA-II [49], [62]. The multi objective technique was used to have an equally better performance in all objectives, contrary to what it is present in the single objective optimization method [50]. A Multi objective optimization technique optimizes a vector (O) of objective functions (in this work Acc, Sen and Spc) and the optimization consists of finding V (in this case, the hyperparameters) which maximizes $O(v)$ represented as

$$O(v) = (O_1(v), O_2(v), \dots, O_k(v)) \quad (7)$$

$$\text{subject to : } y(v) = (y_1(v), y_2(v), \dots, y_j(v)) \quad (8)$$

$$\text{and } v_i^L \leq v_i \leq v_i^U \quad (9)$$

where v is the vector of design variables in V parameter space with N elements with upper bound v_i^U and lower bound v_i^L , $y(v)$ is the objective space and $O(v)$ is the vector representation of the objective functions that has to be maximized [63].

A simplified representation of the implementation strategy is presented in Fig. 2 where all inputs of layers are represented as x and outputs as y . For every generation (Gen) the chromosome of each population (Pop), named P_t , was generated using mutation and crossover with the information needed to create a CNN. Then, it was translated to the CNN structure and parameters using the decoding methods indicated in Table 1. After twofold training and testing, the next generation population (P_{t+1}) was chosen according to Pareto fronts and crowding distance using Acc, Sen and Spc of twofold test.

The implemented technique can be described in 11 steps:

Step 1: A parent population P_0 with the size of N is randomly generated.

Step 2: The system converts the chromosome to a CNN network. A fixed input (3000 neurons for 1 minute, 9000 neurons for 3 minute and 15000 neurons for 5 minute) layer

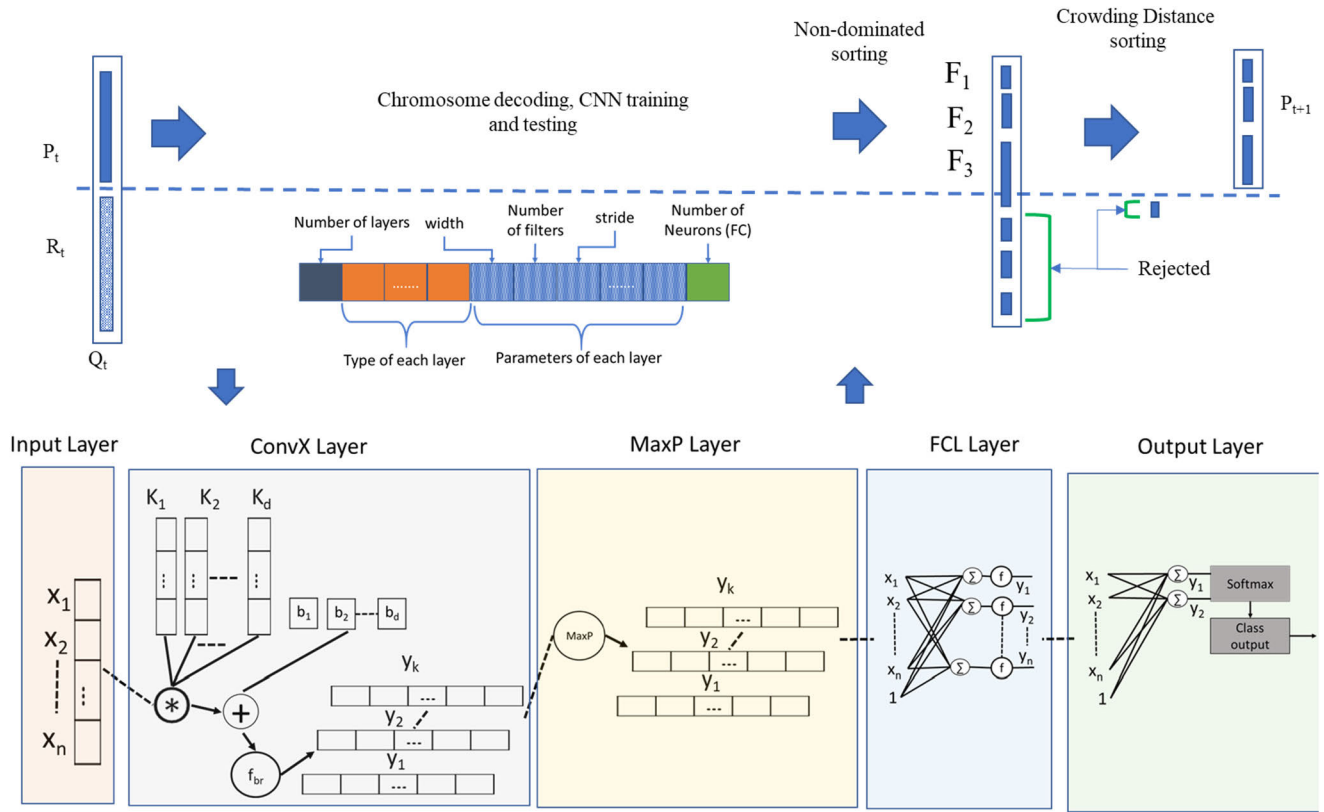


FIGURE 2. Simplified representation of the CNN hyperparameters optimization strategy using multi-objective (NSGA-II [49]) hyperparameter optimization.

and output layer (fully connected layer with 2 outputs, softmax layer and class output) are always presented (fixed), regardless of the structure chosen by the GA algorithm. The GA algorithm was only allowed to choose the number of the layers between fixed layers, type of layers, size of kernels, pooling sizes, stride and number of neurons of a fully connected layer. A real coded chromosome that ranges from 0 to 1 was used in this work. However, different types of parameters for the CNN had different ranges, so proper decoding was done according to Table 1. First the generated chromosomes were scaled between the defined range and then a ceiling function was used to get natural numbers. To reduce the number of possible solutions, hence reducing the simulation time, two different types of layer combinations were used. First was a convolution layer with a ReLU layer and batch normalization defining it as ConvX. Second one was maximum pooling, indicated by MaxP. ConvX layer has three cascaded functions, doing convolution with the input and a defined kernel (k) then a batch normalization and finally a ReLU, indicated together by f_{bd} in the Fig. 1. To prevent losing too much information in each layer, back to back MaxP layer was replaced by a ConvX layer and each parameter has its range defined in Table 1.

Step 3: After generating the CNN structure (hyperparameters), the network was trained using the ADAM algorithm [64] using two fold methods. HuGCDN2008 database has 70 subjects which were divided into 35 subjects in each

TABLE 1. Chromosome decoding techniques and ranges for CNN.

Type	Number of Position	Quantity	Range
1	1	Number of layers	1-10
2	Number of max flexible layers	Type of each layer	1(ConvX)-2(MaxP)
3	Number of highest parameters*Number of max layers	Each layer parameters	
3.1	1	Number of filters	1-15 (ConvX)
3.2	1	Filter size (Width)	3-9(ConvX), 2-5(MaxP)
3.3	1	Stride	1- filter size
4	1	FC neuron number	20-200

train and test set. Subject independence between the training and testing sets was ensured by not mixing the subjects data between the sets. An initial learning rate of 0.001 was employed during the training and each 10 epochs the learning rate drop factor was 0.1. The batch size was 256 and data were shuffled in every epoch. An average of the Acc, Sen and Spc was calculated to be used as objective parameters.

Step 4: Using the objective parameters a non-dominant sorting was performed for sorting the parent population, where $P = P_0$ [49].

Step 5: Simulated binary crossover [65] and polynomial mutation were used to create a new offspring population (Q) of size N .

