Computer Methods and Programs in Biomedicine 226 (2022) 107120

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

ABOSA – Freely available automatic blood oxygen saturation signal analysis software: Structure and validation

Tuomas Karhu[a,b,*], Timo Leppänen[a,b,c], Juha Töyräs[a,c,d], Arie Oksenberg[e], Sami Myllymaa[a,b], Sami Nikkonen[a,b]

[a] Department of Applied Physics, University of Eastern Finland, Kuopio, Finland
[b] Diagnostic Imaging Center, Kuopio University Hospital, Kuopio, Finland
[c] School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Australia
[d] Science Service Center, Kuopio University Hospital, Kuopio, Finland
[e] Sleep Disorders Unit, Loewenstein Hospital-Rehabilitation Center, Raanana, Israel

ARTICLE INFO

Article history:
Received 30 May 2022
Revised 4 August 2022
Accepted 7 September 2022

Keywords:
Pulse oximetry
MATLAB
Hypoxemia
Automatic analysis
Sleep apnea
Resaturation

ABSTRACT

Background and objective: Many sleep recording software used in clinical settings have some tools to automatically analyze the blood oxygen saturation (SpO_2) signal by detecting desaturations. However, these tools are often inadequate for scientific research as they do not provide SpO_2 signal-based parameters which are superior in the estimation of sleep apnea severity and related medical consequences. In addition, these software require expensive licenses and they lack batch analysis tools. Thus, we developed the first freely available automatic blood oxygen saturation analysis software (ABOSA) that provides sophisticated SpO_2 signal-based parameters and enables batch analysis of large datasets.

Methods: ABOSA was programmed with MATLAB. ABOSA automatically detects desaturation and recovery events from the SpO_2 signals (EDF files) and calculates numerous parameters, such as oxygen desaturation index (ODI) and desaturation severity (DesSev). The accuracy of the ABOSA software was evaluated by comparing its desaturation scorings to manual scorings in Kuopio ($n = 1981$) and Loewenstein ($n = 930$) sleep apnea patient datasets. Validation was performed in a second-by-second manner by calculating Matthew's correlation coefficients (MCC) and median differences in parameter values. Finally, the performance of the ABOSA software was compared to two commercial software, Noxturnal and Profusion, in 100 patient subpopulations. As Noxturnal or Profusion does not calculate novel desaturation parameters, these were calculated with custom-made functions.

Results: The agreements between ABOSA and manual scorings were great in both Kuopio (MCC = 0.801) and Loewenstein (MCC = 0.898) datasets. However, ABOSA slightly overestimated the desaturation parameter values. The median differences in ODIs were 0.8 (Kuopio) and 0.0 (Loewenstein) events/h. Similarly, the median differences in DesSevs were 0.02 (Kuopio) and 0.01 (Loewenstein) percentage points. In a second-by-second analysis, ABOSA performed very similarly to Noxturnal and Profusion software in both Kuopio ($MCC_{ABOSA} = 0.807$, $MCC_{Noxturnal} = 0.807$, $MCC_{Profusion} = 0.811$) and Loewenstein ($MCC_{ABOSA} = 0.904$, $MCC_{Noxturnal} = 0.911$, $MCC_{Profusion} = 0.871$) datasets. Based on Noxturnal and Profusion scorings, the desaturation parameter values were similarly overestimated compared to ABOSA.

Conclusions: ABOSA is an accurate and freely available software that calculates both traditional clinical parameters and novel parameters, provides a detailed characterization of desaturation and recovery events, and enables batch analysis of large datasets. These are features that no other software currently provides making ABOSA uniquely suitable for scientific research use.

1. Introduction

Hypoxemia is a condition where blood oxygen saturation (SpO_2) is reduced below the normal level. For example, in sleep apnea patients, apneas and hypopneas during sleep cause transient drops (desaturations) in the SpO_2 [1]. A shunt is a condition, for exam-

* Corresponding author at: Department of Applied Physics, University of Eastern Finland, P.O. Box 1627 (Canthia), Kuopio FI-70211, Finland.
E-mail address: tuomas.karhu@uef.fi (T. Karhu).

https://doi.org/10.1016/j.cmpb.2022.107120

ple, in congenital heart diseases and acute respiratory distress syndrome, where the blood passes from the right side of the heart to the left side without participating in the gas exchange in the lungs causing hypoxemia [2]. The imbalance in ventilation/perfusion ratio in many lung diseases, such as chronic obstructive pulmonary disease leads to prolonged hypoxemia [3]. Hypoxemia also further increases the risk for numerous medical consequences, such as heart failure, metabolic dysfunction, hypertension, and mortality [4–7].

One common non-invasive method to estimate the oxygenation of the blood is to measure the SpO_2 with pulse oximetry commonly placed on a fingertip. Many clinically used software have limited tools to automatically analyze the SpO_2 signal by detecting desaturations and by calculating traditional parameters, e.g., oxygen desaturation index (ODI). However, none of the currently available software provides more sophisticated desaturation parameters [8,9] which have been shown to better describe the severity of sleep apnea [9–15] by considering the duration and depth of the desaturations. In addition, currently available software do not support the analysis of a large number of recordings by one action (i.e., batch analysis) and the analysis must be done manually patient-by-patient. Thus, currently available software are poorly suited for scientific research where large datasets are often required. Furthermore, these software require expensive licenses and may not be compatible with data collected from different manufacturer devices.

To tackle these shortcomings in the currently available software, we developed unique and freely available [16] automatic blood oxygen saturation analysis (ABOSA) software which enables the analysis of a large number of SpO_2 recordings. ABOSA automatically locates desaturation and following recovery events from the SpO_2 signal and produces in-depth information for each event. In addition, ABOSA calculates numerous new SpO_2 signal-based parameters that are useful in the estimation of sleep apnea severity and its consequences [10–12,17,18], assessment of hypoxemia progression over time [13,14,19], and mortality risk [20,21]. These are parameters that no other currently available software provides. In this paper, we describe the structure of the ABOSA software and validate its accuracy to detect desaturations in two large sleep apnea patient datasets.

2. Material and methods

2.1. Software structure

The ABOSA is an automatic blood oxygen saturation scoring software programmed completely with MATLAB (version 2021b, MathWorks, Natick, USA) utilizing App Designer for the creation of the user interface. The ABOSA is freely available [16] and it does not require a MATLAB license to be used. However, MATLAB runtime is required and will be automatically installed before the first use. MATLAB runtime requires at least 3GB of disk space for installation, any x86–64 processor and 4GB of RAM. Specific system requirements for the MATLAB runtime version 2021b are available for Windows [22], Linux [23], and Mac [24]. The ABOSA software has no additional requirements to run apart from MATLAB runtime. The ABOSA takes European Data Format (EDF) files as an input. EDF was chosen for the data format as it is widely used and does not require any specific data formatting by the user. In addition, many sleep recording software support exporting the data to EDF format for which we have developed an automatic export tool [25]. The structure of the ABOSA software is illustrated in Fig. 1 and described in detail below.

2.1.1. Artefact detection

First, major artefacts are detected from the original raw SpO_2 signal (S_{Raw}) by filtering the S_{Raw} signal with 4th order Butterworth high-pass filter with a 1 Hz cutoff. Next, the filtered signal is squared. Values >30 in the squared signal, and values $<50\%$ and $>100\%$ in the S_{Raw} signal are counted as artefacts. Adjacent artefact points are grouped into a single artefact, and artefacts are extended by one second in both directions to add a safety margin. However, artefacts with a duration of ≤ 5 s are linearly interpolated allowing the desaturation and recovery event scoring through these periods (see below sections).

2.1.2. Desaturation event scoring

The scoring of the desaturation events can be divided into multiple steps. In the first step, the S_{Raw} signal is filtered with 2nd order Butterworth low-pass filter with 0.1 Hz cutoff (S_{Low} signal).

