KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Deviaene M., Borzee P., van Gilst M., van Dijk J., Overeem S., Buyse B., Testelmans D., Van Huffel S., Varon C.**Multilevel Interval Coded Scoring to Assess the Cardiovascular Status of Sleep Apnea Patients using Oxygen Saturation Markers** Internal Report 19-106, ESAT-STADIUS, KU Leuven (Leuven, Belgium), 2019. Accepted for publication in IEEE Transactions on Biomedical Engineering.
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	https://ieeexplore.ieee.org/document/8985287
Journal homepage	https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=10
Author contact	your email margot.deviaene@esat.kuleuven.be your phone number *+ 32 (0)16 32 89 60*
Abstract	Objective: Studies have shown an increased cardiovascular risk in obstructive sleep apnea (OSA) patients. In order to prioritize treatment of high risk patients, there is a need for improved cardiovascular OSA phenotyping. This study investigates the use of oxygen saturation (SpO2) parameters for cardiovascular risk assessment of OSA patients. To this end, a novel multilevel interval coded scoring (mICS) algorithm is proposed. Methods: The study includes SpO2 recordings from 1987 overnight polysomnographies, of which 974 are from patients suspected to have OSA, 931 from the general population based Sleep Heart Health Study and 83 from healthy controls. The minimal SpO2 value, SpO2 upslope and amplitude ratio of desaturation over resaturation are extracted for all oxygen desaturations and averaged per patient. These

	three SpO2 parameters are used together with patient demographics to develop a mICS model to predict the probability that a patient had a cardiovascular condition, or had already experienced a cardiovascular event, at the time of the polysomnography. Results: Including the SpO2 parameters in the mICS together with age and BMI improves the model's performance by 2.7% and leads to a test area under the curve (AUC) of 69.5% for the detection of any cardiovascular comorbidity. Moreover, an increase in AUC of 5% was obtained for the detection of cardiovascular events, resulting in an AUC of 93.5%. Conclusions: This study shows that parameters based on SpO2 and the mICS model are useful to predict the cardiovascular comorbidity status of OSA patients. Significance: The proposed model could be used to assist in prioritizing OSA patients for treatment.
IR	/

(article begins on next page)

Multilevel Interval Coded Scoring to Assess the Cardiovascular Status of Sleep Apnea Patients using Oxygen Saturation Markers

Margot Deviaene *Member, IEEE*, Pascal Borzée, Merel van Gilst, Johannes van Dijk *Senior Member, IEEE*, Sebastiaan Overeem, Bertien Buyse, Dries Testelmans, Sabine Van Huffel *Fellow, IEEE*, and Carolina Varon *Member, IEEE*

***Abstract*— Objective: Studies have shown an increased cardiovascular risk in obstructive sleep apnea (OSA) patients. In order to prioritize treatment of high risk patients, there is a need for improved cardiovascular OSA phenotyping. This study investigates the use of oxygen saturation (SpO_2) parameters for cardiovascular risk assessment of OSA patients. To this end, a novel multilevel interval coded scoring (mICS) algorithm is proposed. Methods: The study includes SpO_2 recordings from 1987 overnight polysomnographies, of which 974 are from patients suspected to have OSA, 931 from the general population based Sleep Heart Health Study and 83 from healthy controls. The minimal SpO_2 value, SpO_2 upslope and amplitude ratio of desaturation over resaturation are extracted for all oxygen desaturations and averaged per patient. These three SpO_2 parameters are used together with patient demographics to develop a mICS model to predict the probability that a patient had a cardiovascular condition, or had already experienced a cardiovascular event, at the time of the polysomnography. Results: Including the SpO_2 parameters in the mICS together with age and BMI improves the model's performance by 2.7% and leads to a test area under the curve (AUC) of 69.5% for the detection of any cardiovascular comorbidity. Moreover, an increase in AUC of 5% was obtained for the detection of cardiovascular events, resulting in an AUC of 93.5%. Conclusions: This study shows that parameters based on SpO_2 and the mICS model are useful to predict the cardiovascular comorbidity status of OSA patients. Significance: The proposed model could be used to assist in prioritizing OSA patients for treatment.**

***Index Terms*— Cardiovascular diseases, Clinical decision support system, Obstructive sleep apnea, Oximetry, Phenotyping.**

OBSTRUCTIVE sleep apnea (OSA) is the most common sleep related breathing disorder, characterized by repetitive partial or complete cessations of breathing, caused by a narrowing of the upper airway. These respiratory events often lead to an oxygen desaturation and/or end with an arousal [1]. Although these arousals enable patients to breathe again, they will also cause sleep fragmentation, leading to daytime sleepiness. Moreover, studies have linked these desaturations and arousals, together with intrathoracic pressure changes, to a long-term deterioration of the cardiovascular system and the presence of apnea-associated cardiovascular comorbidities [2]. Timely diagnosis and treatment of OSA is therefore of great importance in order to avoid the further development of these cardiovascular problems. In the study of Marin et al. it was suggested that continuous positive airway pressure (CPAP) treatment decreases the risk of cardiovascular morbidity and mortality [3].

According to current guidelines, diagnosis of OSA is mostly based on an overnight in-hospital polysomnography (PSG) which is manually scored by a sleep technologist [1]. As such, the hypnogram and respiratory event annotations are obtained, and the apnea-hypopnea index (AHI) is computed as the number of respiratory events divided by the hours of sleep. The diagnosis of sleep apnea is made when a patient has an AHI of 5 or higher in combination with symptoms, or an AHI larger than 15, regardless the presence of symptoms [4].

The AHI, however, only counts the number of events, and does not take into account the type of event, nor the severity of the event. The event severity could be defined in terms of duration, degree of oxygen desaturation or presence of an arousal. Another limitation of the AHI is the lack of

This work was supported in part by Bijzonder Onderzoeksfonds KU Leuven (BOF): The effect of perinatal stress on the later outcome in preterm babies: C24/15/036, Prevalentie van epilepsie en slaapstoornissen in de ziekte van Alzheimer: C24/18/097. Agentschap Innoveren en Ondernemen (VLAIO): 150466: OSA+. imec funds 2017. During the first phase of this study, C.Varon was a postdoctoral fellow of the FWO. The HealthBed study was supported by a grant from EIT Health (project no. 18453).

M. Deviaene, S. Van Huffel and C. Varon are with the Department of Electrical Engineering-ESAT, STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics, KU Leuven, Leuven B-3001, Belgium. (e-mail:margot.deviaene@esat.kuleuven.be; sabine. vanhuffel@esat.kuleuven.be; carolina.varon@esat.kuleuven.be). C.Varon is also with the Circuits and Systems (CAS) group, Delft University of Technology, the Netherlands.

M. Van Gilst, J. Van Dijk and S. Overeem are with the Department of Electrical Engineering, TU/e, Eindhoven, The Netherlands and the Kempenhaeghe Foundation, Sleep Medicine Centre, Heeze, The Netherlands. (e-mail: GilstM@Kempenhaeghe.nl; DijkH@Kempenhaeghe.nl; OvereemS@Kempenhaeghe.nl).

D. Testelmans, B. Buyse and P. Borzée are with the Department of Pneumology, Leuven University Centre for Sleep and Wake Disorders, UZ Leuven, Leuven 3000, Belgium (e-mail: dries.testelmans@uzleuven.be; bertien.buyse@uzleuven.be; pascal.borzee@uzleuven.be).