Step 6: A combined population of R_t was created using offspring Q_t and parent population P_t . Thus, the size of R_t becomes $2N$.

Step 7: Fast non-dominated sort was used to sort the entire population in the same way as in Step 4.

Step 8: Calculate the crowding distance using the method defined by Deb *et al.* [49].

Step 9: Combined population, Q_t , was sorted according to a non-dominant sort and crowding distance. If the population size (from first F_1 to last F_l front) was greater than N then a crowded-comparison operator, $<_n$, was used in descending order to populate the population size until N from F_l and others ($F_{>l}$) are discarded. The partial order $<_n$ was given by

$$i <_n j \text{ if } (i_{rank} < j_{rank}) \text{ or } ((i_{rank} == j_{rank}) \text{ and } (i_{distance} > j_{distance}))$$

Step 10: Keep the N number of elements from the sorted list and increased the number of generations.

Step 11: Repeat from Step 5 until the termination condition (50 generations were produced) was met.

E. PERFORMANCE EVALUATOR

The multi-objective method optimized (Eq. 7 where $k = 3$) for Accuracy (Acc), Sensitivity (Sen) and Specificity (Spc) is defined by

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$Sen = \frac{TP}{TP + FP} \quad (11)$$

$$Spc = \frac{TN}{TN + FN} \quad (12)$$

where, TP is the number of apnea minutes classified as apnea minute, TN is the number of normal minutes classified as normal minute, FP is the normal minutes classified as apnea minute, and FN is the number of the normal minute apnea minutes classified as normal minute.

The optimization of CNN hyperparameters was performed using HuGCDN2008 database. The other two databases (UCD and AED) were used for cross-database and transfer-learning performance assessment.

Cross-database: For cross-database testing, the CNN is trained and developed using one database (HuGCDN2008) and testing the performance in another database (UCD and AED).

Transfer-learning: Traditionally, transfer-learning is the process of applying knowledge gain in one field applying in another that is related but in a different problem. In this work, knowledge gain (optimization of CNN) with one database (HuGCDN2008) was implemented in other databases (UCD and AED) with a minimum modification. Practically,

this is achieved by replacing the FC and output layer with new FC and output layer and retain the CNN.

In addition to this, this work investigated the relationship of input sizes, number of layers and performance.

III. RESULTS

The algorithm was implemented in MATLAB and ran in a computer with Intel Core (TM) i7-8700k processor, 64 GB RAM, and two NVIDIA GeForce GTX 1080 Ti GPUs. Two folds were run in parallel in the two GPUs and the average of the obtained results of the objective functions were computed. The optimizations were carried out with three different input sizes, 1 minute with 3000 samples, 3 minute with 9000 samples and 5 minute with 15000 samples, respectively. The termination condition was 50 generations with a population size of 50, which leads to $(50(\text{Gen}) * 50(\text{Pop} = \text{offspring population} = N))$ 2500 different networks and 5000 networks to train for each input size (because a two-fold method was employed). It took, respectively 587.83926 hours (≈ 24.49 days), 832.04399 hours (≈ 34.67 days) and 911.226716 hours (≈ 37.97 days) for 1 minute, 3 minute and 5 minute input to finish the optimization.

A. HYPERPARAMETER OPTIMIZATION

In the first step, the algorithm generates a random population to ensure the diversity of the population and rank them according to the multi objective optimization (Acc, Sen and Spc) [50]. From Fig. 3 it can be seen that for the three simulations (1 minute, 3 minute and 5 minute input) started (1st Gen) with some random solutions all over the problem space. Over each generation, using mutation and crossover, the algorithm was able to reach better solutions as can be verified in Fig. 4, Fig. 5 and Fig. 6. Almost all the solutions of 50th generation were better than the 1st Gen. From the 50th Gen solutions, among three different inputs, it is noticeable that 3 minute and 5 minute inputs have better results compared to 1 minute solutions but in a similar range.

The multi objective algorithm, NSGA-II, ranked the outputted solution according to Pareto front numbers. All the solutions on first Pareto front are valid solutions as NSGA II do not generate a single solution but a set of Pareto non-dominated solutions. One way of assessing the classifiers performance is the receiver Operating Characteristic (ROC) curve (Sen vs 1- Spc). Since the algorithm is using a three dimensional problem space (Fig. 3) a modified version of the ROC curve (with 2 dimension), where all of the first Pareto front solutions are showed in Fig. 7, can allow to compare the solutions. For all three different inputs, CNN's first Pareto front of 50th Gen has better and more solutions than 1st Gen's first Pareto front, as can be seen in Fig. 7 (a), (b), (c). For the 50th Gen it was assessed that 5 minute input has the best solution followed by 3 minute and 1 minute CNN (Fig. 8).

Although the algorithm was trying to solve a multi-objective optimization, NSGA-II treated equally the optimization variables (Acc, Sen, and Spc) due to a restriction

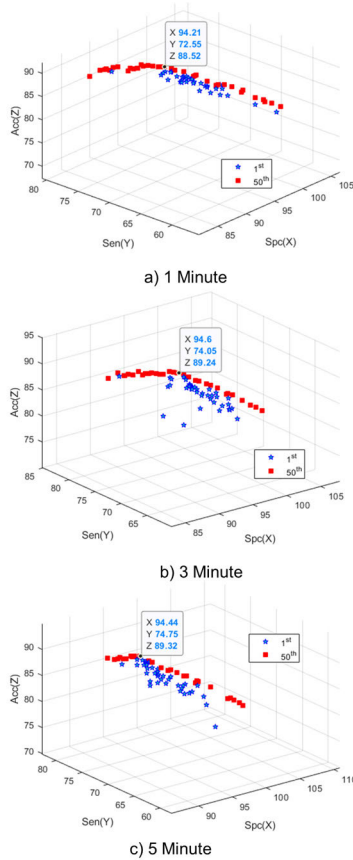


FIGURE 3. The multi objective problem space in percentage for 1st and 50th generation for a) 1 minute b) 3 minute c) 5 minute inputs.

with space problems and constraints such as overlap of apnea and normal events (where the Sen and Spc are dependent and Acc is dependent on Sen and Spc) the solutions do not have equal Sen, Spc, and Acc. Therefore, the final solution was chosen with the highest Acc among all the valid solutions which will also help to compare with other methods presented in the literature. By using Acc as a strategy of choosing one solution over others, three solutions for 1 minute, 3 minute and 5 minute inputs were selected with Acc of 88.2%, 89.24%, and 89.32%; Sen of 72.55%, 74.05%, and 74.75%; and Spc of 94.21%, 94.60%, and 94.44% respectively for two-fold methods. These solutions are marked with a black dot and the values are indicated in a box in Fig 3, Fig 4, Fig 5, Fig 6, Fig. 7 and Fig. 8. The detailed results of two-fold method are shown in Table 2 and the average is shown in Table 5. The numbers of flexible layers (NoFL) was 5, 5 and 7 (Fig. 9 Section III C) resulting into 19, 19 and 21 layers for 1 minute, 3 minute and 5 minute CNN networks, respectively (TABLE 3). The layer sequence in 1 minute and 3 minute CNN is also similar where first Conv layer (L2) and Batchnorm (L3) are the same for both. However, the remaining layers have more kernels or channels. Five minute CNN has two more flexible layers and actual layers are in form of Maximum pooling (L9, L13). It has the same number of conv, ReLU and Batchnorm layers as 1 minute

and 3 minute CNN, except that the number of kernels and channels in the first layer is higher.

