Development and validation of a diagnostic model for obstructive sleep apnea: a Bayesian network approach

Daniela Filipa Ferreira dos Santos

November 2022

Dissertação de candidatura ao grau de Doutor apresentada à Faculdade de Medicina da Universidade do Porto, no âmbito do Programa Doutoral em Investigação Clínica e em Serviços de Saúde.

This PhD thesis has been submitted in fulfilment of the requirements for the PhD degree in Clinical and Health Services Research at the Faculty of Medicine of the University of Porto.

Júri

Exma. Senhora

Mestre Daniela Filipa Ferreira dos Santos

ferreiradossantos.daniela@gmail.com

U.PORTO

v.referência	v.comunicação	n.referência	data
		FOA.26. 3222-2022	2022.09.01

assunto

Provas de Doutoramento

Informo V. Exª. que, por meu despacho de 2022.09.01, proferido no âmbito de delegação reitoral, nomeei o júri proposto para as provas de doutoramento em Investigação Clínica e em Serviços de Saúde, requeridas por V. Exª., com a seguinte constituição:

Presidente: Doutor José Paulo Alves Vieira de Andrade, Professor Catedrático da Faculdade de Medicina da Universidade do Porto

Vogais:

- Doutor Gonzalo César Gutiérrez Tobal, Investigador da University of Valladolid, Spain;

- Doutor Pedro Manuel Henriques da Cunha Abreu, Professor Auxiliar da Universidade de Coimbra;

- Doutora Amélia Maria dos Santos Feliciano, Professora Auxiliar da Universidade Católica Portuguesa;

- Doutor Pedro Pereira Rodrigues, Professor Auxiliar da Faculdade de Medicina da Universidade do Porto;

- Doutora Marta Susana Monteiro Drumond Freitas, Professora Auxiliar Convidada da Faculdade de Medicina da Universidade do Porto;

- Doutora Claudia Camila Rodrigues Pereira Dias, Investigadora Auxiliar da Faculdade de Medicina da Universidade do Porto.

Com os melhores cumprimentos,

A Vice-Reitora,

(Profª. Doutora Fátima Vieira)

(1/6)/PV

UNIVERSIDADE DO PORTO, REITORIA
Praça Gomes Teixeira, 4099-002 Porto
TEL. +351 22 040 8000 FAX +351 22 040 8186/ 8187
URL www.up.pt

Title

Development and validation of a diagnostic model for obstructive sleep apnea: a Bayesian network approach

Candidate

Daniela Filipa Ferreira dos Santos

Adviser

Professor Doutor Pedro Pereira Rodrigues

Professor Auxiliar no MEDCIDS/CINTESIS – Departamento de Medicina da Comunidade, Informação e Decisão em Saúde

Diretor do Programa Doutoral em Ciência de Dados em Saúde

Co-advisers

Professor Doutor Ricardo João Cruz Correia

Professor Auxiliar no MEDCIDS/CINTESIS – Departamento de Medicina da Comunidade, Informação e Decisão em Saúde

Professor Doutor Peter Lucas

Full Professor in University of Twente, Enschede, The Netherlands

Artigo 48o, parágrafo 3º: A Faculdade não responde pelas doutrinas expendidas na dissertação
(Regulamento da Faculdade de Medicina do Porto. Lei nr. 19337, de 29 de Janeiro de 1931)

Esta investigação teve como entidade de acolhimento o Centro de Investigação em Tecnologias e Serviços de Saúde (CINTESIS), e foi financiada por uma bolsa individual de doutoramento da Fundação para a Ciência e Tecnologia (PD/BD/13553/2018 e COVID/BD/152608/2022).

Professores Catedráticos

Patrício Manuel Vieira Araújo Soares Silva
Alberto Manuel Barros Da Silva
José Henrique Dias Pinto De Barros
Maria Fátima Machado Henriques Carneiro
Maria Dulce Cordeiro Madeira
Altamiro Manuel Rodrigues Costa Pereira
Manuel Jesus Falcão Pestana Vasconcelos
João Francisco Montenegro Andrade Lima Bernardes
Maria Leonor Martins Soares David
Rui Manuel Lopes Nunes
José Manuel Pereira Dias De Castro Lopes
Joaquim Adelino Correia Ferreira Leite Moreira
Raquel Ângela Silva Soares Lino
Fernando Manuel Mendes Falcão Dos Reis
Francisco José Miranda Rodrigues Cruz
José Paulo Alves Vieira De Andrade
Jorge Manuel Silva Junqueira Polónia
José Luís Dias Delgado
Isaura Ferreira Tavares
Fernando Carlos De Landér Schmitt
Acácio Agostinho Gonçalves Rodrigues
Maria De Fátima Moreira Martel
João Tiago De Sousa Pinto Guimarães
José Carlos Lemos Machado
José Carlos De Magalhães Silva Cardoso

Professores Catedráticos Jubilados e Aposentados

Alexandre Alberto Guerra Sousa Pinto
Álvaro Jerónimo Leal Machado De Aguiar
António Albino Coelho Marques Abrantes Teixeira
António Carlos De Freitas Ribeiro Saraiva
António José Pacheco Palha
António Manuel Sampaio De Araújo Teixeira
Belmiro Dos Santos Patrício
Cândido Alves Hipólito Reis
Carlos Rodrigo Magalhães Ramalhão
Cassiano Pena De Abreu E Lima
Deolinda Maria Valente Alves Lima Teixeira
Eduardo Jorge Cunha Rodrigues Pereira
Fernando Tavarela Veloso
Francisco Fernando Rocha Gonçalves
Isabel Maria Amorim Pereira Ramos
Jorge Manuel Mergulhão Castro Tavares
José Agostinho Marques Lopes
José Carlos Neves Da Cunha Areias
José Eduardo Torres Eckenroth Guimarães
José Fernando Barros Castro Correia
José Manuel Costa Mesquita Guimarães
José Manuel Lopes Teixeira Amarante
Levi Eugénio Ribeiro Guerra
Luís Alberto Martins Gomes De Almeida
Manuel Alberto Coimbra Sobrinho Simões
Manuel António Caldeira Pais Clemente
Manuel Augusto Cardoso De Oliveira
Manuel Machado Rodrigues Gomes
Manuel Maria Paula Barbosa
Maria Amélia Duarte Ferreira
Maria Da Conceição Fernandes Marques Magalhães
Maria Isabel Amorim De Azevedo
Rui Manuel Almeida Mota Cardoso
Rui Manuel Bento De Almeida Coelho
Serafim Correia Pinto Guimarães
Valdemar Miguel Botelho Dos Santos Cardoso
Walter Friedrich Alfred Osswald

To
Júlio, Rosalina, and Damião

Contents

Acknowledgements

To Professor Pedro, my “mestre”. You were always by my side, and I hope you still will. I truly hope that you are proud of my path and my work. I always aim for more because of you. Thanks for being my guide and my haven.

To Professor Ricardo for always believing in me and my capabilities and for not worrying about anything. I always knew that if I need anything you were always by my side.

To Professor Peter for always answering my emails and all my questions. I really hoped that we could spend more time working face-to-face.

To all the professionals at the physical and digital archive of University Hospital Center of São João as well as the Sleep Laboratories for all the help during the data collection process and their expertise.

To Pedro Amorim for the continuation of the theme and work and for all the help in all the steps and for always being my friend.

To all members of the department and AI4Health group for constant willingness to discuss my work. A special word to Priscila Maranhão, Camila Dias, Matilde Monteiro-Soares, Teresa Henriques, Inês Ribeiro-Vaz, Tiago Martins, João Almeida, and room office ten.

To my parents, Júlio and Rosalina, for always being there by my side, even when they did not understand a word. I hope that you are proud of me as much as I am proud of being your daughter. Finally, to the guy that is always my example in life, and that always have my back – my brother.

I would like to thank Fundação para a Ciência e Tecnologia (FCT) for providing me with a grant that allowed me to live this project to the fullest.

List of publications

This thesis was financed by the Fundação para a Ciência e Tecnologia (FCT) under PhD grants numbers PD/BD/13553/2018 and COVID/BD/152608/2022 from March 2018 up to July 2022. It was conducted in the Center for Health Technology and Services Research (CINTESIS) and Community Medicine, Information and Decision Sciences, Faculty of Medicine of the University of Porto.

Core Research Papers

The 10 papers described below are the core structure of this thesis (7 were already published, 1 is under review, and 2 are waiting for submission). The manuscripts are listed by order of appearance in the thesis.

Enabling early obstructive sleep apnea diagnosis with machine learning: Systematic review
J Med Internet Res, 24(9), 2022. doi: 10.2196/39452
Daniela Ferreira-Santos, Pedro Amorim, Tiago Silva Martins, Matilde Monteiro-Soares, and Pedro Pereira Rodrigues
Journal Citation Reports Impact Factor in 2021: 7.093 (Q1)

Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis
Proceedings of Studies in Health Technology and Informatics. 247: 126–130, 2018. doi:10.3233/978-1-61499-852-5-126
Daniela Ferreira-Santos, Matilde Monteiro-Soares, and Pedro Pereira Rodrigues

Association between co-morbidities and prescribed drugs in obstructive sleep apnea suspected patients: an inductive rule learning approach
(submitted, April 2022)
Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Finding groups in obstructive sleep apnea patients: a categorical cluster analysis
Proceedings of 2018 IEEE 31st Symposium on Computer-Based Medical Systems. 387–392, 2018. doi:10.1109/CBMS.2018.00074
Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Phenotyping obstructive sleep apnea patients: a first approach to cluster visualization
Proceedings of Studies in Health Technology and Informatics. 255: 75–79, 2018. doi:10.3233/978-1-61499-921-8-75
Daniela Ferreira-Santos and Pedro Pereira Rodrigues

A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches

International Journal of Data Science and Analytics. 8: 339–349, 2019. doi:10.1007/s41060-018-0118-x

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Journal Citation Reports Impact Factor in 2019: ? (Q3)

Enhancing obstructive sleep apnea diagnosis with screening through disease phenotypes: algorithm development and validation

JMIR Medical Informatics. 9: 1–16, 2021. doi:10.2196/25124

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Journal Citation Reports Impact Factor in 2021: 3.231 (Q3)

Obstructive sleep apnea: a categorical cluster analysis and visualization.

Pulmonology. 16:53, 2021. doi:10.1016/j.pulmoe.2021.10.003.

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Journal Citation Reports Impact Factor in 2021: 9.216 (Q1)

Prospective validation of a Bayesian network model in the diagnosis of Obstructive Sleep Apnea

(finalizing details)

Pedro Amorim, **Daniela Ferreira-Santos**, Marta Drummond, Pedro Pereira Rodrigues

Validation of OSABayes: a screening tool for obstructive sleep apnea

(finalizing details)

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Some of the articles were first presented as oral or poster communications in national and international conferences.

Oral communications

27º Congresso Português de Cardiopneumologia 2022, Peniche

Machine learning: criação de modelos preditivos clínicos

Special Topic Conference 2018, Zagreb

Phenotyping obstructive sleep apnea patients: a first approach to cluster visualization

Computer-Based Medical Systems (CBMS) 2018, Karlstad

Finding groups in obstructive sleep apnea patients: a categorical cluster analysis

Computer-Based Medical Systems (CBMS) 2018, Karlstad

PhD: Implementation and validation of a diagnostic model in OSA: a Bayesian network approach

Medical Informatics Europe (MIE/SFMI) 2018, Gothenburg

Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis

Poster communications

Encontro com a Ciência e Tecnologia em Portugal 2021, Lisboa

Obstructive sleep apnea: cluster analysis and visualization

e-Poster presented on 2020 European Respiratory Society International Congress

Prospective validation of a Bayesian network model in the diagnosis of Obstructive Sleep Apnea - preliminary results.

Pedro Amorim, Daniela Ferreira-Santos, Marta Drummond, Pedro Pereira Rodrigues

Encontro com a Ciência e Tecnologia em Portugal 2018, Lisboa

A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches

In addition, during the duration of this thesis conduction, the candidate was also author and co-author of other papers. Although these studies were not part of the thesis core structure, they were important to improve the researcher knowledge on the field and/or to present the results to the community. They are listed below:

Identifying common baseline clinical features of COVID-19: a scoping review
BMJ Open. 10: e041079, 2020. doi:10.1136/bmjopen-2020-041079
Daniela Ferreira-Santos, Priscila maranhão, Matilde Monteiro-Soares
Journal Citation Reports Impact Factor in 2021: 3.017 (Q2)

Causality assessment of adverse drug reaction reports using an expert-defined Bayesian network
Artificial Intelligence in Medicine. 91: 12–22, 2018. doi:10.1016/j.artmed.2018.07.005
Pedro Pereira Rodrigues, **Daniela Ferreira-Santos**, Ana Silva, Jorge Polónia, Inês Ribeiro-Vaz
Journal Citation Reports Impact Factor in 2021: 7.011 (Q1)

Abbreviations and Notation

AASM – American Academy of Sleep Medicine

AC – Abdominal circumference

AHI – Apnea-hypopnea index

BN – Bayesian network

CRISP-DM – Cross-Industry standard Process for Data Mining

ICSD-3 – 3rd edition of the International Classification of Sleep Disorders

NB – Naïve Bayes

NC – Neck circumference

OSA – Obstructive sleep apnea

PSG – Polysomnography

ROC – Receiver operator curve

TAN – Tree-augmented Naïve Bayes

List of Figures

Abstract

Obstructive sleep apnea (OSA) is one of the most prevalent sleep problems with various clinical presentations that have not been formally characterized, resulting in missed or delayed diagnosis. Bayesian networks can provide a useful aid in this chronic disease, as published studies, develop to improve diagnosis, have been showing heterogeneous results. As so, the development and evaluation of predictive models are required in OSA.

This thesis has three objectives: summarize the evidence regarding risk and diagnostic factors related to OSA and all data mining techniques applied within this disease, propose Bayesian network models as the elected data mining technique, and implement the model in a non-sleep setting.

Aiming at identifying predictive variables related to OSA, a systematic review was performed. Factors such as body mass index, age, sex, neck circumference and snoring were identified. Two studies reveal maximum sensitivity or specificity, with a traditional statistical approach as the best model for rule-out patients. To propose Bayesian networks as the best model for OSA screening and diagnosis, eight observational studies was conducted: two related to data understanding/preparation and six to modeling/evaluation. Overall results demonstrated that the chosen data mining process has good results and can be applied in OSA. As so, in the last objective, we develop and validate an online form and app, based on the identified factors and technique, achieving an easy and quick tool that can be used in a non-sleep setting.

In conclusion, although clinical predictions algorithms have a low level of accuracy for the diagnosis of OSA, as described in the American Academy of Sleep Medicine latest guideline, they can be used in a non-sleep clinic setting, as Bayesian networks may be helpful at identifying high-risk patients.

Resumo

A apneia obstrutiva do sono (AOS) é uma das mais prevalentes doenças do sono, com múltiplas apresentações clínicas que ainda não foram formalmente caracterizadas, resultando em diagnósticos perdidos ou tardios. As redes Bayesianas podem fornecer um auxílio útil nesta doença crónica, pois os estudos publicados e desenvolvidos para melhorar o seu diagnóstico apresentam resultados heterogéneos. Assim, o desenvolvimento e avaliação de modelos preditivos são necessários na AOS.

Esta tese tem três objetivos: resumir a evidência relativa a fatores de risco e diagnósticos relacionados com a AOS e todas as técnicas de mineração de dados aplicadas nesta doença, propondo as redes Bayesianas como a técnica de eleição, e sua implementação num ambiente fora da consulta do sono.

Com o objetivo de identificar as variáveis preditivas relacionadas com a AOS foi realizada uma revisão sistemática. Fatores como índice de massa corporal, idade, sexo, circunferência do pescoço, e roncopatia foram identificados. Dois estudos revelaram máxima sensibilidade ou especificidade, com uma abordagem estatística tradicional como o melhor modelo a descartar não doentes. De forma a propor as redes Bayesianas como o melhor modelo para o rastreamento e diagnóstico da AOS, oito estudos observacionais foram realizados: dois relacionados com a compreensão e preparação dos dados, e seis com a modelação e avaliação dos modelos. Globalmente, os resultados demonstraram que o processo de mineração escolhido tem bons resultados e pode ser aplicado à AOS. Assim, no último objetivo, desenvolvemos e validamos um formulário online e aplicativo, baseado nos fatores e técnica escolhida, alcançando uma ferramenta de fácil uso e de rápida aquisição de dados, que pode ser usada num ambiente fora da consulta do sono.

Em conclusão, embora os algoritmos clínicos de previsão tenham um baixo nível de precisão para o diagnóstico de AOS, conforme descrito na última *guideline* clínica publicada pela *American Academy of Sleep Medicine,* os mesmos podem ser utilizados num contexto exterior à consulta do sono, sendo que as redes Bayesianas podem ser úteis em reconhecer pacientes com elevado risco da doença.

Outline

The idea of this thesis started in 2013, in a class given by Professor Pedro Pereira Rodrigues. The demonstration of how a Bayesian network works and how it can help the healthcare system fascinated me. As a healthcare professional, seeing that there are other ways of reaching the same goal, which is helping patients, put me on the path that I now finishing. The journey is long but exciting. Along the path, some ideas, discussions, and extra-curricular courses lead me to a framework that is fully integrated into this work.

This thesis is organized as follows:

Chapter 1 presents a brief introduction to obstructive sleep apnea (OSA) and Bayesian networks (BN), with a description of the utilized thesis framework.

Chapter 2 synthetizes the aim and specific objectives of this thesis.

Chapters 3, 4 and 5 present the main results of the different studies developed to achieve this thesis's three main objectives. When a chapter has more than one article with conclusions, a summary is made.

Chapter 3 corresponds to the first phase of the thesis framework with a synthesis of evidence, published in the literature, regarding OSA and machine learning approaches utilized.

Chapter 4 is divided into three sections. The first is related to the second and third phases of the framework, which describe data understanding and preparation. The second corresponds to the fourth phase, which is related to modeling, that in this thesis is demonstrated by applying cluster analysis. The last phase corresponds to the modeling and evaluation, where we explore rule-out approach.

Chapter 5 communicates phase six of the framework, with the implementation and validation of the developed tool.

Attachments include ethical permissions and permissions to reuse the published manuscripts.

Chapter 1: Rationale

Rationale

1.1. Obstructive sleep apnea

To start comprehending the data mining project involved in this thesis, we first need to realize the disease in study. Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by clinical symptoms (e.g., snoring), maintenance of respiratory effort, and at least five events per hour of narrowing (apnea or hypopnea) of the upper airway that impairs normal ventilation during sleep [1]. An apnea consists of a cessation of airflow higher than 90% of the baseline for at least 10 seconds, a hypopnea is a reduction in airflow (30-50%) along with a decreased saturation of 3-4% from pre-event baseline and/or associated with an arousal, and the apnea-hypopnea index (AHI) is the number of such events per hour of sleep, obtained through polysomnography (PSG) [2].

OSA is caused by the relaxation of the muscles in the back of the throat that impairs normal breathing, namely soft palate (back of the roof of the mouth), uvula (the triangular piece of tissue hanging from the soft palate), tonsils, and tongue; basically, a balance between anatomical and neuromuscular responses [3]. Overall, OSA is largely unrecognized and undiagnosed, representing a significant burden to the health care system [4]. Additionally, OSA prevalence has been underestimated as studies varies significantly both in the population being studied, criteria used, and OSA definition (AHI thresholds). The latest study by Benjafield et al [5], utilizing American Academy of Sleep Medicine (AASM) diagnostic criteria and the third edition of the International Classification of Sleep Disorders (ICSD-3) classification, estimated that 936 million adults have OSA; representing 17 % in Portugal. Furthermore, in the primary care context, where the reported prevalence of OSA was 43%, only 3% had a diagnosis, as most primary care physicians do not routinely screen OSA high risk patients [6].

1.1.1. Risk and diagnostic factors

Various groups of variables, such as demographical, physical examination, clinical history, and comorbidities need to be considered when talking about OSA. As referred in Bonsignore et al , OSA is one of the sleep disorders with the most associated comorbidities [7]. Major risk factors included male gender, obesity, aging, increased neck (NC) and abdominal circumferences (AC), craniofacial and upper-airway abnormalities (CFA), alcohol consumption, menopausal status, history of smoking, arterial and pulmonary hypertension, atrial fibrillation, stroke, coronary arterial disease, and even truck driver as a profession needs to be consider [8], [9]. Likewise, when performing an evaluation of a suspicion OSA patient, a sleep-

oriented history and findings identified in PSG are fundamental for stablishing an OSA diagnosis criteria. To do so, physicians need to assess snoring, gasping or choking episodes, daytime sleepiness, nocturia, morning headaches, fragmented sleep, decreased concentration or memory loss, and irritability.

1.2. Bayesian networks

Clinical decision tools are a kind of aid developed for a healthcare professional involved in patient care [10]. Traditionally, a diagnostic model is based in logistic or linear regressions, as this techniques have the advantage of being easily interpretable [11], but lack graphical representation. As so, we need to go beyond traditional biostatistics [12] and utilize all the available evidence, even with uncertainty present. To formalize this uncertainty, we need robust methods which can work with this concept, as Bayesian networks (BN).

A BN is a model of a joint probability distribution of one set of random variables, specifying the assumption of independence between them with the interdependence between variables being represented by a directed acyclic graph [13]. Each variable is represented by a node in the graph and is dependent on the set of variables represented by its ascendant nodes. This dependence is represented by a conditional probability table that describes the probability distribution of each variable, given their ascendant variables. Generally, Bayesian statistical approaches compute the probability that the hypothesis is true by updating the pre-existent hypothesis probability with the new incoming data [16]. This is known as posterior probability and can be easily calculated with Bayes theorem:

$$p\,(A \mid B) = (p\,(B \mid A) \times p\,(A)) \div p\,(B)$$

There are some Bayesian network classifiers, specifically Naïve Bayes (NB), which assumes conditional independence among factors and outcome, and Tree Augmented Naïve Bayes (TAN), which allows for an optional dependence for each factor with an Naïve Bayes extension, up to two dependencies for each factor [12]. The last is a better representation of real-world data, as in the real-world variables are not independent, but related. Both representations include a qualitative and a quantitative model; the first, represents the relationships among variables, and the second, a joint probability distribution represented by the conditional probabilities.

1.2.1. Bayesian network validation

Subsequently the BN development, we need to estimate the performance of the model, namely its capacity to predict cases on an independent sample. This can be done by an independent external sample or internally by applying cross-validation or leave-one-out; the first divide the sample into k exclusive sets of equal size (folds), and the second being a special case of cross-validation, where the number of folds is the number of cases in the sample. The validation result in a confusion matrix, where we have the real result and the prediction (result of the model), which allows us to calculate many measures, such as sensitivity and specificity. Additionally, we can produce a receiver operating characteristics curve (ROC), which assess the discriminatory power of the models by visualizing the relation between model sensitivity and specificity and enabling the selection of an optimal cutoff point (high sensitivity, for example).

Although, Bayesian networks was chosen as the data mining technique to solve underdiagnose OSA, the process in CRISP-DM is iterative, as shown in Figure 1.

1.3. CRoss-Industry Standard Process for Data Mining

CRISP-DM is an industry-independent process and a comprehensive data mining methodology that provides a complete framework, and consists of six phases (Business understanding, Data understanding, Data preparation, Modeling, Evaluation, and Deployment) [15], [16]. The first phase starts with an assessment of the business situation so that we can get an overview of the current situation; that in our case, is the comprehension of what was already done or study in OSA, as well as a compile of the predictive variables for the disease, and even a summary of data mining techniques already performed within this disease. Secondly, we need to collect data from various data sources and afterwards describe the obtained dataset, which in this thesis resulted in two obtained datasets (Vila Nova de Gaia/Espinho and São João Hospitals) with an assessment of data quality performed in both. Thirdly, to construct these datasets, we need to define inclusion and exclusion criteria. The inclusion criteria were based on adult OSA suspicion participants undergoing PSG, while participants already diagnosed, suspected of having another sleep disease, with severe lung or neurological conditions, and pregnant women were excluded. Fourthly, we need to select the best modeling technique based on the problem and the data, as is this that will make possible to make a project plan which elects the best type of data mining and success criteria, that result in predictive data mining, namely classification and prediction analysis, and descriptive data mining – clustering. Fifthly, the result of Phase 4 needs to be evaluated, interpreted and possible define further

actions, which is Phase 5 in the CRISP-DM framework. Sixthly, after compliance with all phases, we need a deployment phase, that in this work resulted not also in the publication of several journal articles, conference proceedings, oral communications, and posters, but in the external validation of a previous published Bayesian network that resulted in an online form and app.

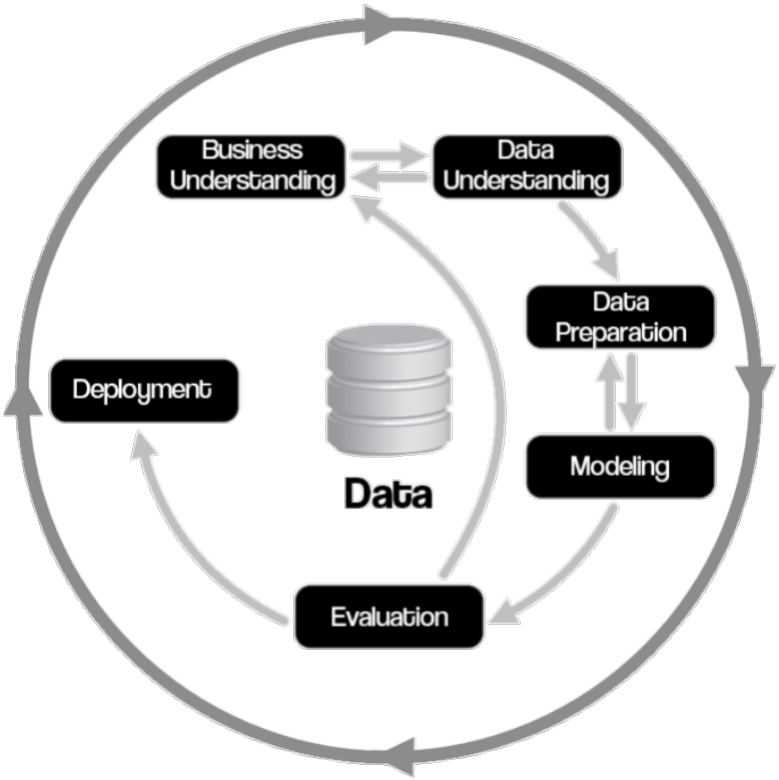


Figure 2 - Phases of the CRISP-DM reference model. Figure extracted from https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png

Chapter 2: Objectives

Objectives

This thesis has three main goals:

Goal 1: Synthesis of evidence (Chapter 3)

- To gather and analyze existing evidence on risk and diagnostic factors for obstructive sleep apnea (OSA) that posteriorly can be collected from Portuguese care centers, as for all available types of models to predict OSA.

Goal 2: Classification and Prediction (Chapter 4)

- To propose diagnostic Bayesian network models based on factors identified, focusing on interpretability.

Goal 3: Implementation (Chapter 5)

- To develop and validate a decision support system, based on the proposed Bayesian network model, to be used in Portuguese health system.

Chapter 3: Synthesis of evidence

Synthesis of evidence

The first aim of this thesis is to summarize evidence regarding obstructive sleep apnea (OSA) and corresponding risk and diagnostic factors, limited to demographical, physical examination, clinical history, and comorbidities. This summary has in consideration the last American Academy of Sleep Medicine (ASSM) guideline [3] for diagnosing OSA, published in 2017, and a literature review performed in 2015. The literature review was not published as an article itself, but is mention in all articles done by the student, especially in the background of the study performed in 2019 [17].

The ASSM guideline recommends that *"clinical tools, questionnaires and prediction algorithms not to be used to diagnose OSA in adult, in the absence of polysomnography (PSG) or home sleep apnea testing",* but also state that *"in non-sleep clinic settings, these tools may be more helpful to identify patients who are at increased risk for OSA";* which lead to a systematic review that compared the accuracy of prediction algorithms against PSG.

One study was conducted and corresponds to Business Understanding – Phase 1 in CRIPS-DM:

3.1. Enabling early obstructive sleep apnea diagnosis with machine learning: Systematic review
J Med Internet Res, 24(9), 2022. doi: 10.2196/39452
Daniela Ferreira-Santos, Pedro Amorim, Tiago Silva Martins, Matilde Monteiro-Soares, and Pedro Pereira Rodrigues
Journal Citation Reports Impact Factor in 2021: 7.093 (Q1)

3.1. Enabling early obstructive sleep apnea diagnosis with machine learning: Systematic review

J Med Internet Res, 24(9), 2022. doi: 10.2196/39452

Daniela Ferreira-Santos, Pedro Amorim, Tiago Silva Martins, Matilde Monteiro-Soares, and Pedro Pereira Rodrigues

Journal Citation Reports Impact Factor in 2021: 7.093 (Q1)

Review

Enabling Early Obstructive Sleep Apnea Diagnosis With Machine Learning: Systematic Review

Daniela Ferreira-Santos[1,2], MSc; Pedro Amorim[1,2,3], MSc; Tiago Silva Martins[2], MSc; Matilde Monteiro-Soares[1,2,4], PhD; Pedro Pereira Rodrigues[1,2], PhD

[1]Department of Community Medicine, Information and Decision Sciences, Faculty of Medicine, University of Porto, Porto, Portugal
[2]Center for Health Technology and Services Research, Porto, Portugal
[3]Sleep and Non-Invasive Ventilation Unit, São João University Hospital, Porto, Portugal
[4]Portuguese Red Cross Health School Lisbon, Lisbon, Portugal

Corresponding Author:
Daniela Ferreira-Santos, MSc
Department of Community Medicine, Information and Decision Sciences
Faculty of Medicine, University of Porto
Rua Dr Plácido da Costa, s/n
Porto, 4200-450
Portugal
Phone: 351 937710766
Email: danielasantos@med.up.pt

Abstract

Background: American Academy of Sleep Medicine guidelines suggest that clinical prediction algorithms can be used to screen patients with obstructive sleep apnea (OSA) without replacing polysomnography, the gold standard.

Objective: We aimed to identify, gather, and analyze existing machine learning approaches that are being used for disease screening in adult patients with suspected OSA.

Methods: We searched the MEDLINE, Scopus, and ISI Web of Knowledge databases to evaluate the validity of different machine learning techniques, with polysomnography as the gold standard outcome measure and used the Prediction Model Risk of Bias Assessment Tool (Kleijnen Systematic Reviews Ltd) to assess risk of bias and applicability of each included study.

Results: Our search retrieved 5479 articles, of which 63 (1.15%) articles were included. We found 23 studies performing diagnostic model development alone, 26 with added internal validation, and 14 applying the clinical prediction algorithm to an independent sample (although not all reporting the most common discrimination metrics, sensitivity or specificity). Logistic regression was applied in 35 studies, linear regression in 16, support vector machine in 9, neural networks in 8, decision trees in 6, and Bayesian networks in 4. Random forest, discriminant analysis, classification and regression tree, and nomogram were each performed in 2 studies, whereas Pearson correlation, adaptive neuro-fuzzy inference system, artificial immune recognition system, genetic algorithm, supersparse linear integer models, and k-nearest neighbors algorithm were each performed in 1 study. The best area under the receiver operating curve was 0.98 (0.96-0.99) for age, waist circumference, Epworth Somnolence Scale score, and oxygen saturation as predictors in a logistic regression.

Conclusions: Although high values were obtained, they still lacked external validation results in large cohorts and a standard OSA criteria definition.

Trial Registration: PROSPERO CRD42021221339; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=221339

(J Med Internet Res 2022;24(9):e39452) doi: 10.2196/39452

KEYWORDS
machine learning; obstructive sleep apnea; systematic review; polysomnography

Introduction

Background

Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by recurrent episodes of partial (hypopnea) or complete (apnea) upper airway obstruction, repeated throughout sleep. Its prevalence varies significantly according to how OSA is defined (methodology, criteria used such as apnea index, apnea-hypopnea index [AHI], or respiratory disturbance index and threshold definitions) and the population being studied [1]. The study by Benjafield et al [2] estimated that worldwide, 936 million adults aged 30 to 69 years have OSA. Despite this high prevalence, many cases remain undiagnosed and untreated, leading to a decrease in patients' quality of life and an increased risk of adverse events, with a high impact on morbidity and mortality [3]. Polysomnography (PSG) is the gold standard test for diagnosing OSA [1]. However, performing PSG is costly, time-consuming, and labor-intensive. Most sleep laboratories face long waiting lists of patients, as PSG is neither a routine clinical practice nor an absolute suitable screening tool [4]. Given these limitations, it would be useful to develop a clinical prediction model that could reliably identify the patients most likely to benefit from PSG, that is, exclude OSA diagnosis when the probability is low, establish a priori probability before considering PSG, and prioritize patients in need of PSG according to the probability of a positive result. This idea was backed up by the American Academy of Sleep Medicine (AASM) in its latest guidelines [1]. Clinical prediction models should be easy to use and easy to calculate. The model must be based on the gold standard and required to be validated, and when used for screening, its purpose depends on whether the path leads to a rule-out or rule-in approach. In the first case, we should have a high-sensitivity model, omitting the need to perform PSG in healthy patients. By contrast, if we chose a rule-in approach, a high-specificity model is needed to select patients with a high probability of having OSA, suitable for undergoing PSG.

Objective

Given these shortcomings, this systematic review aimed to identify, gather, and analyze existing machine learning approaches that are being used for disease screening in adult patients with suspected OSA.

Methods

This systematic review was carried out according to a protocol registered with PROSPERO (International Prospective Register of Systematic Reviews; CRD42021221339).

Search Strategy and Selection Criteria

We searched all evidence available in the MEDLINE database (PubMed) and in Scopus and ISI Web of Knowledge published until June 2020 in English, French, Spanish, or Portuguese. Specific queries were used (with a refresh in October 2021), and a manual search was also performed by using the references of the included studies and pertinent reviews on the topic. In addition, contact with specialists in the field was made to check whether all pertinent information was retrieved. Articles were selected by 3 reviewers independently (blinded to each other's assessment) by applying the criteria to each title and abstract and then assessed fully. Divergent opinions were resolved through consensus. All processes were performed in Rayyan, a web application and mobile app for systematic reviews [5].

Studies including adult patients with suspected OSA (population) that assessed the accuracy of predictive models using known symptoms and signs of OSA (exposure and comparator) and had PSG as the gold standard (outcome) were eligible as per the selection criteria.

Data Extraction

Once the articles were selected, data were extracted into a prespecified Excel spreadsheet and included (1) article information: title, author(s), publication date, country, and journal and (2) methods: study design, setting, study period, type of model, inclusion and exclusion criteria, participant selection, sample size, clinical factors analyzed, diagnostic test analyzed, and potential bias. For each type of model, specific data extraction was created and fulfilled, as demonstrated in the tables in further sections. We have ordered the identified studies by the obtained article results: first, the articles that only developed the algorithm; then the ones that internally validated the algorithm; and finally, the ones that externally validated the prediction algorithm. Within each subsection, we organized the published works by year of publication. Any missing information from the studies is reported in the Results section by "—" (not available), and the best obtained predictive model is marked in italic. Also, if the study applied different machine learning approaches, the clinical factors analyzed, and the discrimination measures are only described for the best obtained model.