Copyright (c) 2017 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending an email to pubs-permissions@ieee.org.

discriminative ability, which was pointed out by the Hypnolaus study [5]. This study observed that in the general population up to 50 % of middle-aged men had an AHI larger than 15. Screening this population would thus lead to over-treatment, since many a-symptomatic subjects with an AHI larger than 15 might not benefit from CPAP treatment.

Therefore, there is a need to phenotype OSA patients beyond the AHI. Many different directions have been proposed to tackle this problem [6]. Studies have for example investigated whether clinical clusters could be defined based on clinical presentation, risk factors and comorbidities [7], or whether a differentiation could be made based on physiological traits causing OSA [8].

This study focused on the detection of OSA patients with an increased cardiovascular risk, in order to prioritize treatment for these patients and to avoid the further development of cardiovascular problems. The use of oxygen saturation (SpO_2) parameters was investigated, since the pulse oximetry signal can easily be acquired at home, and could thus be used for pre-hospital screening. Moreover, several studies have shown that oxygen desaturation severity parameters could be more predictive than the AHI as cardiovascular risk markers in OSA patients [9]–[12]. The study by Sommermeyer et al. [13], used in addition to SpO_2, pulse photoplethysmography (PPG) parameters, which also had a statistically significant association with an increased cardiovascular risk. In our previous study [14], the use of several SpO_2 and PPG parameters, extracted around oxygen desaturations, was investigated for the assessment of cardiovascular comorbidities in OSA patients. The obtained parameter set consisted of the most informative parameters, including 3 SpO_2 and 2 PPG parameters. The addition of PPG parameters, however, did not improve the final cardiovascular comorbidity model. Therefore, only the SpO_2 parameters will be further investigated in this study.

In order to obtain an interpretable and automated clinical decision support system, the interval coded scoring (ICS) [15] model was applied to this problem. Extensions to the ICS model were proposed in order to enable multilevel learning and to constrain the final feature scores to be monotonic. Moreover, statistical tests were used to analyse the differences between patient groups and the correlations between features. A dataset of healthy volunteers was added to have a reference of the SpO_2 parameter values for subjects without any sleep disorder, nor any cardiovascular comorbidity. Finally, the predictive capability of the model was tested on the Sleep Heart Health Study (SHHS) dataset [16], [17].

I. Methods

This section is organized as follows. First, the datasets are introduced, then the extraction of oxygen saturation parameters is presented. After that, the development of the proposed multilevel ICS model is presented. Next, its application to the assessment of cardiovascular comorbidities is described together with the statistical methods used to interpret the results. An overview of all methods can be found in Fig. 1.

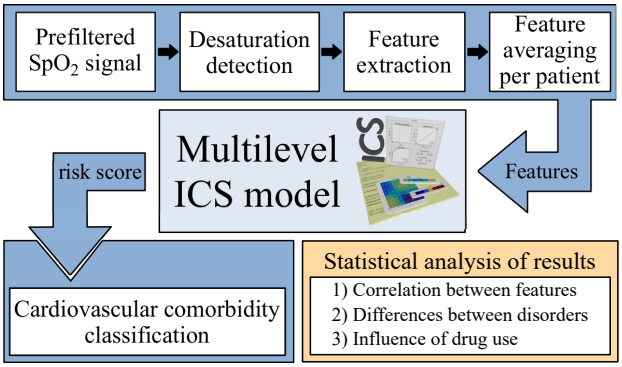


Fig. 1. Overview of the applied methods.

A. Datasets

This study was performed using clinical PSG recordings of 974 subjects suspected to be suffering from sleep apnea, a control dataset of 83 healthy volunteers and a subset of 931 patients from the publicly available SHHS dataset [16], [17]. The clinical dataset (Leuven) was recorded at the sleep laboratory of the University Hospitals Leuven. The healthy control dataset (H) was recorded at the Kempenhaeghe Center for Sleep Medicine in Heeze as part of the HEALTHBED study. These subjects did not have any sleep disorders, nor any cardiovascular comorbidities. The inclusion of the Leuven dataset to this study was approved by the ethical committee of UZ Leuven (S53746, S60319) and the inclusion of the H dataset was approved by the Kempenhaeghe ethical committee and by the medical ethical committee of the Maxima Medical Center Eindhoven (W17.128). Each subject signed an informed consent. The SHHS dataset contains PSG recordings of 5793 persons from the general population. For 2651 of them, also a follow-up PSG is available on average 5 years later, and their cardiovascular disease outcomes were monitored for up to 15 years. In this study, only the 931 subjects which had a follow-up PSG, but had no cardiovascular conditions at baseline were selected.

The SpO_2 signals were extracted from the overnight PSGs. These recordings had a length between 4.0 and 14.3 hours with an average recording length of 8.9 hours. The manually scored respiratory event annotations and hypnograms were available for each PSG. Moreover, the cardiovascular status of each patient was assessed at the time of the PSG. The presence of hypertension, hyperlipidemia, diabetes, atrial fibrillation and peripheral arterial disease was determined. This was done based on either the presence of the condition in the history of the patient, the patient's drug use, or the detection of the condition by means of blood pressure and blood analysis tests. In this study, all these conditions will be grouped in the cardiovascular "condition" (cC) category. The patient history was also examined for the occurrence of more severe cardiovascular "events" (cE), including myocardial infarction, heart failure and stroke. Subjects for whom none of the mentioned cardiovascular conditions were observed will be considered as "controls" (Ct). For the SHHS, the cardiovascular status at follow-up was taken as label, while the SpO_2 signals at baseline were investigated, in order to assess the predictive power of the developed model.

TABLE I
PATIENT DEMOGRAPHICS PER DATASET FOR THE CARDIOVASCULAR CONTROL (CT), CONDITION (CC) AND EVENT (CE) GROUPS.

			#	Age (years) mean ± SD	BMI (kg/m^2) mean ± SD	AHI mean ± SD
Patients	Leuven training	Ct	50	44.7 ± 14.2	29.4 ± 6.0	31.1 ± 23.5
		cC	50	49.5 ± 13.7	34.8 ± 9.4	40.9 ± 25.9
		cE	50	63.0 ± 11.7	32.4 ± 5.6	42.5 ± 22.2
	Leuven	Ct	118	47.1 ± 9.7	28.2 ± 4.8	32.9 ± 20.3
		cC	679	51.2 ± 11.2	31.4 ± 5.9	38.6 ± 21.4
		cE	26	62.9 ± 7.1	30.3 ± 4.1	38.5 ± 18.0
General population	SHHS	Ct	346	54.4 ± 11.1	26.3 ± 4.2	11.4 ± 11.1
		cC	495	61.4 ± 8.5	27.7 ± 4.8	15.8 ± 13.4
		cE	90	68.5 ± 9.5	27.1 ± 4.3	15.5 ± 11.9
Healthy	H	Ct	83	33.7 ± 12.7	23.9 ± 3.1	1.3± 1.3

From the Leuven dataset, 50 patients of each of the three cardiovascular categories were selected as training set (Leuven training) using k-medoids clustering on the age, BMI and gender [18]. The training subjects were selected as the centroids of the clusters. Using this approach, a training set is selected which represents better the underlying structure of the dataset in terms of age, BMI and gender and any bias is avoided. An overview of the patient demographics for each dataset and cardiovascular comorbidity category is given in Table I.