Next, we locate the potential endpoints of the desaturation events. This is done by searching local minimums (L_{min}) from the S_{Low} signal with a minimum peak distance of 5 s and minimum peak prominence of 1 [26]. As L_{min}s are searched from the low-pass filtered signal, the location of a L_{min} does not match the exact minimum value in the S_{Raw} signal. This is corrected by searching the actual minimum value in the S_{Raw} signal within a 10 s window placed around the L_{min} (Fig. 1).

With a similar approach, we locate potential starting points of the desaturation events by searching local maximums (L_{max}) from the S_{Low} signal. However, in the location correction, a maximum value is searched instead of a minimum (Fig. 1). Furthermore, L_{max}-values that are too low to potentially result in proper desaturation events are removed. This is done by checking the maximum difference in S_{Raw} values between two adjacent L_{max}s: if the difference is <3 percentage points, the first L_{max} is removed (for further use, L_{max}s before removing are marked as $L_{max,org}$).

Next, to form a desaturation event, L_{max} is matched with a L_{min}. First, the maximum duration of a desaturation event is limited to 180 s. In addition, the potential L_{max}-L_{min} pair cannot go through another L_{max}. Second, the L_{max} is shifted forward to a point where the fall rate $\leq -0.05\%$/s is reached for the first time. Furthermore, we check whether there is a flat plateau within the potential desaturation event. If a plateau with duration ≥ 30 s exists, L_{max} or L_{min}/L_{min}s are shifted at the end/start of the plateau so that the depth of the potential desaturation is maximized. Next, L_{min}s that would result in a desaturation event with $<3\%$ transient drop are rejected. In addition, the fall rate of the desaturation event is limited between -0.05%/s and -4%/s. At this point, L_{max} could still potentially be matched with more than one L_{min}, although this is rather rare. Thus, we need to choose the most suitable L_{min}. This is done by investigating the SpO_2 signal segment between two adjacent L_{min}s ($L_{min,i}$, L_{min},i+1). If I) this segment includes SpO_2 value that is $\geq SpO_{2min,i}$+2, or II) $\geq 50\%$ of the SpO_2 values within this segment are $\geq SpO_{2min,i}$+1, or III) $SpO_{2min,i}$ is the minimum of all potential endpoints, we choose the $L_{min,i}$ as the ending point of the desaturation. If none of these criteria are met, we investigate the next two adjacent L_{min}s, and repeat. If no suitable L_{min} is found for L_{max}, no desaturation event is formed. Finally, if $>20\%$ of the duration of the formed desaturation is an interpolated artefact, and the total duration of interpolated parts is >5 s, the desaturation is discarded.

2.1.3. Recovery event scoring

Recovery events are searched based on the previously detected desaturation events. This means that a recovery event cannot exist without a desaturation event, but a desaturation event does not need to be followed by a recovery event.

First, the recovery event is forced to start at the same point at which the corresponding desaturation event ends. Second, the

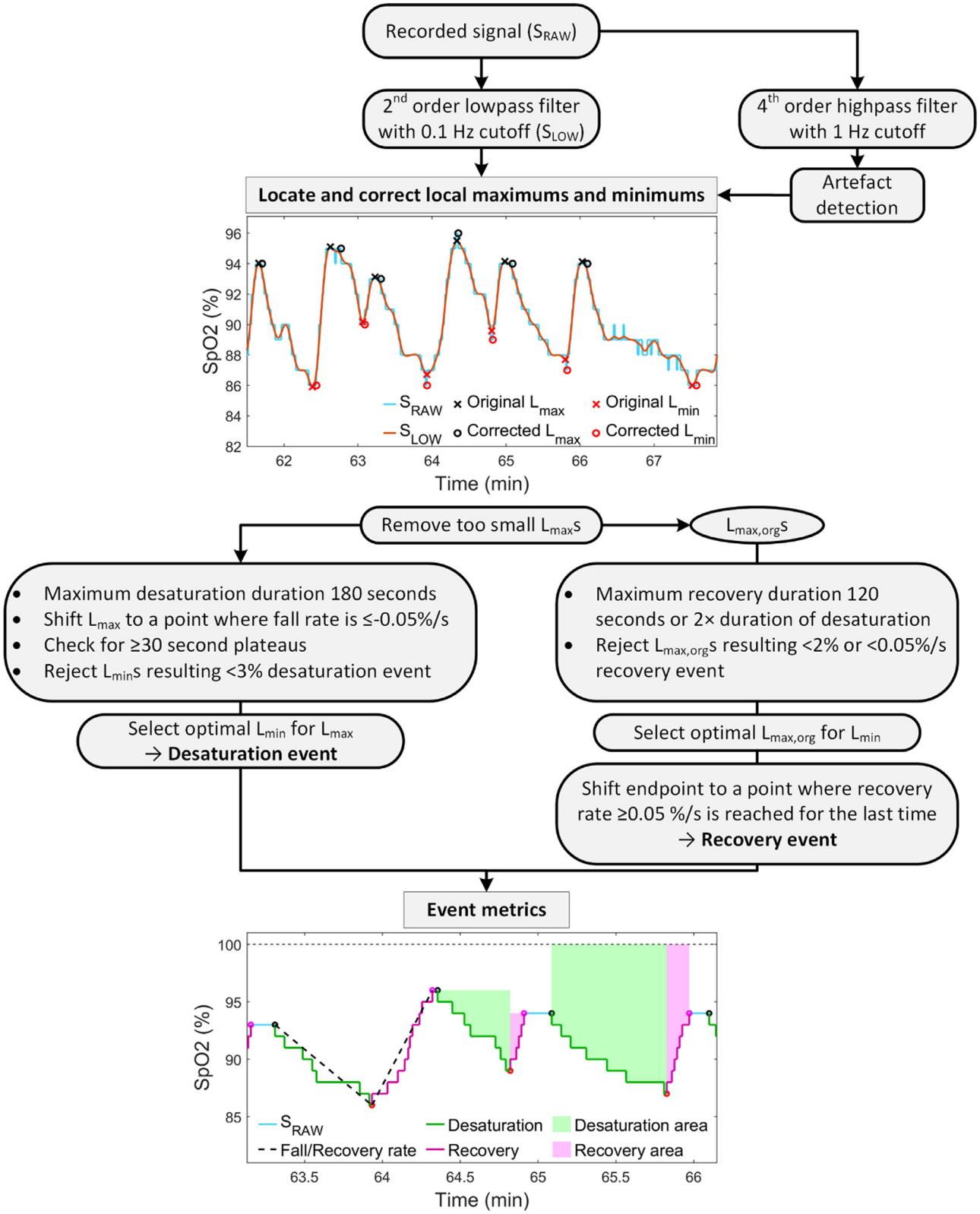

Fig. 1. Structure of the ABOSA software.

maximum recovery event duration is limited to 120 s, or the duration of two times the corresponding desaturation event (minimum of these two options).

Next, to form a recovery event, we search the endpoint for the event. Potential endpoints are the $L_{max,org}$s. $L_{max,org}$s that would result in a recovery event with depth <2% or recovery rate <0.05%/s are rejected. Similarly, as with desaturation events, at this point, more than one endpoint could still exist for the recovery event. Thus, we need to choose only one. This is done similarly as with desaturations: we investigate the segment between two adjacent

Table 1
Formulas and examples of previous findings from the most used desaturation parameters calculated by the ABOSA software. Similar parameters are also calculated for recovery events. Note that ABOSA also calculates other parameters not presented in this table.