Risk of Bias

At 2 points in time, 1 reviewer assessed the risk of bias and applicability by applying the Prediction Model Risk of Bias Assessment Tool (PROBAST) to all the included studies. This is specific for studies developing, validating, or updating diagnostic prediction models. More details are available in the study by Moons et al [6]. An important aspect needs to be referred to, as this tool states that *"if a prediction model was developed without any external validation, and it was rated as low risk of bias for all domains, consider downgrading to high risk of bias. Such a model can only be considered as low risk of bias if the development was based on a very large data set and included some form of internal validation."* This means that the included studies only performing model development will be marked as high risk of bias. For those with internal validation, the risk of bias will depend on the sample size based on the number of events per variable (≥20 ratio between events and variables in development studies and ≥100 participants with OSA for model validation studies). In addition, studies that randomly split a single data set into development and validation are considered as internal validation.

Results

Overview

We retrieved 6769 articles, 1290 being duplicates. From the 5479 articles, we kept 63 studies that fulfilled the inclusion criteria, as shown in Figure 1.

The gold-standard examination—PSG—was performed in all the articles assessed, with one also adding the diagnostic part of the split-night exam [7]. The highest found age was 96 years [8], with 54% (34/63) of studies presenting patients with ages of >18 years. To be certain to include all OSA clinical prediction algorithms, we kept the studies that only reported a mean age and SD, with this value being >42, and SD varying between 6 and 16 years. In addition, 10% (6/63) of studies reported an age group <18 years (>14 and >15 years in 2/6, 33% studies and >16 and >17 in 4/6, 66% others, respectively). Regarding the suspicion of OSA, this description was shown in 65% (41/63) of studies, whereas 32% (20/63) introduced OSA suspicion and any other sleep disorder. In addition, we have a study with healthy patients and patients with suspected OSA [9] and another that does not specifically state this; instead, the authors write that patients already diagnosed with OSA were excluded from the study. The frequency of occurrence of the various clinical factors analyzed in more than 1 study is shown in Table 1.

There were disagreements between the reviewers in both phases, with an overall concordance rate of 78% in the title and abstract screening and 95% in the integral version.

Figure 1. Flow diagram of the study selection process.

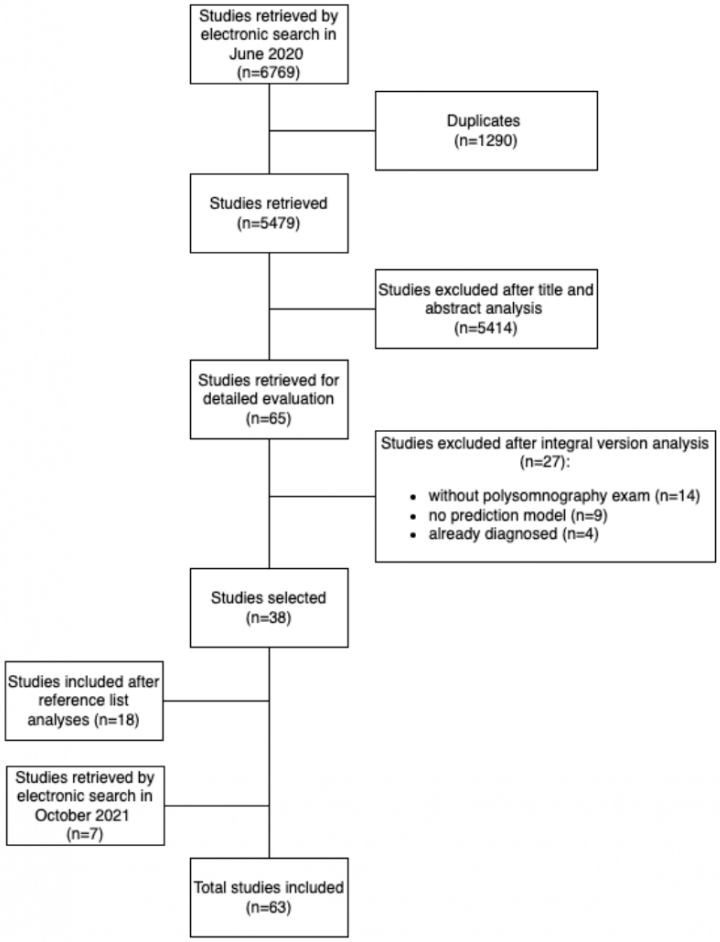


Table 1. The frequency of occurence of the various clinical factors analyzed that appears more than once in all the included studies (n=63).

Clinical factors analyzed	Frequency of occurrence, n (%)
BMI	37 (59)
Age	32 (51)
Sex	29 (46)
Neck circumference	25 (40)
Snoring	14 (22)
Epworth Somnolence Scale	10 (16)
Witnessed apneas	8 (13)
Waist circumference	8 (13)
Breathing cessation	7 (11)
Daytime sleepiness	7 (11)
Hypertension	7 (11)
Gasping	6 (10)
Oxygen saturation	6 (10)
Oxygen desaturation	6 (10)
Blood pressure	5 (8)
Smoking	5 (8)
Tonsil size grading	5 (8)
Modified Mallampati score	4 (6)
Alcohol consumption	3 (5)
Awakenings	3 (5)
Diabetes	3 (5)
Height	3 (5)
Nocturia	3 (5)
Restless sleep	3 (5)
Weight	3 (5)
Craniofacial abnormalities	2 (3)
Driving sleepy	2 (3)
Face width	2 (3)
Friedman tongue score	2 (3)
Snorting	2 (3)

Prediction Models Development

New prediction models were developed in 23 studies, as presented and described in Table 2. The most common approach was regression techniques, with logistic (6/23, 26%), linear (6/23, 26%), logistic and linear (6/23, 26%), and logistic regression compared with decision trees and support vector machines (3/23, 13%). In addition, 4% (1/23) of articles produced a Pearson correlation and another (1/23, 4%) produced a decision tree. The oldest model was developed in 1991 and included sex, age, BMI, and snoring whereas in 2020 the predictive variables included besides these were height, weight, waist size, hip size, neck circumference (NC), modified Friedman score, daytime sleepiness, and Epworth Somnolence Scale score. Only 13% (3/23) studies described the study design and period, with 22% (5/23) being retrospective. Regarding OSA definition by PSG, 4% (1/23) study did not report the cutoff, while 17% (4/23) reported an AHI>10 and 17% (4/23) more reported an AHI≥15. The largest sample size was 953, and the smallest was 96 patients with suspected OSA. An overall prevalence of OSA between 31% and 87% was stated, with 9% (2/23) of studies presenting incorrect percentage values [10,11]. Regarding discrimination measures, although no validation was performed, the best area under the receiver operating characteristic curve (AUC), sensitivity, and specificity were 99%, 100%, and 95%, respectively. It should also be noted that 4% (1/23) has no mention of the best prediction model (not marked in italic in Table 2).

Table 2. Studies' characteristics of prediction model development without internal or external validation with the best obtained model marked as italic in the respective model column.

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Viner et al [12], 1991	Prospective; —[c]	Logistic regression	Sex, age, BMI, and snoring	AHI[d]>10	410	190 (46)	77 (73-82)	28 (—)	95 (—)
Keenan et al [13], 1993	—	Logistic regression	NC[e], age, WA[f], daytime sleepiness, driving sleepy, oxygen desaturation, and heart rate frequency	AHI>15	96	51 (53)	—	20 (—)	5 (—)
Hoffstein et al [14], 1993	—	Linear regression	Subjective impression	AHI>10	594	275 (46)	—	60 (—)	63 (—)
Flemons et al [15] 1994	—; February 1990 to September 1990	Logistic and *linear* regression	NC, hypertension, snoring, and gasping or choking	AHI>10	175	82 (46)	—	—	—
Vaidya et al [16], 1996	—; July 1993 to December 1994	*Logistic* and linear regression	Age, BMI, sex, and total number of symptoms	RDI[g]>10	309	226 (73)	—	96 (—)	23 (—)
Deegan et al [11], 1996	Prospective; —	Logistic and linear regression	Sex, age, snoring, WA, driving sleepy, alcohol consumption, BMI, number of dips ≥4%, lowest oxygen saturation, and NC	AHI≥15	250	135 (54)	—	—	—
Pradhan et al [17], 1996	Prospective; August 1994 to February 1995	Logistic regression	BMI, lowest oxygen saturation, and bodily pain score	RDI>10	150	85 (57)	—	100 (—)	31 (—)
Friedman et al [18], 1999	Prospective; —	Linear regression	Modified Mallampati class, tonsil size grading, and BMI	RDI>20	172	—	—	—	—
Dixon et al [19], 2003	—	*Logistic* and linear regression	BMI, WA, glycosylated hemoglobin, fasting plasma insulin, sex, and age	AHI≥30	99	36 (36)	91 (—)	89 (—)	81 (—)
Morris et al [10], 2008	Prospective; —	Pearson correlation	BMI and snoring severity score	RDI≥15	211	175 (83)	—	97 (—)	40 (—)
Martinez-Rivera et al [20], 2008	—	Logistic regression	Sex, waist-to-hip ratio, BMI, NC, and age	AHI>10	192	124 (65)	—	—	—

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Herzog et al [21], 2009	Retrospective; —	Logistic and *linear* regression	Tonsil size grading, uvula size, dorsal movement during simulated snoring, collapse at tongue level, BMI, and ESS[h] score	AHI>5	622	—	—	Female: 98 (—)	Female: 22 (—)
Yeh et al [22], 2010	Retrospective; April 2006 to December 2007	Linear regression	BMI, NC, and ESS score	AHI≥15	101	83 (82)	—	98 (—)	—
Hukins et al [23], 2010	Retrospective; January 2005 to July 2007	Linear regression	Mallampati class IV	AHI>30	953	297 (31)	—	40 (36-45)	67 (64-69)
Musman et al [24], 2011	—; December 2006 to March 2007	Logistic and *linear* regression	NC, WA, age, BMI, and allergic rhinitis	AHI>5	323	229 (71)	—	—	—
Sareli et al [25], 2011	—; November 2005 to January 2007	Logistic regression	Age, BMI, sex, and sleep apnea symptom score	AHI≥5	342	264 (77)	80 (—)	—	—
Tseng et al [26], 2012	—	Decision tree	Sex, age, preovernight systolic blood pressure, and postovernight systolic blood pressure	AHI≥15	540	394 (73)	—	—	—
Sahin et al [27], 2014	Retrospective; —	Linear regression	BMI, WC[i], NC, oxygen saturation, and tonsil size grading	AHI>5 and symptoms	390	—	—	—	—
Ting et al [28], 2014	Prospective; —	Logistic regression and *decision trees*	Sex, age, and blood pressure	AHI≥15	540	394 (73)	99 (—)	98 (—)	93 (—)
Sutherland et al [29], 2016	—; 2011 to 2012	*Logistic* regression and classification and regression tree	Face width and cervicomental angle	AHI≥10	200	146 (73)	76 (68-83)	89 (—)	28 (—)
Lin et al [4], 2019	Retrospective; —	Linear regression	Sex, updated Friedman tongue position, tonsil size grading, and BMI	AHI≥5	325	283 (87)	80 (74-87)	84 (—)	58 (—)
Del Brutto et al [30], 2020	—	Logistic regression	Neck grasp	AHI≥5	167	114 (68)	62 (54-69)	83 (75-89)	40 (27-54)

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Haberfeld et al [8], 2020	—	Logistic regression and *support vector machine*	Height, weight, WC, hip size, BMI, age, neck size, modified Friedman score, snoring, sex, daytime sleepiness, and ESS score	—	620	357 (58)	Male: 61 (—)	Male: 86 (—)	Male: 70 (—)

[a]OSA: obstructive sleep apnea.

[b]AUC: area under receiver operating characteristic curve.

[c]Not available.

[d]AHI: apnea-hypopnea index.

[e]NC: neck circumference.

[f]WA: witnessed apnea.

[g]RDI: respiratory disturbance index.

[h]ESS: Epworth somnolence scale.

[i]WC: waist circumference.

As stated in the Methods section, given that all these models only performed development with in-sample validation metrics, they were all considered at high risk of bias in the Analysis domain (Table 3). Concerning the Outcome domain, most studies were marked as high risk, as most of them did not have a prespecified or standard outcome definition. In addition, although some were marked as high risk and one as unclear, most included studies were at low risk of bias regarding the Predictors domain, showing that most of the studies did not include predictors after performing PSG. Most studies (15/23, 65%) were identified as unclear for the Participants domain, as almost all studies did not state study design or exclusion criteria. Assessing the applicability aspect of PROBAST, all studies (23/23, 100%) were at low risk of bias for the Participants domain (all studies included patients with suspected OSA), but several were at high risk of applicability for the Outcome domain (OSA definition is not in concordance with current OSA guidelines).

Table 3. Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction model development without internal or external validation.

Study	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Viner et al [12], 1991	? [a]	− [b]	+ [c]	+	−	−	+	+	+
Keenan et al [13], 1993	?	+	−	+	−	?	+	+	+
Hoffstein et al [14], 1993	?	−	+	+	−	−	+	+	+
Flemons et al [15], 1994	+	−	+	+	−	−	+	+	+
Vaidya et al [16], 1996	?	−	+	+	−	−	+	+	+
Deegan et al [11], 1996	?	+	+	+	−	+	−	+	+
Pradhan et al [17], 1996	?	+	+	+	−	?	+	+	+
Friedman et al [18], 1999	?	+	+	+	−	?	+	+	+
Dixon et al [19], 2003	?	+	+	+	−	+	−	+	+
Morris et al [10], 2008	−	−	+	+	−	−	+	+	+
Martinez-Rivera et al [20], 2008	−	−	+	+	−	−	+	+	+
Herzog et al [21], 2009	−	+	+	+	−	+	+	+	+
Yeh et al [22], 2010	?	+	+	+	−	−	−	+	−
Hukins [23], 2010	−	−	+	+	−	−	+	+	+
Musman et al [24], 2011	−	−	+	+	−	−	+	+	+
Sareli et al [25], 2011	+	−	+	+	−	−	−	+	−
Tseng et al [26], 2012	?	−	−	+	−	−	−	+	−
Sahin et al [27], 2014	?	+	−	+	−	?	−	+	?
Ting et al [28], 2014	?	?	−	+	−	−	−	+	−
Sutherland et al [29], 2016	+	+	+	+	−	+	+	+	+
Lin et al [4], 2019	?	−	−	+	−	−	−	+	−
Del Brutto et al [30], 2020	?	+	−	+	−	−	−	+	−
Haberfeld et al [8], 2020	?	−	?	+	−	−	?	+	?

[a]Indicates an unclear risk of bias or concerns regarding applicability.

[b]Indicates a low risk of bias or concerns regarding applicability.

[c]Indicates a high risk of bias or concerns regarding applicability.

Development of Prediction Models With Internal Validation

For purposes of internal validation, we considered studies that performed cross-validation (11/26, 42%), used bootstrapping techniques (4/26, 15%), or used split-data (14/26, 54%) as previously mentioned in the Methods section. The smallest sample size was 83 participants and the highest was 6399, with both presenting validation results for cross-validation. Regarding OSA prevalence, a study had no mention, and another demonstrated an incorrect value [31], whereas others had the lowest value at 30% and the highest at 90%. Different machine learning approaches were used, with the most common being support vector machines (4/26, 15%), followed by logistic regression (3/26, 12%). Moreover, 38% (10/26) of studies described the study type and period, with retrospective design being the most common.

In addition, Table 4 shows different OSA definitions, with 8% (2/26) of studies not reporting cutoff values and the most common definition being AHI≥5 (8/26, 31%), followed by AHI≥15 (5/26, 19%). It should be noted that although the studies indicated that some types of internal validation were performed, some did not present results (10/26, 38%).

Regarding discrimination measures for internal validation, the best AUC, sensitivity, and specificity were 97%, 99%, and 97%, respectively. The model with the best AUC included predictive variables collected from PSG, such as the arousal index, and was also the model with the best specificity. The best sensitivity value was obtained for the neural network model with 19 predictive variables included. A total of 4 studies reported a clinical cutoff, which allows potential clinical threshold importance, with 50% reported in 2 studies and 32% in the other two.

In contrast to Table 3, Table 5 demonstrated that although internal validation was performed, only 8% (2/26) of studies had a low risk of bias in the Analysis domain, the reason being not presenting the relevant calibration or discrimination measures, such as AUC, and using only *P* values to select predictors. Furthermore, in the Participants domain applicability, 8% (2/26) of studies were marked as having a high risk of applicability, as they did not select only patients with suspected OSA.

Table 4. Studies' characteristics of prediction model development with internal validation. If the study applied different machine learning approaches, the clinical factors analyzed and the discrimination measures are only described for the best obtained model, marked as italic in the respective model column.

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Kapuniai et al [9], 1988	—[c]	Discriminant analysis	Breathing cessation, adenoidectomy, BMI, and gasping	AI[d]>5	D_1[e]=43; D_2=53	13 (30)	—	61 (—)	67 (—)
Kirby et al [32], 1999	Retrospective; —	Neural network	Age, sex, frequent awakening, experienced choking, WA[f], observed choking, daytime sleepiness, ESS[g], hypertension, alcohol consumption, smoking, height, weight, BMI, blood pressure, tonsillar enlargement, soft-palate enlargement, crowding of the oral pharynx, and sum of the clinical scores for the binary categorical values	AHI[h]≥10	D_1=255; D_2=150	281 (69)	94 (—)	99 (97-100)	80 (70-90)
Lam et al [33], 2005	Prospective; January 1999 to December 1999	Discriminant analysis	Mallampati score, thyromental angle, NC[i], BMI, age, and thyromental distance	AHI≥5	D_1=120; D_2=119[j]	201 (84)	71 (—)[k]	—	—
Julià-Serdà et al [34], 2006	—	Logistic regression	NC, sex, desaturation, ESS score, and distance between the gonion and the gnathion	AHI≥10	D_1=150; D_2=57	115 (56)	97 (95-99)[k]	94 (—)	83 (—)
Polat et al [35], 2008	Prospective; —	*Decision tree*, neural network, 21 adaptive neuro-fuzzy inference system, and artificial immune recognition system	Arousals index, AHI, minimum oxygen saturation value in stage REM[l], and percentage of sleep time in stage of oxygen saturations intervals bigger than 89%	AHI>5	D_1=41; D_2=42[j]	58 (70)	97 (—)	92 (—)	97 (—)
Chen et al [31], 2008	—; January 2004 to December 2005	Support vector machine	Oxygen desaturation index	AHI≥5	566[j]	491 (87)	—	43 (—)	94 (—)

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Lee et al [36], 2009	Prospective; —	*Logistic regression* and classification and regression tree	Face width, eye width, mandibular length, WA, and modified Mallampati class	AHI≥10	180[j]	114 (63)	87 (—)[k]	85 (—)[k]	70 (—)[k]
Rofail et al [37], 2010	—; July 2006 to November 2007	Logistic regression	Index 1 (snoring, breathing cessation, snorting, gasping), and nasal flow RDI[m]	AHI≥5	D_1=96; D_2=97	139 (72)	89 (81-97)	85 (—)	92 (—)
Chen et al [38], 2011	Retrospective; —	Logistic regression	Desaturation 3%	RDI≥30	D_1=355; D_2=100[j]	307 (86)	95 (—)[k]	90 (—)	90 (—)
Bucca et al [39], 2011	Prospective; January 2004 to December 2005	Linear regression	Age, NC, BMI, FEF50/FIF50[n], COH_B%[o], smoking, F_{eNO}[p], and interaction smoking and F_{eNO}	AHI≥30	201[q]	120 (60)	—	—	—
Bouloukaki et al [40], 2011	Prospective; October 2000 to December 2006	Linear regression	NC, sleepiness severity, BMI, and sex	AHI≥15	D_1=538; D_2=2152	2130 (79)	78 (61-80)[k]	70 (—)[k]	73 (—)[k]
Sun et al [41], 2011	—; February 2009 to June 2009	Logistic regression and *genetic algorithm*	Demographic data, ESS, systemic diseases, snoring, and comorbidities	AHI≥15	D_1=67; D_2=43	53 (48)	—	82 (—)	95 (—)
Laporta et al [42], 2012	Prospective; October 2010 to September 2011	Neural network	Age, weight, sex, height, NC, hypertension, daytime sleepiness, difficulty falling asleep, snoring, breathing cessation, restless sleep, and gasping	AHI≥5	91[q]	68 (75)	93 (85-97)[k]	99 (92-100)[k]	87 (66-97)[k]
Hang et al [43], 2013	Retrospective; January 2005 to December 2006	Support vector machine	Oxygen desaturation index, ESS, or BMI	AHI≥15	D_1=188; D_2=188; D_3=189	—	—	88 (85-90)[k]	90 (87-94)[k]
Hang et al [44], 2015	—; January 2004 to December 2005	Support vector machine	Oxygen desaturation index	AHI>30	1156[j]	285 (46)	D_1: 96 (—)[k]; D_2: 95 (—)[k]	D_1: 87 (—); D_2: 91 (—)[k]	D_1: 93 (—); D_2: 90 (—)[k]
Ustun et al [7], 2016	—; January 2009 to June 2013	Logistic regression, *supersparse linear integer models*, decision tree, and support vector machines	Age, sex, BMI, diabetes, hypertension, and smoking	AHI>5	1922[j]	1478 (77)	79 (—)	64 (—)	23 (—)

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Bozkurt et al [45], 2017	Retrospective; January 2014 to August 2015	Logistic regression, *Bayesian network*, decision tree, random forest, and neural network	Sex, age, BMI, NC, and smoking	AHI≥5	338[j]	304 (90)	73 (—)	86 (—)	85 (—)
Ferreira-Santos [46], 2017	Retrospective; January 2015 to May 2015	Bayesian network	Sex, NC, CFA[r], WA, nocturia, alcohol consumption, ESS, concentration decrease, atrial fibrillation, stroke, myocardial infarction, driver, and daytime sleepiness	AHI≥5	194[j]	128 (66)	76 (73-78)	81 (79-83)	48 (44-51)
Liu et al [47], 2017	—; October 2005 to April 2014 and October 2013 to September 2014	Support vector machine	WC[s], NC, BMI, and age	AHI≥15	6399[j]	3866 (60)	Female: 90 (87-94)	Female: 83 (75-91)	Female: 86 (82-90)
Manoochehri et al [48], 2018	—; 2012 to 2016	Logistic regression and *decision tree*	WC, snoring, sex, sleep apnea, ESS score, and NC	—	D_1=239; D_2=99	208 (62)	—	67 (—)	81 (—)
Manoochehri et al [49], 2018	—; 2012 to 2015	Logistic regression and *support vector machine*	Age, sex, BMI, NC, WC, tea consumption, smoking, hypertension, chronic headache, heart disease, respiratory disease, neurological disease, and diabetes	—	D_1=176; D_2=74	154 (62)	—	71 (—)[k]	85 (—)[k]
Xu et al [50], 2019	—; 2007 to 2016	Nomogram	Age, sex, glucose, apolipoprotein B, insulin, BMI, NC, and WC	AHI>5	4162[q]	3387 (81)	84 (83-86)	77 (76-79)[k]	76 (72-80)[k]
Ferreira-Santos et al [51], 2019	Retrospective; January 2015 to May 2015	Bayesian network	Sex, WA, age, nocturia, CFA, and NC	AHI≥5	194[j]	128 (66)	64 (61-66)	90 (88-92)	24 (20-27)
Keshavarz et al [52], 2020	Retrospective; February 2013 to December 2017	Logistic regression, Bayesian network, *neural network*, k-nearest neighbors, support vector machine, and random forest	Snoring, nocturia, awakening owing to the sound of snoring, snoring, back pain, restless sleep, BMI, and WA	AHI>15	231[j]	152 (66)	75 (—)	86 (—)	53 (—)
Chen et al [53], 2021	Retrospective; September 2015 to January 2020	Nomogram	Age, sex, snoring, type 2 diabetes mellitus, NC, and BMI	AHI≥5	D_1=338; D_2=144[q]	342 (71)	83 (76-90)	69 (63-75)[k]	87 (79-93)[k]

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Hsu et al [54], 2021	—; December 2011 to August 2018	*Logistic regression*, support vector machine, and neural network	Sex, age, and BMI	AHI≥15	D_1=2446; D_2=1049	2539 (73)	82 (—)	73 (—)[k]	77 (—)[k]

[a]OSA: obstructive sleep apnea.
[b]AUC: area under receiver operating characteristic curve.
[c]Not available.
[d]AI: apnea index.
[e]D_1, D_2, and D_3: data set.
[f]WA: witnessed apnea.
[g]ESS: Epworth somnolence scale.
[h]AHI: apnea-hypopnea index.
[i]NC: neck circumference.
[j]cross-validation.
[k]Internal derivation results.
[l]REM: rapid eye movement.
[m]RDI: respiratory disturbance index.
[n]FEF50/FIF50: forced midexpiratory/midinspiratory airflow ratio.
[o]COHB%: carboxyhemoglobin percent saturation.
[p]Fe_{NO}: exhaled nitric oxide.
[q]Bootstrapping.
[r]CFA: craniofacial and upper airway.
[s]WC: waist circumference.

Table 5. Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction model development with internal validation.

Study	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Kapuniai et al [9], 1988	? [a]	+ [b]	+	+	+	+	+	+	+
Kirby et al [32], 1999	?	− [c]	+	+	−	?	+	+	+
Lam et al [33], 2005	−	−	+	?	−	?	+	+	+
Julià-Serdà et al [34], 2006	?	+	+	+	−	?	+	+	+
Polat et al [35], 2008	?	+	+	?	−	+	+	+	+
Chen et al [31], 2008	?	−	+	+	−	?	−	+	?
Lee et al [36], 2009	+	+	+	+	−	?	+	+	+
Rofail et al [37], 2010	−	+	?	+	−	+	−	+	+
Chen et al [38], 2010	+	−	+	?	−	?	+	+	+
Bucca et al [39], 2010	−	+	?	+	−	+	−	+	+
Bouloukaki et al [40], 2011	−	−	−	+	−	−	−	+	−
Sun et al [41], 2011	?	−	+	+	−	?	−	+	?
Laporta et al [42], 2012	?	+	+	+	+	−	−	+	+
Hang et al [43], 2015	?	?	−	+	−	?	−	+	?
Hang et al [44], 2015	?	+	+	−	−	?	+	+	+
Ustun et al [7], 2016	−	?	+	+	−	−	+	+	+
Bozkurt et al [45], 2017	−	−	+	+	−	−	−	+	−
Ferreira-Santos et al [46], 2017	−	−	−	+	−	−	−	+	−
Liu et al [47], 2017	?	−	+	−	−	−	−	+	−
Manoochehri et al [48], 2018	−	−	?	+	−	−	?	+	?
Manoochehri et al [49], 2018	?	−	?	+	−	−	?	+	?
Xu et al [50], 2019	−	+	+	+	−	+	+	+	+
Ferreira-Santos et al [51], 2019	−	−	−	+	−	−	−	+	−
Keshavarz et al [52], 2020	+	?	+	?	−	−	+	+	+
Chen et al [53], 2021	+	−	+	+	−	−	−	+	−
Hsu et al [54], 2021	?	−	+	?	−	−	−	+	−

[a]Indicates an unclear risk of bias or concerns regarding applicability.

[b]Indicates a high risk of bias or concerns regarding applicability.

[c]Indicates a low risk of bias or concerns regarding applicability.

Development of Prediction Models With External Validation

A total of 12 studies performed external validation, as described in Table 6, with 9 (75%) of them choosing logistic regression for the machine learning approach. The other 25% (3/12) elected linear regression, neural networks, or both. Regarding the study design, 3 (25%) studies elected a prospective design for testing and validation and 8% (1/12) of studies for only validation. Similar to the studies that only performed internal validation, the lowest OSA prevalence was 30%, and the highest was 93%, with a sample size varying between 169 and 3432 participants with suspected OSA. The best discriminatory model was logistic regression; it included age, waist circumference, ESS, and minimum oxygen saturation, with an AUC of 0.98 (0.96-0.99), for an OSA definition of AHI≥5. The higher reached sensitivity (100%) was also for a logistic regression but for a cutoff of AHI≥15, including specific respiratory conductance and daytime arterial oxygen saturation. The study also presented a clinical cutoff of 50%. Concerning specificity, the value of 94% was the highest for an AI>10, with self-reporting apneas, NC index, age, and tendency to fall asleep unintentionally as predictive variables.

As shown in Table 7, which aggregates information from the test and validation data sets, most studies were marked as unclear risk of bias in the Participants domain, as the studies referred to the study design for the test population but not for the validation data set. In addition, only 17% (2/12) of studies had a high risk of bias for the Predictors domain, given that the predictors could take time to be assessed or collected. Regarding the Analysis domain, half (6/12, 50%) of the studies were marked as having a low risk of bias, with 33% (4/12) of studies not presenting adequate performance metrics. The applicability in the Predictors domain is unclear in 8% (1/12) of studies, as we cannot assess whether the predictors are available in primary health care.

Table 6. Studies' characteristics of prediction model development with external validation. If the study applied different machine learning approaches, the clinical factors analyzed and the discrimination measures are only described for the best obtained model, marked as italic in the respective model column.

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Crocker et al [55], 1990	—[c]; October 1986 to May 1988	Logistic regression	Age, breathing cessation, BMI, and hypertension	AHI[d]>15	T[e]=100; V[f]=105	62 (30)	—	92 (—)	51 (—)
Pillar et al [56], 1992	—	Logistic regression	WA[g], NC[h] index, age, daytime and sleepiness	AI[i]>10 and symptoms	*t*=86; V_1=50; V_2=105	—	—	V_1=88 (—); V_2=32 (—)	V_1=25 (—); V_2=94 (—)
Maislin et al [57], 1995	—	Logistic regression	BMI, age, sex, index 1 (snoring, breathing cessation, snorting, and gasping), and BMI index 1	RDI[j]≥10	*t*=658; V=193	760 (89)	79 (—)[k]	—	—
Kushida et al [58], 1997	Prospective; 6 months (V)	Linear regression	Palatal height, maxillary intermolar distance, mandibular intermolar distance, overjet, BMI, and NC	RDI≥5	*t*=30; V=300[l,m]	254 (85)	100 (—)[k]	98 (95-99)[k]	100 (92-100)[k]
El-Solh et al [59], 1999	Retrospective (T) and prospective (V); November 1995 to December 1996	*Neural network* and linear regression	Breathing cessation, restless sleep, decreased libido, disturbs bed partner, daytime sleepiness, restless legs, BMI, NC, age, gasping, snoring, and blood pressure	AHI>10	*t*=189[l]; V=80	182 (68)	96 (93-96)	95 (90-98)[k]	65 (50-78)[k]
Zerah-Lancner et al [60], 2000	Retrospective (T) and prospective (V); —	Logistic regression	Specific respiratory conductance and daytime arterial oxygen saturation	AHI≥15	*t*=168; V=101	147 (55)	—	100 (—)	84 (—)
Rodsutti et al [61], 2004	Prospective; February 2001 to April 2003	Logistic regression	Age, sex, BMI, and breathing cessation	AHI≥5	*t*=837; V=243	569 (53)	79 (—)	—	—
Khoo et al [62], 2011	—; December 2005 to December 2007 and March 2008 to June 2008	Logistic regression	Sex, age, NC, and frequent awakening with unrefreshing sleep	AHI≥20	*t*=117; V=52	77 (66)	69 (—)[k]	78 (—)	45 (—)
Zou et al [63], 2013	Retrospective; January 2007 to July 2011	Logistic regression	Age, WC[n], ESS[o], and minimum oxygen saturation	AHI≥5	*t*=2052; V=784	2451 (87)	98 (96-99)	94 (92-96)	86 (79-91)
Karamanli et al [64], 2016	Retrospective; —	Neural network	Sex, age, BMI, and snoring	AHI≥10	*t*=201; V=15	140 (70)	—	—	—

Study	Study design; study period	Machine learning approach	Clinical factors analyzed	OSA[a] definition	Sample size, n	OSA prevalence, n (%)	AUC[b], % (95% CI)	Sensitivity, % (95% CI)	Specificity, % (95% CI)
Tawaranurak et al [65], 2020	Prospective; June 2018 to June 2020	Logistic regression	Sex, choking or apnea, blood pressure, NC, WC, and BMI	AHI≥15	*t*=892; V=374	826 (93)	75 (—)[k]	93 (89-96)	26 (18-35)
Park et al [66], 2021	—; January 2011 to December 2018	Logistic regression	Age, sex, BMI, hypertension, Berlin questionnaire score, and tonsil grade	AHI≥5	*t*=2516; V=916	—	84 (—)	78 (—)	76 (—)

[a]OSA: obstructive sleep apnea.

[b]AUC: area under receiver operating characteristic curve.

[c]Not available.

[d]AHI: apnea-hypopnea index.

[e]T: test data set.

[f]V: validation data set.

[g]WA: witnessed apnea.

[h]NC: neck circumference.

[i]AI: apnea index.

[j]RDI: respiratory disturbance index.

[k]Internal derivation results.

[l]Cross-validation.

[m]Bootstrapping.

[n]WC: waist circumference.

[o]ESS: Epworth Somnolence Scale.

Table 7. Prediction Model Risk of Bias Assessment Tool (PROBAST) for prediction model development with external validation.

Study	Risk of bias				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of bias	Applicability
Crocker et al [55], 1990	? [a]	− [b]	+ [c]	+	−	−	+	+	+
Pillar et al [56], 1994	?	−	+	+	−	−	+	+	+
Maislin et al [57], 1995	?	−	+	−	+	−	+	+	+
Kushida et al [58], 1997	?	+	+	−	+	?	+	+	+
El-Solh et al [59], 1999	−	−	+	?	−	−	+	+	+
Zerah-Lancner et al [60] 2000	?	+	−	+	−	+	−	+	+
Rodsutti et al [61], 2003	?	−	−	−	−	−	−	?	−
Khoo et al [62], 2011	?	−	+	?	+	−	+	+	+
Zou et al [63], 2013	−	−	−	−	−	−	−	−	−
Karamanli et al [64], 2016	?	−	+	+	−	−	+	+	+
Tawaranurak et al [65], 2021	−	−	−	−	−	−	−	−	−
Park et al [66], 2021	−	−	−	−	−	−	−	−	−

[a]Indicates an unclear risk of bias or concerns regarding applicability.

[b]Indicates a low risk of bias or concerns regarding applicability.

[c]Indicates a high risk of bias or concerns regarding applicability.

Prediction Models With External Validation

A total of 2 studies [67,68], one in 2000 and another in 2006, performed the external validation of 5 prediction models. The first was a prospective study that evaluated 4 clinical prediction models [12,15,55,57] for predicting the presence of OSA (AHI≥10). They included 370 patients with suspected OSA who underwent PSG between July 1996 and October 1997. The achieved prevalence of OSA was 67%, and the results are shown in Figure 1 and Table 4 of the original article [67]. The highest AUC, sensitivity, and specificity reached were 74%, 96%, and 54%, respectively. The second study used 80 patients with suspected OSA to evaluate the model described in the study by Kushida et al [58]. The objective was to evaluate the clinical applicability and define a clinical cutoff to differentiate OSA severities. Although the authors stated that the clinical applicability exists, they could not define a threshold for clinical use, and they did not present any discrimination measures.