B. CROSS-DATABASE PERFORMANCE

In order to check the universality of the system, the trained (in HuGCDN2008 database) CNNs were tested in AED [53] and UCD[54] databases. The results are presented in Table 2. The performance of all three networks with UCD [54] database was lower than the originally trained database but were higher in the AED [53]. The highest accuracy, 92.65%, was achieved with 1 minute input in AED [53] database. For HuGCDN2008 and UCD [54] database, the main difficulty was in the detection of short apnea events. This could be related to the fact that some respiratory pauses do not produce a clear pattern in the oximetry signal. This could be related to the hemoglobin dissociation curve, where short events would not be able to decrease the SpO2 percentage because a marked reduction in the partial oxygen pressure did not occur. In addition to that, pH, temperature and 2,3-diphosphoglycerate (2,3-DPG) levels, which are specific to each person, can displace the hemoglobin dissociation curve [23].

TABLE 2. The Result both CNN trained using HuGCDN2008 database in Two-Fold (_F1 and _F2 for fold one and fold two. Different input sizes are indicated as _1, _3 and _5 for 1 minute, 3 minute, and 5 minute, respectively).

Symbol	Test Database	Sen	Spc	Acc
CNN1DF1_1	HuGCDN2008	71.50	95.30	88.50
CNN1DF2_1	HuGCDN2008	73.60	93.10	88.60
CNN1DF1_3	HuGCDN2008	73.70	95.10	88.90
CNN1DF2_3	HuGCDN2008	74.40	94.10	89.50
CNN1DF1_5	HuGCDN2008	75.10	94.90	89.20
CNN1DF2_5	HuGCDN2008	74.40	93.90	89.40
CNN1DF1_1	AED [53]	91.64	93.36	92.65
CNN1DF2_1	AED [53]	87.89	92.54	90.63
CNN1DF1_3	AED [53]	83.77	93.62	89.58
CNN1DF2_3	AED [53]	85.64	93.36	90.20
CNN1DF1_5	AED [53]	79.03	93.62	87.64
CNN1DF2_5	AED [53]	88.58	93.67	91.58
CNN1DF1_1	UCD[54]	56.72	93.32	84.55
CNN1DF2_1	UCD[54]	64.12	90.69	84.33
CNN1DF1_3	UCD[54]	67.35	90.51	84.96
CNN1DF2_3	UCD[54]	82.51	76.30	77.79
CNN1DF1_5	UCD[54]	66.57	90.19	84.53
CNN1DF2_5	UCD[54]	85.37	60.94	66.79

C. EFFECT OF INPUT SIZE

With increasing the input size, the performance of HuGCDN2008 dataset was improved slightly in Acc, from 88.52% to 89.28% and 89.32%. However, between 3 minute and 5 minute, the results were almost the same. The Sen (apnea events) was affected by the input size with an improvement of more than 2% when comparing with the 1 minute and 5 minute input size. The Spc (normal events) remains almost the same. By increasing the Sen and keeping the Spc stable, the classifiers were able to increase the Acc. A possible reason to justify why longer inputs achieve better Sen could be related to the fact that an apnea event could be

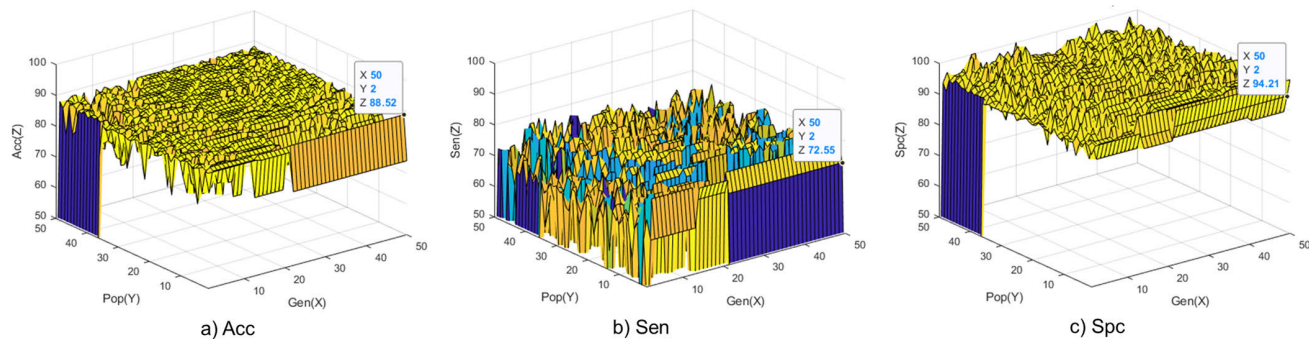


FIGURE 4. The changes of the multi objectives in percentage, a) Acc, b) Sen c) Spc, over the generations (Gen) of the populations (Pop) for 1 minute input.

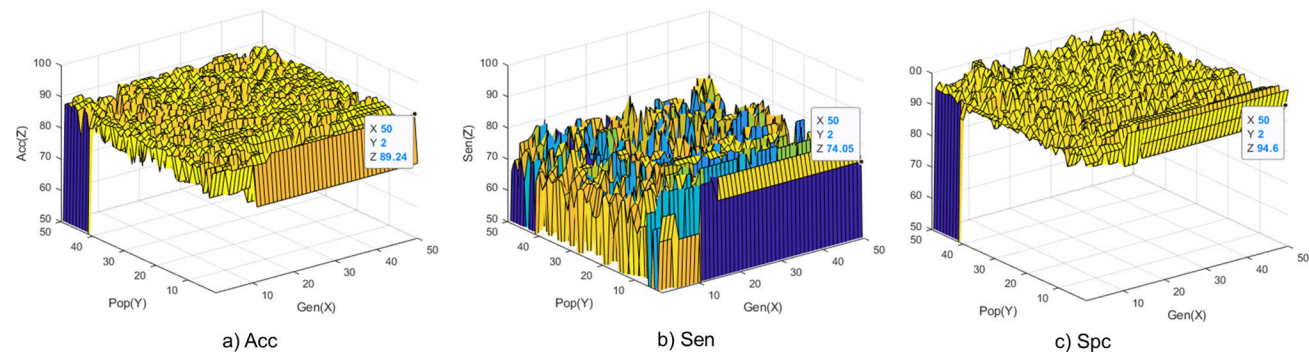


FIGURE 5. The changes of the multi objectives in percentage, a) Acc b) Sen c) Spc, over the generations (Gen) of the populations (Pop) for 3 minute input.

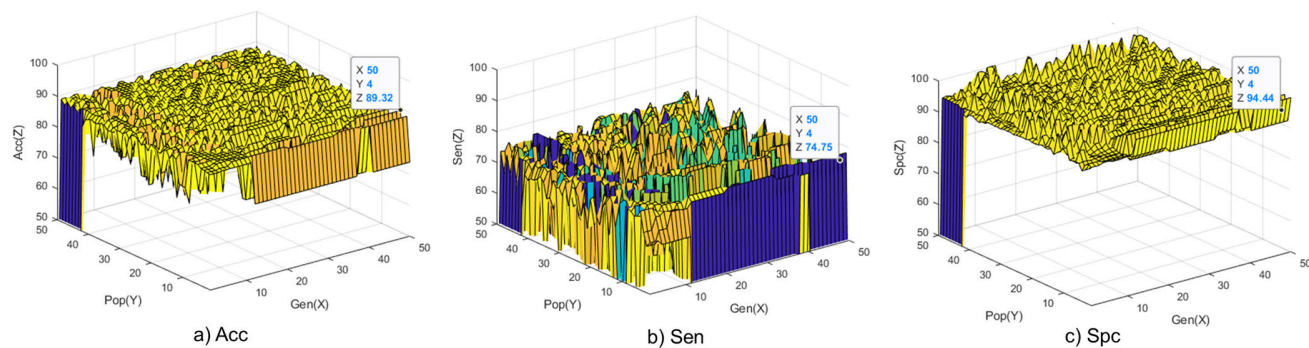


FIGURE 6. The changes of the multi objectives in percentage, a) Acc b) Sen c) Spc, over the generations (Gen) of the populations (Pop) for 5 minute input.

present into different minutes, thus, having the information of longer apnea events increases the detections capabilities. Another reason could be in some work [23] higher (five) minutes spectral features show more relevance. However, in other datasets (AED [53] and UCD [54]) this trend was not consistent. For the AED [53] dataset the highest Acc, 92.65%, and Sen, 91.64%, were achieved by the 1 minute input and for UCD [54] the best results, Acc of 84.96% and Sen of 67.35%, were achieved by the 3 minute input. Sometimes one network of higher input size performs worse than the lower input size. Therefore, the performance parameters are more dependent on the data and train weight than the input size.