Parameter*,†	Formula	Examples of uses and associations in literature‡
ODI (events/h)	$\frac{n_{desaturation}}{TST_{hour}}$	Screening of sleep apnea [29], worsening of hypoxic load over time [14], daytime sleepiness [11,27]
DesSev (%-point)	$\frac{\sum_i DesArea_i}{TST_{second}}$	Worsening of hypoxic load over time [13,14], impaired vigilance [10], daytime sleepiness [11,27], elevated cardiac troponin I [28], cardiovascular disease-related mortality¶ [9], incident heart failure¶ [15]
DesSev100 (%-point)	$\frac{\sum_i DesArea100_i}{TST_{second}}$	Cardiovascular disease mortality¶ [30]
DesDur (%-point)	$\frac{\sum_i DesDur_i}{TST_{second}} \times 100\%$	Worsening of hypoxic load over time [13,14], elevated cardiac troponin I [28], mortality [21]
Average desaturation duration (s)	$\frac{\sum_i DesDur_i}{n_{desaturation}}$	Worsening of hypoxic load over time [13], daytime sleepiness [27], differences in patients with similar sleep apnea severity [8], differences between breathing cessation severities [31]
Average desaturation depth (%)	$\frac{\sum_i DesDepth_i}{n_{desaturation}}$	Impaired vigilance [10], daytime sleepiness [27], differences in patients with similar sleep apnea severity [8], differences between breathing cessation severities [31]
Average desaturation area (s%)	$\frac{\sum_i DesArea_i}{n_{desaturation}}$	Worsening of hypoxic load [13], differences between breathing cessation severities [31]
Average desaturation area from 100% saturation (s%)	$\frac{\sum_i DesArea100_i}{n_{desaturation}}$	
Average fall rate (%/s)	$\frac{\sum_i FallRate_i}{n_{desaturation}}$	Elevated blood pressure [12], daytime sleepiness [32]
Average desaturation/recovery-duration ratio	$\frac{\sum_i (\frac{DesDur_i}{RecDur_i})}{n_{pair}}$	Incident heart failure [33]
Percentual time below 90% saturation (%)	$\frac{t_{<90}}{TST_{second}} \times 100\%$	Impaired vigilance [10], daytime sleepiness [11], elevated cardiac troponin I and T [28], ictal hypoxemia [34], cholesterol and triglyceride levels [35]
Average SpO_2 value during sleep (%)	$\frac{\sum_i SpO_{2i}}{TST_{observation}}$	Impaired vigilance [10], daytime sleepiness [11], cholesterol and triglyceride levels [35], polycythemia [36]
Minimum SpO_2 value during sleep (%)	$\min(SpO_2)$	Daytime sleepiness [11], cholesterol and triglyceride levels [35], carotid plaque burden [37]
The variance of SpO_2 values during sleep ($\%^2$)	$\frac{\sum_i \lvert SpO_{2i} - SpO_{2avg}\rvert^2}{TST_{observation} - 1}$	One of the six features used to screen sleep apnea [38]

$n_{desaturation}$ is the number of desaturations and n_{pair} is the number of desaturation-recovery event pairs. $DesArea_i$, $DesArea100_i$, $DesDur_i$, $DesDepth_i$, and $FallRate_i$ are the area, area from 100% saturation reference level, duration, depth, and fall rate of individual desaturation events, respectively (see Fig. 1). $RecDur_i$ is the duration of recovery event following $DesDur_i$. TST_{hour}, TST_{second}, and $TST_{observation}$ are the total sleep time in hours, seconds, and the number of data points during sleep, respectively. $t_{<90}$ is the time in seconds spent <90% saturation during sleep. SpO_{2i} is the individual saturation value and SpO_{2avg} is the average of the SpO_2 values during sleep. *In addition to averages, median values are calculated. †Percentual times below numerous other thresholds than 90% are also calculated. ‡Note that these are only examples, rather than a comprehensive review. ¶Similar parameter used. Parameters calculated by the ABOSA, but not presented in the table: Average/median of the baseline/minimum values of the desaturation events, maximum saturation value during sleep, and ratios for desaturation-recovery event depths, areas, areas from 100 reference level, and slopes. In addition, total event (i.e., combination of desaturation and recovery event) severity, total event severity from 100% reference, and total event duration parameters are calculated.

$L_{max,org}$s ($L_{max,org,i}$, $L_{max,org,i+1}$). If I) this segment includes SpO_2 value that is < $SpO_{2max,org,i}$-1, or II) ≥50% of the SpO_2 values within this segment are ≤ $SpO_{2max,org,i}$-1, or III) $SpO_{2max,org,i}$ is the maximum of all potential endpoints, we choose the $L_{max,org,i}$ as the ending point of the recovery event. If none of these criteria are met, we investigate the next two adjacent $L_{max,org}$s, and repeat. Next, the selected endpoint is shifted backward to a point where a recovery rate ≥0.05%/s is reached for the last time. If no suitable $L_{max,org}$ is found for L_{min}, no recovery event is formed. Finally, if >20% of the duration of the formed recovery event is an interpolated artefact, and the total duration of interpolated parts is >5 s, the recovery event is discarded.

2.1.4. Event metrics and parameter calculation

For each formed desaturation and recovery event, several metrics are calculated. These metrics consist of duration, depth, area, area from 100% saturation reference level, and slopes (i.e., fall and recovery rate, Fig. 1). Furthermore, ratios for desaturation-recovery event durations, depths, areas, areas from 100% saturation reference level, and slopes are calculated.

The user can decide the criteria for the desaturation scoring by selecting the minimum duration and minimum depth of the desaturations (user interface section below). However, the user cannot influence recovery event scoring. In addition, the user can select the definition for total sleep time (TST) by inputting analysis start/stop times or hypnograms. Based on the inputted criteria, desaturation events, and thus the corresponding recovery events not fulfilling the criteria are ignored. From the remaining events, numerous parameters are calculated (Table 1, Fig. 1). Note that similar parameters are also calculated for recovery events. In this study, we used the conventional ODI parameter, in addition to the desaturation severity (DesSev), DesSev from 100% reference (DesSev100), and desaturation duration (DesDur) parameters, as these parameters have been previously used to describe the severity and medical consequences of sleep apnea [8,10,11,13,14,21,27,28].

2.1.5. User interface

From the user interface (Fig. 2), by pressing the “Select input folder” or by typing in the corresponding text field, the user selects a folder which contains all EDF files to be automatically analyzed. Note that all EDF files must be in a single folder, not in subfolders. Optionally, the user can input a folder that contains the hypnograms (CSV, TXT, XLSX, or XLS files) for each recording. The user can also input a single CSV file consisting of analysis start and stop times for all the recordings. More detailed information and examples on hypnograms and analysis start/stop file formatting is provided in the UserGuide after downloading the software [16].

If hypnograms and/or analysis start/stop times are inputted, the user can change the TST definition. The “Analysis start/stop” option calculates the TST between analysis start and stop timepoints in the SpO_2 signal. The “Sleep onset to offset” calculates TST as the time between the first and last non-wake epochs, based on hypnogram. The “Sleep only” option involves only epochs that are scored as N1, N2, N3, or REM, when determining the TST. The “Whole recording” option uses the duration of the whole recorded SpO_2 signal as TST. TST definition also determines which parts of the