The study of Flemons et al [15], in addition to producing a new prediction model, also applied the 2 equations from studies by Crocker et al [55] and Viner et al [12] to the obtained data set. Although no actual values were presented, the authors stated that the AUCs were very similar.

Furthermore, the study by Flemons et al [15] was externally validated by Khoo et al [62], with 52 patients with suspected OSA, reaching an AUC of 69%. If a clinical threshold of 60% is defined, the model in this independent sample reached 78% sensitivity and 45% specificity.

Discussion

Principal Findings

The AASM guidelines [1] explicitly state that *"clinical prediction algorithms may be used in sleep clinic patients with suspected OSA but are not necessary to substitute the need for PSG,"* whereas *"in non-sleep clinic settings, these tools may be more helpful to identify patients who are at increased risk for OSA."* The evaluation of these tools in a nonsleep clinic setting was not tackled by AASM experts, as it was beyond the guideline scope. Therefore, our work aimed to answer this question by complementing step 1 in the clinical algorithm developed for clinical suspicion of OSA using clinical prediction algorithms in a nonsleep setting. With this, we hope to estimate the probability that OSA is present in a population with suspected OSA that is not yet diagnosed by aggregating information from multivariable prediction models, stating the ones that are best at rule out and rule in.

As such, the studies that only developed a model are the ones that need to gather evidence on whether the model would be helpful to put into clinical practice (high overfitting). To do so,

it is needed to validate the model in a new population data set. One way to do this is by splitting the data set or performing a validity assessment using different techniques, such as cross-validation or bootstrapping, or even better, by applying the algorithm to an independent sample.

Of the 63 included studies, only 14 (22%) performed both development and external validation or only external validation of the algorithm. Most selected studies only developed 36% (23/63) or developed and internally validated 41% (26/63) of prediction models.

The study by Zerah-Lancner et al [60] emerged as the best at rule-out OSA, described a sensitivity value of 100% for an OSA definition of AHI≥15. The predictive variables included were respiratory conductance and oxygen saturation, chosen from an external population of 101 participants. The best at rule-in OSA was the study by Pillar et al [56]; for a validation population of 155 participants, it demonstrated a specificity of 94% for an AI≥10 symptoms, with witnessed apneas, NC, age, and falling asleep easily as predictive variables. Both studies used logistic regression as the machine learning approach. The study by Kushida et al [58] reached maximum specificity, but the authors did not describe whether the obtained results were for testing or external validation, in a 300-participant validation data set. These 2 best models [56,60] were developed and validated in 2000 and 1992, respectively, and presented a high risk of bias and applicability, with none of the studies providing the discriminatory power of the model or metric CIs.

The most recent study by Park et al [66], performed in 2021 with a validation data set of 916 participants (largest sample), only reached values of 78% and 76% for sensitivity and specificity, respectively, when compared with the 2 previous best models. This was also a logistic regression, electing BMI, age, sex, Berlin questionnaire score, and tonsil grade as the clinical factors for an OSA definition of AHI≥5. Although this study continued to lack the reporting of study design or prevalence of OSA, it presented a low risk of bias and applicability. But it only included Asian patients, so it cannot be race generalized, as the authors mention.

Strengths and Limitations

It is important to consider some of the limitations and strengths of our methods and those of the included clinical studies. Although we cannot be sure that we retrieved all published literature, we are confident that our methodology is adequate. Risk was minimized by performing the search in 3 search engines (1 related to health sciences and 2 others with broader spectrums) and in 2 periods.

The PROBAST demonstrated that we face a high risk of bias and applicability, even when only assessing external validation results. Almost all the studies do not report the study design, which can raise problems in generating absolute probabilities or even in terms of inappropriately including or excluding participants. In addition, the definition and measurement of predictors and their association with the outcome were high in the 2 studies, as some of the predictors were not available when the model was intended to be used. Although all outcome definitions were based on PSG, some did not report how the measure was calculated or selected different cutoff values than the ones described in the guidelines. While all studies used appropriate statistical analysis, some lacked a reasonable number of participants with the outcome, in the test or validation data sets. Information regarding exclusion criteria or handling of missing data was not described, and most studies selected predictors based on univariable analysis. Besides all participants who underwent the gold standard exam, some did not have suspected OSA as the only inclusion criterion.

Different approaches have been followed since 1988 with the aim of predicting whether OSA is present in an individual, contributing to unlocking the bottleneck of in-hospital screening or diagnosis. However, assessing the bias or applicability of these approaches is not an easy task, with only 3 studies presenting an overall low risk of bias and applicability [63,65,66]. Furthermore, common missing points need to be pointed out are (1) most studies did not report the study design or period; (2) OSA definition differed within time, guidelines, and studies; (3) OSA prevalence varied from 30% to 93%, with some studies not describing the proportion; (4) needed measures to assess diagnostic value such as sensitivity, specificity, and AUC are not reported, and when reported, did not present CIs; and (5) some studies only create the predictive model and others add the validation task, but external validation is still lacking in all the studies.

Regarding the chosen machine learning approaches, the most common was logistic regression (35/63, 56%), followed by linear regression (16/63, 25%), support vector machine (9/63, 14%), neural networks (8/63, 13%), decision trees (8/63, 13%), Bayesian networks (4/63, 6%), random forest (2/63, 3%), discriminant analysis (2/63, 3%), classification and regression tree (2/63, 3%), nomogram (2/63, 3%), Pearson correlation (1/63, 2%), adaptive neuro-fuzzy inference system (1/63, 2%), artificial immune recognition system (1/63, 2%), genetic algorithm (1/63, 2%), supersparse linear integer models (1/63, 2%), and the k-nearest neighbors algorithm (1/63, 2%).

Conclusions

In summary, this review provides an extensive, comprehensive, and up-to-date synthesis of diagnostic models in OSA. It is possible to predict OSA by only taking into consideration simple and available predictors such as BMI, age, sex, or NC as well as by reaching high levels of sensitivity or specificity, depending on whether we want to elect a rule-out or rule-in approach.

Acknowledgments

DFS acknowledges Fundação para a Ciência e Tecnologia under PhD grants (PD/BD/13553/2018 and COVID/BD/152608/2022) for funding. This paper was supported by National Funds through Fundação para a Ciência e a Tecnologia, I.P., within the Center for Health Technology and Services Research, Research and Development Unit (reference UIDP/4255/2020).

Conflicts of Interest

None declared.

References

1. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American academy of sleep medicine clinical practice guideline. J Clin Sleep Med 2017 Mar 15;13(3):479-504 [FREE Full text] [doi: 10.5664/jcsm.6506] [Medline: 28162150]
2. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MS, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respiratory Med 2019 Aug;7(8):687-698. [doi: 10.1016/s2213-2600(19)30198-5]
3. Kendzerska T, Mollayeva T, Gershon AS, Leung RS, Hawker G, Tomlinson G. Untreated obstructive sleep apnea and the risk for serious long-term adverse outcomes: a systematic review. Sleep Med Rev 2014 Feb;18(1):49-59. [doi: 10.1016/j.smrv.2013.01.003] [Medline: 23642349]
4. Lin H, Lai C, Lin P, Friedman M, Salapatas AM, Chang H, et al. Clinical prediction model for obstructive sleep apnea among adult patients with habitual snoring. Otolaryngol Head Neck Surg 2019 Jul 02;161(1):178-185. [doi: 10.1177/0194599819839999] [Medline: 30935275]
5. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
6. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019 Jan 01;170(1):W1. [doi: 10.7326/m18-1377]
7. Ustun B, Westover MB, Rudin C, Bianchi MT. Clinical prediction models for sleep apnea: the importance of medical history over symptoms. J Clin Sleep Med 2016 Feb;12(2):161-168 [FREE Full text] [doi: 10.5664/jcsm.5476] [Medline: 26350602]
8. Haberfeld C, Sheta A, Hossain MS, Turabieh H, Surani S. SAS mobile application for diagnosis of obstructive sleep apnea utilizing machine learning models. In: Proceedings of the 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2020 Presented at: 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON); Oct 28-31, 2020; New York, NY, USA. [doi: 10.1109/uemcon51285.2020.9298041]
9. Kapuniai LE, Andrew DJ, Crowell DH, Pearce JW. Identifying sleep apnea from self-reports. Sleep 1988 Oct;11(5):430-436. [doi: 10.1093/sleep/11.5.430] [Medline: 3227223]
10. Morris LG, Kleinberger A, Lee KC, Liberatore LA, Burschtin O. Rapid risk stratification for obstructive sleep apnea, based on snoring severity and body mass index. Otolaryngol Head Neck Surg 2008 Nov 24;139(5):615-618. [doi: 10.1016/j.otohns.2008.08.026] [Medline: 18984252]
11. Deegan PC, McNicholas WT. Predictive value of clinical features for the obstructive sleep apnoea syndrome. Eur Respir J 1996 Jan;9(1):117-124 [FREE Full text] [doi: 10.1183/09031936.96.09010117] [Medline: 8834344]
12. Viner S, Szalai JP, Hoffstein V. Are history and physical examination a good screening test for sleep apnea? Ann Intern Med 1991 Sep 01;115(5):356-359. [doi: 10.7326/0003-4819-115-5-356] [Medline: 1863025]
13. Keenan SP, Anderson B, Wiggs B, Ryan CF, Fleetham JA. The predictive accuracy of home oximetry in patients with suspected obstructive sleep apnea. Sleep 1993 Dec;16(8 Suppl):S133-S134. [doi: 10.1093/sleep/16.suppl_8.s133] [Medline: 8178005]
14. Hoffstein V, Szalai JP. Predictive value of clinical features in diagnosing obstructive sleep apnea. Sleep 1993 Feb;16(2):118-122. [Medline: 8446830]
15. Flemons WW, Whitelaw WA, Brant R, Remmers JE. Likelihood ratios for a sleep apnea clinical prediction rule. Am J Respir Crit Care Med 1994 Nov;150(5 Pt 1):1279-1285. [doi: 10.1164/ajrccm.150.5.7952553] [Medline: 7952553]
16. Vaidya AM, Petruzzelli GJ, Walker RP, McGee D, Gopalsami C. Identifying obstructive sleep apnea in patients presenting for laser-assisted uvulopalatoplasty. Laryngoscope 1996 Apr;106(4):431-437. [doi: 10.1097/00005537-199604000-00008] [Medline: 8614217]
17. Pradhan PS, Gliklich RE, Winkelman J. Screening for obstructive sleep apnea in patients presenting for snoring surgery. Laryngoscope 1996 Nov;106(11):1393-1397. [doi: 10.1097/00005537-199611000-00016] [Medline: 8914907]
18. Friedman M, Tanyeri H, La Rosa M, Landsberg R, Vaidyanathan K, Pieri S, et al. Clinical predictors of obstructive sleep apnea. Laryngoscope 1999 Dec;109(12):1901-1907. [doi: 10.1097/00005537-199912000-00002] [Medline: 10591345]
19. Dixon JB, Schachter LM, O'Brien PE. Predicting sleep apnea and excessive day sleepiness in the severely obese: indicators for polysomnography. Chest 2003 Apr;123(4):1134-1141. [doi: 10.1378/chest.123.4.1134] [Medline: 12684304]
20. Martinez-Rivera C, Abad J, Fiz JA, Rios J, Morera J. Usefulness of truncal obesity indices as predictive factors for obstructive sleep apnea syndrome. Obesity (Silver Spring) 2008 Jan;16(1):113-118 [FREE Full text] [doi: 10.1038/oby.2007.20] [Medline: 18223622]

21. Herzog M, Kühnel T, Bremert T, Herzog B, Hosemann W, Kaftan H. The upper airway in sleep-disordered breathing: a clinical prediction model. Laryngoscope 2009 Apr;119(4):765-773. [doi: 10.1002/lary.20153] [Medline: 19266582]
22. Yeh P, Lee Y, Lee W, Chen S, Ho S, Peng W, et al. Clinical predictors of obstructive sleep apnea in Asian bariatric patients. Obes Surg 2010 Jan 12;20(1):30-35. [doi: 10.1007/s11695-009-9854-2] [Medline: 19434465]
23. Hukins C. Mallampati class is not useful in the clinical assessment of sleep clinic patients. J Clin Sleep Med 2010 Dec 15;06(06):545-549. [doi: 10.5664/jcsm.27987]
24. Musman S, Passos VM, Silva IB, Barreto SM. Evaluation of a prediction model for sleep apnea in patients submitted to polysomnography. J Bras Pneumol 2011 Feb;37(1):75-84 [FREE Full text] [doi: 10.1590/s1806-37132011000100012] [Medline: 21390435]
25. Sareli AE, Cantor CR, Williams NN, Korus G, Raper SE, Pien G, et al. Obstructive sleep apnea in patients undergoing bariatric surgery--a tertiary center experience. Obes Surg 2011 Mar 11;21(3):316-327. [doi: 10.1007/s11695-009-9928-1] [Medline: 19669842]
26. Tseng MH, Hsu HC, Chang CC, Ting H, Wu HC, Tang PH. Development of an intelligent app for obstructive sleep apnea prediction on android smartphone using data mining approach. In: Proceedings of the 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing. 2012 Presented at: 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing; Sep 04-07, 2012; Fukuoka, Japan. [doi: 10.1109/uic-atc.2012.89]
27. Sahin M, Bilgen C, Tasbakan MS, Midilli R, Basoglu OK. A clinical prediction formula for apnea-hypopnea index. Int J Otolaryngol 2014;2014:438376-438375 [FREE Full text] [doi: 10.1155/2014/438376] [Medline: 25349613]
28. Ting H, Mai Y, Hsu H, Wu H, Tseng M. Decision tree based diagnostic system for moderate to severe obstructive sleep apnea. J Med Syst 2014 Sep 11;38(9):94. [doi: 10.1007/s10916-014-0094-1] [Medline: 25012477]
29. Sutherland K, Lee RW, Petocz P, Chan TO, Ng S, Hui DS, et al. Craniofacial phenotyping for prediction of obstructive sleep apnoea in a Chinese population. Respirology 2016 Aug 15;21(6):1118-1125. [doi: 10.1111/resp.12792] [Medline: 27083503]
30. Del Brutto OH, Mera RM, Recalde BY, Castillo PR. Assessment of neck grasp as a screening tool for identifying obstructive sleep apnea in community-dwelling older adults. J Prim Care Community Health 2020 Dec 24;11:2150132720984421 [FREE Full text] [doi: 10.1177/2150132720984421] [Medline: 33356814]
31. Chen YF, Chen JH, Lin YJ, Tai CJ. Diagnosis and prediction of patients with severe obstructive apneas using support vector machine. In: Proceedings of the 2008 International Conference on Machine Learning and Cybernetics. 2008 Presented at: 2008 International Conference on Machine Learning and Cybernetics; Jul 12-15, 2008; Kunming. [doi: 10.1109/icmlc.2008.4620964]
32. Kirby SD, Eng P, Danter W, George CF, Francovic T, Ruby RR, et al. Neural network prediction of obstructive sleep apnea from clinical criteria. Chest 1999 Aug;116(2):409-415. [doi: 10.1378/chest.116.2.409] [Medline: 10453870]
33. Lam B, Ip MS, Tench E, Ryan CF. Craniofacial profile in Asian and white subjects with obstructive sleep apnoea. Thorax 2005 Jun 01;60(6):504-510 [FREE Full text] [doi: 10.1136/thx.2004.031591] [Medline: 15923252]
34. Julià-Serdà G, Pérez-Peñate G, Saavedra-Santana P, Ponce-González M, Valencia-Gallardo JM, Rodríguez-Delgado R, et al. Usefulness of cephalometry in sparing polysomnography of patients with suspected obstructive sleep apnea. Sleep Breath 2006 Dec 20;10(4):181-187. [doi: 10.1007/s11325-006-0073-y] [Medline: 17053929]
35. Polat K, Yosunkaya P, Güneş S. Comparison of different classifier algorithms on the automated detection of obstructive sleep apnea syndrome. J Med Syst 2008 Jun 16;32(3):243-250. [doi: 10.1007/s10916-008-9129-9] [Medline: 18444362]
36. Lee RW, Petocz P, Prvan T, Chan AS, Grunstein RR, Cistulli PA. Prediction of obstructive sleep apnea with craniofacial photographic analysis. Sleep 2009 Jan;32(1):46-52 [FREE Full text] [Medline: 19189778]
37. Rofail LM, Wong KK, Unger G, Marks GB, Grunstein RR. The utility of single-channel nasal airflow pressure transducer in the diagnosis of OSA at home. Sleep 2010 Aug;33(8):1097-1105 [FREE Full text] [doi: 10.1093/sleep/33.8.1097] [Medline: 20815193]
38. Chen N, Chen M, Li H, Chen C, Wang P. A two-tier screening model using quality-of-life measures and pulse oximetry to screen adults with sleep-disordered breathing. Sleep Breath 2011 Sep 7;15(3):447-454. [doi: 10.1007/s11325-010-0356-1] [Medline: 20449670]
39. Bucca C, Brussino L, Maule MM, Baldi I, Guida G, Culla B, et al. Clinical and functional prediction of moderate to severe obstructive sleep apnoea. Clin Respir J 2011 Oct;5(4):219-226. [doi: 10.1111/j.1752-699X.2010.00223.x] [Medline: 21801324]
40. Bouloukaki I, Kapsimalis F, Mermigkis C, Kryger M, Tzanakis N, Panagou P, et al. Prediction of obstructive sleep apnea syndrome in a large Greek population. Sleep Breath 2011 Dec 25;15(4):657-664. [doi: 10.1007/s11325-010-0416-6] [Medline: 20872180]
41. Sun LM, Chiu H, Chuang CY, Liu L. A prediction model based on an artificial intelligence system for moderate to severe obstructive sleep apnea. Sleep Breath 2011 Sep 4;15(3):317-323. [doi: 10.1007/s11325-010-0384-x] [Medline: 20602177]
42. Laporta R, Anandam A, El-Solh AA. Screening for obstructive sleep apnea in veterans with ischemic heart disease using a computer-based clinical decision-support system. Clin Res Cardiol 2012 Sep 3;101(9):737-744. [doi: 10.1007/s00392-012-0453-1] [Medline: 22476823]

43. Hang L, Lin H, Cheng C, Chiang JY, Wang H, Chen Y. Diagnosis of severe obstructive sleep apnea with model designed using genetic algorithm and ensemble support vector machine. Appl Math Inf Sci 2015 Feb 1;9(1L):149-157. [doi: 10.12785/amis/091l19]
44. Hang L, Wang H, Chen J, Hsu J, Lin H, Chung W, et al. Validation of overnight oximetry to diagnose patients with moderate to severe obstructive sleep apnea. BMC Pulm Med 2015 Mar 20;15(1):24 [FREE Full text] [doi: 10.1186/s12890-015-0017-z] [Medline: 25880649]
45. Bozkurt S, Bostanci A, Turhan M. Can statistical machine learning algorithms help for classification of obstructive sleep apnea severity to optimal utilization of polysomno graphy resources? Methods Inf Med 2018 Jan 24;56(04):308-318. [doi: 10.3414/me16-01-0084]
46. Ferreira-Santos D, Rodrigues PP. Improving diagnosis in obstructive sleep apnea with clinical data: a Bayesian network approach. In: Proceedings of the IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS). 2017 Presented at: IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS); Jun 22-24, 2017; Thessaloniki, Greece. [doi: 10.1109/cbms.2017.19]
47. Liu W, Wu H, Juang J, Wisniewski A, Lee H, Wu D, et al. Prediction of the severity of obstructive sleep apnea by anthropometric features via support vector machine. PLoS One 2017 May 4;12(5):e0176991 [FREE Full text] [doi: 10.1371/journal.pone.0176991] [Medline: 28472141]
48. Manoochehri Z, Rezaei M, Salari N, Khazaie H, Khaledi Paveh B, Manoochehri S. The prediction of obstructive sleep apnea using data mining approaches. Arch Iran Med 2018 Oct 01;21(10):460-465 [FREE Full text] [Medline: 30415554]
49. Manoochehri Z, Salari N, Rezaei M, Khazaie H, Manoochehri S, paveh B. Comparison of support vector machine based on genetic algorithm with logistic regression to diagnose obstructive sleep apnea. J Res Med Sci 2018;23(1):65. [doi: 10.4103/jrms.jrms_357_17]
50. Xu H, Zhao X, Shi Y, Li X, Qian Y, Zou J, et al. Development and validation of a simple-to-use clinical nomogram for predicting obstructive sleep apnea. BMC Pulm Med 2019 Jan 18;19(1):18 [FREE Full text] [doi: 10.1186/s12890-019-0782-1] [Medline: 30658615]
51. Ferreira-Santos D, Rodrigues PP. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches. Int J Data Sci Anal 2018 Apr 13;8(4):339-349. [doi: 10.1007/s41060-018-0118-x]
52. Keshavarz Z, Rezaee R, Nasiri M, Pournik O. Obstructive sleep apnea: a prediction model using supervised machine learning method. Stud Health Technol Inform 2020 Jun 26;272:387-390. [doi: 10.3233/SHTI200576] [Medline: 32604683]
53. Chen W, Feng J, Wang Y, Wang C, Dong Z. Development and validation of a nomogram for predicting obstructive sleep apnea in bariatric surgery candidates. Nat Sci Sleep 2021 Jun;13:1013-1023. [doi: 10.2147/nss.s316674]
54. Hsu Y, Wang J, Huang P, Chien Y, Chiu C, Lin C. Integrating domain knowledge with machine learning to detect obstructive sleep apnea: snore as a significant bio-feature. J Sleep Res 2022 Apr 21;31(2):e13487. [doi: 10.1111/jsr.13487] [Medline: 34549473]
55. Crocker BD, Olson LG, Saunders NA, Hensley MJ, McKeon JL, Allen KM, et al. Estimation of the probability of disturbed breathing during sleep before a sleep study. Am Rev Respir Dis 1990 Jul;142(1):14-18. [doi: 10.1164/ajrccm/142.1.14] [Medline: 2368960]
56. Pillar G, Peled N, Katz N, Lavie P. Predictive value of specific risk factors, symptoms and signs, in diagnosing obstructive sleep apnoea and its severity. J Sleep Res 1994 Dec;3(4):241-244 [FREE Full text] [doi: 10.1111/j.1365-2869.1994.tb00137.x] [Medline: 10607131]
57. Maislin G, Pack AI, Kribbs NB, Smith PL, Schwartz AR, Kline LR, et al. A survey screen for prediction of apnea. Sleep 1995 Apr;18(3):158-166. [doi: 10.1093/sleep/18.3.158] [Medline: 7610311]
58. Kushida CA, Efron B, Guilleminault C. A predictive morphometric model for the obstructive sleep apnea syndrome. Ann Intern Med 1997 Oct 15;127(8 Pt 1):581-587. [doi: 10.7326/0003-4819-127-8_part_1-199710150-00001] [Medline: 9341055]
59. El-Solh AA, Mador MJ, Ten-Brock E, Shucard DW, Abul-Khoudoud M, Grant BJ. Validity of neural network in sleep apnea. Sleep 1999 Feb 01;22(1):105-111. [doi: 10.1093/sleep/22.1.105] [Medline: 9989371]
60. Zerah-Lancner F, Lofaso F, d'Ortho MP, Delclaux C, Goldenberg F, Coste A, et al. Predictive value of pulmonary function parameters for sleep apnea syndrome. Am J Respir Crit Care Med 2000 Dec;162(6):2208-2212. [doi: 10.1164/ajrccm.162.6.2002002] [Medline: 11112139]
61. Rodsutti J, Hensley M, Thakkinstian A, D'Este C, Attia J. A clinical decision rule to prioritize polysomnography in patients with suspected sleep apnea. Sleep 2004 Jun 15;27(4):694-699. [doi: 10.1093/sleep/27.4.694] [Medline: 15283004]
62. Khoo S, Poh H, Chan Y, Ngerng W, Shi D, Lim TK. Diagnostic characteristics of clinical prediction models for obstructive sleep apnea in different clinic populations. Sleep Breath 2011 Sep 4;15(3):431-437. [doi: 10.1007/s11325-010-0354-3] [Medline: 20440569]
63. Zou J, Guan J, Yi H, Meng L, Xiong Y, Tang X, et al. An effective model for screening obstructive sleep apnea: a large-scale diagnostic study. PLoS One 2013 Dec 2;8(12):e80704 [FREE Full text] [doi: 10.1371/journal.pone.0080704] [Medline: 24312494]

64. Karamanli H, Yalcinoz T, Yalcinoz MA, Yalcinoz T. A prediction model based on artificial neural networks for the diagnosis of obstructive sleep apnea. Sleep Breath 2016 May 19;20(2):509-514. [doi: 10.1007/s11325-015-1218-7] [Medline: 26087718]
65. Tawaranurak K, Kamolphiwong S, Sae-Wong S, Vasupongayya S, Kamolphiwong T, Bumrungsena C, et al. Validity of a new prediction model to identify patients at risk for obstructive sleep apnea hypopnea syndrome. Ear Nose Throat J 2021 Jan 04:145561320986045 [FREE Full text] [doi: 10.1177/0145561320986045] [Medline: 33393817]
66. Park D, Kim J, Park B, Kim HJ. Risk factors and clinical prediction formula for the evaluation of obstructive sleep apnea in Asian adults. PLoS One 2021 Feb 2;16(2):e0246399 [FREE Full text] [doi: 10.1371/journal.pone.0246399] [Medline: 33529265]
67. Rowley JA, Aboussouan LS, Badr MS. The use of clinical prediction formulas in the evaluation of obstructive sleep apnea. Sleep 2000 Nov 01;23(7):929-938. [doi: 10.1093/sleep/23.7.929] [Medline: 11083602]
68. Soares MC, de Azeredo Bittencourt LR, Zonato AI, Gregório LC. Application of the Kushida morphometric model in patients with sleep-disordered breathing. Brazilian J Otorhinolaryngol 2006 Jul;72(4):541-548. [doi: 10.1016/s1808-8694(15)31002-8]

Abbreviations

AASM: American Academy of Sleep Medicine
AHI: apnea-hypopnea index
AUC: area under the receiver operating characteristic curve
NC: neck circumference
OSA: obstructive sleep apnea
PROBAST: Prediction Model Risk of Bias Assessment Tool
PROSPERO: International Prospective Register of Systematic Reviews
PSG: polysomnography

Edited by R Kukafka; submitted 10.05.22; peer-reviewed by M Pičulin, R Damaševičius; comments to author 13.06.22; revised version received 20.06.22; accepted 19.07.22; published 30.09.22

Please cite as:
Ferreira-Santos D, Amorim P, Silva Martins T, Monteiro-Soares M, Pereira Rodrigues P
Enabling Early Obstructive Sleep Apnea Diagnosis With Machine Learning: Systematic Review
J Med Internet Res 2022;24(9):e39452
URL: https://www.jmir.org/2022/9/e39452
doi: 10.2196/39452
PMID:

Chapter 4: Classification and Prediction

Classification and Prediction

The second aim of this thesis is to propose diagnostic Bayesian network models based on the previous identified factors, focusing on interpretability. To do so, some questions like *"Can we increase the identification of the disease?"* or *"Can we diminish the number of unnecessary polysomnography performed every year?"* need to be answered.

Firstly, to reach these answers we needed to collect, curate, and pre-process all medical and/or sleep records from two Sleep Laboratories, namely Vila Nova de Gaia/Espinho Hospital Center and São João University Hospital Center.

Two studies were conducted and correspond to Data Understanding and Preparation – Phase 2 and 3 in CRISP-DM:

4.1. Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis

Proceedings of Studies in Health Technology and Informatics. 247: 126–130, 2018. doi:10.3233/978-1-61499-852-5-126

Daniela Ferreira-Santos, Matilde Monteiro-Soares, and Pedro Pereira Rodrigues

4.2. Association between co-morbidities and prescribed drugs in obstructive sleep apnea suspected patients: an inductive rule learning approach

(submitted, April 2022)

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Both manuscripts intended to investigate data quality. Although they were collected in two different hospitals, the inclusion and exclusion criteria were the same.

The first utilized the data from Vila Nova de Gaia/Espinho, collected between January and May 2015, in the Sleep Laboratory. It intended to analyze the possibility of data imputation by utilizing the 10 nearest neighbors and check if this missing imputation technique is valid in the structure learning of Bayesian networks. The conclusion was promising, which resulted in its application from now on.

The second article refers to the data collection performed on the physical archive and in the Sleep Laboratory of São João. This dataset has all the information related to polysomnography performed in this hospital since there are electronic records. While the first article checks if the data imputation has any

effect on the Bayesian network structure, this one intends to diminish the overall number of missing information in the obtained dataset. As so, association rules were applied, reaching up to 7% of missed diagnoses.

It is our intention to revisit data quality issues, especially in the dataset obtained from São João, and try out new missing data imputation techniques.

4.1. Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis

Proceedings of Studies in Health Technology and Informatics. 247: 126–130, 2018. doi:10.3233/978-1-61499-852-5-126

Daniela Ferreira-Santos, Matilde Monteiro-Soares, and Pedro Pereira Rodrigues

Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth
A. Ugon et al. (Eds.)

doi:10.3233/978-1-61499-852-5-126

Impact of Imputing Missing Data in Bayesian Network Structure Learning for Obstructive Sleep Apnea Diagnosis

Daniela FERREIRA-SANTOS [a,b,1], Matilde MONTEIRO-SOARES [a,b] and Pedro Pereira RODRIGUES [a,b]
[a] *CINTESIS – Centre for Health Technology and Services Research, Portugal*
[b] *MEDCIDS-FMUP – Faculty of Medicine of the University of Porto, Portugal*

Abstract. Numerous diagnostic decisions are made every day by healthcare professionals. Bayesian networks can provide a useful aid to the process, but learning their structure from data generally requires the absence of missing data, a common problem in medical data. We have studied missing data imputation using a step-wise nearest neighbors' algorithm, which we recommended given its limited impact on the assessed validity of structure learning Bayesian network classifiers for Obstructive Sleep Apnea diagnosis.

Keywords. obstructive sleep apnea, Bayesian network, missing data imputation

1. Introduction

Numerous decisions are made every day by healthcare professionals based in an estimated probability that a specific disease or condition is present. In the diagnostic setting, the probability that a particular disease, such as obstructive sleep apnea (OSA), is present can be used to support further testing, request initiate treatment or reassure patients [1]. OSA is one of the most prevalent sleep disorders, affecting approximately 3-7% of men and 2-5% of women worldwide [2]. Despite its high-frequency, OSA remains underdiagnosed and underestimated, with 75-80% of cases remaining unidentified [2]. Its severity is assessed using the apnea-hypopnea index (AHI), stratifying into mild (5-15), moderate (15-30) and severe (higher than 30). Missing data is a relatively common problem in almost all types of studies, having a significant effect on the conclusions that can be drawn from the data. It is defined as the data value that is not stored for a variable in the observation of interest [1]. Its relevance is such that, when reporting a study development or validation of a diagnosis model, TRIPOD checklist has a specific topic for missing data, where it is mandatory to describe how missing data were handled with details [4]. This work objective was to study the impact of missing data imputation, using nearest neighbors (NN), on structure learning of Bayesian network classifiers for OSA diagnosis.

[1] Corresponding Author: Daniela Ferreira-Santos, CINTESIS and MEDCIDS-FMUP, Rua Dr. Plácido da Costa, s/n 4200-450 Porto, Portugal, E-mail: danielasantos@med.up.pt

2. Methods

2.1. Patients

We have included all patients that performed polysomnography at Vila Nova de Gaia/Espinho hospital center sleep laboratory. All medical and/or sleep laboratory records were retrospectively collected between the 1st of January to the 31st of May 2015. Included patients aged more than 18 years old, while patients already diagnosed, patients with severe lung diseases or neurological conditions and pregnant women were excluded. In case of duplicate exams, the best sleep efficiency was selected.

2.2. Variables and pre-processing

A literature review was previously conducted to define the most relevant OSA variables to be collected from administrative records. A total of 48 variables were collected: **demographic variables:** gender, age; **physical examination:** body mass index (BMI), neck (NC) and abdominal circumferences (AC), modified Mallampati, craniofacial/upper-airway abnormalities; **clinical history:** daytime sleepiness, snoring, witnessed apneas, gasping/choking, sleep fragmentation, non-repair sleep, behavior changes, decrease concentration, morning headaches, decreased libido, body position, sleep efficiency, vehicle crashes, drivers, driving sleepiness, nocturia, alcohol consumption, smoking, coffee, sedatives, family history/genetics, Epworth somnolence scale (ESS); **comorbidities:** atrial fibrillation, stroke, myocardial infarction (MI), pulmonary infarction, arterial and pulmonary hypertension, congestive heart failure (CHF), arrhythmias, pacemaker/cardiovector, respiratory alterations,, diabetes, dyslipidemia, renal failure, hypothyroidism, gastroesophageal reflux (GE), insomnia, glaucoma, bariatric surgery, depression/anxiety. The outcome measure was OSA clinical diagnosis, obtained from AHI, categorized into normal (AHI<5) or OSA (AHI≥ 5). We carry out a pre-processing analysis and continuous variables were categorized.

2.3. Imputing missing data

Instead of deleting any case that has missing data, *k*-NN imputation algorithms preserves all cases and replaces the missing data with a value obtained from related cases (*k* similar cases) in the whole set of records [3]. Our strategy followed systematic procedures: a) we observed the percentage of missing data, that ranged from 0% (e.g. gender) to 97% of missing data (e.g. bariatric surgery); b) variables were then ranked for data imputation, starting with outcome-wise statistically significant variables (with no quality problems suspected), followed by the remaining ordered in increasing percentage of missing data; c) 10-nearest neighbors imputation was done for each new included variable; d) odds ratio (OR) were computed to assess the impact of the referred *k*-NN imputation.

2.4. Naïve Bayes and Tree Augmented Naïve Bayes

Globally, a Bayesian network represents a joint distribution of one set of variables, specifying the assumption of independence between them with the interdependence

between variables being represented by a directed acyclic graph. Each variable is represented by a node in the graph, and its dependence on the set of variables is represented by its ascendant nodes. This dependence is represented by a conditional probability table that describes the probability distribution of each variable, given their ascendant variables [5]. Naïve Bayes (NB, which assumes conditional independence among factors) and Tree Augmented Naïve Bayes (TAN, which allows for an optional dependence for each factor) were the Bayesian network classifiers used in this work. Both classifiers structure learning algorithm requires complete cases, so they were built with the imputed dataset, and we assessed also the impact for different number of selected variables. In the first approach, we used the 10 variables that were statistical significant with or without imputation; in the second approach, we augmented the variable set with 6 more variables found significant in the imputed OR calculation.

2.5. Statistical analysis

Variables were selected after performing Chi-square test or Fisher's exact test for categorical variables and student's t-test or Mann-Whitney U test for continuous variables. Variables were selected if presenting an univariate significant association with the outcome, considering a 5% significance level and for which no quality problems were suspected. Model parameters (NB10 and NB16; TAN10 and TAN16) were validated by comparing the AUC in the imputed derivation cohort with those calculated from a leave-one-out, 10 times 2-fold cross validation (for variability assessment with independent training and testing) and the original derivation cohort. We used R software for: a) missing data analysis; b) imputing missing data (package DMwr); c) descriptive and comparative analysis (packages gmodels and epitools); and d) analyzing AUC (package pROC).