B. Oxygen Saturation Parameters

The sampling frequency of the available SpO_2 signals ranged from 1 Hz to 500 Hz, these were all down-sampled to 1 Hz. Afterwards the signals were pre-processed as described in [19]. Artefacts due to sensor disconnections were detected when the SpO_2 signal dropped below 50 %, and were removed. Next, a moving average filter of 3 seconds was applied to the signal to remove sharp changes and ripples due to oversampling. All oxygen desaturations were then detected in this pre-filtered signal using a derivative filter, as explained in [19].

Next, the three parameters selected in [14], which formed the optimal parameter set to detect the presence of cardiovascular comorbidities, were extracted for each of the desaturations. These parameters are:

a) Minimum SpO_2 value: The SpO_2 nadir has been shown to reach lower values for subjects at greater cardiovascular risk, as can also be seen from the example in Fig. 2 on top.

b) PRSA upslope: Using phase rectified signal averaging (PRSA) quasi-periodicities can be detected in the SpO_2 signal. A PRSA averaged resaturation (or desaturation) can be computed and the slope can be extracted from this averaged curve [20]. The method was applied to 5-minute windows around each oxygen desaturation. First all upward points (anchor points, AP) were detected as can be seen in the middle plot of Fig. 2, next segments of 10 seconds around each AP were extracted. These were then aligned, and the PRSA average was computed from these segments, as can be seen in the bottom plot of Fig. 2. In [14], it was shown that OSA patients with a cardiovascular comorbidity have an increased PRSA upward slope, which can also be seen from the example in Fig. 2.

c) Amplitude desat/resat: The ratio between the amplitude of desaturation and the amplitude of resaturation is the last parameter. This parameter is computed by dividing the percentage of SpO_2 drop by the percentage that the SpO_2 increased again after the nadir. An increase in this ratio is seen for patients with a cardiovascular comorbidity, which points to incomplete resaturations to baseline. This can be seen in the middle plot of the example in Fig. 2, where the SpO_2 baseline is plotted with a thick gray line. While the control patient has a stable SpO_2 baseline, the event patient has a decreasing SpO_2 baseline.

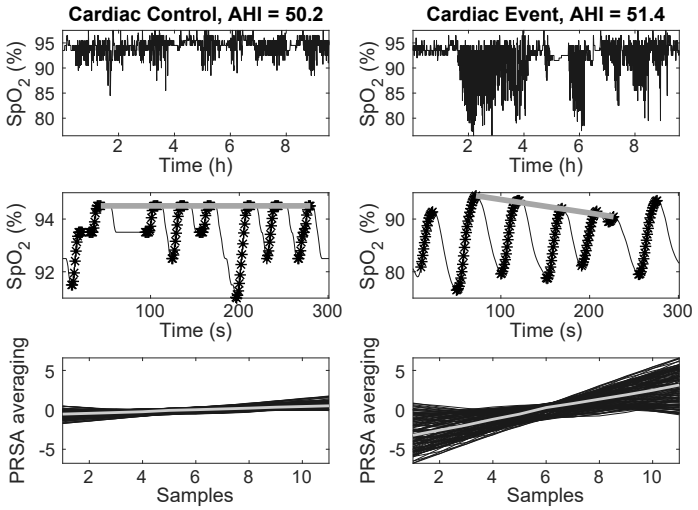


Fig. 2. Comparison of SpO_2 parameters between a cardiovascular control and event patient with similar AHI. On top, the filtered SpO_2 signal is plotted, in the middle a 5 minute segment is shown with the PRSA anchor points indicated by '*'. Additionally, a thick gray line shows the SpO_2 baseline in the middle plots. At the bottom, the aligned PRSA slices are shown for this segment, together with the PRSA averaged curve in light gray.

In order to obtain one value per patient, the median parameter value was taken over all desaturations with an SpO_2 drop from baseline larger than 2 %. In [14], the use of oxygen desaturation thresholds of 2, 3 and 4 % were investigated, and the most discriminative features were obtained with a threshold of 2 %. All desaturations in the recording were taken into account, independently of the sleep stage. The extracted parameters were then used to develop a classification model to assess the cardiovascular status per patient.

C. Interval Coded Scoring

The interval coded scoring model can be used to automatically develop interpretable risk score models for clinical decision support systems [15].

1) Standard ICS model: The algorithm starts by dividing all parameter values in a number of discrete intervals. An optimization is then applied to assign a score to each interval of every parameter. Next, the model is simplified by merging neighbouring intervals with similar scores using an iterative process. During this process, irrelevant parameters will be removed. In the end, the total risk score is obtained by summing the scores of the corresponding intervals for each relevant parameter. This risk score can then be linked to the empirical risk of having the studied condition, in this case an apnea-associated cardiovascular comorbidity.

In this study, the linear programming ICS model of the ICS toolbox [15] was used. This model can also use interactions between two parameters as input. A matrix with discrete intervals for both parameters is then created. Additionally, an elastic net parameter pre-selection method is available, which can remove irrelevant parameters and interactions before computing the ICS model to reduce computation time [15]. The ICS model optimization is a balance between model simplification and performance. The user can define relative area under the receiver operator characteristics (AUC) cut-off values as hyperparameters to guide the parameter pre-selection and interval merging steps. Model simplifications which result in a relative decrease in AUC smaller than these thresholds will be allowed.

2) Multilevel ICS (mICS) model: [1] The standard ICS implementation is based on the soft-margin Support Vector Machine (SVM) classifier optimization [21], but the maximal margin objective of SVM was replaced by total variation minimization [15]. This leads to the following linear programming problem [22]:

$$\min_{\tilde{\mathbf{w}},b,\mathbf{f},\boldsymbol{\varepsilon}} \quad \mathbf{f}^T\mathbf{1} + \gamma\boldsymbol{\varepsilon}^T\mathbf{1}$$
$$s.t. \quad \begin{cases} \mathbf{D}\tilde{\mathbf{w}} - \mathbf{f} \leq \mathbf{0} \\ -\mathbf{D}\tilde{\mathbf{w}} - \mathbf{f} \leq \mathbf{0} \\ \mathbf{Y}(\tilde{\mathbf{Z}}\tilde{\mathbf{w}} + b) \geq \mathbf{1} - \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} \geq \mathbf{0} \end{cases} \tag{1}$$

In this formulation, $\tilde{\mathbf{w}}$ and b represent the coefficients and the bias of the model, respectively. $\boldsymbol{\varepsilon}$ is the vector of slack variables used for regularization controlled by the regularization constant γ and $\mathbf{1} = [1; ...; 1]$. $\mathbf{Y}$ is a diagonal matrix, where $\mathbf{Y}_{ii} = y_i$, with y_i the class label of the i^{th} sample. $\tilde{\mathbf{Z}}$ contains the binary transformed data. $\mathbf{D}$ is a binary matrix which encodes the differences between coefficients of adjacent intervals of the same parameter or parameter interactions, and the vector $\mathbf{f}$ will be forced to be equal to $|\mathbf{D}\tilde{\mathbf{w}}|$.

This standard implementation only supports 2-class classification, with y_i consisting of binary labels. In this study, however, three cardiovascular severity categories were defined. Since there is an ordering between the three severity categories, a 3-class classification is possible by optimizing only two separating hyperplanes, as can be seen from the toy example in Fig. 3. The first hyperplane will, in this study, separate the Ct group from both the cC and cE groups, whereas the second hyperplane will separate the cE group from the Ct and cC groups. Additionally, the proposed mICS approach aims to output one final model which can capture the different levels of severity in the dataset using the same mICS scores. Therefore, the coefficients $\tilde{\mathbf{w}}$ will be kept the same for both hyperplanes, whereas the bias b will determine the difference between the two hyperplanes and two regularization constants γ will be optimized.