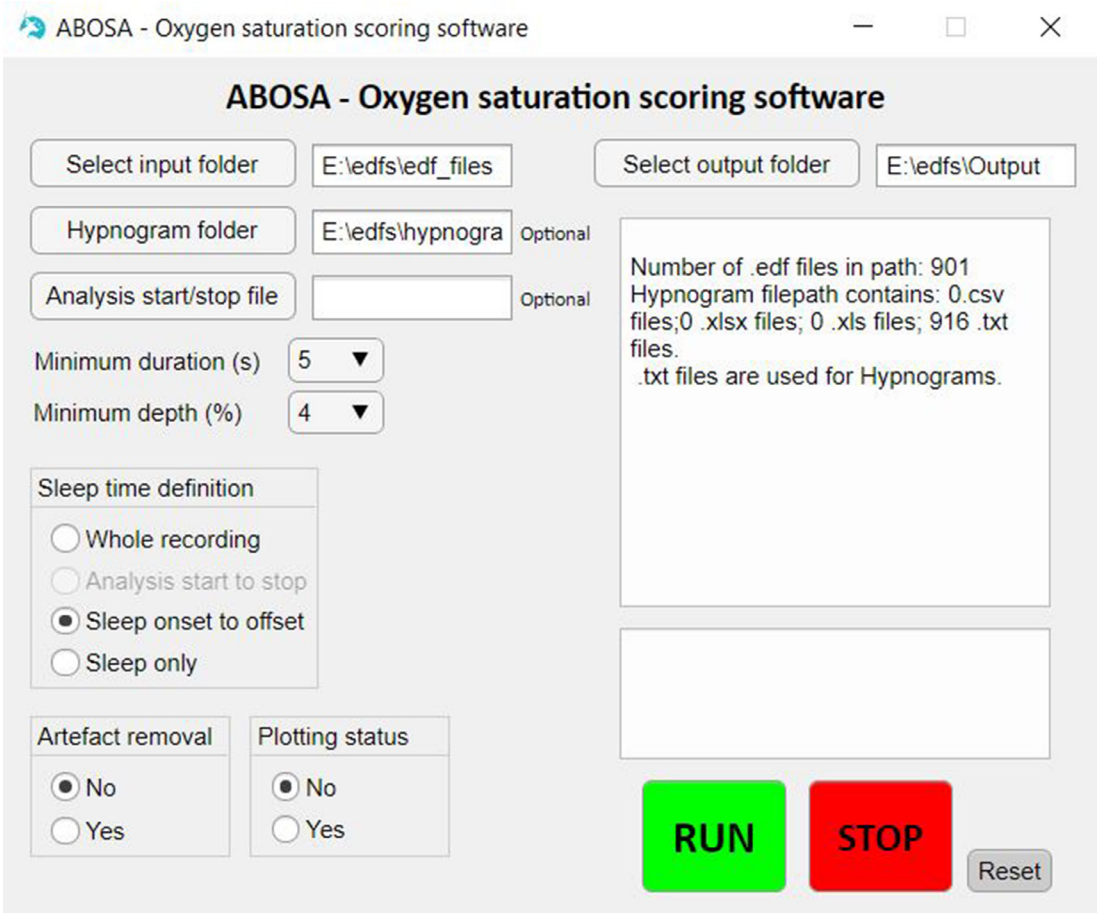

Fig. 2. The ABOSA software user interface.

SpO_2 signal are used for desaturation scoring. For example, if the "Sleep only" is chosen, only desaturations starting in sleep are included. Furthermore, the user can select the criteria for the desaturation scoring by selecting the minimum duration (3 s or longer) and depth (3% or larger) for desaturations. In addition, the user can include or exclude the artefact parts (see Section 2.1.1) of the SpO_2 signal from the TST. Note, that the "Artefact definition" only potentially affects the TST, as no desaturation or recovery events are scored in artefacts with either option.

In the selected output folder, an XLSX file containing all parameter values for all recordings and a TXT file containing potential error messages and information on skipped files, are created. In addition, an XLSX file containing the event metrics data for each recording is created.

Finally, the software is run by pressing the "RUN" button. For the first file, the user must select the label for the SpO_2 signals which is then used for all the EDF files. If no such label exists in some of the EDFs, these files are skipped.

2.2. Validation

2.2.1. Datasets

To validate the accuracy of the ABOSA software, software scorings were compared to manual scorings in two clinical sleep apnea patient datasets: Kuopio and Loewenstein datasets.

The Kuopio dataset was recorded during 1992–2003 at the Kuopio University Hospital (Kuopio, Finland) with a custom-made ambulatory recording device [39]. SpO_2 signal was recorded with Minolta Pulsox-7 finger-pulse oximeter (Konica Minolta, Tokyo, Japan) with a sampling frequency of 4 Hz. Data consists of 1989 recordings which were reanalyzed according to the 2007 American Academy of Sleep Medicine (AASM) guidelines [40] and clinical practices of the Kuopio University Hospital at the time (i.e., 4% desaturation drop criterion). Due to invalid analysis start/end times, eight recordings were excluded from the validation. Therefore, a total of 1981 recordings were included for validation. For the data collection and analysis the positive statement was given by the Ethics Committee of the Hospital District of Northern Savo (127/2004, 24/2013).

Loewenstein polysomnography data was recorded during 2001–2011 at the Loewenstein Hospital – Rehabilitation Center (Raanana, Israel) with REMbrandt Manager System (Medcare, Amsterdam, Netherlands). SpO_2 signal was recorded with Ohmeda Biox 3700 pulse oximeter (Ohmeda, Louisville, USA) and exported with 256 Hz. Data from 990 patients were reanalyzed at the Kuopio University Hospital according to the 2007 AASM guidelines [40]. However, 60 recordings were excluded due to issues in data exportation or due invalid hypnograms. Therefore, a total of 930 recordings were included for validation. Data collection and processing were approved by the Ethical Committee of the Loewenstein Hospital – Rehabilitation Center (0006–17-LOE).

2.2.2. Data analyses

The accuracy of the ABOSA software to detect desaturations was validated by comparing automatically scored desaturations to the manually scored desaturations. The ABOSA desaturations were scored with minimum event duration of 5 s and minimum transient drop of 4% in both datasets. First, SpO_2 signals were segmented into 1 s segments, and each segment was marked as no-event or event. Partial events were marked as an event. Segment-

wise agreements between manual and ABOSA-based desaturation scorings were investigated by calculating Matthew's correlation coefficient (MCC), Cohen's kappa, accuracy, specificity, and sensitivity.

ODI, DesSev, DesSev100, and DesDur parameters calculated from manual and ABOSA scorings were compared in both datasets. The "Analysis start/stop" TST definition was used in Kuopio dataset and the "Sleep only" in Loewenstein dataset. Intraclass correlation coefficients (ICC) were calculated using two-way random effect, single rater, and absolute agreement definition (ICC(2,1) definition [41]).

Furthermore, we compared the scorings in an event-by-event manner. To investigate this, we paired ABOSA scored desaturations to manual events. ABOSA event was paired to a manual event if they overlapped. As multiple ABOSA events may overlap with the same manual event, the first ABOSA event that overlapped was used in the pairing. From the paired events, we calculated differences in desaturation start and end times (ABOSA event–manual event), and percentual differences in durations and areas (difference/manual event).

Finally, we compared the performances of the ABOSA software and two commercial software: Noxturnal (version 5.1.19824, Nox Medical, Reykjavík, Iceland), and Profusion (version 4.5 Build 468, Compumedics ProFusion PSG Lite, Abbortsford, Australia). This was done by randomly selecting 100 patients from both datasets and running software's automatic SpO_2 analyses for them. The 100 patient subpopulations were used, as Noxturnal, or Profusion do not support batch analysis. Therefore, the automatic analyses had to be performed manually patient-by-patient, making the use of whole study populations unfeasible. The automatic settings in Noxturnal and Profusion were selected to correspond to the settings in ABOSA as closely as possible. Therefore, the settings in Noxturnal were a minimum event duration of 5 s, a minimum transient drop of 4%, a maximum plateau duration of 30 s, and no desaturations were scored in parts of the SpO_2 signal with values $<$50%. Settings in Profusion were a minimum transient drop of 4%, a maximum fall rate of 4%/s, and desaturations with $>$50% transient drop were not scored. Note that the Profusion scored the original desaturations from baseline-to-baseline. Thus, these events were split into desaturation and recovery events at nadir, and these split events were used in the comparisons. From these automatic scorings, the same second-by-second comparisons were performed. In addition, ODI, DesSev, DesSev100, and DesDur parameters were calculated and compared to the manual and ABOSA scorings.