3. Results

In the 318 patients included, 211 had OSA. In total, we had 198 males (62%, crude and imputed OR 2.58 [1.65-4.30]), where 148 (70%) had OSA. In 211 patients with OSA, 115 (55%) were categorized as mild, 50 (24%) as moderate and 46 (22%) as severe. Participants had a mean age of 58 (49-67) years old, being higher in the OSA group (61 (53-68), p value ≤ 0.001); age above 45 years presented a higher risk for OSA (crude and imputed OR 3.29 [1.80-6.72]) and 5.93 [3.02-13.44], respectively). BMI median value was 29 (27-32) Kg/m^2; when categorized into normal and obese, we observed higher number of obese patients in the OSA group (83 (54%), crude OR 2.18 [1.30-3.97], imputed OR 1.91 [1.19-3.30]). NC and AC had a mean of 42 (39-44) cm and 106 (100-113) cm in the OSA group, with AC not having statistical significance (p value 0.052). Crude OR of modified Mallampati in the category 4 was significant (4.50 [1.09-38.83]) but when imputing (32% of missing data) it lost significance (3.36 [0.83-27.75]). The same occurred with nocturia (crude OR 2.05 [1.13-4.16], imputed OR 1.32 [0.79-2.40]). In those with craniofacial/upper airway abnormalities we discovered higher number of patients in OSA group (64 (82%), crude OR 1.24 [0.57-3.26], imputed OR 1.31 [0.77-2.42]), without statistical significance; other variables such as snoring, drivers, smoking, use of sedatives, sleep efficiency, gasping/choking, respiratory changes, sleep fragmentation, MI, pulmonary and arterial hypertension, dyslipidemia, anxiety or

depression, pacemaker or cardiovector, vehicle crashes, genetics/family history, hypothyroidism, renal failure, stroke, decreased libido and concentration, behavior changes, pulmonary infarction, glaucoma and bariatric surgery had no statistical significance, also. Subjects in-taking coffee had a higher risk of OSA (133 (86%)) with statistical significance in the imputed OR (0.48 [0.27-0.96]). The same effect was described for CHF, arrhythmias, diabetes, GE and insomnia. OSA group was a higher number of patients with witnessed apneas (109 (64%), crude OR 1.92 [1.18-3.34], imputed OR 2.14 [1.37-3.53]). Additionally, with statistical significance in crude and imputed OR, we found non-repairing sleep, morning headaches, driving sleepiness, alcohol consumption and body position. Daytime sleepiness and ESS presented contradictory results raising data collection quality suspicion, and were not considered for analysis. Impact of imputing missing data was assessed with ROC curves for each model, along with their 95% confidence interval (CI), being presented in Figure 1, demonstrating imputed and original in-sample AUC. Imputed leave-one-out and 10 times 2-fold cross validation values are presented in Table 1. Specific cut-off values were chosen after assessing the AUC of the imputed derivation cohort, aiming at a sensitivity of 95%, to allow a rule-out approach aiming to avoid false negatives. The AUC values of the original and imputed derivation cohort in the four models overlapped, as did imputed leave-one-out and cross-validation.

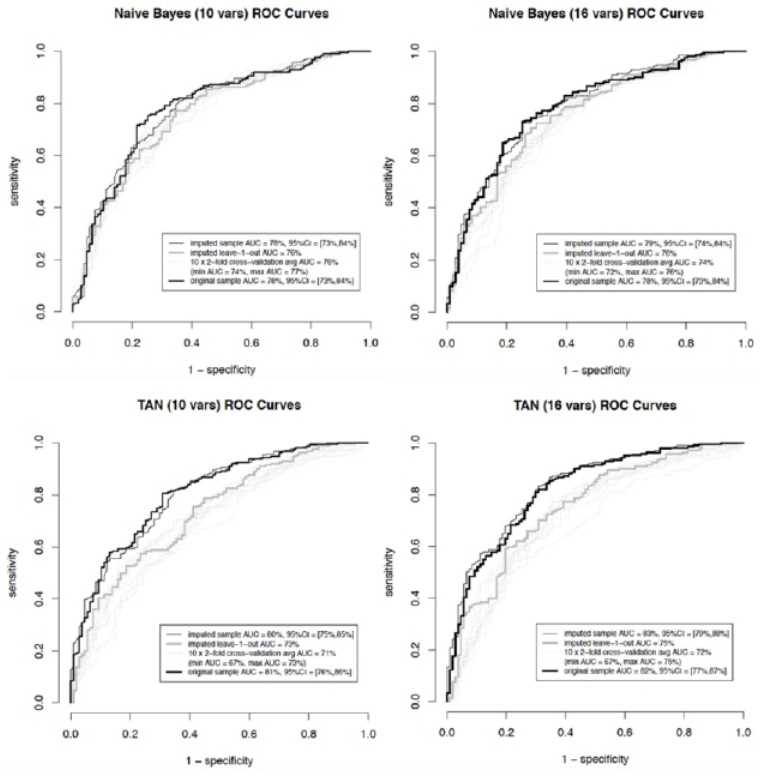


Figure 1. Receiver operating characteristics analyses and area under the curve values for NB10, TAN10, NB16 and TAN16, as well as for the internal validation procedures and the original derivation cohort.

Table 1. Validity assessment [%] estimated from 10 times 2-fold cross validation.

Model	Cut	Accuracy [CI 95%]	Sensitivity [CI 95%]	Specificity [CI 95%]	Precision (+) [CI 95%]	Precision (-) [CI 95%]
NB10	19.56%	70.85% [69.63,72.07]	95.07% [93.72,96.42]	23.09% [21.01,25.17]	70.91% [70.23,71.59]	71.69% [65.29,78.08]
TAN10	34.18%	68.71% [67.51,69.91]	89.05% [87.31,90.80]	28.6% [26.19,31.00]	71.1% [70.41,71.8]	57.70% [53.78,61.61]
NB16	13.17%	70.79% [70.00,71.57]	94.36% [93.00,95.72]	24.29% [22.44,26.13]	71.09% [70.67,71.51]	70.20% [65.62,74.78]
TAN16	23.61%	69.78% [68.39,71.17]	90.33% [88.55,92.11]	29.27% [26.36,32.18]	71.6% [70.73,72.47]	61.48% [56.51,66.46]

NB10, NB16: Naïve Bayes with 10 or 16 variables; **TAN10, TAN16:** Tree Augmented Naïve Bayes with 10 or 16 variables

4. Discussion and Conclusion

The occurrence of missing data is a major concern in several areas, including medical domains such as OSA diagnosis. The work of Hernández-Pereira *et al.* [6] tried to improve detection of apneic events by treating missing data; however, it only addresses numeric values. We have proposed a step-wise *k*-NN imputation approach (instead of more common list-wise deletion), proving to be a far better and valuable solution, with limited impact in structure learning Bayesian network classifiers. Main advantages include: a) imputed values are actually occurring values and not constructed values; b) it makes use of auxiliary information provided by the independent variables, preserving thus the original structure of the data; and c) it is fully non-parametric and does not require explicit models to relate factors and outcomes, being thus less prone to model misspecification.

Acknowledgments

This work has been developed under the scope of project NanoSTIMA [NORTE-01-0145-FEDER-000016], which was financed by the North Portugal Regional Operational Programme [NORTE 2020], under the PORTUGAL 2020 Partner- ship Agreement, and through the European Regional Development Fund [ERDF].

References

[1] H. Kang, The prevention and handling of the missing data, Korean J. Anesthesiol., vol. 64, n. 5, pp. 402-406, 2013.
[2] N. M. Punjabi, The epidemiology of adult obstructive sleep apnea, Proc. Am. Thorac. Soc., vol. 5, n. 2, pp. 136-143, 2008.
[3] L. Beretta e A. Santaniello, Nearest neighbor imputation algorithms: a critical evaluation, BMC Med. Inform. Decis. Mak., vol. 16, n. S3, p. 74, 2016.
[4] G. S. Collins, J. B. Reitsma, D. G. Altman, e K. G. M. Moons, Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement, Eur. Urol., vol. 67, n. 6, pp. 1142-1151, 2015.
[5] T. Mitchell, Machine Learning. McGraw-Hill, 1997.
[6] E. Hernández-Pereira, D. Álvarez-Estévez, e V. Moret-Bonillo, Improving detection of apneic events by learning from examples and treatment of missing data, Stud. Health Technol. Inform., vol. 207, pp. 213-224, 2014.

4.2. Association between co-morbidities and prescribed drugs in obstructive sleep apnea suspected patients: an inductive rule learning approach

(submitted, April 2022)

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Title Page

Title:

Association between co-morbidities and prescribed drugs in obstructive sleep apnea suspected patients: an inductive rule learning approach

Authors and institution(s) of origin:

Daniela Ferreira-Santos

CINTESIS, MEDCIDS, Faculty of Medicine, University of Porto, Porto, Portugal

Pedro Pereira Rodrigues

CINTESIS @ RISE, MEDCIDS, Faculty of Medicine, University of Porto, Porto, Portugal

Corresponding author:

Daniela Ferreira-Santos

E-mail: danielasantos@med.up.pt

Address: MEDCIDS - Departamento de Medicina da Comunidade Informação e Decisão em Saúde; Faculdade de Medicina da Universidade do Porto (CIM - FMUP)

Rua Dr. Plácido da Costa, s/n; 4200-450 Porto; Portugal

Introduction

One way to partially impute missing clinical variables is to find associations with other informative variables described in electronic health records (EHRs). Current evidence suggests that patients with accurate and complete EHRs may receive higher quality care than patients with gaps in their EHRs, as well as being critical for quality measurement and research [1].

This paper explores the automated acquisition of disease-drug associations in documents written daily by any healthcare professional and intends to use the available drug information to help fill missed diagnoses.

Methods

We manually extracted non-structural information relative to diseases and drugs prescribed in each patient's appointment; 26 predictive variables plus prescribed drugs (according to World Health Organization (WHO) Anatomical Therapeutic Chemical (ATC) classification system [2]) and OSA variable itself was gathered, within a consecutive cohort of 619 patients referred to undergo polysomnography at the University Hospital Center of São João, between 2011 and 2019, who had more than 18 years old and suspicion of having OSA.

We defined a strong disease-drug rule as a lift higher than 1 and a confidence higher than 85%, and if the rule was present in both the consequent and antecedent sets (RHS and LHS). After generating the disease-drug association rules, we evaluated the rules with the help of a clinical expert, so we could interpret the clinical significance of each disease-drug pair.

Results

A total of 481 patients with 34 diseases and 50 drugs classified into 2nd ATC-codes (A is related to alimentary tract and metabolism, C is cardiovascular system, and N is nervous system.) had a variation in missing data of 0% to 97%. We found a total of 29 disease-drug associations with a lift higher than 1 (15 rules for LHS and 14 for the RHS), but only 25 had confidence higher than 85%. The LHS and RHS obtained rules are listed in Tables 1 by decreasing order of lift,

alongside with support and confidence, and can visualize in Figure 1, where it is easier to identify all rules with high lift.

Table 1 – Association rules for prescribed drugs and diseases and vice-versa (boldfaced represents strong rules association)

LHS - Drug	*RHS - Disease*	*Lift*	*Support*	*Confidence*
A10	**Diabetes**	**2.05**	**0.40**	**0.91**
C09 & C10	Dyslipidemia	1.30	0.56	0.89
C10	**Dyslipidemia**	**1.28**	**0.60**	**0.87**
C09	**Arterial hypertension**	**1.24**	**0.68**	**0.95**
C09 & A02	Arterial hypertension	1.24	0.45	0.95
C09 & N06	Arterial hypertension	1.24	0.45	0.95
C09 & C10	Arterial hypertension	1.24	0.60	0.95
C09 & C10 & A02	Arterial hypertension	1.24	0.41	0.95
A10	Arterial hypertension	1.21	0.41	0.93
C10 & A02	Arterial hypertension	1.19	0.41	0.92
C10	Arterial hypertension	1.17	0.62	0.90
N06 & C10	Arterial hypertension	1.16	0.40	0.89
LHS - Disease	*RHS – Drug*	*Lift*	*Support*	*Confidence*
Diabetes	**A10**	**2.05**	**0.40**	**0.90**
Diabetes	C10	1.34	0.41	0.93
Arterial hypertension & Dyslipidemia	C10	1.30	0.57	0.90
Obstructive sleep apnea & Arterial hypertension & Dyslipidemia	C10	1.30	0.43	0.90
Arterial hypertension & Dyslipidemia	C09	1.29	0.59	0.93
Dyslipidemia	**C10**	**1.28**	**0.60**	**0.88**
Obstructive sleep apnea & Dyslipidemia	C10	1.27	0.46	0.88
Obstructive sleep apnea & Arterial hypertension & Dyslipidemia	C09	1.27	0.45	0.92
Diabetes	C09	1.25	0.40	0.90
Arterial hypertension	**C09**	**1.24**	**0.68**	**0.89**
Dyslipidemia	C09	1.21	0.60	0.87
Obstructive sleep apnea & Arterial hypertension	C09	1.21	0.53	0.88
Obstructive sleep apnea & Dyslipidemia	C09	1.20	0.45	0.86

A10: drugs used in diabetes; C09: agents acting on the renin-angiotensin system; C10: lipid modifying agents; A02: drugs for acid related disorders; N06: psychoanaleptics

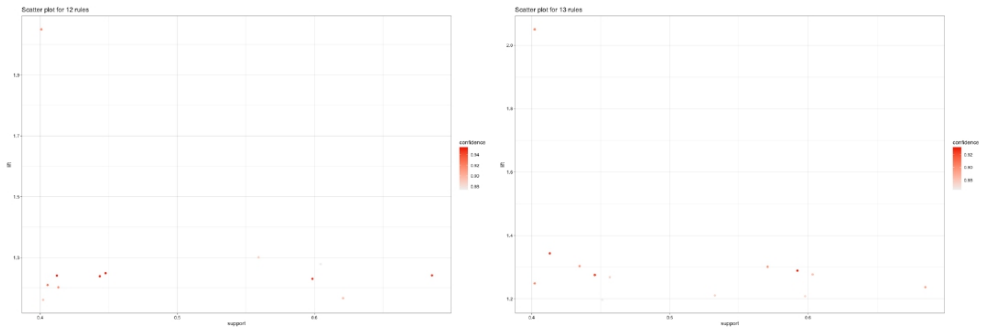


Figure 1 – Association rules for prescribed drugs and diseases (left image) and for diseases and prescribed drugs (right image)

The highest support (68%) was found with the rule that involved C09 and arterial hypertension, whereas the highest confidence obtained was 95% (e.g., of the patients that took C09, 95% had arterial hypertension). Considering our chosen measure, the highest lift value was 2.05 for the rule that involved A10 and diabetes (e.g., knowing that a patient took A10, the probability of having diabetes is 2.05 times higher than the probability of having diabetes not knowing that they are taking A10).

Looking for strong disease-drug rules, we found A10 → diabetes, C10 → dyslipidemia, and C09 → arterial hypertension. Note that the A10, C10 and C09 2nd level ATC-codes also appeared related to other diseases, *e.g.* A10 → arterial hypertension. Nevertheless, these rules are not presented in LHS and RHS.

Regarding missing data in these three strong disease-drug rules, we have an initial percentage of 47% to diabetes, 35% to dyslipidemia, and 23% to arterial hypertension, which evolves to 46%, 31%, and 21% when we applied the rules to our dataset.

Discussion

We hypothesized that disease-drug association rules in an underdiagnosed disease as OSA can be a useful technique for inferring meaningful relations and filling information gaps in EHRs. We found three interesting disease-drug associations; the 1st rule shows that out of the OSA suspected patients taking drugs used in diabetes (A10), 91% had diabetes, while of the ones having diabetes, 90% were prescribed drugs used in diabetes. These confidence levels strengthened our intuition to use this rule for missing data imputation, resulting in a diminishing 1% of all missed diagnoses of diabetes. Rule 2 demonstrates the same strong association between lipid modify agents and dyslipidemia, with a lower confidence level and a lift measure of 1.28, denoting that out of the OSA suspected patients who were prescribed lipid modify agents, the probability of having dyslipidemia is 1.28 times higher than the probability of dyslipidemia not knowing if the patient is taking lipid modify agents. This rule resulted in a 4% diminishing of dyslipidemia missed diagnoses. The 3rd rule connects agents acting on the renin-angiotensin system and arterial

hypertension, with a confidence of 95% in LHS and 89% in RHS, a lift value of 1.24, and a 2% reduction in missed diagnoses. As previously mentioned, all rules were clinically confirmed by an expert on the topic.

We can confidently impute disease information with a lower error of 5% and a higher error of 13% while decreasing missingness up to 4% in missed diagnosis. Future work will measure the impact of using this imputation in classification validity in other datasets.

Acknowledgements

A thank you to Inês Ribeiro-Vaz for expert interpretation of association rules. Daniela Ferreira-Santos acknowledges Fundacão para a Ciência e Tecnologia under PhD grant (PD/BD/13553/2018) and the PhD Program in Clinical and Health Services Research (PD/00003/2013) for funding.

References

1. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. J Biomed Inform. 2010;43: 891–901. doi:10.1016/j.jbi.2010.09.009
2. WHO. World Health Organization - ATC/DDD Index. [cited 26 May 2022]. Available: https://www.whocc.no/atc_ddd_index/

Secondly, we wanted to understand OSA heterogeneity, as the various clinical presentations have not yet been formally characterized. Additionally, the association between OSA and comorbidities is unclear, which poses challenges to its clinical recognition.

Two studies were conducted and correspond to Modelling – Phase 4 in CRISP-DM:

4.3. Finding groups in obstructive sleep apnea patients: a categorical cluster analysis

Proceedings of 2018 IEEE 31st Symposium on Computer-Based Medical Systems. 387–392, 2018. doi:10.1109/CBMS.2018.00074

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

4.4. Phenotyping obstructive sleep apnea patients: a first approach to cluster visualization

Proceedings of Studies in Health Technology and Informatics. 255: 75–79, 2018. doi:10.3233/978-1-61499-921-8-75

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

The objective of both articles was to help define the characteristics related to OSA, as this disease has high heterogeneity. The first article applied k-modes to all collected patients (OSA and healthy ones), while the second only developed and validated clusters for OSA patients; both with three clusters.

The first utilized more predictive variables than the second, with cluster 1 having more patients in both studies. Differences in the number of patients were seen in clusters 2 and 3, with the first article having more in cluster 2 and the second in cluster 3.

Regarding the characteristics, both studies presented a middle-aged women cluster, differing only in neck circumference characteristics. Additionally, one cluster with middle-aged men also appeared, with differences in Mallampati score, if the sleep was repaired or not, and if the person has or not morning headaches. Finally, the third cluster has elderly men with differences in the stage of obesity and Mallampati score.

4.3. Finding groups in obstructive sleep apnea patients: a categorical cluster analysis

Proceedings of 2018 IEEE 31st Symposium on Computer-Based Medical Systems. 387–392, 2018. doi:10.1109/CBMS.2018.00074

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

2018 IEEE 31st International Symposium on Computer-Based Medical Systems

Finding groups in Obstructive Sleep Apnea patients: a categorical cluster analysis

Daniela Ferreira-Santos and Pedro Pereira Rodrigues
CINTESIS - Center for Health Technology and Services Research
MEDCIDS-FMUP - Faculty of Medicine of the University of Porto
Porto, Portugal
{*danielasantos,pprodrigues*}*@med.up.pt*

***Abstract*—Obstructive sleep apnea (OSA) is a significant sleep problem with various clinical presentations that have not been formally characterized. This poses critical challenges for its recognition, resulting in missed or delayed diagnosis. Recently, cluster analysis has been used in different clinical domains, particularly within numeric variables. We applied an extension of *k*-means to be used in categorical variables: *k*-modes, to identify groups of OSA patients. Demographic, physical examination, clinical history, and comorbidities characterization variables (n=46) were collected from 318 patients; missing values were all imputed with k-nearest neighbors (k-NN). Feature selection, through Chi-square test, was executed and 17 variables were inserted in cluster analysis, resulting in three clusters. Cluster 1 having an age between 65 and 90 years (54%), 78% of males, with the presence of diabetes and gastroesophageal reflux, and high OSA prevalence; Cluster 2 presented a lower percentage of OSA (46%), with middle-aged women without comorbidities, but with gastroesophageal reflux; and Cluster 3 was very similar to cluster 1, only differing in age (45-64) and comorbidities were not present. Our results suggest that there are different groups of OSA patients, creating the need to rethink the baseline characteristics of these patients before being sent to perform polysomnography (gold standard exam for diagnosis).**

***Keywords*-categorical data; clinical presentations; cluster analysis; data mining; diagnosis; missing data imputation; obstructive sleep apnea**

I. Introduction

Obstructive sleep apnea (OSA) is a common sleep disorder, characterized by recurrent episodes of collapse of the upper airway during sleep, that occurs in approximately 5-10% of the general population [1], [2]. A diagnosis of OSA is established when a patient has an apnea-hypopnea index (AHI; number of apneas and hypopneas per hour of sleep) ≥ 5 and associated symptoms (e.g., snoring) or an AHI ≥ 15 regardless of associated symptoms [3]. Although commonly observed in clinical practice, the various clinical presentations of OSA have not been formally characterized [4], [5]. Moreover, it is known that there is an association between OSA and the presence of multimorbidity, but it is not clear how much this association is due directly to OSA rather than to other factors often present in the same patients [6]. The lack of knowledge of the heterogeneity of OSA clinical presentations may pose critical challenges to its clinical recognition, resulting in missed or delayed diagnosis [4].

Unsupervised cluster analysis has recently been used to identify subtypes of patients who are diagnosed with a particular disorder, such as asthma, chronic obstructive pulmonary disease, fibromyalgia, and Parkinson's disease [7]–[10]. It is defined as a statistical approach to study the relationships present among groups of patients, or variables, in a large population [11], making possible to identify groups of patients who are similar among themselves but significantly different from others [12].

Aiming to better understanding the heterogeneity of OSA clinical presentations, this study applied categorical cluster analysis (*k*-modes) to identify subgroups of OSA patients who experience distinct combinations of symptoms and comorbidities. This paper is organized as follows: methods section (II) presents the research methodology, section III gives an overview of the achieved results. Following section exposes the related work on the theme, section V and VI interpret the results of the work and provide the main findings and recommendations.

II. Methods

This study was designed according to the Standard for Reporting Diagnostic accuracy studies (STARD) list [13].

A. Patients

We have included all patients that performed polysomnography at Vila Nova de Gaia/Espinho hospital center sleep laboratory, involved in a previous diagnostic test study [14]. All medical and sleep laboratory records were retrospectively collected between January and May, 2015. Inclusion criteria were patients aged more than 18 years old; while patients already diagnosed, patients with severe lung diseases or neurological conditions and pregnant women were excluded. In case of duplicate exams, the one with the best sleep efficiency was selected.

B. Variables and pre-processing

A literature review was previously conducted to define the most relevant OSA variables to be collected from administrative records. A total of 46 variables were collected: **demographic variables:** gender and age; **physical examination:** body mass index (BMI), neck (NC) and abdominal circumferences (AC), modified Mallampati,

2372-9198/18/$31.00 ©2018 IEEE
DOI 10.1109/CBMS.2018.00074

craniofacial/upper-airway abnormalities (CFA); **clinical history:** daytime sleepiness, Epworth somnolence scale (ESS), snoring, witnessed apneas, gasping/choking, nocturia, sleep fragmentation, non-repairing sleep, sleep efficiency, insomnia, morning headaches, behavior changes, decrease concentration, decreased libido, body position, drivers, vehicle crashes, driving sleepiness, alcohol consumption, smoking, coffee intake, use of sedatives, family history/genetics; **comorbidities:** myocardial infarction, congestive heart failure, arrhythmias, pacemaker/cardiovector, arterial and pulmonary hypertension, pulmonary infarction, respiratory alterations, stroke, diabetes, dyslipidemia, renal failure, hypothyroidism, gastroesophageal reflux, glaucoma, anxiety/depression. The outcome measure was OSA clinical diagnosis, obtained from AHI, categorized into normal (AHI<5) or OSA (AHI$\geq$ 5).

We carried out a pre-processing analysis and continuous variables were categorized as follows: BMI (<30kg/m^2: normal, BMI$\geq$ 30kg/m^2: obese); female NC ($\leq$ 37cm: normal, >37cm: increased) and male NC ($\leq$ 42cm: normal, >42cm: increased); female AC ($\leq$ 80cm: normal, > 80cm: increased) and male AC ($\leq$ 94cm: normal, >94cm: increased); age ($<$ 40 years, 40-54 years, 55-69 years, $\geq$ 70 years); smoking (yes, no, ex-smoker); AHI (0-4: normal, 5-14: mild, 15-29: moderate, $\geq$ 30: severe); ESS (0-10: normal, 11-24: daytime sleepiness).

C. Missing data

Instead of deleting any case that has missing data, k-nearest neighbors (k-NN) imputation algorithms preserves all cases and replaces the missing data with a value obtained from related cases (k similar cases) in the whole set of records [15]. As described in previous work [16], our strategy followed systematic procedures: a) we observed the percentage of missing data, that ranged from 0% (e.g., gender) to 92% of missing data (e.g., glaucoma); b) variables were then ranked for data imputation, starting with outcome-wise statistically significant variables (with no quality problems suspected), followed by the remaining ordered in increasing percentage of missing data; c) 10-nearest neighbors imputation was done for each new included variable; d) odds ratios (OR) were computed to assess the impact of the referred k-NN imputation.

D. Feature selection and statistical analysis

Our data set consisted of only categorical variables, which rose a question: what is a good way to distinguish between high influence variables and low or no influence variables? Chi-square analysis involves counting occurrences and comparing these variables to the outcome measure based on the frequencies of occurrences. Variables were selected if presenting an univariate significant association with the outcome, considering a 5% significance level and for which no quality problems were suspected. We used R software to perform descriptive and associative analysis (packages *gmodels* and *epitools*) and k-modes categorical clustering (package *klaR*).

E. Categorical cluster analysis - k-modes

K-means clustering method is well known for its efficiency, however, it only works on numeric data, making it not available in applications where categorical data are involved. The k-modes algorithm [17] extends the k-means paradigm to cluster categorical data by using a) a simple matching dissimilarity measure for categorical objects [18], b) modes instead of means for cluster, and c) a frequency-based method to update modes in the k-means fashion clustering process to minimize the clustering cost function. The dissimilarity measure can be defined by the total mismatches of the corresponding variable categories of the two objects; the smaller the number of mismatches, the more similar the two objects are [18].

III. Results

In the 318 patients included, 211 had OSA (66%); from which 115 (55%) were categorized as mild, 50 (24%) as moderate and 46 (22%) as severe. In total, we had 198 males (62%), where 148 (70%) had OSA. Table I describes the data set obtained from the medical and/or sleep laboratory records.

Patients involved had a mean age of 58 (49-67) years old; the category 20-44 years old presented a higher percentage of patients in the normal group, as oppositely to categories 45-64 and 65-90 years old (p value <0.001). BMI median value was 29 (27-32) kg/m^2; when categorized into normal and obese, we observed higher number of obese patients in the OSA group (85 (40%), p value 0.008). NC and AC had a mean of 42 (39-44) cm and 106 (100-113) cm in the OSA group, with higher number of patients with increased NC in the OSA group, as opposite to AC that presented higher number of patients with increased AC in the normal group. Modified Mallampati categories showed us that our patients have higher total percentages in the inferior levels (1 and 2 in a total of 69%). In those with craniofacial/upper airway abnormalities we discovered higher number of patients in OSA group (174 (82%)), without statistical significance; other variables such as snoring, nocturia, decreased concentration, drivers, family history/genetics, myocardial infarction, arterial hypertension, pacemaker/cardiovector, pulmonary infarction, stroke, renal failure, and hypothyroidism had also no statistical significance. When analyzing gasping/choking we noticed a higher percentage of patients in the normal group (54 (50%)), without statistical significance; other variables presented the same characteristics: sleep fragmentation, behavior changes, decreased libido, vehicle crashes, use of sedatives, respiratory alterations, and anxiety/depression. Also, we had two variables (dyslipidemia and glaucoma) that had the same percentages of patients in both groups, without statistical significance. Daytime sleepiness exhibited

contradictory results (higher percentage of patients in the normal group n=77; 72%) with statistical significance. This was also described for ESS, which presented a contradiction to the literature and the inherent meaning of the variables, and thus was not considered for analysis.

We included a total of 17 variables in the cluster analysis (gender, age, non-repairing sleep, alcohol consumption, BMI, witnessed apneas, modified Mallampati, morning headaches, NC, driving sleepiness, coffee intake, congestive heart failure, pulmonary hypertension, diabetes, arrhythmias, gastroesophageal reflux, and insomnia). Three distinct clusters were identified. Table II summarizes clinical characteristics of the total cohort in the selected variables by cluster, and table III aids to visualize modes in each cluster. Cluster 1 exposed a higher number of elderly males with normal BMI but increased NC. Also, patients in this cluster had repairing sleep, absence of morning headaches, but described insomnia. Additionally, witnessed apneas are reported, as alcohol consumption and coffee. Regarding comorbidities, these patients only reported diabetes and gastroesophageal reflux. Oppositely to cluster 1, cluster 3 exhibited normal NC, non-repairing sleep, morning headaches, no comorbidities and middle-aged males. Cluster 2 included middle-aged females, with normal BMI and NC. This group did not report witnessed apneas and presented only gastroesophageal reflux as comorbidity. Moreover, they had non-repairing sleep, morning headaches and insomnia. In all three clusters, modified Mallampati had a lower value (between 1 and 2), none described driving sleepiness or congestive heart failure. As indicated in table II, no statistical difference was observed in coffee intake, pulmonary hypertension, arrhythmias, and insomnia.

IV. Related work

Ye *et al.* [4] in 2014 included a total of 23 variables in cluster analysis, achieving three clusters: cluster 1 described as disturbed sleep group, cluster 2 as minimally symptomatic group, and cluster 3 as excessive daytime sleepiness group. This study used latent class analysis to cluster subjects into groups based on symptoms and presence of comorbidities, analyzing only moderate-to-severe OSA patients. The study described as main findings 1) identification of subtypes of OSA improves knowledge and awareness of the heterogeneity of OSA, and 2) identifying distinct clinical profiles of OSA creates a foundation for offering more personalized therapies in the future. Another study [19], investigated the possibility of identifying different clusters of patients (three clusters, also). In contrast to the previous study, this applied principal component analysis in the selection of the variables from the initial 20, without the restriction in OSA severity. As main findings, Lacedonia *et al.* [19] referred the knowledge of the main parameters involving in OSA could be useful to stratify the mortality and morbidity risk, and therefore improve the approach to clinical management of these patients. In the same year, Vavougios *et al.* [12] aimed to determine phenotypes of comorbidity in patients that performed polysomnography. This study applied categorical principal component analysis to the comorbidities collected and presented six clusters.

V. Discussion

Understanding different patterns in OSA diagnosis is particularly important. Patients in cluster 1 and 3 were males with different categorized age (elderly vs. middle-aged adults). Analyzing physical examination aspects, we verified that cluster 1 had normal BMI and increased NC, oppositely to cluster 2 and cluster 3 that had normal BMI and NC. Regarding clinical history, cluster 1 reported repairing sleep; cluster 2 and 3 non-repairing sleep; cluster 1 and 3 had positive alcohol, coffee intake and witnessed apneas. Although the three clusters reported insomnia, only cluster 2 and 3 had morning headaches. The only cluster presenting diabetes was cluster 1. Cluster 1 and 2 exhibited gastroesophageal reflux. Moreover, the percentage of the outcome measure was demonstrated in each cluster. We verified that cluster 1 had 82% [74%-87%] of OSA diagnosis, followed by cluster 3 (68% [56%-79%]) and cluster 2, with significantly smaller proportion (43% [33%-53%]). Although OSA severity is widely classified using the AHI, this objective index does not consider the marked heterogeneity of different clinical subgroups of OSA. The work of Ye *et al.* [4] only collected demographic and survey data about sleep-related health issues, identifying three clusters very different from ours. Their cluster 1 presented insomnia-related symptoms, difficulty falling sleep and waking up too early and ours described repairing sleep and no morning headaches. Regarding cluster 2, Ye study demonstrated a higher number of older males; being ours mode the female gender. Cluster 3 is quite different as patients presented a high ESS and our work removed this variable due to quality issues. Although, this cluster has no comorbidities associated like ours. The comparison of our work to Lacedonia *et al.* [19] is much harder once they used instrumental data such as blood gas analysis and spirometry parameters.

VI. Conclusion

To the best of our knowledge, this is the first attempt to explore different clinical presentations of suspected OSA patients using *k*-modes categorical clustering. Our results suggest that there are different clinical subtypes of OSA, helping focus our attention on a detailed description of OSA diagnosis. The major strengths of this study are newly data analysis and the clinical cohort representing OSA patients (all levels of severity) that performed polysomnography. Furthermore, the inclusion of a comprehensive number of risk and diagnostic factors enhances our understanding of OSA diagnosis. Current clinical practice and research, which emphasize only a few ”typical” signs and symptoms such as

snoring and witnessed apneas, may have created a potentially problematic image of a stereotypical OSA patient, that needs to be redefined.

Acknowledgments

DFS acknowledges Fundação para a Ciência e Tecnologia (FCT) under grant number PD/BD/13553/2018. The work of DFS and PPR has been developed under the scope of project NORTE-01-0145-FEDER-000016 (NanoSTIMA), financed by the North Portugal Regional Operational Programme (NORTE-2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] N. Punjabi, "The epidemiology of adult obstructive sleep apnea," *Proceedings of the American Thoracic Society*, vol. 5, no. 2, pp. 136–143, 2008.

[2] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults." *American Journal of Epidemiology*, vol. 177, no. 9, pp. 1006–1014, 2013.

[3] American Academy of Sleep Medicine, "Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine Task Force." *Sleep*, vol. 22, no. 5, pp. 667–689, 1999.

[4] L. Ye, G. W. Pien, S. J. Ratcliffe, E. Björnsdottir, E. S. Arnardottir, A. I. Pack, B. Benediktsdottir, and T. Gislason, "The different clinical faces of obstructive sleep apnoea: A cluster analysis," *European Respiratory Journal*, vol. 44, no. 6, pp. 1600–1607, 2014.

[5] J. B. Dixon, M. E. Dixon, M. L. Anderson, L. Schachter, and P. E. O'brien, "Daytime sleepiness in the obese: not as simple as obstructive sleep apnea." *Obesity*, vol. 15, no. 10, pp. 2504–2511, 2007.

[6] L. Robichaud-Hallé, M. Beaudry, and M. Fortin, "Obstructive sleep apnea and multimorbidity," *BMC Pulmonary Medicine*, vol. 12, no. 1, p. 60, 2012.

[7] W. C. Moore, D. A. Meyers, S. E. Wenzel, W. G. Teague, H. Li, X. Li, R. D'Agostino, M. Castro, D. Curran-Everett, A. M. Fitzpatrick, B. Gaston, N. N. Jarjour, R. Sorkness, W. J. Calhoun, K. F. Chung, S. A. Comhair, R. A. Dweik, E. Israel, S. P. Peters, W. W. Busse, S. C. Erzurum, and E. R. Bleecker, "Identification of asthma phenotypes using cluster analysis in the severe asthma research program," *American Journal of Respiratory and Critical Care Medicine*, vol. 181, no. 4, pp. 315–323, 2010.