D. EFFECT OF LAYERS

Due to the success of big (deeper) networks, one can assume that more layers give better results in case of deep learning. But this assumption is not always true. The number of chosen layers for each solution can be seen in Fig. 9. Analyzing the figures, it is possible to assess that the algorithm was trying different layer sizes to solve the problem and a better solution did not have the highest number of layers (NoFL). A Similar conclusion was presented by Urtnasan *et al.* [30], [34] that have the occurrence of optimum six layered CNN while testing from 3 to 9 layers.

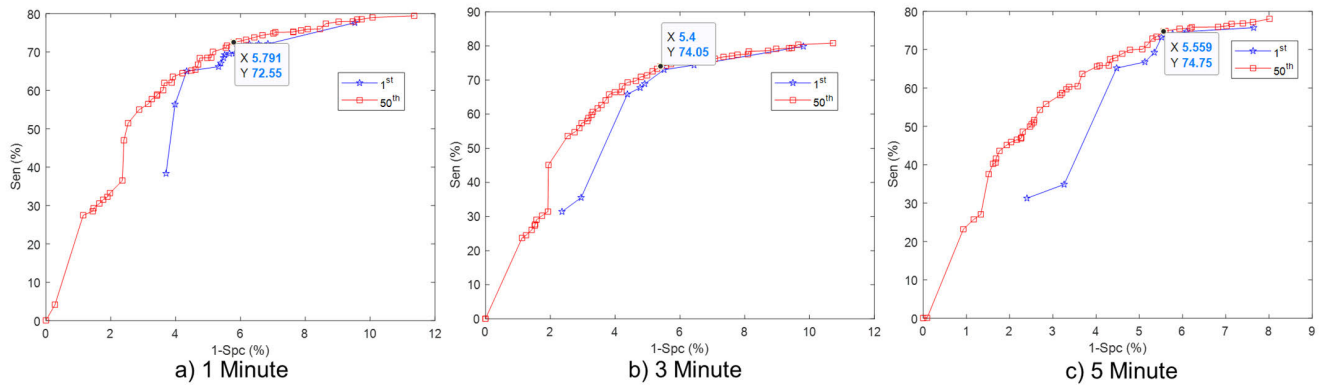


FIGURE 7. The solutions for first (1st) and last (50th) generation of a) 1 minute b) 3 minute c) 5 minute SpO2 signal. The solution of first generation is marked with star and 50th is marked with box.

TABLE 3. Chosen CNN’s layers and hyperparameters.

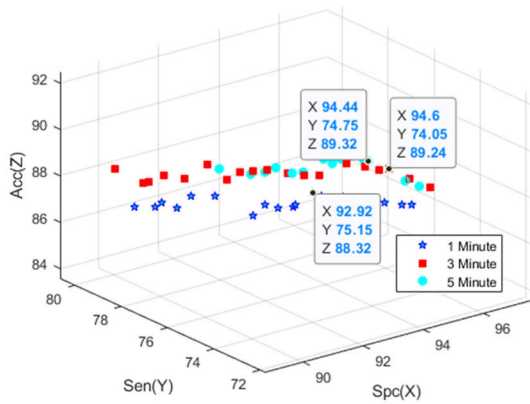
1 minute (60 seconds)			3 minute (180 seconds)		5 minute (300 seconds)	
No.	Layer	Layer Parameters	Layer	Layer Parameters	Layer	Layer Parameters
L1	Input	1x3000x1	Input	1x9000x1	Input	1x15000x1
L2	Conv_1	15@1x9x1_1x5	Conv_1	15@1x9x1_1x 5	Conv	6@1x7x1_1x3
L3	Batchnorm_1	15 channels	Batchnorm_1	15 channels	Batchnorm_1	6 channels
L4	ReLU_1		ReLU_1		ReLU_1	
L5	MaxP_1	1x5_1x5	MaxP_1	1x5_1x5	MaxP_1	1x3_1x3
L6	Conv_2	8@1x8x15_1x3	Conv_2	12@1x8x15_1x3	Conv_2	14@1x9x6_1x9
L7	Batchnorm_2	8 channels	Batchnorm_2	12 channels	Batchnorm_2	14 channels
L8	ReLU_2		ReLU_2		ReLU2	
L9	Conv_3	8@1x8x12_1x3	Conv_3	9@1x8x12_1x3	MaxP_2	1x3_1x2
L10	Batchnorm_3	8 channels	Batchnorm_3	9 channels	Conv_3	12@1x7x14_1x1
L11	ReLU3		ReLU3		Batchnorm_3	12 channels
L12	Conv_4	13@1x4x8_1x2	Conv_4	14@1x4x9_1x2	ReLU_3	
L13	Batchnorm_4	13 channels	Batchnorm_4	14 channels	MaxP_3	1x3_1x 2
L14	ReLU_4		ReLU4		Conv_4	15@1x5x12_1x3
L15	FC1	117 FCL	FC1	124 FCL	Batchnorm_4	15 channels
L16	ReLU_5		ReLU_5		ReLU_4	
L17	FC2	2 FC2	FC2	2 FC2	FC1	99 FCL
L18	Softmax		Softmax		ReLU_5	
L19	Classoutput		Classoutput		FC2	2 FC2
L20					Softmax	
L21					Classoutput	

Layer parameters such as the input size, size of the filter, number of the filters represented as a form of [number of filter]@ [vertical width of filter]x [horizontal width of filter]x [number of Channels of filter]_ [vertical width of stride]x[horizontal width of stride]

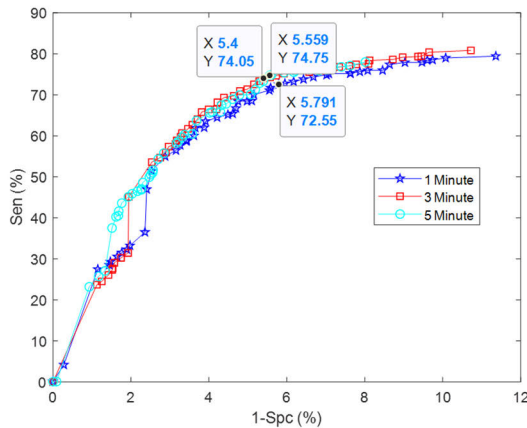
E. TRANSFER LEARNING PERFORMANCE

Transfer learning could be useful apply the information learned from one problem on others. This work mainly focused on OSA detection, so transfer learning performance was analyzed with AED [53] and UCD [54] where the main network was trained using HuGCDN2008 database. The last three layers (L 17-19 for 1 minute and 3 minute, L 19-21 for 5 minute, indicated in Table 3) were removed and replaced with a similar types of layers. Afterwards, it was retrained with leave one out methods (due to their low number of subjects). There were two different weighted networks for

each CNN input network (generated using two-fold methods implemented in HuGCDN2008 dataset). Because the actual train data for these transfer learning networks were coming from different dataset there is no need for averaging the two networks results (like two fold methods in HuGCDN2008). Table 4 presents the results from two networks and best results are summarized in Table 5. It was verified that in all of the cases the transfer learned networks have a better accuracy. However, there were database dependencies. In some of the cases, it was not the same original network trained with HuGCDN2008 (e.g. CNN1DF2_1 second fold network for



a) 50th Gen in 3D problem space.



b) 50th Gen in 2D problem space.