These adaptations will result in the following linear programming problem:

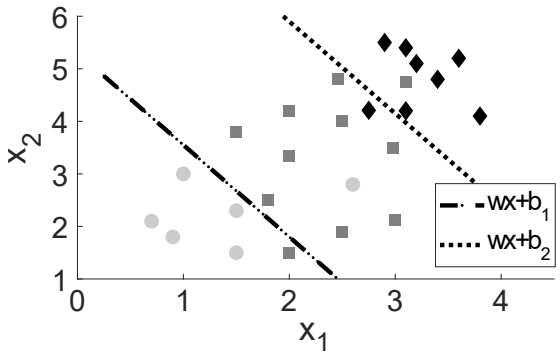


Fig. 3. Example of finding the optimal hyperplanes for a multilevel problem.

$$\min_{\tilde{\mathbf{w}},b_1,b_2,\mathbf{f},\boldsymbol{\varepsilon_1},\boldsymbol{\varepsilon_2}} \quad \mathbf{f}^T\mathbf{1} + \gamma_1\boldsymbol{\varepsilon_1}^T\mathbf{1} + \gamma_2\boldsymbol{\varepsilon_2}^T\mathbf{1}$$
$$s.t. \quad \begin{cases} \mathbf{D}\tilde{\mathbf{w}} - \mathbf{f} \leq \mathbf{0} \\ -\mathbf{D}\tilde{\mathbf{w}} - \mathbf{f} \leq \mathbf{0} \\ \mathbf{Y_1}(\tilde{\mathbf{Z}}\tilde{\mathbf{w}} + b_1) \geq \mathbf{1} - \boldsymbol{\varepsilon_1} \\ \mathbf{Y_2}(\tilde{\mathbf{Z}}\tilde{\mathbf{w}} + b_2) \geq \mathbf{1} - \boldsymbol{\varepsilon_2} \\ \boldsymbol{\varepsilon_1} \geq \mathbf{0} \\ \boldsymbol{\varepsilon_2} \geq \mathbf{0} \end{cases} \tag{2}$$

3) mICS model with monotonicity constraint: [1] In order to ensure monotonicity of the mICS interval scores per parameter, the constraints of the mICS weight optimization were adapted.

To impose monotonicity of the coefficients per parameter, the first two constraints of this problem will be adapted to ensure that the differences between coefficients always have the right sign. These constraints then become:

$$\begin{cases} \mathbf{D}\tilde{\mathbf{w}}\mathbf{S} \leq \mathbf{0} \\ -\mathbf{D}\tilde{\mathbf{w}}\mathbf{S} - \mathbf{f} = \mathbf{0} \end{cases} \tag{3}$$

The new vector $\mathbf{S}$ contains -1's for parameters that have a positive correlation with an increasing risk, and +1's if the parameters are negatively correlated with the risk.

4) Application to cardiovascular screening in OSA patients: The multilevel ICS model was trained using the Leuven training subset. The rest of the data was used as independent test set. The training was performed with and without inclusion of parameter interaction terms and parameter pre-selection, the AUC cut-offs for model simplification were varied from 85 % to 97.5 %. In order to obtain easy-to-interpret models, the constraints of the mICS optimization were adapted to ensure the monotonicity of interval scores per parameter. mICS models were trained with different parameter sets including the derived SpO_2 parameters and patient information, such as age, BMI, gender and AHI. Moreover, an estimation of the AHI based on automatic processing of the SpO_2 signal by the algorithm provided in [19] was also considered as input parameter for the mICS model.

The resulting mICS risk scores were then used to classify the subjects in the cardiovascular comorbidity categories. For this classification, the optimal mICS thresholds were defined as the point were the multiplication of sensitivity and specificity reached its maximum. This optimization was chosen in order to balance the sensitivity and specificity performance

[1]Both the multilevel and monotonicity constraint extension were added to the ICS toolbox [15] and will be made available upon publication.

measures, so that subjects from the positive and negative class have a similar chance to be correctly classified.

Since the Framingham risk score [23] was available for 843 subjects of the Leuven dataset, the results of the obtained mICS model were also compared with a classification based on this risk score.

D. Statistical Analysis

The obtained mICS risk scores and the extracted SpO_2 parameters were further analysed using statistical tests on the Leuven dataset. The correlations between the different parameters and patient demographics were quantified using the Spearman's rank correlation coefficient. This analysis was performed in order to study the influence that the patient demographics have on the extracted SpO_2 parameters.

Moreover, to have an idea about differences between groups of patients having different cardiovascular conditions the Kruskal-Wallis test with Bonferroni correction, with $\alpha = 0.05$, was performed between the different groups.

Many cC subjects were taking drugs for hypertension, diabetes or hyperlipidemia at the time of the PSG. In order to estimate the influence of these drugs on the extracted SpO_2 parameters, the differences between patients taking and not taking drugs were analysed, per condition, using the Kruskal-Wallis test with $\alpha = 0.05$. For 100 patients of the Leuven dataset, no information on drug usage was available, these patients were thus excluded from this analysis.

II. Results

A. Extracted SpO_2 parameters

Boxplots of the SpO_2 parameters per dataset and per cardiovascular comorbidity group are shown in Fig. 4. For the Leuven training and test set, the SpO_2 parameters follow the expected trends. Within the SHHS dataset, however, the differences between the cardiovascular groups are not visible, and most values are within the range of the Leuven Ct groups. The healthy population (H) is clearly separated from the clinical datasets and for all parameters these differences are statistically significant.

The correlation between features was tested using the Spearman's rank correlation coefficient. Only for the min SpO_2 parameter a weak correlation with the age was found, with a correlation coefficient of -0.28. This correlation, however, dropped when it was computed within each cardiovascular category, a value of -0.20 was, for example, found for the cC group. The BMI, on the other hand, had a weak correlation with the AHI (0.29) and the three SpO_2 parameters; the strongest correlation was found with the min SpO_2 (-0.39). The AHI also had moderate correlations with the SpO_2 parameters, having the strongest correlation with the PRSA upslope (0.60). Within the different SpO_2 parameters, the strongest correlation occurred between min SpO_2 and PRSA upslope (-0.64).

B. mICS Model

mICS models trained using different subsets of the extracted parameters and patient demographics were compared. When interaction terms were included in the model, they were removed during the model simplification step as they came out irrelevant. Therefore, further model optimization was performed without the inclusion of interaction terms. Since only 8 parameters were considered, the use of parameter pre-selection did not significantly reduce the computation time, nor increased the performance, therefore parameter pre-selection was not used. Different AUC cut-off values for model simplification were tested, setting this parameter to 97.5 % resulted in the best test set performances.

An overview of the training and test set performances to classify cC and cE subjects from Ct, on the different datasets for different subsets of features can be seen in Table II.

When only age is considered, cC and cE subjects can be detected with an Leuven test AUC of 56.0 % and a sensitivity of only 17.9 %. If only cE and Ct subjects are considered, an AUC_{cE} of 74.0 % can be achieved. This model is very simple, since only two age intervals remained after the optimization procedure. As a result, patients older than 63 years were classified in the cardiovascular event group. However, due to the simplicity of the model, a poor generalization was achieved and no differentiation of mICS risk threshold could be observed between the cC and cE groups. The performance improved when the BMI was added to the model, an overall Leuven test AUC of 66.8 % was obtained and the sensitivity increased to 40.7 %.