[8] J. Garcia-Aymerich, F. P. Gomez, M. Benet, E. Farrero, X. Basagana, A. Gayete, C. Pare, X. Freixa, J. Ferrer, A. Ferrer, J. Roca, J. B. Galdiz, J. Sauleda, E. Monso, J. Gea, J. A. Barbera, A. Agusti, and J. M. Anto, "Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes," *Thorax*, vol. 66, no. 5, pp. 430–437, 2011.

[9] R. Erro, C. Vitale, M. Amboni, M. Picillo, M. Moccia, K. Longo, G. Santangelo, A. De Rosa, R. Allocca, F. Giordano, G. Orefice, G. De Michele, L. Santoro, M. T. Pellecchia, and P. Barone, "The Heterogeneity of Early Parkinson's Disease: A Cluster Analysis on Newly Diagnosed Untreated Patients," *PLoS ONE*, vol. 8, no. 8, pp. 1–8, 2013.

[10] E. Docampo, A. Collado, G. Escaramís, J. Carbonell, J. Rivera, J. Vidal, J. Alegre, R. Rabionet, and X. Estivill, "Cluster Analysis of Clinical Data Identifies Fibromyalgia Subgroups," *PLoS ONE*, vol. 8, no. 9, pp. 1–7, 2013.

[11] C. Gallo and V. Capozzi, "Clustering Techniques for Revealing Gene Expression Patterns," *Encyclopedia of Information Science and Technology, Third Edition*, pp. 438–447, 2015.

[12] G. D. Vavougios, G. Natsios, C. Pastaka, S. G. Zarogiannis, and K. I. Gourgoulianis, "Phenotypes of comorbidity in OSAS patients: Combining categorical principal component analysis with cluster analysis," *Journal of Sleep Research*, vol. 25, no. 1, pp. 31–38, 2016.

[13] P. M. Bossuyt, J. B. Reitsma, D. E. Bruns, C. A. Gatsonis, P. P. Glasziou, L. Irwig, J. G. Lijmer, D. Moher, D. Rennie, H. C. W. De Vet, H. Y. Kressel, N. Rifai, R. M. Golub, D. G. Altman, L. Hooft, D. A. Korevaar, and J. F. Cohen, "STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies," 2015.

[14] D. Ferreira-Santos and P. P. Rodrigues, "Improving diagnosis in Obstructive Sleep Apnea with clinical data: a Bayesian network approach," in *2017 IEEE 30th International Symposium on Computer-Based Medical Systems*, 2017, pp. 612–617.

[15] L. Beretta and A. Santaniello, "Nearest neighbor imputation algorithms: a critical evaluation," *BMC Medical Informatics and Decision Making*, vol. 16, no. S3, p. 74, 2016.

[16] D. Ferreira-Santos, M. Monteiro-Soares, and P. Rodrigues, "Impact of imputing missing data in bayesian network structure learning for obstructive sleep apnea diagnosis." *Studies in health technology and informatics*, vol. 247, pp. 126–130, 2018.

[17] Z. Huang, "A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining," *Research Issues on Data Mining and Knowledge Discovery*, pp. 1–8, 1997.

[18] L. Kaufman and P. J. Rousseuw, "Finding Groups in Data: An Introduction to Cluster Analysis." *Wiley*, vol. 47, no. 2, p. 788, 1990.

[19] D. Lacedonia, G. E. Carpagnano, R. Sabato, M. M. lo Storto, G. A. Palmiotti, V. Capozzi, M. P. F. Barbaro, and C. Gallo, "Characterization of obstructive sleep apnea-hypopnea syndrome (OSA) population by means of cluster analysis," *Journal of Sleep Research*, vol. 25, no. 6, pp. 724–730, 2016.

Table I
DESCRIPTIVE ANALYSIS OF THE COHORT, N(%) (P-VALUES RESULT OF CHI-SQUARE TEST, UNLESS OTHERWISE SPECIFIED).

	Total (n=318)	**Normal** (n=107)	**OSA** (n=211)	*p-value*
Male gender	198 (62)	50 (47)	148 (70)	<0.001
Age				
20-44	49 (15)	31 (29)	18 (9)	<0.001
45-64	168 (53)	55 (51)	113 (54)	
65-90	101 (32)	21 (20)	80 (38)	
Obese	112 (35)	27 (25)	85 (40)	0.008
Increased NC	158 (50)	33 (31)	125 (59)	<0.001
Increased AC	206 (65)	73 (68)	133 (63)	0.360
Mallampati				
1	77 (24)	28 (26)	49 (23)	0.019#
2	143 (45)	57 (53)	86 (41)	
3	85 (27)	21 (20)	64 (30)	
4	13 (4)	1(1)	12 (6)	
CFA	257 (81)	83 (78)	174 (82)	0.295
Daytime sleepiness	196 (62)	77 (72)	119 (56)	0.007
ESS				
Daytime sleepiness	133 (42)	57 (53)	76 (36)	0.003
Snoring	286 (90)	94 (88)	192 (91)	0.378
Witnessed apneas	175 (55)	45 (42)	130 (62)	0.001
Gasping/Choking	142 (45)	54 (50)	88 (42)	0.138
Nocturia	252 (79)	81 (76)	171 (81)	0.267
Sleep fragmentation	225 (71)	77 (72)	148 (70)	0.736
Non-repairing sleep	164 (52)	69 (64)	95 (45)	0.001
Sleep efficiency				
Bad	200 (63)	71 (66)	129 (61)	0.363
Good	118 (37)	36 (34)	82 (39)	
Insomnia	275 (86)	86 (80)	189 (90)	0.023
Morning headaches	163 (51)	69 (64)	94 (45)	0.001
Behavior changes	284 (89)	97 (91)	187 (89)	0.580
Concent. Decrease	294 (92)	97 (91)	197 (93)	0.387
Decreased libido	265 (83)	95 (89)	170 (81)	0.063
Body position				
Decubitus	17 (5)	9 (8)	8 (4)	0.062
Left lateral	55 (17)	15 (14)	40 (19)	
Right lateral	155 (49)	59 (55)	96 (45)	
Supine	91 (29)	24 (22)	67 (32)	
Drivers	19 (6)	5 (5)	14 (7)	0.485
Vehicle crashes	17 (5)	7 (7)	10 (5)	0.500
Driving sleepiness	39 (12)	22 (21)	17 (8)	0.001
Alcohol consumption	207 (65)	56 (52)	151 (72)	0.001
Smoking				
Yes	41 (13)	17 (16)	24 (11)	0.294
Ex-smoker	105 (33)	30 (28)	75 (36)	
Coffee intake	255 (80)	93 (87)	162 (77)	0.032
Use of sedatives	75 (24)	31 (29)	44 (21)	0.107
Family history/Genetics	16 (5)	2 (2)	14 (7)	0.066
MI	29 (9)	9 (8)	20 (9)	0.755
Arterial hypertension	275 (86)	91 (85)	184 (87)	0.595
Cong. Heart Fail.	34 (11)	5 (5)	29 (14)	0.013
Arrhythmias	20 (6)	2 (2)	18 (9)	0.021
Pacemaker/Cardiovector	11 (3)	1 (1)	10 (5)	0.107#
Pulmonary infarction	3 (1)	0 (0)	3 (1)	0.553#
Pulm. Hypertension	17 (5)	2 (2)	15 (7)	0.050
Respiratory alterations	146 (46)	55 (51)	91 (43)	0.162
Stroke	202 (64)	65 (61)	137 (65)	0.464
Diabetes	122 (38)	31 (29)	91 (43)	0.014
Dyslipidemia	273 (86)	92 (86)	181 (86)	0.962
Renal failure	52 (16)	15 (14)	37 (18)	0.423
Hypothyroidism	32 (10)	7 (7)	25 (12)	0.137
Gastroesoph. reflux	223 (70)	67 (63)	156 (74)	0.037
Glaucoma	6 (2)	2 (2)	4 (2)	0.657#
Anxiety/Depression	273 (86)	95 (89)	178 (84)	0.285

Fisher's exact test
OSA: obstructive sleep apnea; NC: neck circumference; AC: abdominal circumference; CFA: craniofacial and upper-airway abnormalities; ESS: Epworth somnolence scale; Concent. Decrease: concentration decrease; MI: myocardial infarction; Cong. Heart Fail.: congestive heart failure; Pulm. Hypertension: pulmonary hypertension; Gastroesph. reflux: gastroesophageal reflux.

Table II
CLINICAL CHARACTERISTICS OF THE COHORT BY THE DEFINED CLUSTERS, % [95%CI]

	Cluster1 (n=147)	**Cluster2** (n=102)	**Cluster3** (n=69)	*p-value*
Male gender	78 [71-84]	23 [15-32]	87 [76-93]	<0.001
Age				<0.001
20-44	7 [3-12]	25 [17-34]	20 [12-32]	
45-64	39 [31-47]	60 [50-69]	72 [60-82]	
65-90	54 [46-63]	16 [10-25]	7 [3-17)	
Non-repairing sleep	29 [22-37]	75 [65-82]	65 [53-76]	<0.001
Alcohol consumption	86 [79-91]	25 [18-35]	80 [68-88]	<0.001
Obese	48 [40-57]	34 [25-44]	9 [4-19]	<0.001
Witnessed apneas	65 [57-72]	21 [13-30]	84 [73-91]	<0.001
Mallampati				<0.001
1	18 [13-26]	16 [10-25]	49 [37-61]	
2	47 [39-55]	57 [47-67]	23 [14-35]	
3	30 [23-38]	23 [15-32]	26 [17-38]	
4	5 [2-10]	5 [2-12]	1 [0-8]	
Morning headaches	29 [22-37]	73 [63-81]	68 [56-79]	<0.001
Increased NC	76 [68-83]	37 [28-47]	12 [5-22]	<0.001
Driving sleepiness	9 [5-15]	11 [6-19]	22 [13-34]	0.023
Coffee intake	84 [76-89]	77 [68-85]	77 [65-86]	0.350
Cong. Heart Fail.	16 [11-24]	6 [2-13]	6 [2-15]	0.011
Pulm. Hypertension	7 [4-13]	5 [2-12]	1 [0-9]	0.213#
Diabetes	71 [63-78]	12 [6-20]	9 [4-19]	<0.001
Arrhythmias	9 [5-15]	4 [1-10]	4 [1-13]	0.255#
Gastroesph. reflux	82 [75-88]	71 [61-79]	43 [32-56]	<0.001
Insomnia	90 [83-94]	84 [75-90]	83 [71-90]	0.262
OSA	82 [74-87]	43 [33-53]	68 [56-79]	<0.001

Fisher's exact test
CI: confidence interval; BMI: body mass index; NC: neck circumference; Cong. Heart Fail.: congestive heart failure; Pulm. Hypertension: pulmonary hypertension; Gastroesph. reflux: gastroesophageal reflux.

Table III
MODES OF THE CLUSTERED CLINICAL PRESENTATIONS OF OBSTRUCTIVE SLEEP APNEA

Cluster1	**Cluster2**	**Cluster3**
	Demographics	
Male	Female	Male
65-90	45-64	45-64
	Physical Exam	
Normal BMI	Normal BMI	Normal BMI
Increased NC	Normal NC	Normal NC
Mallampati: 2	Mallampati: 2	Mallampati: 1
	Clinical History	
Repairing sleep	Non-repairing sleep	Non-repairing sleep
Alcohol intake	No alcohol intake	Alcohol intake
Witnessed apneas	No witnessed apneas	Witnessed apneas
No morning headaches	Morning headaches	Morning headaches
No driving sleepiness	No driving sleepiness	No driving sleepiness
Coffee intake	Coffee intake	Coffee intake
Insomnia	Insomnia	Insomnia
	Comorbidities	
No cong. heart fail.	No cong. heart fail.	No cong. heart fail.
Diabetes	No diabetes	No diabetes
Gastroesph. reflux	Gastroesph. reflux	No gastroesph. reflux
No arrhytymias	No arrhytymias	No arrhytymias
No pulm. hypertension	No pulm. hypertension	No pulm. hypertension

BMI: body mass index; NC: neck circumference; Cong. Heart Fail.: congestive heart failure; Pulm. Hypertension: pulmonary hypertension; Gastroesph. reflux: gastroesophageal reflux.

4.4. Phenotyping obstructive sleep apnea patients: a first approach to cluster visualization

Proceedings of Studies in Health Technology and Informatics. 255: 75–79, 2018. doi:10.3233/978-1-61499-921-8-75

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Decision Support Systems and Education
J. Mantas et al. (Eds.)

doi:10.3233/978-1-61499-921-8-75

Phenotyping Obstructive Sleep Apnea Patients: A First Approach to Cluster Visualization

Daniela FERREIRA-SANTOS[a,b,1] and Pedro PEREIRA RODRIGUES[a,b]
[a] *CINTESIS – Centre for Health Technology and Services Research*
[b] *MEDCIDS-FMUP – Faculty of Medicine of the University of Porto*

Abstract. The varied phenotypes of obstructive sleep apnea (OSA) poses critical challenges, resulting in missed or delayed diagnosis. In this work, we applied k-modes, aiming to identify groups of OSA patients, based on demographic, physical examination, clinical history, and comorbidities characterization variables (n=41) collected from 318 patients. Missing values were imputed with *k*-nearest neighbours (*k*-NN) and chi-square test was held. Thirteen variables were inserted in cluster analysis, resulting in three clusters. Cluster 1 were middle-aged men, while Cluster 3 were the oldest men and Cluster 2 mainly middle-aged women. Cluster 3 weighted the most, whereas Cluster 1 weighted the least. The same effect was described in increased neck circumference. The percentages of variables driving sleepiness, congestive heart failure, arrhythmias and pulmonary hypertension were very low (<20%) and OSA severity was more common in mild level. Our results suggest that it is possible to phenotype OSA patients in an objective way, as also, different (although not considered innovative) visualizations improve the recognition of this common sleep pathology.

Keywords. Categorical data, cluster analysis, data visualization, obstructive sleep apnea, phenotypes

1. Introduction

In obstructive sleep apnea (OSA), respiratory effort is preserved but ventilation decreases/fades due to partial/total occlusion of the upper airway. A diagnosis is established when a patient has an apnea-hypopnea index (AHI) ≥ 5 with associated symptoms or an AHI ≥ 15 regardless of associated symptoms [1]. OSA affects about 4% of men and at least 2% of women worldwide [1]; nonetheless the diverse phenotypes of OSA have not yet been formally characterized, posing critical challenges to its clinical recognition, resulting in missed or delayed diagnosis [2]. Nowadays, cluster analysis has been used to identify subtypes of patients who are diagnosed with a particular disorder, such as asthma [3]. The *k*-modes algorithm [4] extends the *k*-means paradigm to cluster categorical data by using a simple matching dissimilarity measure for categorical objects [5], modes instead of means for clustering, and a frequency-based method to update modes in the *k*-means fashion clustering process. The dissimilarity measure can be defined by the total mismatches of the corresponding variable categories of the two

[1] Corresponding Author, Rua Dr. Plácido da Costa, s/n, 4200-450, Porto, Portugal; E-mail: danielasantos@med.up.pt

objects: the smaller the number of mismatches, the more similar the two objects are [5]. As we know, the primary goal of data visualization is to communicate information clearly and efficiently via statistical or information graphics. Each visualization intends to help users analysing and reasoning about data and evidence, that is why our intention was to phenotype OSA patients, applying categorical cluster analysis (*k*-modes) to identify groups, based on risk and diagnostic factors, and performed two different clusters visualizations.

2. Methods

In a previous diagnostic test study [6], we included all patients who undertook polysomnography (PSG) at Vila Nova de Gaia/Espinho hospital centre. All administrative records were retrospectively collected between January-May, 2015 and inclusion criteria was patients aged more than 18 years old; patients already diagnosed, patients with severe lung diseases or neurological conditions and pregnant women were excluded. We performed a pre-processing analysis and continuous variables were categorized; *k*-nearest neighbours (*k*-NN) imputation was conducted aiming to preserve all cases and missing data was replaced with a value obtained from related cases from the complete set of records [7]. A literature review helped define the most relevant OSA variables to be collected, in a total of 41 variables: demographic variables (e.g., gender); physical examination (e.g., body mass index (BMI)); clinical history (e.g., snoring); and comorbidities (e.g., stroke). Our dataset portrayed only categorical variables, which rose a question: what is the best way to distinguish high influence variables from low or no influence variables? Variables were selected, after chi-square analysis, if presenting a univariate significant association with the outcome (AHI), considering a 5% significance level. We used R software to perform descriptive and associative analysis (packages *gmodels* and *epitools*) and *k*-modes categorical clustering (package *klaR*) and to create standard barplot (*ggplot2*) and heatmap (*gplots*).

3. Results

From the 318 patients covered, 211 had OSA (66%); from which, 115 (55%) were categorized as mild, 50 (24%) as moderate and 46 (22%) as severe. In total, we had 148 males (70%) with OSA, presenting a mean age of 61 (53-68) years old; the category 20-44 presented a lower percentage of patients in the OSA group, oppositely to categories 45-64 and 65-90 (p value <0.001). Focusing our attention only on the group with the pathology, BMI median value was 30 (27-30) kg/m^2 (p value 0.008); neck circumference (NC) and abdominal circumference (AC) had a mean of 42 (39-44) cm and 107 (100-113) cm. Modified Mallampati categories showed us that our patients have higher total percentages in the inferior levels (1 and 2 in a total of 64%). In those with craniofacial/upper airway abnormalities (CFA) we discovered higher number of patients in the OSA group (195 (92%)) without statistical significance; other variables such as snoring, nocturia, sleep fragmentation, insomnia, drivers, family history, myocardial infarction, arterial hypertension, pacemaker, stroke, renal failure, dyslipidaemia, and hypothyroidism had also no statistical significance. When analysing gasping/choking, we noticed a higher percentage of patients in the normal group (54 (50%)), without statistical significance; other variables present the same characteristic: behaviour

changes, decreased libido, vehicle crashes, coffee, use of sedatives, respiratory alterations, and anxiety/depression. We included a total of 13 variables in the cluster analysis (gender, age, BMI, NC, modified Mallampati, witnessed apneas, non-repairing sleep, morning headaches, driving sleepiness, alcohol, congestive heart failure, arrhythmias, and pulmonary hypertension). Three distinct clusters were identified.

Table 1. Clinical characteristics of the cohort by the defined clusters, % [95%CI]

	Cluster 1 (n=122)	**Cluster 2 (n=44)**	**Cluster 3 (n=55)**	**p-value**	**P(C|F) (c_1, c_2, c_3)**
Male gender	85 [77-91]	20 [10-36]	80 [67-89]	**<0.001**	(.64, .06, .30)
Age				**<0.001***	
20-44	6 [3-13]	11 [4-25]	11 [5-23]		(.39, .28, .33)
45-64	65 [56-74]	57 [41-71]	27 [14-41]		(.65, .22, .33)
65-90	29 [21-38]	32 [19-48]	62 [48-74]		(.40, .18, .42)
Obese	21 [14-29]	43 [29-59]	76 [63-86]	**<0.001**	(.27, .23, .50)
Increased NC	30 [22-40]	77 [62-88]	96 [86-99]	**<0.001**	(.28, .28, .44)
Mallampati				**<0.001***	
1	22 [15-31]	11 [4-25]	36 [24-50]		(.50, .10, .40)
2	48 [39-58]	61 [46-75]	5 [1-16]		(.64, .32, .04)
3	26 [18-35]	23 [12-38]	47 [34-61]		(.45, .15, .40)
4	4 [1-9]	5 [1-17]	11 [5-23]		(.33, .17, .50)
Witnessed apneas	64 [55-73]	30 [17-45]	75 [61-85]	**<0.001**	(.57, .10, .33)
Non-repairing sleep	46 [36-55]	70 [55-83]	29 [18-43]	**<0.001**	(.52, .32, .16)
Morning headaches	45 [35-54]	84 [69-93]	18 [10-31]	**<0.001**	(.52, .38, .10)
Driving sleepiness	13 [7-20]	5 [1-17]	2 [0-11]	**0.045**	(.82, .12, .06)
Alcohol	85 [77-91]	20 [10-36]	91 [79-97]	**<0.001**	(.62, .06, .32)
Cong. Heart Fail.	10 [5-17]	14 [6-28]	18 [10-31]	0.309	(.41, .22, .37)
Arrhythmias	5 [2-12]	14 [6-28]	11 [5-23]	0.153	(.34, .33, .33)
Pulm. hypertension	4 [1-9]	9 [3-23]	13 [6-25]	0.073	(.27, .27, .46)
OSA				0.495	
Mild	56 [47-66]	61 [46-75]	45 [32-59]		(.55, .23, .22)
Moderate	24 [17-33]	20 [10-36]	25 [15-39]		(.54, .18, .28)
Severe	20 [13-28]	18 [9-33]	29 [18-43]		(.48, .17, .35)

*Fisher's exact test; CI: confidence interval; P(C|F): probability of belonging to a cluster given the presence of a factor; NC: neck circumference; Cong. Heart Fail.: congestive heart failure; Pulm. Hypertension: pulmonary hypertension

Table 1 summarizes the clinical characteristics of the total cohort in the selected variables by cluster. Figure 1 and 2 visually synthesizes the information obtained from Clusters 1 to 3. As shown, patients in Cluster 1 were middle-aged men, weighted the least, with non-increased NC and lower percentages in the lowest levels of modified Mallampati. This cluster had the second higher percentage of witnessed apneas, non-repairing sleep, morning headaches, and alcohol consumption. Driving sleepiness had the high percentage in this cluster; oppositely to congestive heart failure, arrhythmias and pulmonary hypertension. These patients presented a severity category of mild, the second higher percentage. Cluster 2 was comprised mostly of middle-aged women, with the second higher percentage of increased NC and BMI. Most of the patients had a level 2 modified Mallampati. This cluster had the lowest percentage of witnessed apneas, alcohol consumption, and very low percentages of comorbidities. Regarding OSA severity they were also mostly categorized in the mild stage. Cluster 3 had the oldest men, weighted the most and presented an increased NC in 96% of the patients. This cluster showed a higher percentage of the modified Mallampati in the higher levels (3 and 4), and a higher number of patients reporting witnessed apneas. Regarding non-repairing sleep, morning headaches, and driving sleepiness, Cluster 3 had the lower percentages, in contrast with alcohol consumption, congestive heart failure, and

pulmonary hypertension. This cluster presented a lower value in the mild severity category and a higher value in the severe level.

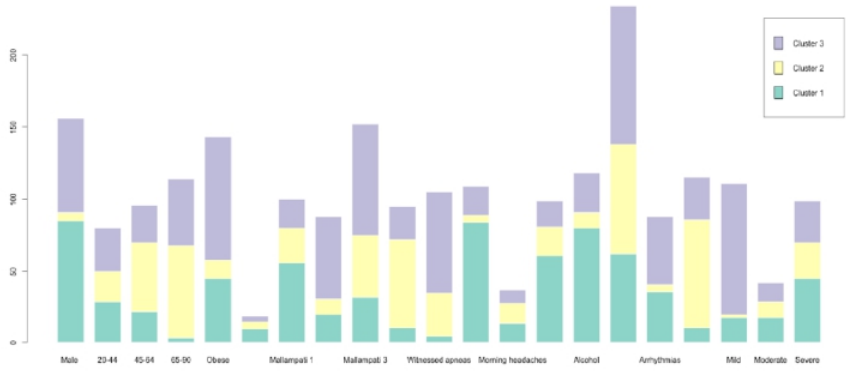


Figure 1. Clinical characteristics of the cohort in Cluster 1, 2 and 3

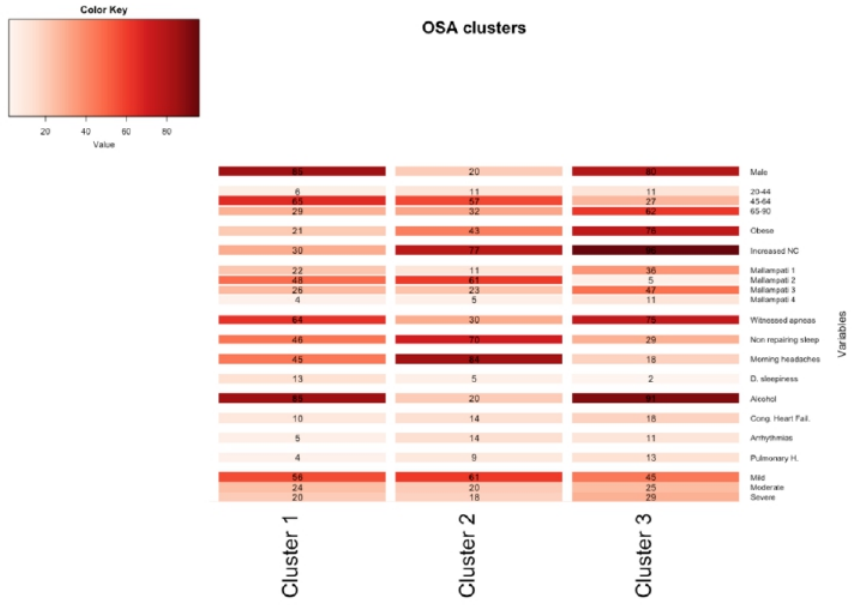


Figure 2. Percentages of each clinical characteristics by cluster

4. Discussion and Conclusion

Understanding different phenotypes in OSA diagnosis is particularly important. Patients in Cluster 1 and 3 were males with different categorized age (middle-aged adults vs.

elderly). Analysing physical examination aspects, we verified that Cluster 1 had lower BMI and NC, oppositely to Cluster 2 and 3. Regarding clinical history, the overall percentage of modified Mallampati was in the lower levels. Cluster 2 reported lower percentage of witnessed apneas, but higher values in non-repairing sleep and morning headaches. The global percentages of driving sleepiness and all selected comorbidities (congestive heart failure, arrhythmias, and pulmonary hypertension) were very low and without statistical significance. Moreover, the percentage of the outcome measure was demonstrated in each cluster; Cluster 2 had 61% [46%-75%] of mild severity, followed by Cluster 1 (56% [47%-66%]) and Cluster 3, with significantly smaller proportion (45% [32%-59%]). To the best of our knowledge, this is the first attempt to phenotype OSA patients using k-modes categorical clustering. Our results suggest that there are different clinical sub-types of OSA, helping focus our attention on a detailed description of OSA diagnosis. The major strengths of this study are newly data analysis, applying standard visualizations to the data, and the clinical cohort representing OSA patients (all levels of severity) that performed PSG. Furthermore, the inclusion of a comprehensive number of risk and diagnostic factors enhances our understanding of OSA diagnosis. This first approach to cluster visualization only intended to summarize the available options and prepare the path for future and more complex visualizations, like circle packing or sunburst.

Acknowledgements

DFS acknowledges Fundação para a Ciência e Tecnologia (FCT) under grant number PD/BD/13553/2018. The work has been developed under the scope of project NORTE-01-0145- FEDER-000016 (NanoSTIMA), financed by the North Portugal Regional Operational Programme (NORTE-2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

[1] American Academy of Sleep Medicine, Sleep-related breathing disorders in adults: recommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine Task Force, *Sleep* **22** (1999), 667–689.

[2] L. Ye, G. Pien, S. Ratcliffe, E. Björnsdottir, E. Arnardottir, A. Pack, B. Benediktsdottir, T. Gislason, The different clinical faces of obstructive sleep apnoea: A cluster analysis, *European Respiratory Journal* **44** (2014), 1600–1607. doi:10.1183/09031936.00032314

[3] W. Moore, D. Meyers, S. Wenzel, W. Teague, H. Li, X. Li, R. D'Agostino, M. Castro, D. Curran-Everett, A. Fitzpatrick, B. Gaston, N. Jarjour, R. Sorkness, W. Calhoun, K. Chung, S. Comhair, R. Dweik, E. Israel, S. Peters, W. Busse, S. Erzurum, E. Bleecker, Identification of asthma phenotypes using cluster analysis in the severe asthma research program, *American Journal of Respiratory and Critical Care Medicine* **181** (2010), 315–323.

[4] Z. Huang, A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining, *Research Issues on Data Mining and Knowledge Discovery* (1997), 1–8. doi:10.1.1.6.4718

[5] L. Kaufman, P. J. Rousseuw, Finding Groups in Data: An Introduction to Cluster Analysis, *Wiley* **47** (1990), 788. doi:10.2307/2532178

[6] D. Ferreira-Santos, P. P. Rodrigues, Improving Diagnosis in Obstructive Sleep Apnea with Clinical Data: A Bayesian Network Approach, *2017 IEEE 30th International Symposium on Computer-Based Medical Systems* (2017), 612–617. doi:10.1109/CBMS.2017.19

[7] D. Ferreira-Santos, M. Monteiro-Soares, P. P. Rodrigues, Impact of Imputing Missing Data in Bayesian Network Structure Learning for Obstructive Sleep Apnea Diagnosis, *Studies in Health Technology and Informatics* **247** (2018), 126–130.

Thirdly, we sought to explore rule-out approaches, as the consequences of untreated OSA and underdiagnosed patients justify screening. As so, we aimed at a high sensitivity tool, as false negatives need to be avoided. Also, we wanted a tool capable of supporting the decision to undergo PSG, based on easily available predictive variables, such as signs or symptoms.

Three studies were conducted and correspond to Modelling and Evaluation – Phase 4 and 5 in CRISP-DM:

4.5. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches

International Journal of Data Science and Analytics. 8: 339–349, 2019. doi:10.1007/s41060-018-0118-x

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

4.6. Enhancing obstructive sleep apnea diagnosis with screening through disease phenotypes: algorithm development and validation

JMIR Medical Informatics. 9: 1–16, 2021. doi:10.2196/25124

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

4.7. Obstructive sleep apnea: a categorical cluster analysis and visualization.

Pulmonology. 16:53, 2021. doi:10.1016/j.pulmoe.2021.10.003

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

The first and second studies intended to establish a new clinical prediction algorithm. The first, besides using Bayesian networks, also developed a risk matrix. The second, also used Bayesian networks, adding the use of clusters in its structure with the creation of a dissimilarity measure based on the odds ratio.

Some differences need to be referred: a) pre-processing for age and body mass index variables; b) outcome definition; c) missing data imputation; d) p-value for significant association with the outcome; e) number of included patients; f) number of included predictive variables; and g) type of Bayesian network classifier used. Both studies included gender, age, and nocturia, and reached a sensitivity above 90%.

The third article is an improvement of article 4.4., by using the previously created OSA clusters and performing external validation, presenting good results.

4.5. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches

International Journal of Data Science and Analytics. 8: 339–349, 2019. doi:10.1007/s41060-018-0118-x

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

International Journal of Data Science and Analytics (2019) 8:339–349
https://doi.org/10.1007/s41060-018-0118-x

REGULAR PAPER

A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches

Daniela Ferreira-Santos[1,2] · Pedro Pereira Rodrigues[1,2]

Received: 16 October 2017 / Accepted: 23 March 2018 / Published online: 13 April 2018

Abstract
In obstructive sleep apnea, respiratory effort is maintained but ventilation decreases/disappears due to upper-airway partial/total occlusion. This condition affects about 4% of men and 2% of women worldwide. This study aimed to define an auxiliary diagnostic method that can support the decision to perform polysomnography, based on risk and diagnostic factors. Our sample performed polysomnography between January and May 2015. Two Bayesian classifiers were used to build the models: Naïve Bayes and Tree Augmented Naïve Bayes, using 38 variables identified by literature review or just a selection of 6. Area under the ROC curve, sensitivity, specificity and predictive values were evaluated using leave-one-out and cross-validation techniques. From a total of 241 patients, only 194 fulfilled the inclusion criteria, 123 (63%) were male, with a mean age of 58 years, 66 (34%) patients had a normal result and 128 (66%) a diagnosis of obstructive sleep apnea. The cross-validated AUCs for each model were: NB38: 69.2%; TAN38: 69.0%; NB6: 74.6% and TAN6: 63.6%. Regarding risk matrix, female gender presented a starting rate of 8%, comparing to 20% in male gender, almost 3 times higher. The high (34%) proportion of normal results confirms the need for a pre-evaluation prior to polysomnography, making the search for a validated model to screen patients with suspicion of obstructive sleep apnea essential, especially at primary care level.

Keywords Obstructive sleep apnea · Risk factors · Diagnosis · Bayesian network · Clinical model · Sensitivity · Specificity

1 Introduction

The substantial medical, social, and economic consequences of untreated obstructive sleep apnea (OSA), the overwhelming number of patients who have escaped clinical detection, and the likelihood of successful treatment strongly justify screening [6]. In this clinical outcome, diagnostic models need to have a high sensitivity, as false negatives should be avoided, to prevent excluding patients with moderate or severe OSA from performing polysomnography (PSG) [24,34], the standard test for OSA final diagnosis.

This study aimed to define auxiliary diagnostic methods that can support the decision to perform PSG, based on risk and diagnostic factors by means of interactive models or risk matrix. The secondary objectives were to describe the population; develop and validate a Bayesian network-based risk matrix in the study population; optimize the need to perform PSG and produce a Bayesian network model for daily use in the primary care setting.

This paper is organized as follows. The background section exposes the related work on the theme. Following section presents the research methodology. Section 4 gives an overview of the achieved results, and Sects. 5 and 6 interpret the results of the work and provide the main findings and recommendations for the work.

2 Background

Apnea is defined as the complete cessation of airflow for at least 10 s, while a hypopnea is a reduction in airflow (30–50%) that is followed by an awakening or a decrease in oxyhemoglobin saturation (3–4%) [6,22]. There are 3 types

✉ Daniela Ferreira-Santos
danielasantos@med.up.pt

[1] Center for Health Technology and Services Research (CINTESIS), Faculty of Medicine of the University of Porto, Porto, Portugal

[2] Community Medicine, Information and Health Decision Sciences (MEDCIDS) Department, Faculty of Medicine of the University of Porto, Rua Dr. Placido da Costa, s/n, 4200-450 Porto, Portugal

of apneas: central, mixed and obstructive. Central sleep apnea is a reduction in the respiratory effort resulting in reduced or absent ventilation, while mixed apnea begins with central apneas that leads to obstructive events [5,22]. In OSA, respiratory effort is maintained but ventilation decreases or disappears because of partial/total occlusion of the upper airway [6,22,23,28,32].

OSA severity is assessed with apnea–hypopnea index (AHI), obtained through PSG, which is the number of apneas and hypopneas per hour of sleep [22]. Recommendations from the American Academy of Sleep Medicine state that OSA is present when AHI ≥ 5. It can be classified as mild (AHI: 5–15), moderate (AHI: 15–30), or severe (AHI $\geq$ 30) [6,7,13,22]. Approximately 30% of the general public is affected by a significant sleep problem, often of long standing [39]. OSA affects about 4% of men and at least 2% of women worldwide [5,7,13,18,39]. The signs, symptoms, and consequences of OSA are a direct result of upper-airway repetitive collapse. This leads to sleep fragmentation, hypoxemia, hypercapnia, marked swings in intrathoracic pressure, and increased sympathetic activity.