FIGURE 8. Comparison of three inputs in 50th generation a) in 3D (Acc, Sen, and Spc) problem space. b) in 2D (Sen and 1-Spc) problem space.

1 minute in Table 2) who performed the best in other two datasets (e.g. CNN1DF1_1 and CNN1DF1_1 in Table 4).

F. COMPARISON WITH STATE OF THE ART WORK

The closest match for the comparison with this work was developed by Ravelo-García *et al.* [23] where the same database, HuGCDN2008, was used. A shallow classifier,

TABLE 4. Transfer learning (F1 and F2 for fold one and fold two classifiers from HuGCDN2008 database. Different input sizes are indicated as _1, _3, and _5 for 1 minute, 3 minute and 5 minute, respectively).

Symbol	Database	Sen	Spc	Acc
CNN1DF1_1	AED [53]	92.04	95.78	94.24
CNN1DF2_1	AED [53]	89.25	94.65	92.42
CNN1DF1_3	AED [53]	92.79	94.56	93.83
CNN1DF2_3	AED [53]	89.87	96.78	93.93
CNN1DF1_5	AED [53]	91.49	95.52	93.86
CNN1DF2_5	AED [53]	87.76	96.61	92.96
CNN1DF1_1	UCD[54]	54.39	94.14	84.52
CNN1DF2_1	UCD[54]	58.32	93.32	84.85
CNN1DF1_3	UCD[54]	60.02	93.93	85.73
CNN1DF2_3	UCD[54]	60.38	93.90	85.79
CNN1DF1_5	UCD[54]	60.42	93.43	85.44
CNN1DF2_5	UCD[54]	60.34	93.54	85.51

linear discriminant analysis (LDA), was employed with SpO2 signal and combination of SpO2 and HRV. The proposed work achieved 89.32% Acc using 5 minute window with only SpO2 compared to 86.5% and 86.9% with mixed of 1 minute and 5 minute window using SpO2, and SpO2 and HRV signals. The proposed implementation was also able to keep the same performance level with 3 minute window and not sacrificing parameters. Even the one-minute window has better ACC and Sen compared to the other works in Table 5.

For the AED [53] dataset the proposed optimized CNN achieved 92.65% Acc, 93.36% Spc and 91.64% Sen. Though the Acc was not the best among other implementations it has one of the best Sen only surpassed by long short-term memory (LSTM) [66] and deep auto encoder network (DAE) [26]. However, neither of these works, [26], [66], were subject independent. For UCD [54] the transfer learning approach has achieved the highest accuracy compared to the other works except the DAE [26] that was also not subject independent. In both databases, transfer learning increases the performance parameters.

If the comparison only includes deep learning, the proposed networks achieved the best accuracy among all

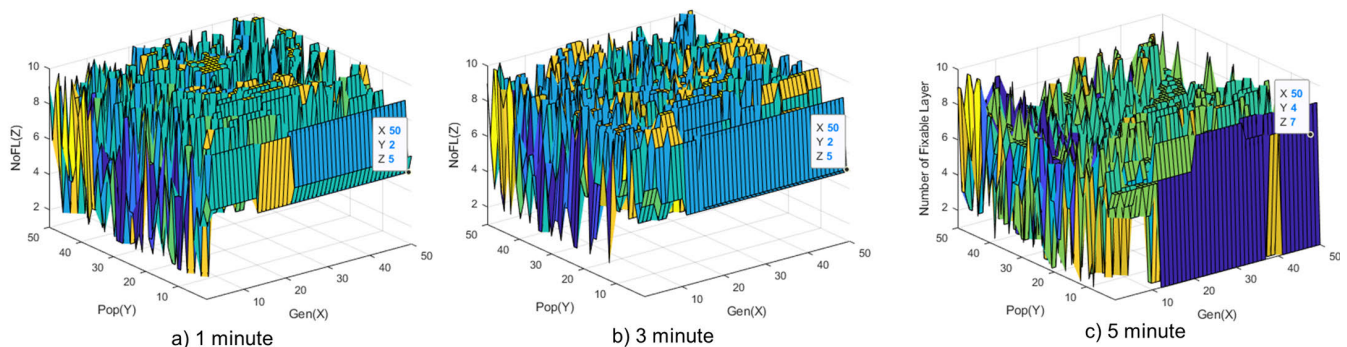


FIGURE 9. The changes of the numbers of flexible layers (NoFL) over the generations (Gen) of the populations (Pop) for inputs.

TABLE 5. Comparison with literature.

Ref.	Signal	Classifier	Database	Recordings or Patients	Input size (Seconds)	No of features	Sen	Spc	Acc
[7]	SpO2	SVM	Own	40	150	7	-	-	90
[69]	SpO2 +ECG	Bagging.RepTree	UCD [54]	25	60	39	79.75	85.89	84.80
[69]	SpO2	Bagging.RepTree	UCD [54]	25	60	39	78.23	84.25	82.79
[70]	SpO2	ANN	AED [53]	8	60	3	87.5	100	90.3
[23]	Spo2	LDA	HuGCDN2008	70	60,300	19	75.6	91.00	86.5
[23]	Spo2+HRV	LDA	HuGCDN2008	70	60,300	33	73.4	92.3	86.9
[25]	SpO2	SVM-L	AED [53]	8	60	50	83.76	97.03	96.89
[25]	SpO2	SVM-L	AED [53]	8	60	20	84.57	97.28	97.38
[25]	SpO2	ANN	UCD [54]	25	60	2	43.31	95.03	81.95
[25]	SpO2	LD	UCD [54]	25	60	9	61.78	91.03	83.27
[26]	SpO2	ANN	AED [53]	8	60	7	96.5	98.5	97.7
[71]	SpO2	DAE	AED [53]	8	60	-	78.75	95.89	97.64
[71]	SpO2	DAE	UCD [54]	25	60	-	60.36	91.71	85.26
[66]	SpO2+IHR	LSTM	AED [53]	8	60	-	84.7	-	92.1
[66]	SpO2	LSTM	AED [53]	8	60	-	92.9	-	95.5
[33]	SpO2+oronasal airflow+ribcage and abdomen movements.	CNN2D	UCD[54]	23	1	-	-	-	79.6
[67]	SpO2+airflow+respiration.	RCNN	MGH	10000	1	-	-	-	88.2
[67]	SpO2+airflow+respiration.	RCNN	SHHS	5804	1	-	-	-	80.2
P	SpO2	CNN1D	HuGCDN2008	70	60	-	72.55	94.21	88.52
P	SpO2	CNN1D	HuGCDN2008	70	180	-	74.05	94.60	89.24
P	SpO2	CNN1D	HuGCDN2008	70	300	-	74.75	94.44	89.32
P ^a	SpO2	CNN1D	AED [53]	8	60	-	91.64	93.36	92.65
P ^a	SpO2	CNN1D	AED [53]	8	180	-	85.64	93.36	90.20
P ^a	SpO2	CNN1D	AED [53]	8	300	-	88.58	93.67	91.58
P ^a	SpO2	CNN1D	UCD[54]	25	60	-	56.72	93.32	84.55
P ^a	SpO2	CNN1D	UCD[54]	25	180	-	67.35	90.51	84.96
P ^a	SpO2	CNN1D	UCD[54]	25	300	-	66.57	90.19	84.53
TL ^a	SpO2	CNN1D	AED [53]	8	60	-	92.04	95.78	94.24
TL ^a	SpO2	CNN1D	AED [53]	8	180	-	89.87	96.78	93.93
TL ^a	SpO2	CNN1D	AED [53]	8	300	-	87.76	96.61	92.96
TL ^a	SpO2	CNN1D	UCD[54]	25	60	-	58.32	93.32	84.85
TL ^a	SpO2	CNN1D	UCD[54]	25	180	-	60.38	93.90	85.79
TL ^a	SpO2	CNN1D	UCD[54]	25	300	-	60.34	93.54	85.51