The model trained only on the SpO_2 parameters has some predictive value with an overall test AUC of 61.1 % and an AUC_{cE} of 81.7 %. But adding the age to this model gave increases in performance of, respectively, 4 % and 9.5 %. Adding only the BMI to the SpO_2 parameters did not improve the performance. The best model performance was obtained when combining the SpO_2 parameters with the age and BMI. In this case, the overall Leuven test AUC increased to 69.5 % and cE subjects could be detected with an AUC_{cE} of 93.5 %. Next to AUC_{cE}, Table II also provides the AUC_{cC} for the classification between cC and Ct subjects, these are about 20 % lower than AUC_{cE} for the Leuven test set.

Classifying subjects based on the AHI or the SpO_2 based estimate of the AHI gave poor results, AUCs of, respectively, 51.4 and 52.2 % were obtained. Adding the AHI to the models mentioned above also did not improve performances. In the last rows of Table II, the classification results based on the Framingham risk score are presented for the Leuven datasets. An optimal risk score cut-off value of 12.2 was obtained based on the training data. With a test AUC of 80.9 %, classification based on the Framingham score outperforms our proposed model, especially when focusing on the cC subjects. But the computation of the Framingham score already includes some of the cardiovascular conditions; the presence of diabetes is included as well as the cholesterol and blood pressure levels. Moreover, if the necessary parameters would be available to compute the Framingham score, it can be combined with the proposed mICS risk model. Classification based on the multiplication of the mICS risk score and the Framingham score resulted in an increase in overall test AUC to 81.1 % and a AUC_{cE} of 96.9 %.

Table II also includes the performance scores on the SHHS

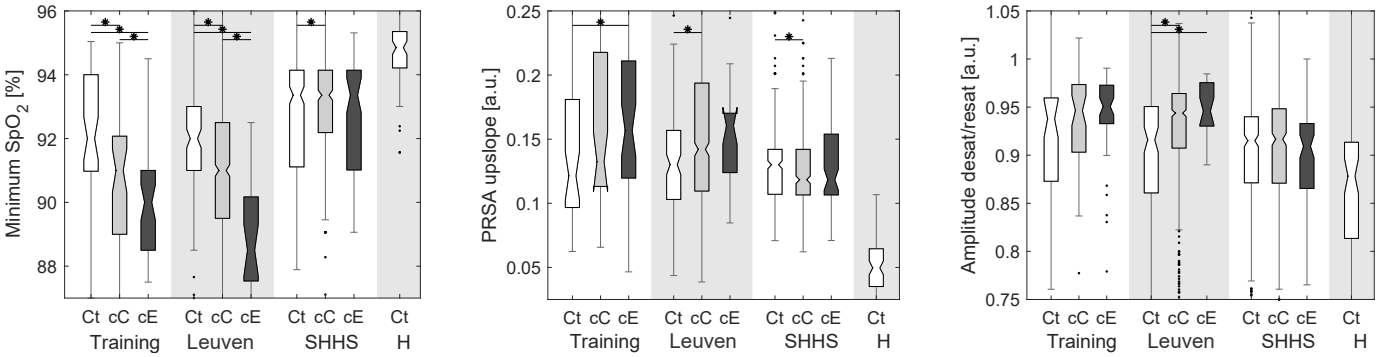


Fig. 4. Overview of the extracted SpO_2 parameters for the different cardiovascular comorbidity groups (cardiovascular controls (Ct), conditions (cC) and events (cE)). Statistically significant differences between groups are marked with * $p < 0.05$.

TABLE II

CLASSIFICATION PERFORMANCE OF THE DIFFERENT MICS MODELS AND FRAMINGHAM SCORE FOR THE DETECTION OF CARDIOVASCULAR COMORBIDITIES (CT VS CC + CE). AUC_{cC} AND AUC_{cE} DEPICT, RESPECTIVELY, THE AUC OF THE SEPARATION BETWEEN THE CT AND CC CLASSES, AND THE CT AND CE CLASSES.

	Training					Independent Test Sets									
						Leuven					SHHS				
	Se	Sp	AUC	AUC_{cC}	AUC_{cE}	Se	Sp	AUC	AUC_{cC}	AUC_{cE}	Se	Sp	AUC	AUC_{cC}	AUC_{cE}
Age	33.0	90.0	61.5	52.0	71.0	17.9	94.1	56.0	55.3	74.0	45.0	76.3	60.6	58.4	73.2
Age+BMI	62.0	86.0	78.1	68.7	87.6	40.7	87.3	66.8	66.0	88.5	58.3	70.8	69.5	68.2	76.5
Age+BMI+AHI	44.0	88.0	73.7	65.6	81.8	25.5	92.4	64.1	63.4	82.7	39.7	84.4	65.3	64.3	70.7
SpO_2	48.0	82.0	70.2	63.2	77.2	37.9	83.9	61.9	61.1	81.7	28.4	82.9	52.9	53.7	48.6
SpO_2+Age	70.0	82.0	81.6	73.9	89.3	45.0	82.2	65.9	64.9	91.2	46.8	76.3	65.2	64.3	70.0
SpO_2+Age+BMI	**72.0**	**78.0**	**82.8**	**74.9**	**90.7**	**47.5**	**80.5**	**69.5**	**68.6**	**93.5**	**44.6**	**78.3**	**67.1**	**66.3**	**72.0**
SpO_2+Age+BMI+AHI	72.0	78.0	83.0	75.5	90.4	48.5	80.5	69.0	68.1	93.3	47.9	76.3	68.9	67.5	76.9
AHI	100.0	14.0	57.0	57.0	57.0	97.7	5.1	51.4	51.4	52.5	75.2	39.0	57.1	56.8	59.0
AHI SpO_2	96.0	22.0	59.0	59.0	59.0	93.5	11.0	52.2	52.1	55.5	81.2	32.4	56.8	56.4	59.0
Framingham	65.0	88.0	81.8	77.0	87.1	52.8	92.4	80.9	80.4	94.9	Not available				
Framingham+mICS	69.0	88.0	84.6	78.6	91.1	53.2	90.7	81.1	80.5	96.9					

dataset, a clear drop in sensitivity and AUC can be seen for the SpO_2 based model. This could be expected, since the SpO_2 features did not show any differences in Fig. 4.

The final model including SpO_2 parameters, age and BMI is shown in Fig. 5. All three SpO_2 parameters came out the mICS model optimization as relevant and were included in the final model.

The resulting mICS risk scores for the different cardiovascular groups and datasets are shown in Fig. 6. A clear separation can be seen between the three groups for the Leuven training and test sets. The Ct, cC and cE groups have, respectively, median values of 0.29, 0.43-0.57 and 0.70-0.81. For the SHHS dataset the Ct and cC values lie within similar ranges as the Leuven data, but the cE group does not have an increased risk with respect to cC. This lack of increased risk could also be extracted from the similar AUC_{cC} and AUC_{cE} in Table II. The H group has very small risk scores, with a median value of 0.07.

The correlation between the mICS risk score and the Framingham risk score was tested. A correlation of 0.66 was obtained, which was mainly due to the fact that both risk scores highly depend on the age.

The resulting confusion matrix for 3 class classification on the Leuven test set is plotted in Fig. 7. Again a clear separation between the Ct and cE groups is seen, but the cC overlaps with both groups.