Reported risk factors for developing OSA include different groups of variables, such as demographic, clinical history, comorbidities and even factors collected during the consultation, for example, male gender [1,10,16,20–22,27,30,36,42–45], aging [1,16,20–22,27,30,36,42–45], obesity [1,10,12,16,20–22,30,36,42,43,45], history of smoking [1,10,12,22,30,45], increased neck [1,10,16,20–22,27,30,36,43,45] and abdominal [30,45] circumferences, arterial [1,12,16,20,22,30,36,43–45] and pulmonary [12,30] hypertension, atrial fibrillation [30,36], stroke [1,22,30,45], myocardial infarction [1,30,45], and high-risk driving populations (such as truck drivers) [10,12,30,42].

Diagnostic criteria for OSA are based on clinical signs and symptoms determined during a sleep consultation, which includes a sleep oriented history, physical examination, and findings identified in an objective exam [13,28]. It should also include an evaluation, for example, of snoring, witnessed apneas, gasping/choking episodes, daytime sleepiness severity with the Epworth Sleepiness Scale (ESS), nocturia, and morning headaches [22]. The diagnostic methods available are PSG and home testing with portable monitors (tends to underestimate the severity) [3,6,7,13,18,22,39]. PSG is time consuming, labor intensive, limited to urban areas, costly and faces long waiting lists [18,22], so many studies have been trying to tackle the problem that comes with it. Rodsutti et al. [35] conducted a study to develop and validate a decision rule (based on risk factors) that would allow prioritization on the waiting list, using univariate analysis and multiple logistic regression, achieving a scoring scheme or color-coded tables for easy clinical application. Sun et al. [40] used three questionnaires to improve sensitivity and specificity for discrimination of moderate to severe OSA, based on a genetic algorithm. Montoya et al. [37] based their work on several epidemiological and clinical variables, sought to find alternatives to PSG, using logistic regression analysis and multivariate logistic regression to determine the best model for distinguishing OSA patients from the healthy ones. In the end, the work produced a algorithm to calculate the prediction of AHI in a new patient.

All the previous have focused on traditional simpler methods for decision support. Nowadays, prediction models are generated by artificial intelligence, using decision trees, neural networks, support vector machines, and Bayesian networks [9]. All should have good performance, good ability to handle data entry errors or omissions, transparency of diagnostic knowledge, ability to explain decisions, and the algorithms should be able to reduce the number of tests needed for making a reliable diagnosis [26]. While searching, we found studies that attempted to apply these new techniques in OSA. One study [14] tackled the tedious and time-consuming task of analyzing PSG records, automatizing both the detection and classification of sleep apneas, through analysis of wavelets and Bayesian neural networks. The other [19] classified patients with possible diagnosis of OSA into groups according to the severity of the disease using a decision tree producing algorithm based on nonlinear analysis of three respiratory signals instead of full PSG. However, none of the found approaches addressed only clinical and demographic variables that could be used earlier in the healthcare process flow, as they require diagnostic data from PSG.

In fact, technologies exist to tackle the problem at later stages, e.g., disease management [31], but there is a clear lack of solutions for the early diagnosis of OSA. Bayesian networks have been used in several clinical domains, especially given their balance between accurate predictions and their specific interpretability in the clinical domain, resembling the human reasoning in a probabilistic way, along with their ability to produce predictions with missing values, presenting high performance in areas like pneumonia and breast cancer [26].

3 Methods

This study was designed according to the Standard for Reporting Diagnostic accuracy studies (STARD) list [4], updated in 2015. Its guiding principle was to select items that, when reported, would help readers to judge the potential for bias in the study, to appraise the applicability of the study findings and the validity of conclusions and recommendations. STARD guidelines have been generally used for adequately validating diagnostic tools.

3.1 Patients

We have included all patients referred to perform a polysomnography at Vila Nova de Gaia/Espinho hospital center sleep laboratory, between January and May 2015. Inclusion criteria were defined as follows: patients aged more than 18 years old and with suspicion of OSA. On the other hand, patients already diagnosed (performing therapeutic studies), patients with suspicion of another sleep disease, patients with severe lung diseases or neurological conditions, and pregnant women were excluded. In case of duplicate exams, the best sleep efficiency was selected.

3.2 Variables

A literature review on PubMed (April 19, 2015) was performed to define the most relevant variables to be collected from medical and/or sleep laboratory records. The search contained "risk factors", "sleep apnea, obstructive" and "diagnosis" as Mesh terms, obtaining 1397 articles, from which 48 were used for variable definition.[1] A total of 38 predictive variables were collected: **demographic variables:** gender, race and age; **physical examination:** body mass index (BMI), neck (NC) and abdominal circumferences (AC) and craniofacial and upper-airway abnormalities (CFA); **clinical history:** daytime sleepiness, snoring, witnessed apneas, choking/gasping, refreshing sleep, restless sleep, humor alterations, concentration decrease, morning headaches, decreased libido, motor vehicle crashes, drivers, nocturia, alcohol, smoking, coffee, sedatives, family history/genetics and ESS; **comorbidities:** atrial fibrillation (AF), stroke, myocardial infarction (MI), arterial and pulmonary hypertension (PHT), congestive heart failure (CHF), diabetes, dyslipidemia, renal failure, hypothyroidism, gastroesophageal reflux, depression and anxiety.

3.3 Data collection and preprocessing

Medical and/or sleep laboratory records were retrospectively collected between January 1, 2015 to May 31, 2015. Clinical information of each patient (39 variables) was extracted from the central clinical records along with the sleep laboratory data, making all the clinical files available. We screened for missing information but, although we had all the records, some predictive variables were not present or described. This rose a problem in the construction of the Bayesian network models, creating the need to make an assumption: when learning the networks' structure (and only then), if the variable was not present in the records, we assumed it was absent hence imputing category "No". Our models development and validation were performed in this sample. The outcome measure was the clinical diagnosis, obtained from AHI, categorized into normal (AHI < 5) or OSA (AHI≥ 5).

We performed a preprocessing of the data and all the continuous variables were categorized:

- BMI ($< 30\,\mathrm{Kg/m^2}$: normal, BMI $\geq 30\,\mathrm{Kg/m^2}$: obese);
- female NC (≤ 37 cm: normal, > 37 cm: increased); male NC (≤ 42 cm: normal, > 42 cm: increased);
- female AC (≤ 80 cm: normal, > 80 cm: increased); male AC (≤ 94 cm: normal, > 94 cm: increased);
- age (< 40, 40–54, 55-69, ≥ 70 years);
- smoking (yes, no, ex-smoker);
- ESS (0–10: normal, 11–24: daytime sleepiness);
- AHI (0–4: normal, 5–14: mild, 15–29: moderate, ≥ 30: severe).

3.4 Bayesian networks

Generally, a Bayesian network represents a joint distribution of one set of variables, specifying the assumption of independence between them with the interdependence between variables being represented by a directed acyclic graph [29]. Each variable is represented by a node in the graph and is dependent on the set of variables represented by its ascendant nodes. This dependence is represented by a conditional probability table that describes the probability distribution of each variable, given their ascendant variables. Naïve Bayes (NB), which assumes conditional independence among factors, and Tree Augmented Naïve Bayes (TAN)[17], which allows for an optional dependence for each factor, were the Bayesian network classifiers used to build our models. They were chosen given their previous results in other clinical domains [11,25]. Four models were evaluated and compared, differing on the classifier (NB and TAN) and the number of predictive factors (38 or 6).

3.5 Statistical analysis

Diagnostic models were defined using a Bayesian network built over the set of available variables and area under the curve (AUC) was performed. Model parameters were validated by comparing the AUC in the derivation cohort with those calculated from a leave-one-out and a 10 times twofold stratified cross-validation (for variability assessment with independent training and testing). We used R software to: (a) perform descriptive and associative analysis, using packages *gmodels* [15], *epitools* [2] and *MASS* (modern applied statistics with S) [41]; (b) learn and validate the models, using packages *bnlearn* [38] and *gRain*; and (c) analyze AUC, using package *pROC* [33]. *SamIam* [8] software was used to visually consult the conditional probabilities, given the outcome.

[1] Full review description and references used in this phase are not shown for space purposes but can be provided upon request.

The application of the selected models generated in this work can be visualized by means of (a) Bayesian inference (TAN6) and (b) appropriately defined risk matrix (NB6). The models with selected variables (NB6 & TAN6) were built after performing Chi-square test (unless otherwise specified) to all the 38 variables. The selected variables were chosen as those with a univariate significant association with the outcome, considering a 5% significance level, or a 10% significance level if at least 5 patients were observed in each outcome category (Normal or OSA), and for which no quality problems were suspected.

In order to choose which variables should be included in the risk matrix, we evaluated the variables with statistical significance, the odds ratios obtained in the multivariate logistic regression and those which had higher clinical relevance were chosen as factor. Each cell of the matrix represents the marginal posterior outcome probability estimate for that subgroup of patients. The precision of such estimates is given by a 95% credible interval, computed from a Monte Carlo simulation of one hundred thousand samples from the derived joint probability model (i.e., the NB6). The risk values in each cell of the matrix represent the expected risk for a patient in that subgroup, while the credible interval encloses 95% of risk estimates for patients in that subgroup (i.e., only 5% of patients in that subgroup have a risk estimate outside the credible interval). We believe that this approach is more interesting from the clinical point of view than the usual one, in which a confidence interval (CI) of the expected risk of all patients in each subgroup is computed and presented. To assess the discriminative ability of the risk matrix for the outcome, specific cutoff values were chosen after assessing the AUC of the derivation cohort, aiming at a sensitivity of 95%, to allow a rule-out approach aiming to avoid false negatives.

This was approved by the Ethics Commission of Vila Nova de Gaia/Espinho hospital center, following the Declaration of Helsinki.

4 Results

4.1 Population characteristics and analyzed outcome

We considered for inclusion 241 patients, being 47 excluded for several reasons: 7 duplicates; 8 missing clinical file; 19 under eighteen years old, and 13 therapeutic studies. In the 194 patients included, 123 (63%) were male (mean age 58 years); sixty-six patients (34%) had a normal result with a mean age of 50 years. Of the 128 patients with OSA (66%) (mean age 62 years), 63 (33%) were categorized as mild, 32 (16%) as moderate, and 33 (17%) as severe.

Table 1 describes the dataset obtained from the medical and/or sleep laboratory records.

In those with craniofacial and upper-airway abnormalities, described as part of physical examination, the percentage of patients in OSA group was higher than in the group without OSA (97 vs. 79%, p value ≤ 0.05). Other variables describe the same effect, such as witnessed apneas, nocturia, alcohol consumption, atrial fibrillation, stroke, myocardial infarction and driver. The opposite effect emerges in daytime sleepiness (94 vs. 77% in the OSA group, p value ≤ 0.05), concentration decreased and ESS, which presented a contradiction to the literature and the inherent meaning of the variables.

4.2 Bayesian diagnostic models

In order to unveil the interdependent relationships between the analyzed outcome (OSA) and the 38 variables considered, Bayesian network-based models were built. R software was used to learn the probabilities obtained from the dataset described in Table 1, using the aforementioned imputation strategy for structure learning. This resulted in four models: NB38 and TAN38, NB6 and TAN6 with TAN structures presented in Figs. 1 and 2 and NB structures presented in supplementary figures S1 and S2. The ROC curves of each model are presented in Fig. 3, demonstrating in sample AUC of 82% [76–88%] for NB38, 90% [86–94%] for TAN38, 79% [73–86%] for NB6, and 79% [73–86%] for TAN6.

Clinically speaking, interpreting TAN38 is a hard task, given the time to apply it in a primary care consultation. Thus, TAN6 contains only relevant variables, that is, gender, witnessed apneas, age, nocturia, CFA, and NC. Although AC, concentration decrease, ESS, and daytime sleepiness would be eligible for the final model, categorized AC lost significance, while the remaining three presented contradictory results raising data collection quality suspicion.

4.3 Model validation

The Bayesian models were validated following an internal approach, which consisted of two different tests (leave-one-out and 10 times twofold stratified cross-validation). ROC analysis was performed independently for the derivation cohort and the respective AUC, along with their 95% CIs, illustrated in Fig. 3. The AUC values of the leave-one-out nearly overlapped those of the cross-validation. Furthermore, the overall discrimination power was high for both strategies, using leave-one-out. Based on NB6, the best model according to the validation results (Table 2), the following cutoff was determined: values above 32.0% were considered to be a positive test result, i.e., outcome presence. Table 2 presents the performance of the chosen cutoffs for all models, presenting sensitivities of 90.0% [88.2–91.8%] for NB38, 81.9% [77.7–86.0%] for TAN38, 94.1% [92.9–95.4%] for NB6, and 90.2% [88.0–92.4%] for TAN6.

Table 1 Descriptive analysis of the derivation cohort before imputation (absolute and relative frequencies are presented and p values result of Chi-square test, unless otherwise specified). Variables with high number of missing values (e.g., snoring) are described with two proportions: the proportion in the real dataset, and the proportion in the imputed dataset

	Total ($n = 194$)	OSA ($n = 128$)	Normal ($n = 66$)	Crude OR [CI 95%]	p value	Adjusted OR [CI 95%]
Male gender	123 (63)	92 (72)	31 (47)	2.72 [1.55–5.28]	0.001	3.72 [1.81–7.90]
Race						
Caucasian	191 (99)	126 (98)	65 (99)		NA	
African	3 (2)	2 (2)	1 (2)	0.51 [0.11–6.68]		
Age						
Mean (IQR)	58 (50–67)	62 (54–70)	50 (41–61)	1.08 [1.05–1.11]	< 0.001*	
< 40	19 (10)	5 (4)	14 (21)	(Ref)	< 0.001	(Ref)
40–54	56 (29)	29 (23)	27 (41)	2.42 [0.93–8.59]		2.26 [0.66–8.57]
55–69	79 (41)	59 (46)	20 (30)	6.56 [2.54–23.04]		7.18 [2.17–27.13]
≥ 70	40 (21)	35 (27)	5 (8)	13.61 [4.50–64.37]		17.83 [4.18–91.19]
BMI						
Median (IQR)	29 (26–34)	30 (26–34)	29 (26–33)	1.02 [0.98–1.08]	0.249^{+}	
Obese	91 (47)	64 (50)	27 (41)	1.37 [0.79–2.60]	0.229	
NC						
Median (IQR)	42 (39–45)	42 (40–46)	40 (37–44)	1.13 [1.05–1.22]	0.003^{+}	
Increased	107 (55)	77 (60)	30 (46)	1.72 [0.99–3.27]	0.051	2.01 [1.00–4.09]
AC						
Median (IQR)	107 (99–113)	108 (100–114)	105 (97–111)	1.02 [1.00–1.05]	0.031^{+}	
Increased	180 (93)	120 (94)	60 (91)	1.31 [0.52–4.43]	$0.560^{\#}$	
Snoring	175 (90)–(98)	114 (97)	61 (100)	0.00 [0.01–5.23]	$0.552^{\#}$	
Witnessed apneas	104 (54)–(70)	72 (75)	32 (60)	1.83 [0.96–3.99]	0.063	1.22 [0.59–2.54]
Gasping/choking	76 (39)–(59)	49 (59)	27 (60)	0.90 [0.46–2.01]	0.916	
Vehicle crashes	6 (3)–(8)	2 (4)	4 (16)	0.16 [0.04–1.20]	$0.083^{\#}$	
Refreshing sleep	45 (23)–(29)	31 (32)	14 (25)	1.30 [0.67–2.89]	0.363	
Humor alterations	2 (1)	1 (100)	1 (100)	NA	NA	
Nocturia	70 (36)–(63)	53 (69)	17 (50)	2.00 [0.96–4.94]	0.058	1.51 [0.71–3.27]
Restless sleep	76 (39)–(77)	48 (75)	28 (80)	0.68 [0.29–2.06]	0.573	
Decreased libido	18 (9)–(95)	13 (100)	5 (83)	2.17 [0.26–210.02]	$0.316^{\#}$	
Morning headaches	66 (34)–(59)	44 (61)	22 (55)	1.19 [0.59–2.78]	0.529	
Alcohol consumption	96 (50)–(61)	65 (66)	31 (53)	1.63 [0.89–3.30]	0.102	
Smoking						
Yes	32 (17)	20 (17)	12 (19)	0.78 [0.39–1.85]	0.525	
Ex-smoker	56 (29)–(30)	40 (33)	16 (25)			
Sedatives	50 (26)–(93)	27 (96)	23 (89)	1.69 [0.37–19.98]	$0.342^{\#}$	
Driver	16 (8)–(9)	13 (12)	3 (5)	1.96 [0.70–7.97]	0.130	
Coffee	120 (62)–(87)	76 (88)	44 (85)	1.23 [0.52–3.70]	0.525	
Daytime sleepiness	114 (59)–(84)	65 (77)	49 (94)	0.20 [0.07–0.79]	0.010	
Genetics/family	2 (1)–(67)	0 (0)	2 (100)	0.00 [0.00–5.49]	$0.333^{\#}$	
ESS						
Median (IQR)	11 (5–15)	10 (3–13)	13 (7–16)	0.91 [0.87–0.96]	$< 0.001^{+}$	
Daytime sleepiness	102 (53)	57 (45)	45 (68)	0.36 [0.20–0.71]	0.002	
Concent. Decrease	42 (22)–(54)	20 (43)	22 (71)	0.28 [0.12–0.81]	0.014	
CFA	43 (22)–(92)	32 (97)	11 (79)	4.00 [0.87–50.10]	$0.073^{\#}$	2.03 [0.86–5.12]

Table 1 continued

	Total ($n = 194$)	OSA ($n = 128$)	Normal ($n = 66$)	Crude OR [CI 95%]	p value	Adjusted OR [CI 95%]
Atrial fibrillation	15 (8)–(22)	15 (31)	0 (0)	8.82 [1.08–334.34]	0.003#	
Stroke	17 (9)–(23)	14 (29)	3 (12)	2.14 [0.73–9.46]	0.109	
MI	14 (7)–(20)	12 (26)	2 (9)	2.22 [0.67–12.47]	0.197#	
Pulm. Hypertension	1 (1)–(25)	1 (50)	0 (0)	1.00 [0.11–220.62]	NA	
Cong. Heart Fail.	20 (10)–(25)	16 (28)	4 (17)	1.52 [0.56–5.83]	0.277	
Diabetes	47 (24)–(72)	33 (72)	14 (74)	0.79 [0.29–3.02]	0.873	
Dyslipidemia	99 (51)–(93)	76 (95)	23 (89)	1.90 [0.58–11.03]	0.359#	
Renal failure	3 (2)–(30)	2 (33)	1 (25)	0.60 [0.11–15.16]	NA	
Hypothyroidism	2 (1)–(33)	1 (50)	1 (25)	0.75 [0.12–45.17]	NA	
Gastroesoph. Reflux	3 (2)–(21)	2 (22)	1 (20)	0.50 [0.10–10.35]	NA	
Hypertension	123 (63)–(94)	91 (96)	32 (89)	2.21 [0.72–11.04]	0.214#	
Depression	53 (27)–(95)	28 (97)	25 (93)	1.08 [0.23–15.11]	0.605#	
AHI						
Mild		63 (33)				
Moderate		32 (16)				
Severe		33 (17)				

OSA obstructive sleep apnea, *OR* odds ratio, *CI* confidence interval, *IQR* interquartile range, *BMI* body mass index, *NC* neck circumference, *AC* abdominal circumference, *ESS* Epworth somnolence scale, *Concent. Decrease* concentration decrease, *CFA* craniofacial and upper-airway abnormalities, *MI* myocardial infarction, *Pulm. Hypertension* pulmonary hypertension, *Cong. Heart Fail.* congestive heart failure, *AHI* apnea–hypopnea index

#Fisher's exact test, *Independent T test, +Mann–Whitney *U* test

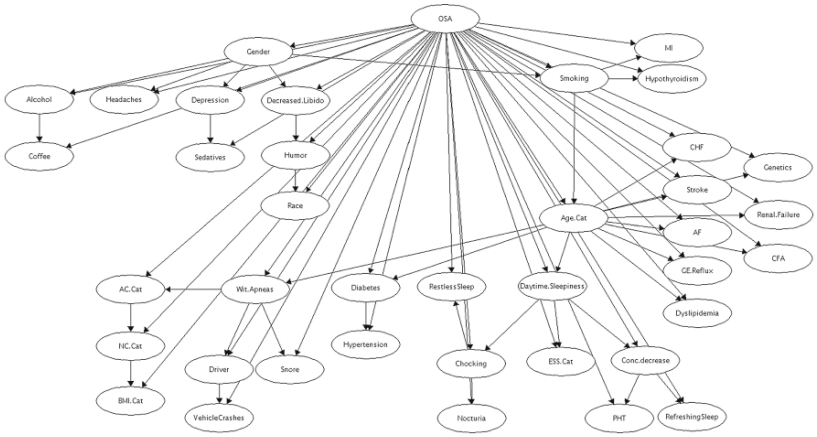


Fig. 1 Tree Augmented Naïve Bayes with 38 variables (TAN38) representing the relationship between the outcome (OSA) and each variable, and relationships between predictive factors

4.4 Risk matrix

To determine which variables should be included in the risk matrix, a multivariate logistic regression was carried out using all the independent variables considered. Those that were clinically relevant were selected and included in the final matrix as risk factors for OSA: gender (OR 3.72 [1.81–7.90%]), age (< 40: OR ref, 40–54: OR 2.26 [0.66–8.57%], 55–69: OR 7.18 [2.17–27.13%], ≥ 70 years: OR 17.83 [4.18–91.19%]), neck circumference (OR 2.01 [1.00–

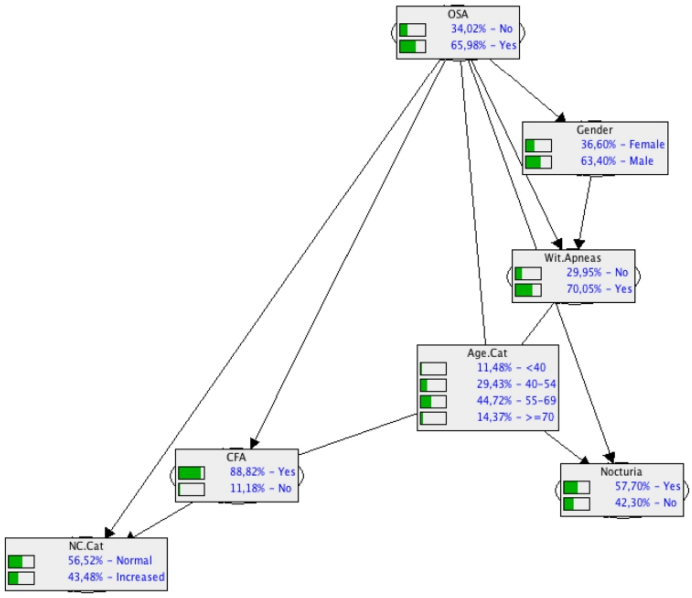


Fig. 2 Tree Augmented Naïve Bayes with 6 variables (TAN6) representing the relationship between the outcome (OSA) and each variable, and relationships between predictive factors. The bars within each variable represent the prior marginal probabilities for each variable's category. Arrows represent association between variables, but do not convey any causal relationship, the association between the outcome and each of the remaining variables being imposed on the model

4.09%]), and witnessed apneas (OR 1.22 [0.59–2.54%]). The risk matrix (Table 3) conveys the risk of having OSA stratified by relevant factors. The highest value (95%) was observed for male patients, age above seventy, increased neck circumference and witnessed apneas. The lowest value (8%) was observed in female patients, age under forty, with normal neck circumference, and no witnessed apneas.

5 Discussion

To our knowledge, no one has attempted to analyze risk and diagnostic factors for OSA the way this study does. Our study is one of the first to build and validate risk models for OSA based solely on clinical and demographic variables, which have the key advantage of being easily available and quickly acquired. We focused on the most important risk and diagnostic factors, being aware of the clinical definition of OSA. We obtained a proportion of normal results in 66 patients (34%), revealing a large number of unnecessary exams that are performed every day in Vila Nova de Gaia/Espinho hospital center, with the possibility of this result being generalized to the different hospital centers in the country.

Male gender was more prevalent (63%), agreeing with the literature [1,10,16,20–22,27,30,36,42–45]. Possible explanations are the higher prevalence of craniofacial and upper-airway abnormalities (21%), and also snoring (90%). Also, higher age (> 55 years old) was linking to a higher OSA prevalence. The strata of 55–69 and more than 70 years old had a total of 94 patients (73%), followed by 29 patients (23%) in the 40–54 years old. In previous work, age was not included due to the fact that it was dominating the remaining dependencies, so we have been working in a new approach for dealing with this variable. We performed a new analysis with consequent alterations in the preprocessing levels, making the model more robust and clinically accurate, as in the literature this variable is described as one of the most important risk factors for OSA.

During the physical examination, body mass index, neck, and abdominal circumferences were collected. Regarding obesity, described as an important risk factor for OSA, we found that the percentage of patients, in the pathology group, having a normal BMI is equal to the percentage of the patients with obesity (50%). This could explain why in our study obesity is not included in the selected variables. Analyzing neck and abdominal circumferences, unadjusted for gender, we saw higher percentages of increased level, 60 vs. 46% in NC and 94 vs. 91% in AC. In some studies, neck circumference is described as one of the most important risk factors, but not always considered by several medical doctors, creating the need to perform more studies with this variable. Craniofacial and upper-airway abnormalities are

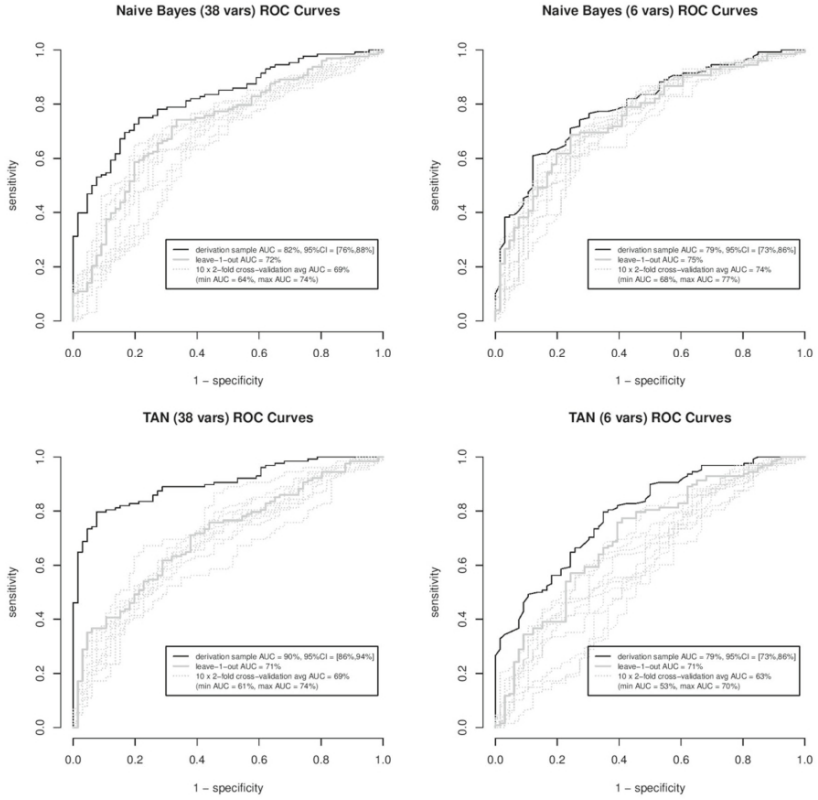


Fig. 3 Receiver operating characteristics analyses and area under the curve values for NB38, TAN38, NB6 and TAN6, as well as for the internal validation procedures

Table 2 Validity assessment [%] estimated from 10 times twofold cross-validation

Model	Cut	Accuracy	Sensitivity	Specificity	Precision(+)	Precision(−)	AUC
NB38	14.7	67.1 [65.6–68.7]	90.0 [88.2–91.8]	22.9 [19.0–26.8]	69.4 [68.4–70.5]	53.9 [48.3–59.5]	69.2 [66.6–71.8]
TAN38	67.8	66.9 [64.5–69.3]	81.9 [77.7–86.0]	37.9 [31.3–44.5]	72.1 [70.2–74.0]	53.1 [48.1–58.1]	69.0 [66.4–71.6]
NB6	32.0	70.2 [69.2–71.3]	94.1 [92.9–95.4]	23.8 [19.9–27.7]	70.6 [69.7–71.6]	68.4 [64.2–72.5]	74.6 [72.9–76.2]
TAN6	38.5	67.5 [66.3–68.6]	90.2 [88.0–92.4]	23.5 [20.0–27.0]	69.6 [68.8–70.4]	56.6 [52.7–60.5]	63.6 [60.9–66.3]

NB38, NB6: Naïve Bayes with 38 or 6 variables; TAN38, TAN6: Tree Augmented Naïve Bayes with 38 or 6 variables

important in the pathogenesis of OSA, particularly in non-obese patients, and our study demonstrated it. In patients with normal weight, we found that 26 (96%) patients had craniofacial and upper-airway abnormalities. The differences in craniofacial morphology may explain some of the variation in risk of OSA. Only 3% of OSA patients reported not snoring during the night, while 100% of the patients in the group without OSA report snoring. This might explain why snoring cannot continue to be an important risk factor. Depression or anxiety is highly related to sleep problems, becoming nowadays one of the most studied factors, but only 53 (27%) patients have the diagnosis.

Table 3 Risk matrix showing the probability [%] of having obstructive sleep apnea [95% credible interval] (color table online)

Age		<40		40-54		55-69		>=70	
NC		Normal	Increased	Normal	Increased	Normal	Increased	Normal	Increased
♂	👁	33 [4-46]	47 [7-61]	60 [12-72]	73 [35-82]	80 [45-87]	88 [59-93]	91 [66-94]	95 [93-97]
	🙈	20 [2-30]	31 [4-44]	43 [6-56]	58 [21-70]	68 [29-78]	79 [43-87]	83 [49-89]	90 [87-94]
♀	👁	15 [2-23]	24 [3-35]	34 [4-47]	48 [8-61]	59 [21-71]	72 [34-81]	77 [40-85]	86 [54-91]
	🙈	8 [1-13]	14 [1-21]	21 [2-31]	32 [4-45]	42 [6-55]	57 [20-69]	63 [13-74]	76 [38-90]

♂, male; ♀, female; 👁, apneas witnessed; 🙈, apneas not witnessed; NC, neck circumference; low risk (< 25%—green); medium risk (between 25% and 50%—yellow); high risk (between 50% and 75%—orange); very high risk (>75%—red)

The clinical definition of OSA includes several diagnostic factors, such as daytime sleepiness. One way to confirm its presence is with ESS. Even though this questionnaire is not specific for OSA, it has been often used in Vila Nova de Gaia/Espinho hospital center. When we analyzed the OSA group, we discovered a median of 10 (3–13), demonstrating a normal result for this group (cutoff in 10 points). However, when we analyzed the median in patients without OSA we found a higher value (13 (7–16)). This highlights the possibility that ESS is not adjusted for the pathology. Another common diagnostic factor is witnessed apneas. Men have a percentage of 60% against 43% in women. This can be explained by female bed partners having a lower threshold for symptom perceptions and reporting it less than male bed partners.

For visual inference Bayesian network we chose TAN6. Even though NB6 presents a higher value of sensitivity (94.1% [92.9–95.4%]), in clinical settings, the variables presented in our study are not independent, so we should assume the relation among variables, especially when not all variables are available to the physician. The TAN6 has a sensitivity of 90.2% [88.0–92.4%], meaning that 90% of OSA patients would perform PSG, rejecting 10% of patients who would be referred to follow-up consultation. Further studies are needed to optimize the model and also to externally validate it preferable with a prospective validation cohort.

Visually, the network model TAN6 is very intuitive. It is a simple and friendly model that can be easily accessed and filled in, helping the diagnosis of this pathology in the primary care centers or other facilities that have a gap in its diagnosis. The substructures that raised from this model were also interesting. Gender is related to witnessed apneas, aspect described by physicians and specialists; likewise, witnessed apneas are associated with age. Additionally, age is associated with nocturia (aging increases the need to use the bathroom more often) and with craniofacial and upper-airway abnormalities (possibility of accidents in the adult life and also obesity development). Another common relation described in care is the association between craniofacial and upper-airway abnormalities and increased neck circumference, which was also present in our model.

The final risk results were arranged into a color-coded and user-friendly matrix that constitutes a preliminary but useful tool that can be used by primary care physicians and others in the diagnostic decision-making process. In the case of female gender, the rate started at 8% (under than 40 years, normal neck circumference and without witnessed apneas) to 86% (higher than 70 years old, neck circumference increased and witnessed apneas). In the case of male gender, the initial rate is almost 3 times higher than for female gender. It started at 20% (under than 40 years, normal neck circumference and without witnessed apneas) to 95% (more than 70 years, with neck circumference increased and witnessed apneas). Low risk of having OSA is related with the female gender under 40 years. In contrast, high risk is prevalent in the male gender above 50 years old.

One limitation to fully use the set of predictive variables was the lack of representativeness of some factors, such as vehicle crashes, humor alterations, decreased libido, pulmonary hypertension, congestive heart failure, renal failure, hypothyroidism, gastroesophageal reflux and genetics, which might have led to a bias in the 38 variables models. Also, we acknowledge that the retrospective nature of the derivation cohort is a limitation to the study and, most of all, that the low specificity of the resulting models make them somewhat limited. However, we manage to provide a sensitive tool (sensitivity > 90%) which nonetheless prevents 1 out of 5 healthy individuals from unnecessarily performing PSG exam (specificity > 20%), improving from current clinical practice. This improvement could also result in financial benefits for the healthcare system. Using a simple back of the envelope calculation, the relatively small district hospital where our cohort was recruited could potentially perform 466 PSG (yearly extrapolation of 194 PSG in 5 months). Given a 20% specificity of our model, from the 158 normal PSG (yearly extrapolation of 66 normal PSG in 5 months), 32 would not have been referred to perform PSG. Considering Portuguese mandamus number 207/2017, each PSG is priced at €939,14 leading to a total saving of €30.052,48. More-

over, 31 patients with OSA (yearly extrapolation of 10% of 128 OSA patients in 5 months) would also have their PSG delayed by our approach, representing additional savings of €29.113,34, for a grand total of €59.165,82. Certainly, these 31 patients should have performed PSG, thus diminishing the clinical benefit of our proposal. However, limited inspection on the data supports our belief that these would have been mild or moderate OSA patients, who could perhaps be rescheduled for follow-up consultation and PSG in the subsequent months without serious harm. Either way, given the current wait list situation, the 63 PSG vacancies would most likely be filled with patients, with the expected savings corresponding to one and a half months worth of PSG work. Nevertheless, we defer this discussion to future work, where we will perform a cost-effectiveness analysis (using Monte Carlo simulations) to assess the expected overall impact of our approach.

6 Conclusion

Our study added important knowledge to the state of art regarding OSA. Moreover, this knowledge is delivered in the form of an intuitive and user-friendly model (TAN6), which can be used by any physician, and in the form of a risk matrix that physicians can use to quantify the probability for having OSA. We present as main risk factors: gender, age, neck circumference, and craniofacial and upper-airway abnormalities, and as diagnostic factors: witnessed apneas and nocturia. Naïve Bayes was considered a better classifier, while TAN showed advantages for visual inference, proving the great advantages of Bayesian network models (dealing with missing information and simple graphical representation, showing not only the probabilities given the patient characteristics but also the relationship between variables).