P is for proposed networks optimized by GA and trained using HuGCDN2008 database. TL indicates transfer learning where the proposed networks were retrained using respective database.

^aBetween two networks best one is showed. For details please see Table II and Table IV.

subject independent implementation. Even compared to some implementation where more signals were employed, such as a combination of SpO2, airflow and respiration [67] or the combination of SpO2, oronasal airflow and movements (ribcage and abdomen) [33].

IV. CONCLUSION, LIMITATIONS AND FUTURE WORK

The goal of the work was to develop and test a novel fully automated hyperparameters optimization algorithm for CNN and significant results were attained.

Three different window sizes were also tested and is was verified that there is almost no difference between 3 minute and 5 minute window sizes. In some cases, the 1 minute outperformed the 3 minute and 5 minute inputs. Compared to shallow networks, the developed CNNs were able to achieve a better performance parameter with smaller input size and without the need for feature extraction.

It was also verified that the performance of the almost a similar structure networks was more sensitive to train and data than the hyperparameters choice. Also, it was verified

that transfer leaning has a strong potential for implementation in similar domains.

One of the limitations of the work is the fact that multi objective optimization was only applied to hyperparameter optimization and not used for the training. So, when the transfer learning concept was implemented, the network was sacrificing Sen to achieve a better Acc. The second limitation is the population number, only 50 which cannot ensure that the network had a strong diversity to start with. However, this issue was mitigated by the use of mutation. This work was not designed to be optimized for the layer size. So, even the 50th generation has a substantially different sized networks. One way of solving this issue would be running for more generations until stable solutions was found. Another way of doing it could be involving the number of layers as one of the objectives which are under consideration for future research. It was verified in the literature, that increasing the number of signals [37] or selecting a recurrent neural network (RNN) [68] could possibly improve the results [34], [30]. So, this could be investigated in the future increasing the number of signals.

ACKNOWLEDGMENT

The authors acknowledge “NVIDIA Corporation” for their GPU donation.