3. Results

The agreements between the ABOSA software and manual scorings in the second-by-second comparisons were high in both datasets. Agreements were higher in the Loewenstein dataset compared to the Kuopio dataset (Table 2). In addition, ICCs were high for all investigated parameters in both datasets (Table 3).

Parameter values calculated from the ABOSA scorings were slightly higher compared to the manual scorings in the Kuopio dataset but very similar in the Loewenstein dataset (Table 3). For example, median differences in ODIs were 0.8 and 0.0 events/h in Kuopio and Loewenstein datasets, respectively. Bland-Altman plots (Fig. 3) also illustrate differences between manual and ABOSA scored ODIs.

The ABOSA correctly detected majority of the desaturation events scored by the manual scorers. A total of 160,934 and 111,591 events were paired in the event-by-event analysis for Kuopio and Loewenstein data, respectively. Of the manually scored desaturations, 4.5% in the Kuopio dataset and 4.1% in the Loewenstein dataset were not paired with ABOSA scored events. Moreover, 14.9% (Kuopio) and 4.5% (Loewenstein) of the ABOSA scored desaturations were not paired with manually scored desaturations. Out of all ABOSA scored desaturations, 94.7% (Kuopio) and 96.8% (Loewenstein) were followed by recovery events.

In the paired desaturation events, only minor differences were observed in event start and end times (Fig. 4). Median differences in desaturation event start times were 0.75 s (Kuopio) and 0.19 s (Loewenstein). Similarly, differences in desaturation event end times were small: 1.00 s (Kuopio) and 0.03 s (Loewenstein). Positive values indicate that ABOSA events start/end later. As a re-

Table 2
1-second segment-based comparisons between manual and ABOSA software scorings. The percentage of 1-s segments scored as desaturation are presented in brackets.

Dataset	The number of 1 s segments scored as desaturation in manual scoring	The number of 1 s segments scored as desaturation in ABOSA scoring	Matthew's correlation coefficient (MCC)	Cohen's kappa	Accuracy	Sensitivity	Specificity
Kuopio	5,223,203 (9.9%)	5,584,390 (10.6%)	0.801	0.801	0.963	0.849	0.976
Loewenstein	3,195,793 (17.0%)	3,182,962 (17.0%)	0.898	0.898	0.971	0.913	0.983

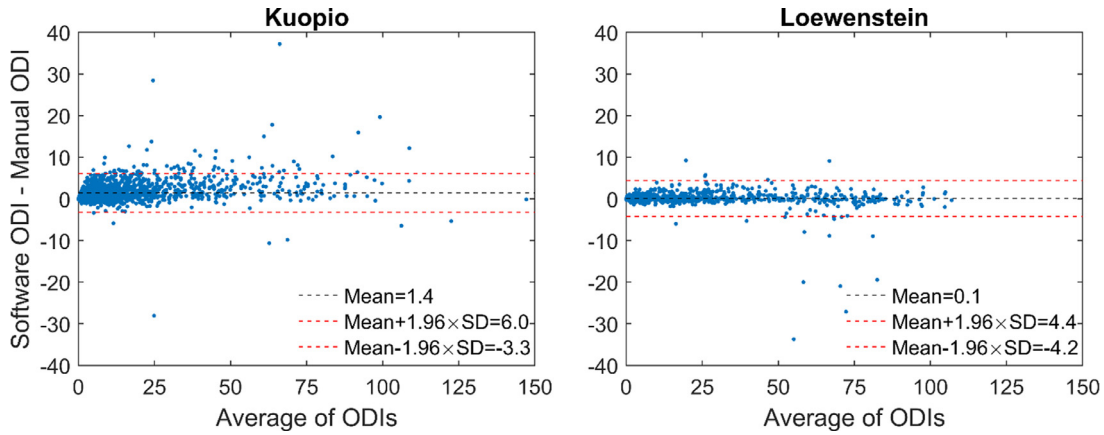

Fig. 3. Comparison between manual and ABOSA software scored oxygen desaturation index (ODI) using Bland-Altman plots. The most extreme outliers were due to inconsistencies in manual scorings. For example, for a few patients in the Loewenstein dataset's manual scorings, there was a significant number of desaturations scored in the parts of the oxygen saturation signal with values $<$50% into which the ABOSA does not score events.

Table 3
Median (interquartile range) parameter values from manual and ABOSA software scorings and their median difference.

Dataset/Parameter	Manual	ABOSA	Difference	ICC
Kuopio (n = 1981)				
ODI (events/h)	4.4 (1.1, 13.8)	5.5 (1.8, 15.6)	0.8 (0.2, 2.0)	0.989
DesSev (%-point)	0.11 (0.03, 0.39)	0.13 (0.04, 0.44)	0.02 (0.00, 0.06)	0.989
DesSev100 (%-point)	0.30 (0.07, 0.99)	0.33 (0.09, 1.02)	0.02 (−0.02, 0.11)	0.988
DesDur (%-point)	4.0 (1.0, 12.1)	4.4 (1.4, 12.6)	0.4 (−0.3, 1.4)	0.981
Loewenstein (n = 930)				
ODI (events/h)	11.2 (2.7, 31.6)	11.3 (2.9, 32.4)	0.0 (−0.2, 0.6)	0.996
DesSev (%-point)	0.26 (0.05, 0.90)	0.30 (0.06, 1.00)	0.01 (0.00, 0.06)	0.983
DesSev100 (%-point)	0.61 (0.13, 1.97)	0.61 (0.14, 2.01)	0.00 (−0.05, 0.04)	0.959
DesDur (%-point)	8.6 (2.3, 24.5)	8.8 (2.3, 25.0)	0.0 (−0.7, 0.7)	0.993

Intraclass correlation coefficients (ICC) were calculated using two-way random effect, single rater, and absolute agreement definition (ICC(2,1) definition [41]). Differences are calculated as ABOSA parameter value – manual parameter value. %-point = percentage point, ODI = oxygen desaturation index, DesSev = desaturation severity, DesSev100 = desaturation severity from 100% saturation reference level, DesDur = desaturation duration.