According to our cross-validated evaluation, we expected around 30% of false positives which, although unwanted, is nonetheless an improvement if we compare with all patients at risk being referred to sleep consultation and polysomnography. Nonetheless, we were able to rule out 25% of healthy patients, which, in our understanding, would alleviate health services by reducing the burden of unneeded consultations and the wait lists for polysomnography, while identifying more than 90% of patients with OSA.

Portugal, like many other countries, does not have a validated method to screen patients with suspicion of OSA, before performing polysomnography, so we think that our models (TAN6 and risk matrix) consist in valid methods. Current work is focused on bringing the model toward primary care facilities.

Acknowledgements The authors would like to thank the sleep laboratory team of Vila Nova de Gaia/Espinho hospital center, especially Liliana Leite, and its informatics department in the name of Domingos Pereira and Joaquim Pereira, and Matilde Monteiro-Soares for critical review of the manuscript. DFS acknowledges "Fundação para a Ciência e Tecnologia (FCT)", Portugal under grant number PD/BD/13553/2018. The work of DFS and PPR has been developed under the scope of project NORTE-01-0145-FEDER-000016 (NanoSTIMA), financed by the North Portugal Regional Operational Programme (NORTE-2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF).

References

1. Al Lawati, N.M., Patel, S.R., Ayas, N.T.: Epidemiology, risk factors, and consequences of obstructive sleep apnea and short sleep duration. Prog. Cardiovasc. Dis. **51**(4), 285–293 (2009). https://doi.org/10.1016/j.pcad.2008.08.001
2. Aragon, T. J.: epitools: epidemiology tools. R Package (2012)
3. Blondet, M., Yapor, P., Latalladi-Ortega, G., Alicea, E., Torres-Palacios, A., Rodríguez-Cintrón, W.: Prevalence and risk factors for sleep disordered breathing in a Puerto Rican middle-aged population. Sleep Breath. **13**(2), 175–180 (2009). https://doi.org/10.1007/s11325-008-0216-4
4. Bossuyt, P.M., Reitsma, J.B., Bruns, D.E., Gatsonis, C.A., Glasziou, P.P., Irwig, L., Lijmer, J.G., Moher, D., Rennie, D., De Vet, H.C.W., Kressel, H.Y., Rifai, N., Golub, R.M., Altman, D.G., Hooft, L., Korevaar, D.A., Cohen, J.F., for the STARD Group: STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin Chem **61**, 1446–1452 (2015). https://doi.org/10.1373/clinchem.2015.246280
5. Broström, A., Sunnergren, O., Årestedt, K., Johansson, P., Ulander, M., Riegel, B., Svanborg, E.: Factors associated with undiagnosed obstructive sleep apnoea in hypertensive primary care patients. Scand. J. Primary Health Care **30**(2), 107–113 (2012). https://doi.org/10.3109/02813432.2012.675563
6. Chung, S., Jairam, S., Hussain, M.R.G., Shapiro, C.M.: How, what, and why of sleep apnea. Perspectives for primary care physicians. Can. Fam. Phys. **48**, 1073–80 (2002)
7. Corral-Peñafiel, J., Pepin, J.L., Barbe, F.: Ambulatory monitoring in the diagnosis and management of obstructive sleep apnoea syndrome. Eur. Respir. Rev. **22**(129), 312–24 (2013). https://doi.org/10.1183/09059180.00004213
8. Darwiche, A.: Modeling and Reasoning with Bayesian networks. Cambridge University Press, New York (2009)
9. Darwiche, A.: Bayesian networks. Commun. ACM **53**(12), 80–90 (2010). https://doi.org/10.1145/1859204.1859227
10. Davies, R.J., Ali, N.J., Stradling, J.R.: Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome. Thorax **47**(2), 101–5 (1992)
11. Dias, C.C., Rodrigues, P.P., Coelho, R., Santos, P.M., Fernandes, S., Lago, P., Caetano, C., Rodrigues, Â., Portela, F., Oliveira, A., Ministro, P., Cancela, E., Vieira, A.I., Barosa, R., Cotter, J., Carvalho, P., Cremers, I., Trabulo, D., Caldeira, P., Antunesl, A., Rosa, I., Moleiro, J., Peixe, P., Herculano, R., Gonçalves, R., Gonçalves, B., Sousa, H.T., Contente, L., Morna, H., Lopes, S., Magroc, F.: Development and validation of risk matrices for Crohn's disease outcomes in patients who underwent early therapeutic interven-

tions. J. Crohn's Colitis **11**(4), 445–453 (2017). https://doi.org/10.1093/ecco-jcc/jjw171
12. Doghramji, P.P.: Recognition of obstructive sleep apnea and associated excessive sleepiness in primary care. J. Fam. Pract. **57**(8 Suppl), S17–23 (2008)
13. Epstein, L.J., Kristo, D., Strollo, P.J., Friedman, N., Malhotra, A., Patil, S.P., Ramar, K., Rogers, R., Schwab, R.J., Weaver, E.M., Weinstein, M.D.: Adult Obstructive Sleep Apnea Task Force of the American Academy of Sleep Medicine: clinical guideline for the evaluation, management and long-term care of obstructive sleep apnea in adults. J. Cin. Sleep Med. **5**(3), 263–276 (2009)
14. Fontenla-Romero, O., Guijarro-Berdiñas, B., Alonso-Betanzos, A., del Rocío Fraga-Iglesias, A., Moret-Bonillo, V.: A Bayesian Neural Network Approach for Sleep Apnea Classification, pp. 284–293. Springer, Berlin (2003)
15. Gregory Warnes, A.R., Bolker, B., Lumley, T., Johnson, R.C.: Various R programming tools for model fitting. R Package (2015)
16. Hoffstein, V., Szalai, J.P.: Predictive value of clinical features in diagnosing obstructive sleep apnea. Sleep **16**(2), 118–22 (1993)
17. Huang, K., King, I., Lyu, M.R.: Constructing a large node Chow-Liu tree based on frequent itemsets. In: ICONIP 2002—Proceedings of 9th International Conference on Neural Information Processing Computing Intelligent E-Age, vol. 1, pp. 498–502 (2002)
18. Jennum, Pc, Riha, R.L.: Epidemiology of sleep apnoea/hypopnoea syndrome and sleep-disordered breathing. Eur. Respir. J. **33**(4), 907–914 (2009). https://doi.org/10.1183/09031936.00180108
19. Kaimakamis, E., Bratsas, C., Sichletidis, L., Karvounis, C., Maglaveras, N.: Screening of patients with obstructive sleep Apnea syndrome using C4.5 algorithm based on non linear analysis of respiratory signals during sleep. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 3465–3469. IEEE (2009). https://doi.org/10.1109/IEMBS.2009.5334605
20. Kapur, V.: Obstructive sleep apnea: diagnosis, epidemiology, and economics. Respir. Care **55**(9), 1155–1167 (2010)
21. Kohler, M.: Risk factors and treatment for obstructive sleep apnea amongst obese children and adults. Curr. Opin. Allergy Clin. Immunol. **9**(1), 4–9 (2009). https://doi.org/10.1097/ACI.0b013e32831d8184
22. Lam, J.C.M., Sharma, S.K., Lam, B.: Obstructive sleep apnoea: definitions, epidemiology & natural history. Indian J. Med. Res. **131**, 165–170 (2010)
23. Lee, W., Nagubadi, S., Kryger, M., Mokhlesi, B.: Epidemiology of obstructive sleep apnea: a population-based perspective. Expert Rev. Respir. Med. **2**(3), 349–364 (2008). https://doi.org/10.1586/17476348.2.3.349.Epidemiology
24. Leite, L., Costa-Santos, C., Rodrigues, P.P.: Can we avoid unnecessary polysomnographies in the diagnosis of obstructive sleep apnea? A Bayesian network decision support tool. In: 2014 IEEE 27th International Symposium on Computer-Based Medical Systems, pp. 28–33. IEEE (2014). https://doi.org/10.1109/CBMS.2014.30
25. Libânio, D., Dinis-Ribeiro, M., Pimentel-Nunes, P., Dias, C., Rodrigues, P.: Predicting outcomes of gastric endoscopic submucosal dissection using a Bayesian approach: a step for individualized risk assessment. Endosc. Int. Open **05**(07), E563–E572 (2017). https://doi.org/10.1055/s-0043-106576
26. Lucas, P.J.F., van der Gaag, L.C., Abu-Hanna, A.: Bayesian networks in biomedicine and health-care. Artif. Intell. Med. **30**(3), 201–14 (2004). https://doi.org/10.1016/j.artmed.2003.11.001
27. Manber, R., Armitage, R.: Sex, steroids, and sleep: a review. Sleep **22**(5), 540–55 (1999)
28. Mansfield, D.R., Antic, N.A., McEvoy, R.D.: How to assess, diagnose, refer and treat adult obstructive sleep apnoea: a commentary on the choices. Med. J. Aust. **199**(8), S21–6 (2013). https://doi.org/10.5694/mja13.10909
29. Mitchell, T.: Machine Learning. McGraw-Hill, Singapore (1997)
30. Pagel, J., Hirshkowitz, M., Doghramji, P., Ballard, R.: Obstructive sleep apnea: recognition and management in primary care. Suppl. J. Fam. Pract. **348**(5), S1–S31 (2008). https://doi.org/10.1056/NEJM200301303480520
31. Rafael-Palou, X., Steblin, A., Vargiu, E.: Remotely supporting patients with obstructive sleep apnea at home. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering, LNICST (2016)
32. Robichaud-Hallé, L., Beaudry, M., Fortin, M.: Obstructive sleep apnea and multimorbidity. BMC Pulm. Med. **12**(1), 60 (2012). https://doi.org/10.1186/1471-2466-12-60
33. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.C., Müller, M.: pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform. **12**(1), 77 (2011). https://doi.org/10.1186/1471-2105-12-77
34. Rodrigues, P.P., Santos, D.F., Leite, L.: Obstructive sleep apnea diagnosis: the Bayesian network model revisited. In: 2015 IEEE 28th International Symposium on Computer-Based Medical Systems, pp. 115–120. IEEE (2015). https://doi.org/10.1109/CBMS.2015.47
35. Rodsutti, J., Hensley, M., Thakkinstian, A., D'Este, C., Attia, J.: A clinical decision rule to prioritize polysomnography in patients with suspected sleep apnea. SLEEP **27**(4), 694–699 (2004)
36. Romero, E., Krakow, B., Haynes, P., Ulibarri, V.: Nocturia and snoring: predictive symptoms for obstructive sleep apnea. Sleep Breath. **14**(4), 337–343 (2010). https://doi.org/10.1007/s11325-009-0310-2
37. Santaolalla Montoya, F., Iriondo Bedialauneta, J.R.R., Aguirre Larracoechea, U., Martinez Ibargüen, A., Sanchez Del Rey, A., Sanchez Fernandez, J.M., Martinez Ibarguen, A., Sanchez Del Rey, A., Sanchez Fernandez, J.M.: The predictive value of clinical and epidemiological parameters in the identification of patients with obstructive sleep apnoea (OSA): a clinical prediction algorithm in the evaluation of OSA. Eur. Arch. Oto-Rhinolaryngol. **264**(6), 63743 (2007). https://doi.org/10.1007/s00405-006-0241-5
38. Scutari, M.: Learning Bayesian networks with the bnlearn R Package. J. Stat. Softw. **35**(3), 1–22 (2010). https://doi.org/10.18637/jss.v035.i03
39. Stores, G.: Clinical diagnosis and misdiagnosis of sleep disorders. J. Neurol. Neurosurg. Psychiatry **78**(12), 1293–1297 (2007). https://doi.org/10.1136/jnnp.2006.111179
40. Sun, L.M., Chiu, H.W., Chuang, C.Y., Liu, L.: A prediction model based on an artificial intelligence system for moderate to severe obstructive sleep apnea. Sleep Breath. **15**(3), 317–323 (2011). https://doi.org/10.1007/s11325-010-0384-x
41. Venables, W.N., Ripley, B.D.: Modern applied statistics with S. Technometrics **45**(1), 111–111 (2003). https://doi.org/10.1198/tech.2003.s33
42. Wall, H., Smith, C., Hubbard, R.: Body mass index and obstructive sleep apnoea in the UK: a cross-sectional study of the over-50s. Primary Care Respir. J. **21**(4), 371–376 (2012). https://doi.org/10.4104/pcrj.2012.00053
43. Young, T.: Predictors of sleep-disordered breathing in community-dwelling adults. The Sleep Heart Health Study. Arch. Intern. Med. **162**(8), 893 (2002). https://doi.org/10.1001/archinte.162.8.893
44. Young, T., Evans, L., Finn, L., Palta, M.: Estimation of the clinically diagnosed proportion of sleep apnea syndrome in middle-aged men and women. Sleep **20**(9), 705–6 (1997)
45. Young, T., Skatrud, J., Peppard, P.E.: Risk factors for obstructive sleep apnea in adults. J. Am. Med. Assoc. **291**(16), 2013–2016 (2004). https://doi.org/10.1001/jama.291.16.2013

4.6. Enhancing obstructive sleep apnea diagnosis with screening through disease phenotypes: algorithm development and validation

JMIR Medical Informatics. 9: 1–16, 2021. doi:10.2196/25124

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Original Paper

Enhancing Obstructive Sleep Apnea Diagnosis With Screening Through Disease Phenotypes: Algorithm Development and Validation

Daniela Ferreira-Santos[1,2], BSc, MSc; Pedro Pereira Rodrigues[1,2], BSc, MSc, PhD

[1]MEDCIDS-FMUP – Community Medicine, Information and Decision Sciences, Faculty of Medicine of the University of Porto, Porto, Portugal
[2]CINTESIS – Center for Health Technology and Services Research, Porto, Portugal

Corresponding Author:
Daniela Ferreira-Santos, BSc, MSc
MEDCIDS-FMUP – Community Medicine, Information and Decision Sciences
Faculty of Medicine of the University of Porto
Rua Dr. Plácido da Costa, s/n
Porto, 4200-450
Portugal
Phone: 351 22 551 3622
Email: danielasantos@med.up.pt

Abstract

Background: The American Academy of Sleep Medicine guidelines suggest that clinical prediction algorithms can be used in patients with obstructive sleep apnea (OSA) without replacing polysomnography, which is the gold standard.

Objective: This study aims to develop a clinical decision support system for OSA diagnosis according to its standard definition (apnea-hypopnea index plus symptoms), identifying individuals with high pretest probability based on risk and diagnostic factors.

Methods: A total of 47 predictive variables were extracted from a cohort of patients who underwent polysomnography. A total of 14 variables that were univariately significant were then used to compute the distance between patients with OSA, defining a hierarchical clustering structure from which patient phenotypes were derived and described. Affinity from individuals at risk of OSA phenotypes was later computed, and cluster membership was used as an additional predictor in a Bayesian network classifier (model B).

Results: A total of 318 patients at risk were included, of whom 207 (65.1%) individuals were diagnosed with OSA (111, 53.6% with mild; 50, 24.2% with moderate; and 46, 22.2% with severe). On the basis of predictive variables, 3 phenotypes were defined (74/207, 35.7% low; 104/207, 50.2% medium; and 29/207, 14.1% high), with an increasing prevalence of symptoms and comorbidities, the latter describing older and obese patients, and a substantial increase in some comorbidities, suggesting their beneficial use as combined predictors (median apnea-hypopnea indices of 10, 14, and 31, respectively). Cross-validation results demonstrated that the inclusion of OSA phenotypes as an adjusting predictor in a Bayesian classifier improved screening specificity (26%, 95% CI 24-29, to 38%, 95% CI 35-40) while maintaining a high sensitivity (93%, 95% CI 91-95), with model B doubling the diagnostic model effectiveness (diagnostic odds ratio of 8.14).

Conclusions: Defined OSA phenotypes are a sensitive tool that enhances our understanding of the disease and allows the derivation of a predictive algorithm that can clearly outperform symptom-based guideline recommendations as a rule-out approach for screening.

(JMIR Med Inform 2021;9(6):e25124) doi: 10.2196/25124

KEYWORDS

obstructive sleep apnea; screening; risk factors; phenotypes; Bayesian network classifiers

Introduction

Background

Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by clinical symptoms (eg, daytime sleepiness) and at least five events per hour of narrowing (apnea or hypopnea) of the upper airway that impairs normal ventilation during sleep [1]. An apnea consists of a cessation of airflow higher than 90% of the baseline, a hypopnea is a reduction in airflow along with a decreased saturation of 3% from pre-event baseline and/or associated with an arousal, and the apnea-hypopnea index (AHI) is the number of such events per hour of sleep. OSA prevalence has been underestimated, with studies varying significantly, both in the population being studied and in OSA definition. A study using a simpler hypopnea definition (4% desaturation) estimated a prevalence of 14% in men and 5% in women [2]. In 2 other studies, the prevalence was substantially higher but was estimated for specific populations, such as patients being evaluated for bariatric surgery [3] or patients who have had a transient ischemic attack or stroke [4], reaching values of 70% and 72%, respectively. The latest study by Benjafield et al [5] estimated that 936 million adults have OSA; in Portugal, it represents 17%, and approximately 74% have moderate to severe OSA. Overall, this disease is largely unrecognized and undiagnosed, representing a significant burden to the health care system [6], especially for patients who remain untreated or at an increased risk of developing cardiovascular disease, metabolic dysregulation, or diabetes [1,7-11]. The failure to clinically recognize OSA leads to significant morbidity and mortality, making it essential to anticipate its recognition, diagnosis, and treatment [1]. OSA diagnosis, for which a comprehensive sleep evaluation (sleep history and physical examination) plus polysomnography (PSG) is the gold standard [1], can effectively decrease health care utilization and costs, whereas timely treatment can improve quality of life, lower the rates of motor vehicle crashes, and reduce the risk of chronic health consequences [12].

In 2017, a new clinical practice guideline for diagnostic testing for adults with OSA was issued by the American Academy of Sleep Medicine (AASM) [1], updating 2 previous AASM guidelines from 2005 [8] and 2007 [13]. Of the 9 PICO (patient, population or problem, intervention, comparison, and outcome) questions raised in this new guideline, the task force reported insufficient evidence to directly address the first one: "In adult patients with suspected OSA, do clinical prediction algorithms accurately identify patients with a high pretest probability for OSA compared to history and physical exam?," as no studies comparing the efficacy of clinical prediction algorithms with clinical history and physical examination were identified. Therefore, they compared the efficacy of clinical prediction algorithms with PSG, crafting recommendation 1: "We recommend that clinical tools, questionnaires and prediction algorithms not to be used to diagnose OSA in adults, in the absence of PSG," affirming that clinical prediction algorithms can, however, be used in patients with suspected OSA, as long as not to establish the need for PSG or to become a substitute for PSG. Rather, these tools can be more helpful, in specialties other than sleep-oriented ones, to identify patients with an increased risk for OSA.

Objective

In this study, we aim to establish a new clinical prediction algorithm to allow OSA screening (high pretest probability for OSA) based on demographics, physical examination, clinical history, and comorbidities, using standard OSA definition (AHI ≥5 plus symptoms), extending traditional approaches that assess only preestablished symptoms, such as snoring, witnessed apneas, and excessive daytime sleepiness.

Methods

Overview

Using retrospective data from a cohort of patients who underwent PSG, after proper referral by a physician, significant predictive variables were selected and used to compute distances among patients with OSA, which supported a clustering algorithm to derive patient phenotypes from resulting clusters, with missing data being analyzed and imputed as needed. To assess the consistency of our phenotypes, each healthy individual was also tested against the clustering structure, and the resulting phenotyping was analyzed. Then, to assess the benefit of this phenotyping strategy, cluster membership was used as an additional predictive variable and included in a Bayesian network classifier, with validity compared with an equivalent classifier without phenotype information, following the 2015 STARD (Standards for Reporting Diagnostic Accuracy Studies) guideline.

Patients

Data from patients referred to undergo PSG at Vila Nova de Gaia and Espinho Hospital Center Sleep Laboratory were retrospectively collected. Patients who underwent PSG between January and May 2015 were included if they were aged >18 years and were suspected of having OSA. Nonetheless, exclusion criteria included patients already diagnosed (performing positive airway pressure therapies), patients suspected of having another sleep disease, patients with severe lungs or neurological conditions, and pregnant women. In case of multiple examinations of the same patient, the one with the best sleep efficiency was selected. This study was approved by the Ethics Commission of Vila Nova de Gaia and Espinho Hospital Center, in accordance with the Declaration of Helsinki.

Predictive Variables

An author-performed literature review on PubMed (April 19, 2015) supported the definition of the relevant variables to be collected from medical and/or sleep laboratory records in which the presence or absence of each information was assessed by a physician, resulting in a total of 47 predictive variables, all in accordance with the current and previous OSA guidelines. The search contained "risk factors," "sleep apnea, obstructive," and "diagnosis" as MeSH terms, obtaining 1397 articles, of which 47 were used for variable definition (full review description and references used in this phase are not shown for space purposes but can be provided on request). Selected variables included basic demographic data (gender and age), physical examination

(BMI, neck and abdominal circumferences, modified Mallampati classification, and craniofacial and upper airway abnormalities), clinical history (daytime sleepiness, snoring, witnessed apneas, gasping and/or choking, sleep fragmentation, nonrepairing sleep, behavior changes, decreased concentration, morning headaches, decreased libido, sleeping body position, sleep efficiency, participation in vehicle crashes, truck driver activity, driving sleepiness, nocturia, alcohol consumption, smoking, coffee intake, use of sedatives before sleep, family history or genetic evidence, and Epworth Sleepiness Scale), and comorbidities and cointerventions (stroke, myocardial infarction, pulmonary infarction, arterial or pulmonary hypertension, congestive heart failure, arrhythmias, respiratory changes, diabetes, dyslipidemia, renal failure, hypothyroidism, gastroesophageal reflux, anxiety and/or depression, insomnia, glaucoma, pacemaker or implantable cardioverter-defibrillator, and bariatric surgery).

Data Set Description

Clinical data from each patient (47 predictive variables plus the outcome) were extracted from the central clinical data registry (all records were fulfilled by a physician) along with sleep laboratory data and adequately anonymized to ensure patient privacy. Original files included structured demographic data, structured PSG reports, and unstructured textual annotations from the medical records, with many abbreviations and short-form text. The outcome measure was obtained from the AHI, categorized as mild (AHI between 5 and 14), moderate (AHI between 15 and 29), and severe (AHI >30). Given the categorical characteristic of our modeling strategies, all continuous variables were discretized, and the following common cutoffs were extracted from the literature: (1) age (20-44 years, 45-64 years, and 65-90 years), (2) BMI (<25 kg/m^2 as normal weight, 25-30 kg/m^2 as overweight, and ≥30 kg/m^2 as obese), (3) female neck circumference (≤37 cm as normal and >38 cm as increased), (4) male neck circumference (≤41 cm as normal and >42 cm as increased), (5) female abdominal circumference (≤80 cm as normal and >81 cm as increased), (6) male abdominal circumference (≤94 cm as normal and >95 cm as increased), (7) Epworth Sleepiness Scale (0-10 as normal and 11-24 as excessive daytime sleepiness), and (8) AHI (0-4 as normal, 5-14 as mild, 15-29 as moderate, and ≥30 as severe).

Missing Data Imputation

Although we had all the electronic clinical records from the included patients, after screening all unstructured text reports, some predictive variables were not fully present or described, as physicians normally do not mention the absence of a disease or it could only be noted in paper records (missing data proportions ranged from 0% for gender to 97% for bariatric surgery). In our previous study [14], we studied the impact of missing data imputation, using nearest neighbor (NN) strategies, on the structure learning of Bayesian network classifiers for OSA diagnosis, concluding that it can expand the body of evidence for modeling without compromising validity. In this study, we followed the same strategy: (1) variables with more than 80% missing values were removed from the analysis (ie, behavior changes, decreased libido, decreased concentration, pulmonary infarction, glaucoma, and bariatric surgery); (2) remaining variables were ranked by the proportion of missing values; (3) data imputation started using only complete and outcome-wise statistically significant variables (P<.20), imputing incomplete likewise significant variables; and (4) remaining incomplete variables were then imputed stepwise by increasing the proportion of missing values per variable. All imputations were performed using majority voting from the 10 NNs/patients.

Clinical Prediction Algorithm

Aspiring to a more personalized approach to evaluate patients with OSA and targeting to recognize high pretest probability for OSA, cluster analysis (a statistical approach for studying the relationship present among groups of patients or variables [7]) was applied to distinguish whether there are different subgroups of patients with different clinical presentations, that is phenotypes. Clustering has been widely used in health research, particularly in the analysis of gene expression [15], asthma [16], chronic obstructive pulmonary disease [17], fibromyalgia [18], Parkinson disease [19], and sleep apnea [20-22]. The aim is to identify clusters of patients who are similar among themselves, although significantly different from patients of other clusters [7]. As expected, different clusters created from predictive variables express different disease risks, hence defining risk-aligned phenotypes.

Connectivity-Based Clustering

In this study, we applied a hierarchical clustering algorithm to obtain a hierarchy of possible solutions, ranging from one single group with all patients to having every single patient separated from each other. This process, where a cluster hierarchy is created, is based on the distance between data observations (ie, patients), giving as output a dendrogram (a tree diagram that presents different clustering definitions for all possible numbers of clusters, from which the user might choose the desired number of clusters after inspecting the intracluster and intercluster distances of each possible cut point). Therefore, the definition of the distance function is a crucial step in the application of this technique, especially in categorical data, as an incorrect distance can easily lead to biased results with potentially serious consequences to the conclusions drawn.

In this study, we computed the distance measure between 2 patients, a and b, based only on significant variables (univariate significant association with the outcome, for a 20% significance level in both the original and imputed data sets, using chi-square and Fisher exact tests), and each variable was weighted according to the corresponding crude odds ratio for the severe level, as follows:

$$d_{ab} = \frac{\sum_i X_{i\,(a,b)} * W_i}{\sum_i W_i}, \text{ where } X_{i\,(a,b)} = \begin{cases} 1, & a_i = b_i \\ 0, & a_i := b_i \end{cases} \text{ and } W_i = |\ln(or_i)| + 1$$

This distance encoded the similarity between patients weighted by the contribution of each variable toward the outcome, regularized for significant variables only, and was subsequently used in hierarchical clustering with Ward linkage, leading to a complete dendrogram. Afterward, the obtained OSA clusters were defined by inspecting the outcome proportion by cluster and the corresponding 95% CIs.

Phenotypes Consistency

To assess whether predetermined phenotypes would also help in segmenting healthy patients, each healthy patient was assigned to the closest phenotype using the aforementioned distance measure and the same significant variables, determining the distance between each healthy patient and obtained OSA cluster. The resulting clustering definition was then described and analyzed, as was done for the cohort of patients with OSA.

Phenotypes Predictive Value

To assess whether the phenotypes could encode any predictive value, Bayesian network classifiers were built with and without cluster information as a predictive variable. First, a naïve Bayesian network classifier was induced using the selected variables. Then, assigned cluster was also included in the model as a parent node of all independent variables. Validity was then assessed and compared using leave-one-out and 10 times twofold cross-validation strategies, comparing validity measures, such as sensitivity, specificity, accuracy, predictive values, area under the receiver operating characteristic (ROC) curve, likelihood ratios, posttest odds and posttest probabilities, and diagnostic odds ratio.

Statistical Software

R 3.2.2 (R Development Core Team) software [23] was used on every statistical step of this work: discretization of continuous variables (package car [24]), descriptive and comparative analyses (packages gmodels [25] and epitools [26]), missing data analysis (package summarytools [27]), missing data imputation (package DMwR [28]), hierarchical clustering (package stats [23]), Bayesian network inference (packages bnlearn [29] and gRain [30]), and ROC curve analysis (package pROC [31]). Bayesian networks were visually inspected using SamIam software (developed by the University of California, Los Angeles) [32].

Results

Baseline Characteristics

Of the 318 patients included, 207 (65.1%) had OSA. Of these 207 patients, 111 (53.6%) were classified as mild, 50 (24.2%) as moderate, and 46 (22.2%) as severe. Baseline characteristics of patients with OSA and the proportion of missing values for each predictive variable are described below in Table 1 (original data) and in Multimedia Appendix 1 (for the curated data, after missing data imputation).

Patients with OSA had a mean age of 61 (SD 11) years, being slightly older in the moderate subgroup (24/50, 48%; aged >65 years), whereas the proportion of males was higher in the moderate (40/50, 80%) and severe (35/46, 76%) subgroups. Beyond these 2 variables, only sleep efficiency was found to be complete (no missing data), and no differences were found across OSA levels (P=.65). For the remaining variables, distributions were computed before and after data imputation.

The presence of witnessed apneas (109/169, 64.5%), nonrepairing sleep (93/183, 50.8%), nocturia (99/136, 72.8%), stroke (23/44, 52%), arterial hypertension (136/159, 85.5%), diabetes (62/99, 63%), and dyslipidemia (125/148, 84.5%) were more prevalent in patients with OSA than in healthy patients, whereas the opposite was observed in family history (14/77, 18%), pulmonary hypertension (15/117, 12.8%), congestive heart failure (26/138, 18.8%), arrhythmias (17/99, 17%), pacemaker or implantable cardioverter-defibrillator (10/91, 11%), and respiratory changes (81/185, 43.8%). After data imputation, the same variables remained different across OSA levels, except for family history. Only variables significantly associated with the outcome (P<.20) on both the original and curated data sets were further considered for the clustering process.

Table 1. Descriptive analysis of patients with obstructive sleep apnea (absolute and relative frequencies are presented, and *P* values are the results of chi-square tests unless otherwise specified).

Characteristic	Mild (n=111), n (%)	Moderate (n=50), n (%)	Severe (n=46), n (%)	Total (N=207), n (%)	*P* value	Missing, n (%)
Gender (male)	72 (64.9)	40 (80.0)	35 (76.1)	147 (71.0)	*.10* [a]	207 (0.0)
Age (years)					*.18* [b]	207 (0.0)
20-44	7 (6.3)	5 (10.0)	6 (13.0)	18 (8.7)		
45-64	67 (60.4)	21 (42.0)	24 (52.2)	112 (54.1)		
65-90	37 (33.3)	24 (48.0)	16 (34.8)	77 (37.2)		
BMI (kg/m^2)					*.05* [b]	169 (18.4)
Normal weight	14 (15.6)	1 (2.6)	1 (2.5)	16 (9.5)		
Overweight	34 (37.8)	21 (53.8)	16 (40.0)	71 (42.0)		
Obesity	42 (46.7)	17 (43.6)	23 (57.5)	82 (48.5)		
Increased neck circumference	50 (64.1)	23 (67.6)	19 (70.4)	92 (66.2)	.82	139 (32.9)
Increased abdominal circumference	48 (87.3)	23 (95.8)	21 (100.0)	92 (92.0)	.22[b]	100 (51.7)
Modified Mallampati					.44	142 (31.4)
Class I	19 (23.2)	3 (10.0)	5 (16.7)	27 (19.0)		
Class II	29 (35.4)	15 (50.0)	9 (30.0)	53 (37.3)		
Class III	29 (35.4)	9 (30.0)	12 (40.0)	50 (35.2)		
Class IV	5 (6.1)	3 (10.0)	4 (13.3)	12 (8.5)		
Craniofacial and upper airway abnormalities	42 (84.0)	15 (83.3)	6 (66.7)	63 (81.8)	.49[b]	77 (62.8)
Daytime sleepiness	61 (55.5)	27 (60.0)	21 (50.0)	109 (55.3)	.64	197 (4.8)
Snoring	103 (92.8)	43 (93.5)	41 (93.2)	187 (93.0)	>.99[b]	201 (2.9)
Witnessed apneas	55 (58.5)	30 (76.9)	24 (66.7)	109 (64.5)	*.12*	169 (18.4)
Gasping and/or choking	39 (45.3)	12 (36.4)	16 (45.7)	67 (43.5)	.65	154 (25.6)
Sleep fragmentation	55 (73.3)	22 (68.8)	19 (73.1)	96 (72.2)	.88	133 (35.7)
Nonrepairing sleep	47 (47.5)	27 (62.8)	19 (46.3)	93 (50.8)	*.20*	183 (11.6)
Morning headaches	34 (46.6)	14 (48.3)	17 (53.1)	65 (48.5)	.83	134 (35.3)
Body position					.36[b]	201 (2.9)
Decubitus	5 (4.5)	0 (0.0)	1 (2.3)	6 (3.0)		
Left lateral	20 (18.2)	8 (17.0)	8 (18.2)	36 (17.9)		
Right lateral	56 (50.9)	22 (46.8)	16 (36.4)	94 (46.8)		
Supine	29 (26.4)	17 (36.2)	19 (43.2)	65 (32.3)		
Bad sleep efficiency	68 (61.3)	28 (56.0)	30 (65.2)	126 (60.9)	.65	207 (0.0)
Vehicle crashes	7 (20.6)	0 (0.0)	3 (20.0)	10 (16.4)	.28[b]	61 (70.5)
Truck driver	5 (4.7)	5 (10.4)	4 (9.5)	14 (7.1)	.32[b]	197 (4.8)
Driving sleepiness	5 (8.9)	4 (17.4)	4 (18.2)	13 (12.9)	.38[b]	101 (51.2)
Nocturia	47 (64.4)	20 (69.0)	32 (94.1)	99 (72.8)	*.005*	136 (34.3)
Alcohol consumption	61 (66.3)	29 (70.7)	29 (74.4)	119 (69.2)	.64	172 (16.9)
Smoking					.74	204 (1.4)
Yes	11 (10.0)	7 (14.6)	5 (10.9)	23 (10.9)		
Ex-smoker	38 (34.5)	20 (41.7)	17 (37.0)	75 (36.8)		

Characteristic	Mild (n=111), n (%)	Moderate (n=50), n (%)	Severe (n=46), n (%)	Total (N=207), n (%)	*P* value	Missing, n (%)
Coffee intake	77 (87.5)	30 (83.3)	25 (86.2)	132 (86.3)	.85[b]	153 (26.1)
Use of sedatives	23 (22.8)	13 (29.5)	7 (16.3)	43 (22.9)	.34	188 (9.2)
Family history	8 (18.2)	1 (5.9)	5 (31.2)	14 (18.2)	*.18* [b]	77 (62.8)
Epworth Sleepiness Scale	33 (37.5)	17 (44.7)	10 (29.4)	60 (37.5)	.41	160 (22.7)
Stroke	9 (37.5)	8 (80.0)	6 (60.0)	23 (52.3)	*.08* [b]	44 (78.7)
Myocardial infarction	9 (12.7)	3 (9.7)	6 (20.0)	18 (13.6)	.52[b]	132 (36.2)
Arterial hypertension	67 (79.8)	32 (91.4)	37 (92.5)	136 (85.5)	*.09*	159 (23.2)
Pulmonary hypertension	5 (8.3)	3 (10.0)	7 (25.9)	15 (12.8)	*.08* [b]	117 (43.5)
Congestive heart failure	7 (9.9)	6 (17.6)	13 (39.4)	26 (18.8)	*.002*	138 (33.3)
Arrhythmias	5 (10.0)	4 (16.7)	8 (32.0)	17 (17.2)	*.06* [b]	99 (52.2)
Pacemaker and/or cardioverter	3 (6.1)	2 (9.5)	5 (23.8)	10 (11.0)	*.09* [b]	91 (56.0)
Respiratory changes	43 (43.0)	15 (32.6)	23 (59.0)	81 (43.8)	*.05*	185 (10.6)
Diabetes	28 (51.9)	12 (60.0)	22 (88.0)	62 (62.6)	*.008*	99 (52.2)
Dyslipidemia	63 (78.8)	28 (90.3)	34 (91.9)	125 (84.5)	*.11*	148 (28.5)
Renal failure	10 (27.0)	6 (50.0)	7 (36.8)	23 (33.8)	.33	68 (67.1)
Hypothyroidism	12 (25.5)	6 (37.5)	6 (35.3)	24 (30.0)	.58	80 (61.4)
Gastroesophageal reflux	22 (48.9)	10 (71.4)	7 (53.8)	39 (54.2)	.34	72 (65.2)
Anxiety and/or depression	41 (78.8)	23 (92.0)	17 (77.3)	81 (81.8)	.31[b]	99 (52.2)
Insomnia	25 (71.4)	10 (76.9)	10 (90.9)	45 (76.3)	.48[b]	59 (71.5)

[a]*P*<.20 are italicized.
[b]Fisher exact test.