REFERENCES

- [1] M. J. Sateia, “International classification of sleep disorders-third edition,” *Chest*, vol. 146, no. 5, pp. 1387–1394, Nov. 2014.
- [2] L. J. Findley, J. T. Barth, D. C. Powers, S. C. Wilhoit, D. G. Boyd, and P. M. Suratt, “Cognitive impairment in patients with obstructive sleep apnea and associated hypoxemia,” *Chest*, vol. 90, no. 5, pp. 686–690, Nov. 1986.
- [3] L. Grote, T. Ploch, J. Heitmann, L. Knaack, T. Penzel, and J. H. Peter, “Sleep-related breathing disorder is an independent risk factor for systemic hypertension,” *Amer. J. Respiratory Crit. Care Med.*, vol. 160, no. 6, pp. 1875–1882, Dec. 1999.
- [4] T. Moee, K. A. Franklin, K. Holmström, T. Rabben, and U. Wiklund, “Sleep-disordered breathing and coronary artery disease: Long-term prognosis,” *Amer. J. Respiratory Crit. Care Med.*, vol. 164, no. 10, pp. 1910–1913, Nov. 2001.
- [5] V. Mohsenin, “Sleep-related breathing disorders and risk of stroke,” *Stroke*, vol. 32, no. 6, pp. 1271–1278, Jun. 2001.
- [6] T. Young, M. Palta, J. Dempsey, J. Skatrud, S. Weber, and S. Badr, “The occurrence of sleep-disordered breathing among middle-aged adults,” *New England J. Med.*, vol. 328, no. 17, pp. 1230–1235, Apr. 1993.
- [7] J. Zhang, Q. Zhang, Y. Wang, and C. Qiu, “A real-time auto-adjustable smart pillow system for sleep apnea detection and treatment,” in *Proc. 12th Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, 2013, pp. 179–190.
- [8] R. Agarwal and J. Gotman, “Computer-assisted sleep staging,” *IEEE Trans. Biomed. Eng.*, vol. 48, no. 12, pp. 1412–1423, Dec. 2001.
- [9] D. R. Hillman, A. S. Murphy, and L. Pezzullo, “The economic cost of sleep disorders,” *Sleep*, vol. 29, no. 3, pp. 299–305, 2006.
- [10] N. Alghanim, V. R. Comondore, J. Fleetham, C. A. Marra, and N. T. Ayas, “The economic impact of obstructive sleep apnea,” *Lung*, vol. 186, no. 1, pp. 7–12, 2008.
- [11] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, “Devices for home detection of obstructive sleep apnea: A review,” *Sleep Med. Rev.*, vol. 41, pp. 149–160, Oct. 2018.
- [12] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, “A review of obstructive sleep apnea detection approaches,” *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 825–837, Mar. 2019.
- [13] S. S. Mostafa, F. Mendonça, Ravelo-García, and Morgado-Dias, “A systematic review of detecting sleep apnea using deep learning,” *Sensors*, vol. 19, no. 22, p. 4934, Nov. 2019.
- [14] M. B. Uddin, C. M. Chow, and S. W. Su, “Classification methods to detect sleep apnea in adults based on respiratory and oximetry signals: A systematic review,” *Physiol. Meas.*, vol. 39, no. 3, Mar. 2018, Art. no. 03TR01.
- [15] D. Álvarez, R. Hornero, D. Abásolo, F. D. Campo, and C. Zamarrón, “Non-linear characteristics of blood oxygen saturation from nocturnal oximetry for obstructive sleep apnoea detection,” *Physiol. Meas.*, vol. 27, no. 4, pp. 399–412, Apr. 2006.
- [16] J. V. Marcos, R. Hornero, D. Álvarez, F. D. Campo, and M. Aboy, “Automated detection of obstructive sleep apnoea syndrome from oxygen saturation recordings using linear discriminant analysis,” *Med. Biol. Eng. Comput.*, vol. 48, no. 9, pp. 895–902, Sep. 2010.
- [17] F. Roche, V. Pichot, E. Sforza, I. Court-Fortune, D. Duverney, F. Costes, M. Garet, and J.-C. Barthelemy, “Predicting sleep apnoea syndrome from heart period: A time-frequency wavelet analysis,” *Eur. Respiratory J.*, vol. 22, no. 6, pp. 937–942, Dec. 2003.
- [18] L. Chen, X. Zhang, and C. Song, “An automatic screening approach for obstructive sleep apnea diagnosis based on single-lead electrocardiogram,” *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 106–115, Jan. 2015.
- [19] R. Haidar, I. Koprinska, and B. Jeffries, “Sleep apnea event detection from nasal airflow using convolutional neural networks,” in *Proc. Int. Conf. Neural Inf. Process. (ICONIP)*, 2017, pp. 819–827.
- [20] M. M. Paul and M. Amithab, “SAHS detection based on ANFIS using single channel airflow signal,” *Int. J. Innov. Res. Sci. Eng. Technol.*, vol. 5, no. 7, pp. 13053–13061, 2016.
- [21] T. Praydas, B. Wongkittisuksa, and S. Tanthanuch, “Obstructive sleep apnea severity multiclass classification using analysis of snoring sounds,” in *Proc. 2nd World Congr. Electr. Eng. Comput. Syst. Sci.*, Jul. 2016, pp. 16–20.
- [22] D. Patil, V. M. Wadhai, S. Gujar, K. Surana, P. Devkate, and S. Waghmare, “APNEA Detection on Smart Phone,” *Int. J. Comput. Appl.*, vol. 59, no. 7, pp. 15–19, 2012.
- [23] A. Ravelo-García, J. Kraemer, J. Navarro-Mesa, E. Hernández-Pérez, J. Navarro-Esteve, G. Juliá-Serdá, T. Penzel, and N. Wessel, “Oxygen saturation and RR intervals feature selection for sleep apnea detection,” *Entropy*, vol. 17, no. 5, pp. 2932–2957, May 2015.
- [24] T. M. Cover, “The best two independent measurements are not the two best,” *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-4, no. 1, pp. 116–117, Jan. 1974.
- [25] S. S. Mostafa, F. Morgado-Dias, and A. G. Ravelo-García, “Comparison of SFS and mRMR for oximetry feature selection in obstructive sleep apnea detection,” *Neural Comput. Appl.*, 2018.
- [26] S. S. Mostafa, J. P. Carvalho, F. Morgado-Dias, and A. Ravelo-García, “Optimization of sleep apnea detection using SpO2 and ANN,” in *Proc. XXVI Int. Conf. Inf., Commun. Autom. Technol. (ICAT)*, Oct. 2017, pp. 1–6.
- [27] G. E. Hinton, S. Osindero, and Y. W. Teh, “A fast learning algorithm for deep belief nets,” *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
- [28] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, “Deep learning for visual understanding: A review,” *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.
- [29] S. Kiranyaz, T. Ince, and M. Gabbouj, “Real-time patient-specific ECG classification by 1-D convolutional neural networks,” *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, Mar. 2016.
- [30] E. Urtnasan, J.-U. Park, E.-Y. Joo, and K.-J. Lee, “Automated detection of obstructive sleep apnea events from a single-lead electrocardiogram using a convolutional neural network,” *J. Med. Syst.*, vol. 42, no. 6, p. 104, Jun. 2018.
- [31] P. Lakhan, A. Dithapron, N. Banluesombatkul, and T. Wilaiprasitporn, “Deep neural networks with weighted averaged overnight airflow features for sleep apnea-hypopnea severity classification,” 2018, *arXiv:1808.10845*. [Online]. Available: <http://arxiv.org/abs/1808.10845>
- [32] S. McCloskey, R. Haidar, I. Koprinska, and B. Jeffries, “Detecting hypopnea and obstructive apnea events using convolutional neural networks on wavelet spectrograms of nasal airflow,” in *Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining (PAKDD)*, vol. 2018, pp. 361–372.
- [33] L. Cen, Z. L. Yu, T. Kluge, and W. Ser, “Automatic system for obstructive sleep Apnea events detection using convolutional neural network,” in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 3975–3978.
- [34] E. Urtnasan, J.-U. Park, and K.-J. Lee, “Multiclass classification of obstructive sleep Apnea/Hypopnea based on a convolutional neural network from a single-lead electrocardiogram,” *Physiol. Meas.*, vol. 39, no. 6, Jun. 2018, Art. no. 065003.
- [35] D. Dey, S. Chaudhuri, and S. Munshi, “Obstructive sleep apnoea detection using convolutional neural network based deep learning framework,” *Biomed. Eng. Lett.*, vol. 8, no. 1, pp. 95–100, Feb. 2018.
- [36] S. H. Choi, H. Yoon, H. S. Kim, H. B. Kim, H. B. Kwon, S. M. Oh, Y. J. Lee, and K. S. Park, “Real-time Apnea-Hypopnea event detection during sleep by convolutional neural networks,” *Comput. Biol. Med.*, vol. 100, pp. 123–131, Sep. 2018.
- [37] R. Haidar, S. McCloskey, I. Koprinska, and B. Jeffries, “Convolutional neural networks on multiple respiratory channels to detect hypopnea and obstructive apnea events,” in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–7.
- [38] S. Albelwi and A. Mahmood, “A framework for designing the architectures of deep convolutional neural networks,” *Entropy*, vol. 19, no. 6, p. 242, 2017.
- [39] Z. You and Y. Pu, “The genetic convolutional neural network model based on random sample,” *Int. J. u-e-Service, Sci. Technol.*, vol. 8, no. 11, pp. 317–326, Nov. 2015.
- [40] E. Real, S. Moore, A. Selle, S. Saxena, Y. Leon Suematsu, J. Tan, Q. Le, and A. Kurakin, “Large-scale evolution of image classifiers,” 2017, *arXiv:1703.01041*. [Online]. Available: <http://arxiv.org/abs/1703.01041>
- [41] L. Xie and A. Yuille, “Genetic CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1388–1397.
- [42] S. R. Young, D. C. Rose, T. Johnston, W. T. Heller, T. P. Karnowski, T. E. Potok, R. M. Patton, G. Perdue, and J. Miller, “Evolving deep networks using HPC,” in *Proc. Mach. Learn. HPC Environ. (MLHPC)*, 2017.
- [43] F. Assunção, N. Lourenço, P. Machado, and B. Ribeiro, “Evolving the topology of large scale deep neural networks,” in *Proc. 21st Eur. Conf. Genetic Program.*, 2018, pp. 19–34.