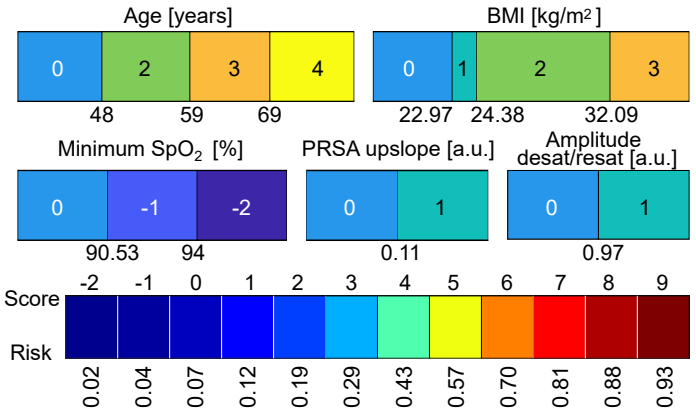


Fig. 5. Resulting mICS model including SpO_2 parameters, age and BMI.

C. Differences between cardiovascular disorders

The obtained mICS risk scores were compared between the different cardiovascular disorders in order to investigate whether differences between disorders could be seen. Firstly, the cardiovascular conditions were compared, 6 groups of patients were formed: 156 subjects which only had hypertension were compared with 120 with only hyperlipidemia, 307 which had both hypertension and hyperlipidemia, 95 subjects which had diabetes amongst other conditions, 17 subjects with

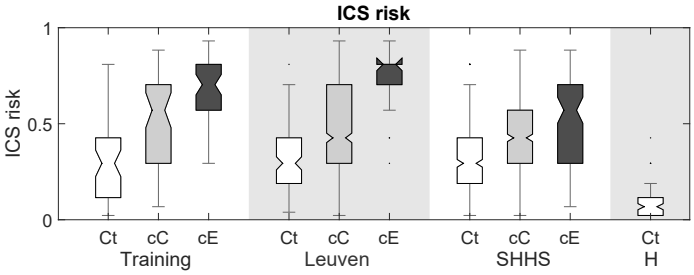


Fig. 6. mICS model risks for the different datasets and cardiovascular groups; control (Ct), condition (cC) and event (cE).

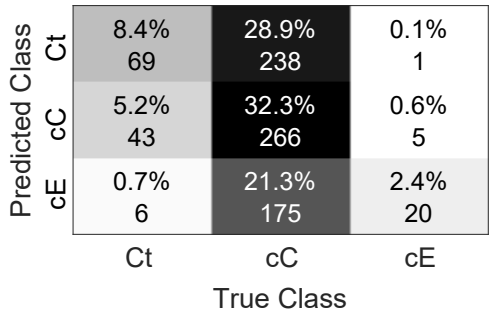


Fig. 7. Confusion matrix of classifying Leuven test subjects using the mICS model including SpO_2 parameters, age and BMI.

arterial disease and 19 with atrial fibrillation (AF). Statistically significant differences in mICS risk scores were found between these groups of patients. The diabetic and AF patients, on the one hand, had an increased risk correlating to a statistically significant decrease in min SpO_2. The other two SpO_2 parameters also display a tendency towards higher values for these patients, but these differences were not statistically significant. Diabetic patients also had a significant increase in BMI with respect to the other groups, whereas the AF population had an increased age. The patients with hyperlipidemia, on the other hand, had a decreased mICS risk, which was caused by a statistically significant decrease in age. The SpO_2 parameters did not differ between the hyperlipidemia, hypertension and hyperlipidemia plus hypertension groups. Finally, a tendency towards an increased risk for the arterial disease group could be observed, which was caused by an increase in age.

Within the cardiovascular events, 4 groups could be compared: 36 subjects had a myocardial infarction, 19 a heart failure, 15 a stroke and 6 subjects had more than one type of event. No statistical significant differences could, however, be observed between these four groups.

D. Influence of drug use

The influence of drug use on the mICS risk score model was assessed by comparing those who did not take cardiovascular drugs to those with the same disorder, who were taking drugs. A statistically significant increase in age was found for both hypertension and hyperlipidemia patients taking drugs. This increase was also observed for patients suffering from both disorders which were taking drugs for either disorder. These age increases resulted in an increased mICS risk. For the hyperlipidemia group, no statistically significant differences in SpO_2 parameters could be observed between both groups. Subjects taking anti-hypertensive drugs, however, had a statistically significant decrease in minimum SpO_2 values, independently of the fact whether they were taking drugs for hyperlipidemia. This analysis could not be performed for anti-diabetic drugs, since in this group many patients were taking multiple types of drugs.

III. Discussion

This study developed an interpretable risk score model to assess the cardiovascular status of OSA patients based on SpO_2 parameters and patient demographics. An extension to the ICS model was proposed, in which multilevel classification is possible and monotonicity of the mICS interval scores per parameter is imposed. These extensions can be used in other applications including different levels of severity.

The novel SpO_2 parameters, proposed by [14] for the assessment of cardiovascular comorbidities in OSA patients, and the obtained results, were further analysed using statistical tests. A set of healthy recordings, was added to the analysis in order to get a reference on how the values of these SpO_2 parameters behave in the general population. The model was also applied to a subset of the SHHS in order to have an idea on the predictive capabilities of the mICS model. This section will further discuss all of the obtained results.

A. SpO_2 parameters and cardiovascular comorbidities

The correlation analysis of the SpO_2 parameters and the patient demographics revealed that the minimum SpO_2 value was mainly correlated with the BMI. This relationship also became clear in the analysis of the different conditions, in which patients having diabetes had both an increased BMI and decreased minimum SpO_2 values. This relationship between the BMI and the severity of oxygen desaturation was previously shown by Peppard et al. [24]. The PRSA upslopes, on the other hand, were more correlated with the AHI. This relationship can be expected since the PRSA searches for quasi-periodicities in the SpO_2 signal, and the PRSA averaged curve will be more pronounced if multiple oxygen desaturations occur within the 5 minute segment under investigation.

The analysis on the influence of drug use showed a statistically significant decrease in minimum SpO_2 for patients taking anti-hypertensive drugs. This group of patients, however, also had a non-significant increase in BMI, which could have affected the minimum SpO_2 instead of the drug use, since the BMI is correlated with the minimum SpO_2. Therefore, no conclusive effects of drug use on the SpO_2 parameters, independently of changes in patient demographics were found in this study. Another conclusion that came out of this analysis, is that patients treated for cardiovascular conditions tended to be older. This is in accordance to the Hypertension and Diabetes Risk Screening and Awareness study [25], which observed that older hypertensive patients tended to be more often diagnosed and treated than younger hypertensive patients. The severity of the disorder, measured by blood pressure and blood analysis, could not be linked to this difference, since these measurements were only available at the time of the PSG,

when patients were already taking drugs. Therefore, patients on drugs had a lower observed cardiovascular disease severity.

Studies have shown that OSA patients have more severe oxygen desaturations during rapid eye movement (REM) sleep [26]. The presented results were, however, obtained by averaging the SpO_2 parameters over all desaturations, independent of the sleep stage. An additional analysis of the influence of sleep stages, nevertheless, had shown no increase in performance when REM and non-REM sleep were separated. It is thus not necessary to differentiate between REM and non-REM desaturations and no sleep staging is required for the algorithm. Additionally, the ratio of the number of REM over non-REM events was also not significantly different between the cardiovascular comorbidity categories.