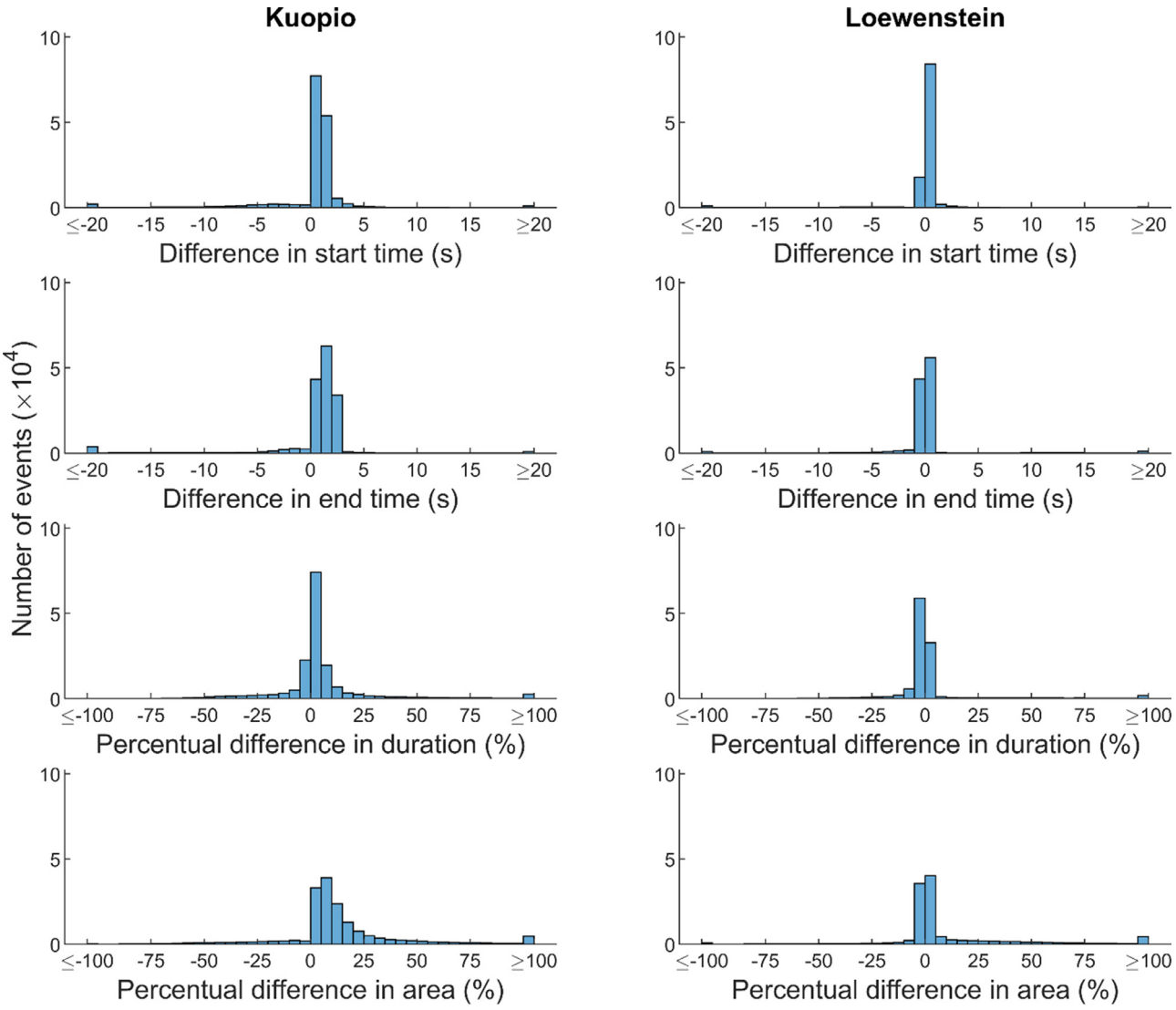

Fig. 4. Histograms of the differences (ABOSA event – manual event) in desaturation event start and end times, and percentual differences (difference/manual event) in durations and areas between manual and ABOSA software scored events.

Table 4
1-second segment-based comparisons between manual scorings and automatic ABOSA, Nocturnal, and Profusion software scorings in subpopulations of 100 randomly selected patients from Kuopio and Loewenstein datasets. The percentage of 1-s segments scored as desaturation are presented in brackets.

Dataset/Software	The number of 1 s segments scored as desaturation in software scoring	Matthew's correlation coefficient (MCC)	Cohen's kappa	Accuracy	Sensitivity	Specificity
Kuopio (n = 100)						
ABOSA	266,205 (10.0%)	0.807	0.806	0.967	0.870	0.976
Noxturnal	257,917 (9.6%)	0.807	0.806	0.967	0.856	0.978
Profusion	277,466 (10.4%)	0.811	0.809	0.967	0.892	0.974
Loewenstein (n = 100)						
ABOSA	381,997 (18.6%)	0.904	0.904	0.971	0.922	0.982
Noxturnal	366,782 (17.9%)	0.911	0.911	0.973	0.909	0.988
Profusion	381,058 (18.6%)	0.871	0.871	0.961	0.894	0.976

The number of 1-s segments manually scored as desaturation was 239,660 (9.0%) and 381,957 (18.6%) in Kuopio and Loewenstein datasets, respectively.

Table 5
Median (interquartile range) parameter values calculated from manual scorings and automatic ABOSA, Noxturnal, and Profusion software scorings in subpopulations of 100 randomly selected patients from Kuopio and Loewenstein datasets.

Dataset/Parameter	Manual	ABOSA	Noxturnal	Profusion
Kuopio (n = 100)				
ODI (events/h)	3.4 (0.9, 13.3)	4.9 (1.9, 14.7)	4.9 (1.7, 14.7)	4.1 (1.4, 13.8)
DesSev (%-point)	0.08 (0.02, 0.38)	0.11 (0.04, 0.46)	0.09 (0.04, 0.41)	0.11 (0.04, 0.49)
DesSev100 (%-point)	0.25 (0.05, 1.00)	0.26 (0.09, 1.13)	0.29 (0.07, 1.15)	0.32 (0.09, 1.22)
DesDur (%-point)	3.0 (0.9, 11.8)	3.5 (1.4, 12.8)	3.6 (1.2, 12.8)	3.5 (1.4, 14.0)
Loewenstein (n = 100)				
ODI (events/h)	12.1 (2.5, 45.7)	12.2 (2.9, 46.0)	12.2 (2.9, 45.4)	10.1 (2.5, 43.0)
DesSev (%-point)	0.27 (0.05, 1.35)	0.34 (0.06, 1.56)	0.32 (0.06, 1.43)	0.36 (0.05, 1.51)
DesSev100 (%-point)	0.76 (0.11, 2.41)	0.77 (0.12, 2.59)	0.74 (0.12, 2.44)	0.82 (0.12, 2.53)
DesDur (%-point)	9.4 (2.0, 33.2)	9.9 (2.4, 35.5)	9.1 (2.1, 33.0)	9.8 (1.8, 34.8)

Note that neither the Noxturnal nor Profusion does not provide DesSev, DesSev100, or DesDur parameter values. Therefore, all parameter values were calculated with separate custom-made functions utilizing "Analysis start/stop" and "Sleep only" total sleep time criteria in the Kuopio and Loewenstein datasets, respectively. %-point = percentage point, ODI = oxygen desaturation index, DesSev = desaturation severity, DesSev100 = desaturation severity from 100% saturation reference level, DesDur = desaturation duration.

sult, ABOSA scored events were slightly longer and had higher areas (Fig. 4).

ABOSA, Noxturnal, and Profusion software performed very similarly in the second-by-second analysis in the Kuopio dataset subpopulation (n = 100, Table 4). In the Loewenstein dataset subpopulation (n = 100), ABOSA and Noxturnal performed very similarly, while Profusion was slightly worse. In addition, based on ABOSA and Noxturnal scorings, very similar ODI, DesSev, DesSev100, and DesDur parameter values were obtained in the Loewenstein dataset compared to the manual scorings. In the Kuopio dataset all three software slightly overestimated the parameter values compared to the manual scorings (Table 5).

Furthermore, a total of 7753 (ABOSA), 7654 (Noxturnal), and 7446 (Profusion) desaturations were paired between manual and software scorings in the Kuopio dataset. Similarly, 13,539 (ABOSA), 13,599 (Noxturnal), and 12,820 (Profusion) desaturations were paired in the Loewenstein dataset. Of the manually scored desaturations, 3.8% (ABOSA), 5.0% (Noxturnal), and 7.6% (Profusion) were not paired with software-based desaturations in the Kuopio dataset, and 3.0% (ABOSA), 2.6% (Noxturnal), and 8.2% (Profusion) in the Loewenstein dataset. Furthermore, 14.7% (ABOSA), 15.8% (Noxturnal), and 12.6% (Profusion) of the software scored desaturations were not paired with manually scored desaturations in the Kuopio dataset, and 3.8% (ABOSA), 4.8% (Noxturnal), and 2.9% (Profusion) in the Loewenstein dataset.

4. Discussion

In this paper, we introduced the ABOSA software for the automatic analysis of SpO_2 signals. The ABOSA is the first software that provides several sophisticated desaturation and recovery parameters, produces detailed event-specific characteristics, and enables batch-like analysis of large datasets. These are features that no other software provides. We observed that ABOSA slightly overestimated desaturation parameter values compared to manual scoring. This was not only due to a higher event count, but also because the ABOSA scored desaturations were slightly longer. However, the ABOSA missed only a small fraction of manually scored desaturations. The ABOSA also performed very similarly to the Noxturnal and Profusion software in the automatic desaturation event detection.