- [44] A. Baldominos, Y. Saez, and P. Isasi, "Evolutionary convolutional neural networks: An application to handwriting recognition," *Neurocomputing*, vol. 283, pp. 38–52, Mar. 2018.
- [45] A. Baldominos, Y. Saez, and P. Isasi, "Evolutionary design of convolutional neural networks for human activity recognition in sensor-rich environments," *Sensors*, vol. 18, no. 4, p. 1288, Apr. 2018.
- [46] A. Martín, R. Lara-Cabrera, F. Fuentes-Hurtado, V. Naranjo, and D. Camacho, "EvoDeep: A new evolutionary approach for automatic deep neural networks parametrisation," *J. Parallel Distrib. Comput.*, vol. 117, pp. 180–191, Jul. 2018.
- [47] H. Tian, S. Pouyanfar, J. Chen, S.-C. Chen, and S. S. Iyengar, "Automatic convolutional neural network selection for image classification using genetic algorithms," in *Proc. IEEE Int. Conf. Inf. Reuse Integr. (IRI)*, Jul. 2018, pp. 444–451.
- [48] I. D. Falco, G. D. Pietro, G. Sannino, U. Scafuri, E. Tarantino, A. D. Cioppa, and G. A. Trunfio, "Deep neural network hyper-parameter setting for classification of obstructive sleep apnea episodes," in *Proc. IEEE Symp. Comput. Commun. (ISCC)*, Jun. 2018, pp. 01187–01192.
- [49] K. Deb, S. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, Apr. 2002.
- [50] S. S. Mostafa, N. Horta, A. G. Ravelo-García, and F. Morgado-Dias, "Analog active filter design using a multi objective genetic algorithm," *AEU-Int. J. Electron. Commun.*, vol. 93, pp. 83–94, Sep. 2018.
- [51] M. Li, S. Liu, L. Zhang, H. Wang, F. Meng, and L. Bai, "Non-dominated sorting genetic algorithms-II based on multi-objective optimization model in the water distribution system," *Procedia Eng.*, vol. 37, pp. 309–313, 2012.
- [52] S. H. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [53] T. Penzel, G. Moody, R. Mark, A. Goldberger, and J. Peter, "The apnea-ECG database," in *Proc. Comput. Cardiol.*, Sep. 2000, pp. 255–258.
- [54] *St. Vincent's University Hospital/University College Dublin Sleep Apnea Database*. Accessed: Jul. 18, 2018. [Online]. Available: <https://physionet.org/pn3/ucddb/>
- [55] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
- [56] J. S. J. Ren and L. Xu, "On vectorization of deep convolutional neural networks for vision tasks," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 1840–1846.
- [57] D. Stutz, "Understanding convolutional neural networks," *Nips*, vol. 2016, no. 3, pp. 1–23, 2014.
- [58] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.
- [59] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Proc. 20th Int. Conf. Artif. Neural Netw. (ICANN)*, Sep. 2010, pp. 92–101.
- [60] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 448–456.
- [61] S. Haykin, *Neural Networks: A Comprehensive Foundation*, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
- [62] A. Seshadri. (2009). *Multi-Objective Optimization using Evolutionary Algorithms*. MathWorks. Accessed: Sep. 20, 2017. [Online]. Available: <https://www.mathworks.com/matlabcentral/fileexchange/10351-multi-objective-optimizaion-using-evolutionary-algorithm>
- [63] S. M. R. Rafiei, M. H. Kordi, G. Griva, and H. Yassami, "Multi-objective optimization based optimal compensation strategies study for power quality enhancement under distorted voltages," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jul. 2010, pp. 3284–3291.
- [64] D. P. Kingma and J. L. Ba, "ADAM: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.
- [65] K. Deb and R. B. Agrawal, "Simulated binary crossover for continuous search space," *Complex Syst.*, vol. 9, no. 2, pp. 115–148, 1995.
- [66] R. K. Pathinarupothi, E. S. Rangan, E. A. Gopalakrishnan, R. Vinaykumar, and K. P. Soman, "Single sensor techniques for sleep apnea diagnosis using deep learning," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, Aug. 2017, pp. 524–529.
- [67] S. Biswal, H. Sun, B. Goparaju, M. B. Westover, J. Sun, and M. T. Bianchi, "Expert-level sleep scoring with deep neural networks," *J. Amer. Med. Inform. Assoc.*, vol. 25, no. 12, pp. 1643–1650, Dec. 2018.
- [68] E. Urtnasan, J.-U. Park, and K.-J. Lee, "Automatic detection of sleep-disordered breathing events using recurrent neural networks from an electrocardiogram signal," *Neural Comput. Appl.*, vol. 32, no. 9, pp. 4733–4742, May 2020.
- [69] B. Xie and H. Minn, "Real-time sleep apnea detection by classifier combination," *IEEE Trans. Inf. Technol. Biomed.*, vol. 16, no. 3, pp. 469–477, May 2012.
- [70] L. Almazaydeh, M. Faezipour, and K. Elleithy, "A neural network system for detection of obstructive sleep apnea through SpO2 signal features," *Int. J. Adv. Comput. Sci. Appl.*, vol. 3, no. 5, pp. 7–11, 2012.
- [71] S. S. Mostafa, F. Mendonca, F. Morgado-Dias, and A. Ravelo-Garcia, "SpO2 based sleep apnea detection using deep learning," in *Proc. IEEE 21st Int. Conf. Intell. Eng. Syst. (INES)*, Oct. 2017, pp. 91–96.



SHEIKH SHANAWAZ MOSTAFA (Graduate Student Member, IEEE) was born in Khulna, Bangladesh, in 1988. He received the B.S.Engg. degree from the University of Khulna, Bangladesh, in 2010, and the M.S. degree from the Khulna University of Engineering and Technology, Bangladesh, in 2012. He is currently pursuing the Ph.D. degree with the Instituto Superior Técnico, University of Lisbon, Portugal.

He worked in different research projects related to biomedical signal and image processing and hardware implementation. His research interests include biomedical signal processing, artificial neural networks, and hardware implementations.

Mr. Mostafa is a Graduate Student Member of the IEEE Signal Processing Society and the IEEE Circuits and Systems Society. He is also a member of the International Association of Engineers (IAENG) and the Bangladesh Computer Society.



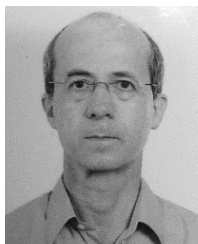
FÁBIO MENDONÇA was born in Funchal, Portugal, in 1990. He received the B.S. and M.S. degrees in electrical and telecommunications engineering from the University of Madeira, Portugal, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with the Instituto Superior Técnico, University of Lisbon, Portugal.

He was a Lecturer with the University of Madeira. He was a Researcher with the Madeira Interactive Technologies Institute. He is also a Researcher with the Laboratory for Robotics and Engineering Systems. His research interests include sleep monitoring, machine learning, and renewable energy.



ANTONIO G. RAVELO-GARCIA was born in Las Palmas de Gran Canaria, Spain, in 1973. He received the M.S. and Ph.D. degrees in telecommunication engineering from the Universidad de Las Palmas de Gran Canaria (ULPGC), Spain.

He is currently a Docent and a Researcher with the Department of Signal and Communications (DSC), Institute for Technological Development and Innovation in Communications (IDeTIC), ULPGC. He is also an International Collaborator Member of ITI/LARSyS (Portugal). He has participated in multiple research projects and publications. His current research interests include biomedical signal processing, nonlinear signal analysis, data mining, and intelligent systems in multimedia, communications, environment, health, and wellbeing.



GABRIEL JULIÁ-SERDÁ received the degree in medicine from the Autonomous University of Barcelona, the specialty of pulmonology from the Bellvitge Hospital, Barcelona, and the Ph.D. degree from the University of Las Palmas de Gran Canaria. He spent three years as a Postdoctoral Fellow with the Toronto General Hospital and the Mount Sinai Hospital, Toronto, ON, Canada, where he investigated issues related to airway physiology. He has worked as a Pulmonologist in several hospitals in Spain. He has been an Associate Professor of medicine with the University of Las Palmas. He is currently with the Dr. Negrín University Hospital of Las Palmas de Gran Canaria. He is also with the Perpetuo Socorro Clinic, Las Palmas. His research interests include physiology and respiratory diseases. He has made several publications both nationally and internationally on these topics.



FERNANDO MORGADO-DIAS (Member, IEEE) was born in Coimbra, Portugal, in 1971. He received the B.S. degree from the University of Aveiro, Portugal, in 1994, the M.S. degree from University Joseph Fourier, Grenoble, France, in 1995, and the Ph.D. degree from the University of Aveiro, in 2005.

He was a Lecturer with the Technical University of Setúbal, Portugal. He is currently an Assistant Professor with the Universidade da Madeira. He is also a Researcher with the Madeira Interactive Technologies Institute. He is also the Director of the Ph.D. degree in automation and instrumentation. His research interests include sleep monitoring, renewable energy, artificial neural networks, and FPGA implementations.

Dr. Morgado-Dias is also a member of the Fiscal Council of the Portuguese Association of Automatic Control.

...