B. mICS model for cardiovascular comorbidity assessment

The final mICS model obtained an overall Leuven test AUC of 69.5 % for the differentiation between patients without any cardiovascular disease Ct and those with a cardiovascular comorbidity. Adding the SpO_2 parameters gave an improvement of 2.7 % compared with the model using only age and BMI. For the detection of cE patients, an improvement in performance of 5 % was obtained, whereas for the detection of cC patients, the increase in performance was only 2.6 % in AUC_{cC}. This improvement in performance might seem limited, but the SpO_2 parameters showed a clear correlation with the cardiovascular comorbity classes for the Leuven dataset in Fig. 4, while only weak correlations between these parameters and age or BMI were found. These results indicate that the SpO_2 parameters have an added value compared to the age and BMI. Since the cardiovascular comorbidity classes are not matched for age and BMI, the obtained model is highly dependent on those parameters. If these mICS models would be tested on a more balanced population in terms of age and BMI, the SpO_2 parameters might provide larger increases in performance. From a cardiovascular prevention view, the focus would ideally go to the youngest OSA subjects. In this subpopulation it is, however, hard to find subjects with detectable cardiovascular comorbidities.

The age dependency of the model also leads to the fact that, in general, misclassified Leuven Ct patients were significantly older than correctly classified Leuven controls. Especially, it should be noted that subjects older than 69 years and with a BMI higher than 32.1 have a minimum mICS score of 5, resulting in a predicted risk of 57%. Therefore, all these subjects were classified in the cE group. In order to have a better differentiation within this demographic group, more Ct patients should be included in this range. On the other hand, the model will always classify subjects younger than 48 years and with a BMI smaller than 24.4 in the Ct group. For this age and BMI range, no cE subjects were available. The performance to classify cE from Ct subjects was recomputed for all subjects lying outside these two demographic ranges, an overall Leuven test AUC of 67.2 % was obtained, with an AUC_{cE} and AUC_{cC} of, respectively, 92.3 and 66.2 %. Also an increase in AHI can be seen for misclassified controls and they tended to reach lower SpO_2 values.

Moreover, when analysing the signals of 7 Ct patients in the Leuven dataset, a large number of ectopic beats could be observed in the electrocardiogram signals of these patients. Although these patients were labeled as cardiovascular controls, they could already have been far in the development of a cardiovascular comorbidity, which was still unknown at the moment the PSG was recorded. These subjects had mICS risk scores ranging from 0.12 to 0.70, with a median value of 0.57, while the Ct group had for this dataset only a median risk of 0.29. An increased cardiovascular risk could thus already be detected.

In total there were 6 cardiovascular event patients misclassified, one myocardial infarction patient was classified in the Ct group. The other five patients were classified in the cC group, two had a myocardial infarction, two heart failure and one subject had a stroke.

C. Predictive power of the obtained mICS model

A drawback of the Leuven datasets, is the fact that the patients only had an assessment of cardiovascular comorbidities at the time of the PSG. No follow-up information was available, this means that cardiovascular control patients could have developed a cardiovascular comorbidity shortly after their PSG, as could have been the case for the 7 controls with ectopic beats as discussed above. Therefore, the publicly available Sleep Heart Health Study (SHHS) was included in this study. This test had an overall AUC of 67.1 % to predict whether cardiovascular control subjects were going to develop a cardiovascular comorbidity or not. This performance is, however, lower than the one on the Leuven set. Moreover, when the SpO_2 parameters were investigated, no clear separation for the SHHS groups could be observed. Especially, when looking to the minimum SpO_2 parameter, for which the three cardiovascular groups, Ct, cC and cE had median values of 93 %. This parameter gave the clearest separation for the Leuven dataset, but for the SHHS it does not provide any separation. These values are all in the range of the Leuven Ct groups. This points to the fact that the subjects included in the SHHS have less severe apnea related oxygen desaturations, and thus have no increased cardiovascular risk reflected on the SpO_2 parameters. This could be due to the fact that the cardiovascular comorbidity was not present at baseline yet, and that the SpO_2 parameters might therefore still be similar between the different groups. The first cardiovascular event could have occurred up to 15 years after the PSG which was used to extract the SpO_2 parameters. It was tested whether limiting the time between the PSG and the occurrence of cardiovascular events improved the results. No significant differences were, however, observed when only the 24 cE subjects were included which had a cardiovascular event within 5 years of their baseline PSG.

It should also be taken into account that the SHHS dataset contains subjects from the general population, which often did not have any sleep related complaints. Even when the population with an AHI larger than 15 is considered, the minimum SpO_2 values do not decrease. Moreover, the mean Epworth sleepiness scale (ESS) [27] is only 7.7 for this subgroup,

while only a score above 10 is considered abnormal. For comparison, the average ESS for the Leuven dataset is 10.4. Therefore, the developed cardiovascular comorbidities might be less related to the sleep apnea events in the SHHS than in the Leuven population, thus resulting in no abnormal SpO_2 parameters. Additionally, the SHHS subjects are significantly older than the Leuven population. The average age at baseline was 54.4 years for the controls of the SHHS, which is 10 years older than in the Leuven dataset. This will further complicate the differentiation between the cardiovascular groups since studies suggest that although the overall cardiovascular risk increases with age, the cardiovascular risk associated with OSA decreases with age [28]. This decreased association might be explained by a healthy survivor effect, in which OSA subjects with a high cardiovascular risk are unlikely to be alive or without any cardiovascular condition from certain ages on.

These results show that one should be careful with diagnosing OSA based on AHI alone, in order to avoid over-diagnosis and over-treatment of patients without complaints that might not have an increased long-term risk. The proposed parameters could therefore assist to prioritize treatment for those patients which have more severe SpO_2 features, linked to a higher cardiovascular risk.

D. Influence of training set selection on the performance

The training set was selected based on the age, BMI and gender of the subjects, it was tested which performances could be obtained when the selection of training subjects was performed at random. 10 iterations were run, and the detection AUC on the Leuven test set ranged from 59.5 to 70.4 %, with an average of 68.8 %. Moreover, for the AUC_{cE} a maximum of 90.1 % was obtained, while our classifier obtains an AUC_{cE} of 93.5 %. This shows that the intelligent selection of training samples using k-medoids clustering increases the performance of the model. Including subjects from the SHHS in the training set resulted in a decrease in performance, so this was not considered further.

Moreover, 10-fold cross validation was tested on the Leuven dataset. Within each fold, an equal amount of Ct, cC and cE subjects were selected for training using the k-medoids algorithm, in order to ensure class balance. An averaged AUC and AUC_{cE} of, respectively, 67.0 and 85.8 % were obtained when using the age, BMI and all SpO_2 features. When only taking into account the age and BMI these performances dropped to 65.1 and 81.5 %. These results have a drop in performance, but show similar trends when compared to the results presented in Table II using a fixed test set. Additionally, when using cross validation slight changes in the model were observed between different folds. Therefore, the manuscript was focused on the final ICS model obtained with a fixed training set. If more data would be available, the cross validation models would become more stable.

E. Limitations of the study

One of the drawbacks of the used Leuven dataset is the lack of patient follow-up, as discussed above. Nevertheless, this dataset represents the patients that are referred to the sleep lab, which have OSA-like complaints. This is the most interesting population to develop a cardiovascular phenotyping of OSA patients. Since studies have shown that symptomatic OSA patients have a higher risk to develop cardiovascular events, when compared to a general population dataset, as the SHHS, in which subjects can have a-symptomatic OSA with an AHI larger than 15 and a cardiovascular problem, but these two problems might be unrelated [10], [29]. Moreover, future follow-up studies could also take into account cardiovascular conditions and treatments that are present at baseline and track if the cardiovascular comorbidites worsen. When doing so, the comorbidities and treatments present at baseline could be added as input to the classification model.