One factor to explain the observed differences between manual and all three software-based desaturation scorings is the absence of standardized desaturation scoring rules. The only currently implemented rule for desaturation scoring is the 3% minimum transient drop criterion used in sleep apnea diagnostics when scoring hypopnea events [40]. Therefore, when a disagreement between manual and automatic software scoring is present, it is challenging to distinguish which scoring is correct, if either. Another factor that can explain the higher number of desaturations scored by the ABOSA compared to manual analysis, is the possibility of missing desaturations in the manual analysis due to human error. How-

ever, for recovery events, no scoring rules exist at all, as currently recovery events are not used in clinical practice. Thus, it should be noted that some of the scoring criteria implemented in the ABOSA software are somewhat subjective, and the selected criteria were based on visual inspection (Sections 2.1.2 and 2.1.3). In case more specific desaturation and recovery event scoring rules will be published in the future, these rules could be implemented in the ABOSA software. Furthermore, no commercial software scores recovery events, and thus,cannot provide any parameters from them. Due to these reasons, the use of recovery events has not been studied extensively. However, in addition to in-depth desaturation event analysis, recovery events could enhance especially the sleep apnea research [38,42,43] even further as they could indicate how well and fast the body can recover from a hypoxic state. For this, the ABOSA software is an excellent tool.

One limitation in the ABOSA software is the artefact detection which is relatively simple and only intended to identify obvious sharp peak artefacts and periods where the pulse oximeter is most probably completely off or disconnected. Thus, slow trend-like artefacts where the SpO_2 values get suspiciously low, but still plausible (e.g., 60–70%), due to improper pulse oximeter contact or poor perfusion [44], are not detected. However, slow trend drifts which do not fulfill the desaturation event criteria can also be a physiological phenomenon increasing the nocturnal hypoxic load even in the absence of sleep apnea. Such drifts can be troublesome as they have been shown to increase the mortality risk [45] while being resistant to sleep apnea treatments [46]. Thus, classifying whether a drift is an artefact, or a real physiological feature is challenging. Although the ABOSA does not locate these drifts, the lowered SpO_2 values do reflect into smaller mean and median values and increased t90 (and t85, t80, etc.) values during the recording which ABOSA does provide. Furthermore, as the analysis is purely based on the SpO_2 signal, motion artefacts can be misidentified as events. However, to mitigate the effect of signal noise on the event scorings, the detection of event start and end times is conducted from the lowpass filtered signal. The ABOSA was also robust on SpO_2 signals recorded with different devices.

Another limitation is the input format as the ABOSA only allows EDF files as input. EDF was chosen as it is the most common and a standardized format, thus including all relevant information for the automatic analysis, such as sampling frequency and labeling. Other data formats would require file-specific manual inputting of the needed features or specific data formatting by the user. As the aim of the ABOSA software was to enable a batch analysis of large datasets without reformatting the data, the input format was limited to the EDF at this stage of the software development. Additionally, there are multiple EDF libraries freely available for different programming languages [47–49] and even standalone solutions with graphical user interfaces [50] that do not require programming knowledge which can be used to convert other filetypes to EDF [51]. These solutions could be used to preprocess unsupported filetypes before using the ABOSA software. However, also building this functionality directly into ABOSA and thus easily allowing other input formats, such as .txt, ASCII, .xlsx, .csv, or .mat, is a potential future step to take in the development of the ABOSA software.

In summary, the ABOSA software was developed to be used in scientific research as the interest in using simple oxygen saturation-based methods to estimate the severity of sleep apnea and its numerous health issues has increased in recent years [10–14,17–20], and as the currently available clinical software do not provide these parameters. The ABOSA provides several desaturation and recovery event parameters, in-depth event-specific characterization, and it enables batch analysis of large datasets. In addition, the ABOSA is completely free to use [16] and does not require any additional software or licenses to be used. These are features that are lacking in the commercial software designed for clinical use. We showed that compared to two commercially available software, Noxturnal and Profusion, the ABOSA performed very similarly in the automatic SpO_2 signal scoring. Therefore, the ABOSA can be considered a reliable software, and it could be used to reanalyze datasets for further research use, especially in the field of sleep research.

Funding

This work was funded by the European Union's Horizon 2020 Research and Innovation Programme (Grant 965417), NordForsk (NordSleep project 90458) through Business Finland (5133/31/2018), the Academy of Finland (323536), the Research Committee of the Kuopio University Hospital Catchment Area for the State Research Funding (Projects 5041767, 5041768, 5041787, 5041794, 5041797, and 5041809), Tampere Tuberculosis Foundation, Instrumentarium Science Foundation, Finnish Cultural Foundation, Foundation of the Finnish Anti-Tuberculosis Association, and Research Foundation of the Pulmonary Diseases.

Declaration of Competing Interest

None declared.

References

[1] J.A. Dempsey, S.C. Veasey, B.J. Morgan, et al., Pathophysiology of sleep apnea, Physiol. Rev. 90 (2010) 47–112.
[2] P.D. Wagner, The physiological basis of pulmonary gas exchange: implications for clinical interpretation of arterial blood gases, Eur. Respir. J. 45 (2015) 227–243.
[3] B.D. Kent, P.D. Mitchell, W.T. McNicholas, Hypoxemia in patients with COPD: cause, effects, and disease progression, Int. J. COPD 6 (2011) 199–208.
[4] N.A. Dewan, F.J. Nieto, V.K. Somers, Intermittent hypoxemia and OSA: implications for comorbidities, Chest 147 (2015) 266–274.
[5] P.S. Chen, W.T. Chiu, P.L. Hsu, et al., Pathophysiological implications of hypoxia in human diseases, J. Biomed. Sci. 27 (2020) 63.
[6] O. Oldenburg, B. Wellmann, A. Buchholz, et al., Nocturnal hypoxaemia is associated with increased mortality in stable heart failure patients, Eur. Heart J. 37 (2016) 1695–1703.
[7] J.S. Floras, Sleep apnea and cardiovascular disease an enigmatic risk factor, Circ. Res. 122 (2018) 1741–1764.
[8] A. Kulkas, P. Tiihonen, K. Eskola, et al., Novel parameters for evaluating severity of sleep disordered breathing and for supporting diagnosis of sleep apnea-hypopnea syndrome, J. Med. Eng. Technol. 37 (2013) 135–143.
[9] A. Azarbarzin, S.A. Sands, K.L. Stone, 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, Eur. Heart J. 40 (2019) 1149–1157.
[10] S. Kainulainen, B. Duce, H. Korkalainen, et al., Severe desaturations increase psychomotor vigilance task-based median reaction time and number of lapses in obstructive sleep apnoea patients, Eur. Respir. J. 55 (2020) 1901849.
[11] S. Kainulainen, J. Töyräs, A. Oksenberg, et al., Severity of desaturations reflects OSA-related daytime sleepiness better than AHI, J. Clin. Sleep Med. 15 (2019) 1135–1142.
[12] N. Wang, Z. Meng, N. Ding, et al., Oxygen desaturation rate as a novel intermittent hypoxemia parameter in severe obstructive sleep apnea is strongly associated with hypertension, J. Clin. Sleep Med. 16 (2020) 1055–1062.
[13] T. Karhu, S. Myllymaa, S. Nikkonen, et al., Longer and deeper desaturations are associated with the worsening of mild sleep apnea: the sleep heart health study, Front. Neurosci. 15 (2021) 657126.
[14] T. Karhu, S. Myllymaa, S. Nikkonen, et al., Diabetes and cardiovascular diseases are associated with the worsening of intermittent hypoxaemia, J. Sleep Res. 31 (2022) e13441.
[15] A. Azarbarzin, S.A. Sands, L. Taranto-Montemurro, et al., The sleep apnea-specific hypoxic burden predicts incident heart failure, Chest 158 (2020) 739–750.
[16] T. Karhu, T. Leppänen, J. Töyräs, et al. Automatic oxygen saturation signal analysis software, 2022. [Online]. Available: https://zenodo.org/record/6962129. Accessed: September 9, 2022.
[17] A. Kulkas, B. Duce, T. Leppänen, et al., Severity of desaturation events differs between hypopnea and obstructive apnea events and is modulated by their duration in obstructive sleep apnea, Sleep Breath. 21 (2017) 829–835.
[18] S. André, F. Andreozzi, C. Van Overstraeten, et al., Cardiometabolic comorbidities in obstructive sleep apnea patients are related to disease severity, nocturnal hypoxemia, and decreased sleep quality, Respir. Res. 21 (2020) 35.
[19] C.Y. Lin, T.S. Shih, S.H. Liou, et al., Predictors for progression of sleep disordered breathing among public transport drivers: a 3-year follow-up study, J. Clin. Sleep Med. 11 (2015) 419–425.