Despite the large number of recordings that were available for this study, there was a large gap in patient demographics, such as age and BMI, between cardiovascular control and event patients. Patient matching would have greatly impacted the size of our dataset, and this dataset represents the patient population that is recorded at the sleep lab in clinical practice. Therefore, it was decided not to correct for patient demographics, but to take them into account in the definition of the risk score, in order to obtain a model that is useful in daily clinical practice. The H subjects used as reference also differed substantially in age and BMI from the Leuven datasets.

As future work, the addition of novel PPG features could be investigated as well as new measurement and artefact detection techniques to improve the reliability of PPG features. The mICS model could be compared with other interpretable classification models as well. Moreover, if more data would be available, an indepth analysis towards the influences of cardiovascular diseases and drug use on the risk score could be performed.

The obtained cardiovascular risk model for OSA patients showed good performances, but should be tested further on clinical datasets with cardiovascular follow-up, before it can be used in clinical practice for OSA phenotyping. Since the used parameters are extracted from the SpO_2 signal, which can be easily acquired at home, multiple night testing could also be investigated in the future.

IV. Conclusion

This study showed that parameters based on SpO_2 are of added value to predict the cardiovascular comorbidity status of OSA patients. A multilevel interval coded scoring model was constructed based on age, BMI and these SpO_2 parameters. This model was able to predict whether a subject had any cardiovascular comorbidity with a test AUC of 69.5 %, and if (s)he suffered from a cardiovascular event or not with an AUC of 93.5 %. The resulting mICS risk scores of cardiovascular condition patients showed an increased risk, but lower than those of subjects who experienced a cardiovascular event. This model could be useful in clinical practice to estimate the cardiovascular risk of sleep apnea patients, and to do a screening to prioritize patients for treatment.

REFERENCES

[1] R. B. Berry et al, "Rules for scoring respiratory events in sleep: update of the 2007 aasm manual for the scoring of sleep and associated events," *J Clin Sleep Med*, vol. 8, no. 5, pp. 597–619, 2012.

[2] T. D. Bradley and J. S. Floras, "Obstructive sleep apnoea and its cardiovascular consequences," *The Lancet*, vol. 373, no. 9657, pp. 82–93, 2009.

[3] J. M. Marin et al, "Long-term cardiovascular outcomes in men with obstructive sleep apnoea-hypopnoea with or without treatment with continuous positive airway pressure: an observational study," *The Lancet*, vol. 365, no. 9464, pp. 1046–1053, 2005.

[4] M. J. Sateia, "International classification of sleep disorders," *Chest*, vol. 146, no. 5, pp. 1387–1394, 2014.

[5] R. Heinzer et al, "Prevalence of sleep-disordered breathing in the general population: the hypnolaus study," *The Lancet Respiratory Medicine*, vol. 3, no. 4, pp. 310–318, 2015.

[6] A. V. Zinchuk et al, "Phenotypes in obstructive sleep apnea: a definition, examples and evolution of approaches," *Sleep medicine reviews*, vol. 35, pp. 113–123, 2017.

[7] S. Bailly et al, "Obstructive sleep apnea: a cluster analysis at time of diagnosis," *PLoS One*, vol. 11, no. 6, p. e0157318, 2016.

[8] D. J. Eckert, "Phenotypic approaches to obstructive sleep apnoea-new pathways for targeted therapy," *Sleep Medicine Reviews*, 2016.

[9] E. Watanabe et al, "Prognostic importance of novel oxygen desaturation metrics in patients with heart failure and central sleep apnea," *Journal of Cardiac Failure*, 2016.

[10] T. Kendzerska et al, "Obstructive sleep apnea and risk of cardiovascular events and all-cause mortality: a decade-long historical cohort study," *PLoS medicine*, vol. 11, no. 2, p. e1001599, 2014.

[11] J. D. Gottlieb et al, "Hypoxia, not the frequency of sleep apnea, induces acute hemodynamic stress in patients with chronic heart failure," *Journal of the American College of Cardiology*, vol. 54, no. 18, pp. 1706–1712, 2009.

[12] A. Azarbarzin et al, "The hypoxic burden of sleep apnoea predicts cardiovascular disease-related mortality: the osteoporotic fractures in men study and the sleep heart health study," *European heart journal*, 2018.

[13] D. Sommermeyer et al, "The use of overnight pulse wave analysis for recognition of cardiovascular risk factors and risk: a multicentric evaluation," *Journal of hypertension*, vol. 32, no. 2, pp. 276–285, 2014.

[14] M. Deviaene et al, "Pulse oximetry markers for cardiovascular disease in sleep apnea," in *2019 Computing in Cardiology (CinC)*. IEEE, 2019, pp. 1–4.

[15] L. Billiet et al, "Interval coded scoring: a toolbox for interpretable scoring systems," *PeerJ Computer Science*, vol. 4, p. e150, 2018.

[16] S. F. Quan et al, "The sleep heart health study: design, rationale, and methods," *Sleep*, vol. 20, no. 12, pp. 1077–1085, 1997.

[17] S. Redline et al, "Methods for obtaining and analyzing unattended polysomnography data for a multicenter study," *Sleep*, vol. 21, no. 7, pp. 759–767, 1998.

[18] C. Varon et al, "Noise level estimation for model selection in kernel pca denoising," *IEEE transactions on neural networks and learning systems*, vol. 26, no. 11, pp. 2650–2663, 2015.

[19] M. Deviaene et al, "Automatic screening of sleep apnea patients based on the spo 2 signal," *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 607–617, 2019.

[20] A. Bauer et al, "Phase-rectified signal averaging detects quasi-periodicities in non-stationary data," *Physica A: Statistical Mechanics and its Applications*, vol. 364, pp. 423–434, 2006.

[21] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013.

[22] L. Billiet, "Clinical decision support: Interpretability and applications in patient monitoring," 2018.

[23] R. B. D'agostino et al, "General cardiovascular risk profile for use in primary care," *Circulation*, vol. 117, no. 6, pp. 743–753, 2008.

[24] P. E. Peppard et al, "The impact of obesity on oxygen desaturation during sleep-disordered breathing," *American journal of respiratory and critical care medicine*, vol. 180, no. 8, pp. 788–793, 2009.

[25] A. M. Sharma et al, "High prevalence and poor control of hypertension in primary care: cross-sectional study," *Journal of hypertension*, vol. 22, no. 3, pp. 479–486, 2004.

[26] J. Krieger et al, "Respiratory effort during obstructive sleep apnea: role of age and sleep state," *Chest*, vol. 112, no. 4, pp. 875–884, 1997.

[27] M. W. Johns, "A new method for measuring daytime sleepiness: the epworth sleepiness scale," *sleep*, vol. 14, no. 6, pp. 540–545, 1991.

[28] D. J. Gottlieb et al, "Prospective study of obstructive sleep apnea and incident coronary heart disease and heart failure the sleep heart health study," *Circulation*, vol. 122, no. 4, pp. 352–360, 2010.

[29] Y. Endeshaw et al, "Snoring, daytime sleepiness, and incident cardiovascular disease in the health, aging, and body composition study," *Sleep*, vol. 36, no. 11, pp. 1737–1745, 2013.