[20] G. Labarca, J. Jorquera, J. Dreyse, et al., Hypoxemic features of obstructive sleep apnea and the risk of mortality: a cluster analysis, Sleep Breath. 25 (2021) 95–103.
[21] E. Watanabe, K. Kiyono, S. Matsui, et al., Prognostic importance of novel oxygen desaturation metrics in patients with heart failure and central sleep apnea, J. Card. Fail. 23 (2017) 131–137.
[22] Mathworks. System requirements - Release 2021b - Windows. [Online]. Available: https://www.mathworks.com/content/dam/mathworks/mathworks-dot-com/support/sysreq/files/system-requirements-release-2021b-windows.pdf. Accessed: September 9, 2022.
[23] MathWorks. System requirements - Release 2021b - Linux. [Online]. Available: https://www.mathworks.com/content/dam/mathworks/mathworks-dot-com/support/sysreq/files/system-requirements-release-2021b-linux.pdf. Accessed: September 9, 2022.
[24] MathWorks. System requirements - Release 2021b - Mac. [Online]. Available: https://www.mathworks.com/content/dam/mathworks/mathworks-dot-com/support/sysreq/files/system-requirements-release-2021b-macintosh.pdf. Accessed: September 9, 2022.
[25] S. Nikkonen, H. Korkalainen, J. Töyräs, et al. STAR sleep recording export software, 2022. [Online]. Available: https://zenodo.org/record/7056195. Accessed: September 9, 2022.
[26] MathWorks. MATLAB “findpeaks” documentation. [Online]. Available: https://www.mathworks.com/help/signal/ref/findpeaks.html. Accessed: September 9, 2022.
[27] S. Kainulainen, J. Töyräs, A. Oksenberg, et al., Power spectral densities of nocturnal pulse oximetry signals differ in OSA patients with and without daytime sleepiness, Sleep Med. 73 (2020) 231–237.
[28] F.D. Sigurdardottir, C.T. Øverby, S. Nikkonen, et al., Novel oxygen desaturation parameters are associated with cardiac troponin I: data from the Akershus sleep apnea project, J. Sleep Res. (2022) e13581.
[29] N.H.A. Rashid, S. Zaghi, M. Scapuccin, et al., The value of oxygen desaturation index for diagnosing obstructive sleep apnea: a systematic review, Laryngoscope 131 (2021) 440–447.
[30] P. De Chazal, N. Sadr, H. Dissanayake, et al., Predicting cardiovascular outcomes using the respiratory event desaturation transient area derived from overnight sleep studies, in: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, 2021, pp. 5496–5499.
[31] A. Kulkas, B. Duce, T. Leppänen, et al., Gender differences in severity of desaturation events following hypopnea and obstructive apnea events in adults during sleep, Physiol. Meas. 38 (2017) 1490–1502.
[32] S. Zhang, Z. Meng, X. Zhang, et al., The rate of decrease in oxygen desaturation during severe obstructive sleep apnea syndrome is correlated with subjective excessive daytime sleepiness, Sleep Breath. 25 (2021) 1285–1291.
[33] D. Linz, K.A. Loffler, P. Sanders, et al., Low prognostic value of novel nocturnal metrics in patients with OSA and high cardiovascular event risk: post Hoc analyses of the save study, Chest 158 (2020) 2621–2631.
[34] E. Bruno, G. Maira, A. Biondi, et al., Ictal hypoxemia: a systematic review and meta-analysis, Seizure 63 (2018) 7–13.
[35] Y. Toyama, K. Chin, Y. Chihara, et al., Association between sleep apnea, sleep duration, and serum lipid profile in an urban, male, working population in Japan, Chest 143 (2013) 720–728.
[36] N. Li, H.P. Li, P. Wang, et al., Nocturnal mean oxygen saturation is associated with secondary polycythemia in young adults with obstructive sleep apnea, especially in men, Nat. Sci. Sleep 11 (2019) 377–386.
[37] S.I. Gunnarsson, P.E. Peppard, C.E. Korcarz, et al., Minimal nocturnal oxygen saturation predicts future subclinical carotid atherosclerosis: the Wisconsin sleep cohort, J. Sleep Res. 24 (2015) 680–686.
[38] M. Deviaene, D. Testelmans, B. Buyse, et al., Automatic screening of sleep apnea patients based on the SpO_2 signal, IEEE J. Biomed. Health Inform. 23 (2019) 607–617.
[39] A. Muraja-Murro, K. Eskola, T. Kolari, et al., Mortality in middle-aged men with obstructive sleep apnea in Finland, Sleep Breath. 17 (2013) 1047–1053.
[40] R.B. Berry, R. Brooks, C.E. Gamaldo, et al., The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications - Version 2.4, American Academy of Sleep Medicine, 2017.
[41] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychol. Bull. 86 (1979) 420–428.
[42] D. Sánchez-Morillo, M.A. López-Gordo, A. León, Novel multiclass classification for home-based diagnosis of sleep apnea hypopnea syndrome, Expert Syst. Appl. 41 (2014) 1654–1662.
[43] D.S. Morillo, N. Gross, Probabilistic neural network approach for the detection of SAHS from overnight pulse oximetry, Med. Biol. Eng. Comput. 51 (2013) 305–315.
[44] E.D. Chan, M.M. Chan, M.M. Chan, Pulse oximetry: understanding its basic principles facilitates appreciation of its limitations, Respir. Med. 107 (2013) 789–799.
[45] M. Baumert, S.A. Immanuel, K.L. Stone, et al., Composition of nocturnal hypoxaemic burden and its prognostic value for cardiovascular mortality in older community-dwelling men, Eur. Heart J. 41 (2020) 533–541.
[46] D. Linz, M.V. Malfertheiner, N. Werner, et al., Nocturnal hypoxemic burden during positive airway pressure treatment across different central sleep apnea etiologies, Sleep Med. 79 (2021) 62–70.
[47] PyEDFlib -EDF/BDF Toolbox in Python. [Online]. Available: https://pyedflib.readthedocs.io/en/latest/. Accessed: September 9, 2022.
[48] EDFplusLibraryR. [Online]. Available: https://github.com/diegoalvareze/EDFplusLibraryR. Accessed: September 9, 2022.
[49] Luna. [Online]. Available: https://zzz.bwh.harvard.edu/luna/#things-luna-aims-to-do. Accessed: September 9, 2022.
[50] EDFbrowser. [Online]. Available: https://www.teuniz.net/edfbrowser/. Accessed: September 9, 2022.
[51] Free EDF(+) software and datafiles. [Online]. Available: https://www.edfplus.info/downloads/. Accessed: September 9, 2022.