OSA Clusters

Using the 14 variables significantly associated with the outcome, a hierarchical clustering structure was derived, where, given the resulting clustering structure, a 10-cluster cutoff point was chosen (following the hierarchical structure of the clustering in the dendrogram). The resulting clusters had median AHI values of 8, 10 (4 clusters), 12, 13, 14, 31, and 34. As 10 clusters are difficult to interpret in a medical context, we chose to aggregate the 10 created clusters into 3 clusters according to their median values: (1) clusters with median 8 and 10, (2) clusters with median 12, 13, and 14, and (3) clusters with median 31 and 34.

The OSA cluster characteristics of the 14 predictive variables are described below and listed in Table 2. The witnessed apneas variable was also statistically significant in both the original and the curated data but was not considered for the cluster hierarchy, as it depends on third-party reporting, which might create a strong bias in the analysis.

Table 2. Clinical characteristics of the obstructive sleep apnea cohort by the defined clusters (*P* values are the results of chi-square tests unless otherwise specified).

Characteristics	Cluster 1 (n=74)		Cluster 2 (n=104)		Cluster 3 (n=29)		*P* value
	Patient, n (%)	95% CI	Patient, n (%)	95% CI	Patient, n (%)	95% CI	
Gender (male)	51 (68.9)	57-79	72 (69.2)	60-78	24 (82.8)	64-93	.32
Age (years)							<.001[a]
20-44	6 (8.1)	3-17	12 (11.5)	6-20	0 (0.0)	0-15	
45-64	46 (62.2)	50-73	59 (56.7)	47-66	7 (24.1)	11-44	
65-90	22 (29.7)	20-42	33 (31.7)	23-42	22 (75.9)	56-89	
BMI (kg/m^2)							<.001[a]
Normal weight	13 (17.6)	10-29	3 (2.9)	1-9	0 (0.0)	0-15	
Overweight	15 (20.3)	12-32	50 (48.1)	38-58	9 (31.0)	16-51	
Obesity	46 (62.2)	50-73	51 (49.0)	39-59	20 (69.0)	49-84	
Nonrepairing sleep	34 (45.9)	34-58	57 (54.8)	45-65	10 (34.5)	19-54	.13
Nocturia	14 (18.9)	11-30	104 (100.0)	96-100	29 (100.0)	85-100	<.001
Stroke	34 (45.9)	34-58	88 (84.6)	76-91	27 (93.1)	76-99	<.001
Arterial hypertension	54 (73.0)	61-82	96 (92.3)	85-96	29 (100.0)	85-100	<.001[a]
Pulmonary hypertension	9 (12.2)	6-22	2 (1.9)	0-8	6 (20.7)	9-40	.002[a]
Congestive heart failure	4 (5.4)	2-14	1 (1.0)	0-6	23 (79.3)	60-91	<.001[a]
Arrhythmias	4 (5.4)	2-14	1 (1.0)	0-6	12 (41.4)	24-61	<.001[a]
Pacemaker and/or cardioverter	0 (0.0)	0-6	4 (3.8)	1-10	6 (20.7)	9-40	<.001[a]
Respiratory changes	35 (47.3)	36-59	28 (26.9)	19-37	18 (62.1)	42-79	.001
Diabetes	37 (50.0)	39-61	69 (66.3)	56-75	29 (100.0)	85-100	<.001
Dyslipidemia	57 (77.0)	66-86	94 (90.4)	83-95	29 (100.0)	85-100	.003[a]
Apnea-hypopnea index							<.001
Mild	51 (68.9)	57-79	54 (51.9)	42-62	6 (20.7)	9-40	
Moderate	18 (24.3)	15-36	24 (23.1)	16-33	8 (27.6)	13-48	
Severe	5 (6.8)	3-16	26 (25.0)	17-35	15 (51.7)	33-70	

[a]Fisher exact test.

As shown in Table 2, 68.9% (51/74) of the patients in cluster 1 (74/207, 35.7%) were male, 62.2% (46/74) were aged between 45 and 64 years, and 62.2% (46/74) were obese. Nonrepairing sleep was reported in almost half of the patients, and only 18.9% (14/74) reported nocturia. The occurrence of stroke (34/74, 45.9%) did not reach half of the patients, whereas arterial hypertension (54/74, 73.0%) and dyslipidemia (57/74, 77.0%) surpassed it. Pulmonary hypertension, congestive heart failure, arrhythmias, and pacemaker or implantable cardioverter-defibrillator had percentages lower than 15%. The median AHI was 10 (range 7-17), the lowest AHI value, with 69.8% (44/169) reporting witnessed apneas.

Cluster 2 (104/207, 50.2%) had 69.2% (72/104) of males (the same as cluster 1), and only 2.9% (3/104) had normal weight. In contrast to cluster 1, 100.0% (104/104) of patients reported nocturia, 84.6% (88/104) reported stroke, 92.3% (96/104) reported arterial hypertension, and 90.4% (94/104) reported dyslipidemia. Similar to cluster 1, pulmonary hypertension, congestive heart failure, arrhythmias, and pacemaker or implantable cardioverter-defibrillator had percentages lower than 15%. Respiratory changes were reported in 26.9% (28/104) of the patients, and diabetes was reported in 66.3% (69/104) of the patients, compared with cluster 1. Regarding the clinical outcome, this cluster had a median AHI of 14 (range 8-30). Concerning witnessed apneas, cluster 2 had a percentage of 57.8% (52/169), the lowest value of all 3 clusters.

Cluster 3 (29/207, 14.0%) included the highest percentage of men (24/29, 82.8%). None of the patients were aged between 20 and 44 years or had normal weight. This cluster had the lowest proportion of patients aged between 45 and 64 years; nevertheless, it reached the highest proportion of all clusters in patients aged between 65 and 90 years. Although it had one of the lowest proportions of overweight patients, this cluster had the highest percentage (20/29, 69.0%) of patients with obesity.

In contrast to cluster 1, but in concordance with cluster 2, nocturia was described in all patients in cluster 3. In addition, arterial hypertension, diabetes, and dyslipidemia were observed in all the patients. The median AHI was 31 (range 21-60); therefore, it was the highest in all 3 clusters. Witnessed apneas were found with the highest proportion of all clusters (13/169, 81.2%).

Age strata and BMI were found to be different among clusters (P<.001). Comorbidities, such as stroke, arterial hypertension, diabetes (P<.001), and dyslipidemia (P=.003), were increasingly more prevalent from cluster 1 to clusters 2 and 3. Only male sex (P=.32) and nonrepairing sleep (P=.13) were not found to be significantly different.

On the basis of the description of clusters mentioned earlier, the OSA phenotypes can be defined. We classified patients into low (cluster 1), medium (cluster 2), and high (cluster 3) severity phenotypes, as their median AHI corresponded to mild, moderate, and severe levels respectively, defined in PSG for OSA diagnosis. The low severity phenotype includes age >45 years, a fair distribution in normal and overweight patients, accentuating obesity, and low prevalence of symptoms and comorbidities, except for dyslipidemia and arterial hypertension. The medium severity phenotype has almost the same distribution in age as the low severity phenotype, but less normal-weight patients and more overweight patients. Symptoms and comorbidities were higher, with stroke, arterial hypertension, dyslipidemia, and nocturia appearing in more than 85% of the patients with this phenotype. The high severity phenotype presents older and obese patients, with additional comorbidities (congestive heart failure and diabetes) beyond those present in the medium severity phenotype. The foremost difference between our phenotypes and AHI alone is that we considered the risk and diagnostic factors associated with the patient and not only a single value or a counting of events.

Affinity Between Healthy Patients and OSA Phenotypes

Given that our data set included patients who are healthy and with OSA (a total of 318 individuals), we focused our attention on exploring whether the determined OSA phenotypes could also help to segment healthy patients. To do so, we computed the aforementioned distance measure between 2 individuals using the same 14 significant variables. Table 3 describes the baseline characteristics of healthy patients for each OSA phenotype.

As expected, a high severity phenotype was less common in healthy patients (7/111, 6.3%), including older (P<.001), females (P=.49), and obese individuals (P=.50), with a lower proportion of individuals reporting nonrepairing sleep (P=.36). This phenotype also presented the highest proportion of reported nocturia, stroke, arterial hypertension, congestive heart failure, and diabetes (P<.001); pulmonary hypertension and arrhythmias (P=.01); and respiratory changes (P=.11). The medium severity phenotype had the highest proportion of overweight males aged between 45 and 64 years. Although comorbidities such as pulmonary hypertension, congestive heart failure, arrhythmias, and pacemaker or implantable cardioverter-defibrillator do not reach proportions higher than 1%, others such as stroke, arterial hypertension, diabetes, and dyslipidemia present proportions higher than 70%. The low severity phenotype is similar to the medium severity phenotype in terms of the proportion of overweight males, but individuals are younger. Nocturia, pulmonary hypertension, congestive heart failure, arrhythmias, pacemaker or implantable cardioverter-defibrillator, and diabetes have not been reported in this phenotype. Dyslipidemia was the most common comorbidity (16/25, 64%), followed by arterial hypertension (14/25, 56%) and respiratory changes (7/25, 28%).

Table 3. Clinical characteristics of the healthy cohort by the predefined obstructive sleep apnea phenotypes (*P* values are the result of chi-square test, unless otherwise specified).

Characteristics	Low OSA[a] (n=25)		Medium OSA (n=79)		High OSA (n=7)		*P* value[b]
	Patient, n (%)	95% CI	Patient, n (%)	95% CI	Patient, n (%)	95% CI	
Gender (male)	10 (40)	22-61	39 (49)	38-61	2 (29)	5-70	.49
Age (years)							<.001
20-44	16 (64)	43-81	15 (19)	11-30	0 (0)	0-44	
45-64	7 (28)	13-50	47 (59)	48-70	2 (29)	5-70	
65-90	2 (8)	1-28	17 (22)	13-32	5 (71)	30-95	
BMI							.50
Normal weight	4 (16)	5-37	6 (8)	3-16	1 (14)	1-58	
Overweight	11 (44)	25-65	40 (51)	39-62	2 (29)	5-70	
Obesity	10 (40)	22-61	33 (42)	31-53	4 (57)	20-88	
Nonrepairing sleep	18 (72)	50-87	51 (65)	53-75	3 (43)	12-80	.36
Nocturia	0 (0)	0-17	54 (68)	57-78	6 (86)	42-99	<.001
Stroke	1 (4)	0-22	66 (84)	73-91	6 (86)	42-99	<.001
Arterial hypertension	14 (56)	35-75	78 (99)	92-100	7 (100)	56-100	<.001
Pulmonary hypertension	0 (0)	0-17	1 (1)	0-8	2 (29)	5-70	.01
Congestive heart failure	0 (0)	0-17	0 (0)	0-6	7 (100)	56-100	<.001
Arrhythmias	0 (0)	0-17	1 (1)	0-8	2 (29)	5-70	.01
Pacemaker and/or cardioverter	0 (0)	0-17	1 (1)	0-8	0 (0)	0-44	>.99
Respiratory changes	7 (28)	13-50	34 (43)	32-55	5 (71)	30-95	.11
Diabetes	0 (0)	0-17	55 (70)	58-79	7 (100)	56-100	<.001
Dyslipidemia	16 (64)	43-81	75 (95)	87-98	6 (86)	42-99	.001

[a]OSA: obstructive sleep apnea.

[b]Fisher exact test.

Beyond OSA Phenotypes

OSA is a systemic disorder that remains underdiagnosed. Physicians, particularly nonspecialists in sleep disorders, urgently need a simple yet complete tool that allows them to identify a high pretest probability for OSA. This ability, which could enhance current screening, could lead to personalized treatment by additionally improving the understanding of OSA mechanisms and the risk for adverse events.

Our clinical prediction algorithm, that is, previously described OSA phenotypes, is a new way to screen patients, extending traditional approaches. To implement this new strategy, we need a simple, understandable, and updatable tool that can be used daily and that takes into account the knowledge of experts, the literature evidence, and the clinical data.

Belief or Bayesian networks [33] are probabilistic graphical models used to represent knowledge about an uncertain domain; each node represents a random variable, whereas directed edges between the nodes represent probabilistic dependencies among the corresponding variables. Bayesian networks are both mathematically rigorous and intuitively understandable, as they reflect a simple conditional independence statement, that is, each variable is independent of its nondescendants in the graph, given the state of its parents. The Bayesian network thus consists of both a qualitative model (which shows the relationship among variables) and a quantitative model (the joint probability distribution is expressed as conditional probabilities).

Initially, we created the simplest Bayesian classifier (naïve Bayes; Figure 1, Model A), which assumes independence among predictive variables and conditional independence, given the outcome. Subsequently, we extended the model (Figure 2, Model B), adding the defined phenotypes as a parent node of all predictors, thereby adjusting the model by capturing possible interactions among them, expressed by the corresponding phenotype associated with the tested individual. To evaluate the benefits of including OSA phenotypes in the clinical risk assessment tool, it was necessary to estimate the overall performance of each model. The ROC curves of each model (for both leave-one-out and cross-validation estimates) are presented in Figure 3, assessing the discriminative power of both models. As shown in Table 4, the derivation sample (area under the curve [AUC]) improved from 72% (95% CI 66-78) for model A to 84% (95% CI 80-89) for model B. The validity assessment confirmed the improvement achieved by the inclusion of OSA phenotypes, with leave-one-out estimates of 68% to 78%, respectively, from model A to model B and with

10 times twofold cross-validation averaging 67% and 77%, respectively. In addition, the diagnostic odds ratio, as a measure of the effectiveness of a diagnostic test, was 3.55 for model A and 2 times more for model B

Figure 1. Naïve Bayesian network representation of the relationships between the outcome (obstructive sleep apnea) and each of the 14 significant predictive variables. The bars within each variable represent the prior marginal probabilities for the category of each variable. CDI: implantable cardioverter-defibrillator; CHF: congestive heart failure; OSA: obstructive sleep apnea.

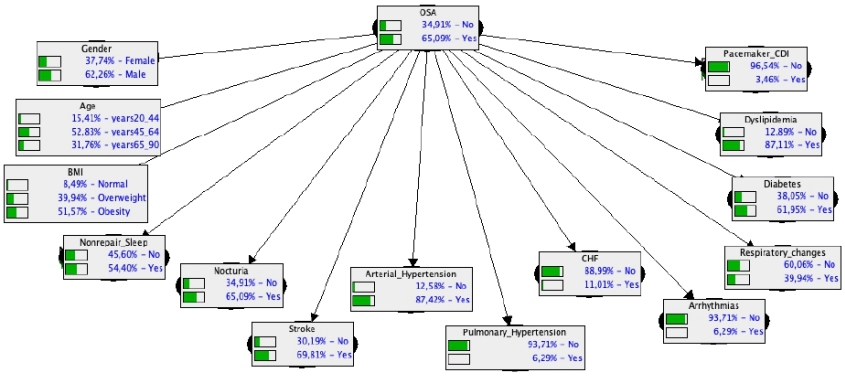


Figure 2. Naïve Bayesian network representation with additional node obtained from predefined obstructive sleep apnea phenotypes. The bars within each variable represent the prior marginal probabilities for the category of each variable. CDI: implantable cardioverter-defibrillator; CHF: congestive heart failure; OSA: obstructive sleep apnea.

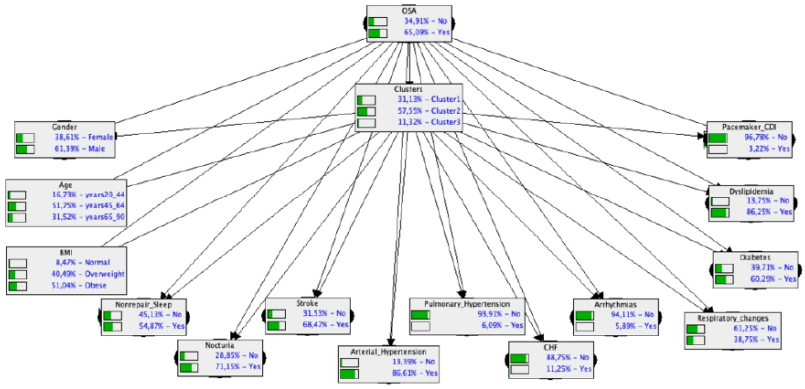


Figure 3. Receiver operating characteristic analyses and AUCs for models A (top) and B (bottom) as well for the internal validation procedures. AUC: area under the curve.

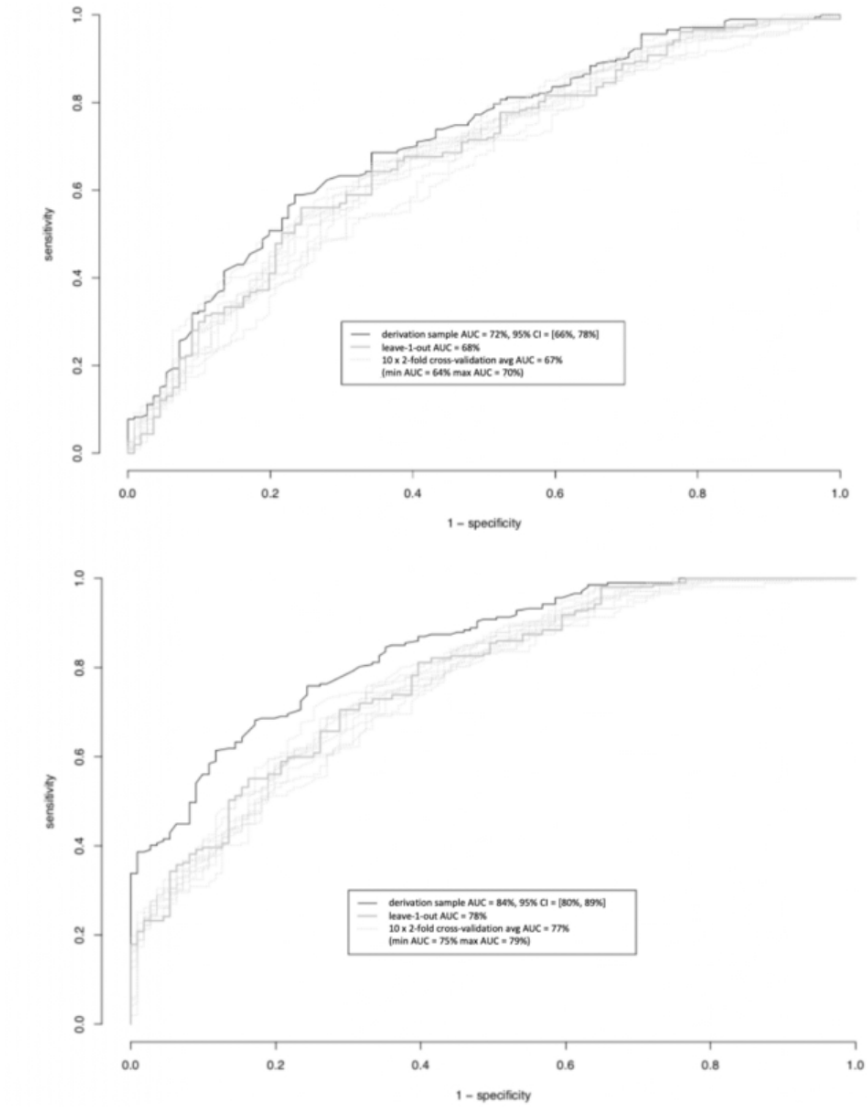


Table 4. Validity assessment estimated from 10 times twofold cross-validation.

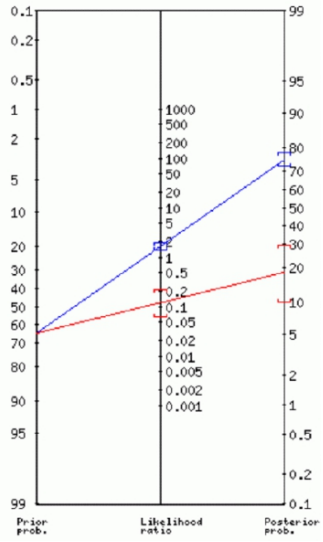
Variables	Obstructive sleep apnea	
	Model A	Model B
Cutoff point	30	22
Accuracy, % (95% CI)	69 (67-70)	74 (72-75)
Sensitivity, % (95% CI)	91 (89-94)	93 (91-95)
Specificity, % (95% CI)	26 (24-29)	38 (35-40)
Positive predictive value, % (95% CI)	70 (69-70)	73 (73-74)
Negative predictive value, % (95% CI)	64 (58-70)	75 (70-80)
Area under the curve, % (95% CI)	67 (67-70)	77 (76-78)
Positive likelihood ratio (95% CI)	1.32 (1.17-1.49)	1.63 (1.39-1.91)
Negative likelihood ratio (95% CI)	0.17 (0.09-0.34)	0.12 (0.06-0.22)
Positive odds posttest (95% CI)	2.45 (2.02-3.01)	3.02 (2.43-3.78)
Negative odds posttest (95% CI)	0.32 (0.19-0.56)	0.22 (0.12-0.38)
Posttest probability (95% CI)	71 (66-76)	75 (70-80)

Aiming at a 95% sensitivity target (screening strategies look for rule-out approaches), cutoff points were defined based on the derivation sample ROC curve, and the corresponding validity assessment results for cross-validation are displayed in Table 4, presenting an increase of specificity (26%-38%) for the desired level of sensitivity and presenting a posttest odds of 3 to 1 for the positive result and almost 1 to 5 for the negative result.

On the basis of the model with OSA phenotypes, OSA probabilities >22% were considered a positive result. The application of this cutoff resulted in a sensitivity value of 93% (95% CI 91-95) and 73% (95% CI 73-74) of positive predictive value, managing to provide a sensitive tool that prevents 1 out of 5 healthy individuals from unnecessarily undergoing PSG.

In our sample, the pretest probability was 65%, whereas the posttest probability increased to 75% using model B, with a posttest negative probability of 18%, as shown in Figure 4. These results highlight the value of using defined OSA phenotypes as predictors of OSA risk in referred individuals.

Figure 4. Fagan nomogram for model B. The blue and red lines represent the positive and negative posttest probability, respectively.



Discussion

Principal Findings

Understanding OSA patterns is important, particularly in the diagnosis of OSA. The AASM task force affirmed that the evaluation with clinical tools, such as clinical prediction algorithms, was less burdensome to the patient and physicians when compared with PSG. However, their low levels of accuracy and the likelihood of misdiagnosis must be weighted. Therefore, they proposed a clinical algorithm for the implementation of clinical practice guidelines for OSA. In the second step of this algorithm, the increased risk of moderate to severe OSA is measured by the presence of excessive daytime sleepiness and at least two of the following 3 criteria: habitual loud snoring, witnessed apnea or gasping or choking, or diagnosed hypertension. When we applied this moderate to severe risk in our data set (n=318), we found a sensitivity of 29%, a specificity of 68%, a positive predictive value of 50%, and a positive likelihood ratio of 0.875, showing possible benefits for a rule-out approach. However, considering the target of moderate to severe OSA identification, this approach revealed a very low level of sensitivity for a rule-in approach, which would be expected in this case.

To the best of our knowledge, this study is the first attempt to explore different clinical phenotypes of patients with OSA using categorical cluster analysis combined with Bayesian networks. We applied a hierarchical clustering procedure using Ward linkage on 14 significant predictive variables (out of the tested 47) that were grouped into 3 clusters: low, medium, and high severity phenotypes. These phenotypes were then used to expand a clinical prediction algorithm based on Bayesian networks, creating a simple but complete and updatable tool for OSA screening that can deal with missing information, based only on clinical and demographic variables, which have the main advantage of being easily available and quickly acquired by physicians.

Cluster analysis has been used in many medical conditions aiming to identify clinical phenotypes, as in the case of patients with asthma [16], where 5 clinical phenotypes illustrated the heterogeneity of the disease and relevant differences in treatment. Regarding OSA, clustering had been discussed as a possible helpful tool back in 1992, where the work of Tsuchiya et al [34] tried to apply cluster analysis in patients with OSA to overcome the stated overemphasis regarding obesity, which may have caused some physicians to overlook other potential factors that predispose this condition. They considered the apnea index (the standard at the time) and applied hierarchical clustering with average linkage, resulting in 2 clusters. The authors highlighted the controversy on the number of clusters, stating that "it should be essential to determine the number of clusters in a realistic way, and also to interpret the structures of clusters from a biologic standpoint." Ye et al [35] collected demographic and survey data about sleep-related health issues (using numeric predictive variables) identifying 3 clusters: cluster 1 as *disturbed sleep group*, cluster 2 as *minimally symptomatic group*, and cluster 3 as *excessive daytime sleepiness group*. Although we have studied predictive variables related to daytime sleepiness, none were considered statistically significant; therefore, it is difficult to compare the results of the study by Ye et al [35] with the results of this study. Lacedonia et al [7] developed the work of Ye et al [35], enhancing the results using instrumental data, such as blood gas analysis and spirometry parameters (unavailable to us), to identify clinical presentations of patients with OSA. The authors used 2 approaches: a first one with hierarchical clustering revealing 3 clusters and the second one expanding it to 8 clusters with local optimization through principal component analysis.

Other studies are recently being developed, namely, the broad one in sleep apnea from the Sleep Apnea Network or European Sleep Apnea Database (ESADA) group. In 2016, Saaresranta et al [22] hypothesized that distinct OSA phenotypes should be present when discussing comorbidities and adherence to nasal continuous positive airway pressure (CPAP) therapy. This study has 3 main differences from ours: the ESADA database accepted PSG and cardiorespiratory polygraphy, whereas we only accepted PSG results; they accepted CPAP therapy and divided their patients into categories based only on subjective daytime sleepiness and nocturnal complaints. Regarding this last aspect, in our study, both subjective excessive daytime sleepiness and Epworth Sleepiness Scale were not considered in the cluster analysis. In 2020, a study by Bailly et al [21] applied latent class analysis to identify OSA phenotypes while reflecting geographical variations, resulting in 8 distinct clusters that were divided into 2 main categories: gender-based phenotypes (clusters 2 and 6 with only men and clusters 7 and 8 with only women) and men with various combinations (clusters 1, 3, 4, and 5), with which we can compare results. Cluster 3 of the study by Bailly et al [21] is described as obese comorbid patients, being the most similar to our low severity OSA cluster, presenting almost the same percentage of males (69% vs 73%) and higher levels of metabolic comorbidities.

Our results suggest 3 OSA phenotypes that can help in the screening, diagnosis, and later treatment of patients with OSA, capturing the full OSA spectrum of patients, focusing our attention on a detailed description of patients with OSA and not on a stereotypical one, where only a few *typical* symptoms such as snoring or daytime sleepiness are analyzed. To augment awareness of this prevalent disease, we even analyzed healthy patients to determine whether we could use the created phenotypes as identifiers of precursors of OSA.

Strengths and Limitations

This study had a modest number of patients, mainly because of the short period for data collection, which was performed in a small district hospital. Nevertheless, we believe that the procedure and the results are relevant. We also acknowledge that our phenotypes are not fully in accordance with the clinical phenotyping experience, particularly those regarding upper airway morphology. We suppose that the inclusion of other relevant outcome data could create a more robust analysis of the determined phenotypes. The inclusion of more patients and even dissociating variables, such as craniofacial upper airway abnormalities, could benefit future research.

The major strengths of this study are the study of a clinical cohort representing patients with OSA with all levels of severity

and the inclusion of a comprehensive number of risk and diagnostic factors that enhance our understanding of OSA diagnosis, with an overall cross-validated discriminative power of AUC of 77%, improving the specificity of a (designed) 95% sensitivity rule-out clinical prediction algorithm (3 to 1 odds for a positive result and 1 to 5 odds for a negative result). In addition, a diagnostic odds ratio higher than 1 was observed for models A and B, supporting the effectiveness of both models, with model B (inclusion of the disease phenotypes) doubling the diagnostic model performance. To assess the validity of our approach, we evaluated a logistic regression model in the derivation cohort, with and without predefined clusters, which highlighted the added discrimination value of using OSA phenotypes as a predictive variable (81% vs 83%). Moreover, we are aware that several clinical questionnaires (Berlin, STOP-BANG [snoring, tiredness, observed apnea, blood pressure, body mass index, age, neck circumference and gender], and NoSAS [neck, obesity, snoring, age, sex]) are helpful in identifying patients who are at risk of OSA. The Berlin questionnaire, when applied to the general population, reaches values of 37% for sensitivity and 84% for specificity, whereas when applied to primary care patients, the values are 86% and 77% [36], respectively. If we look at the STOP-BANG questionnaire, validation was performed in preoperative patients; the sensitivity and specificity values are 84% and 39%, respectively, for OSA diagnosis [37]. Finally, the NoSAS score was validated for the general population; the sensitivity values varied between 79% and 85%, the specificity varied between 69% and 77%, and AUC varied between 74% and 81% [38]. Comparing these results with our results, we can see that our sensitivity has the highest value, as we aim to establish a rule-out approach. On the other hand, our values for specificity and AUC were lower, only comparable with the value obtained for STOP-BANG.

Conclusions

We can affirm that using OSA phenotypes as predictors allows the creation of sensitive tools, with the defined phenotypes being a reflection of the early expression and the natural history of OSA. Nevertheless, OSA and individual responses are not static and evolve with time, creating the need for further studies on evaluating the phenotyping fluctuations and determining their long-term diagnosis implications.

Acknowledgments

DFS acknowledges Fundacão para a Ciência e Tecnologia under PhD grant (PD/BD/13553/2018) and the PhD Program in Clinical and Health Services Research (PD/00003/2013).

Authors' Contributions

DFS and PPR designed the study. DFS extracted the data. All authors screened the article, analyzed and interpreted the data, produced and revised all important intellectual content, gave their final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive analysis of patients with obstructive sleep apnea after missing data imputation (absolute and relative frequencies are presented, and <italic>P</italic> values are the result of chi-square tests unless otherwise specified). Footnote a: <italic>P</italic><.20; these values are italicized. Footnote b: Fisher exact test. Footnote c: the odds ratio was calculated for moderate and severe levels combined because of the absence of patients with normal abdominal circumference at the severe level. [PNG File , 240 KB-Multimedia Appendix 1]

References

1. Kapur VK, Auckley DH, Chowdhuri S, Kuhlmann DC, Mehra R, Ramar K, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American Academy of Sleep Medicine Clinical Practice guideline. J Clin Sleep Med 2017 Mar 15;13(3):479-504 [FREE Full text] [doi: 10.5664/jcsm.6506] [Medline: 28162150]
2. Peppard PE, Young T, Barnet JH, Palta M, Hagen EW, Hla KM. Increased prevalence of sleep-disordered breathing in adults. Am J Epidemiol 2013 May 01;177(9):1006-1014 [FREE Full text] [doi: 10.1093/aje/kws342] [Medline: 23589584]
3. Ravesloot MJ, van Maanen JP, Hilgevoord AA, van Wagensveld BA, de Vries N. Obstructive sleep apnea is underrecognized and underdiagnosed in patients undergoing bariatric surgery. Eur Arch Otorhinolaryngol 2012 Jul 5;269(7):1865-1871 [FREE Full text] [doi: 10.1007/s00405-012-1948-0] [Medline: 22310840]
4. Johnson KG, Johnson DC. Frequency of sleep apnea in stroke and TIA patients: a meta-analysis. J Clin Sleep Med 2010 Apr 15;06(02):131-137. [doi: 10.5664/jcsm.27760]
5. Benjafield AV, Ayas NT, Eastwood PR, Heinzer R, Ip MS, Morrell MJ, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med 2019 Aug;7(8):687-698. [doi: 10.1016/s2213-2600(19)30198-5]

6. Kapur V, Blough D, Sandblom R, Hert R, de Maine JB, Sullivan SD, et al. The medical cost of undiagnosed sleep apnea. Sleep 1999 Sep 15;22(6):749-755. [doi: 10.1093/sleep/22.6.749] [Medline: 10505820]
7. Lacedonia D, Carpagnano GE, Sabato R, Storto MM, Palmiotti GA, Capozzi V, et al. Characterization of obstructive sleep apnea-hypopnea syndrome (OSA) population by means of cluster analysis. J Sleep Res 2016 May 18;25(6):724-730. [doi: 10.1111/jsr.12429] [Medline: 27191534]
8. Kushida CA, Littner M, Morgenthaler T, Alessi CA, Bailey D, Coleman J, et al. Practice parameters for the indications for polysomnography and related procedures: an update for 2005. Sleep 2005 Apr;28(4):499-521. [doi: 10.1093/sleep/28.4.499] [Medline: 16171294]
9. Campos-Rodriguez F, Martinez-Garcia MA, Cruz-Moron I, Almeida-Gonzales C, Catalan-Serra P, Montserrat JM. Cardiovascular mortality in women with obstructive sleep apnea with or without continuous positive airway pressure treatment. Ann Intern Med 2012 Jun 05;156(2):115-122. [doi: 10.7326/0003-4819-156-2-201201170-00006] [Medline: 22250142]
10. Shi Q, Rodrigues P. Monitoring the effectiveness of clinical guidelines: is the recommendation still valid? In: Proceedings of the IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). 2018 Jun Presented at: IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS); June 18-21, 2018; Karlstad, Sweden p. 304-309 URL: https://www.computer.org/csdl/proceedings-article/cbms/2018/606001a304/12OmNxGAKWQ [doi: 10.1109/cbms.2018.00060]
11. Kent BD, Grote L, Ryan S, Pépin J, Bonsignore MR, Tkacova R, et al. Diabetes mellitus prevalence and control in sleep-disordered breathing. Chest 2014 Oct;146(4):982-990. [doi: 10.1378/chest.13-2403] [Medline: 24831859]
12. Kakkar RK, Berry RB. Positive airway pressure treatment for obstructive sleep apnea. Chest 2007 Sep;132(3):1057-1072. [doi: 10.1378/chest.06-2432] [Medline: 17873201]
13. Nickerson J, Lee E, Nedelman M, Aurora RN, Krieger A, Horowitz CR. Feasibility of portable sleep monitors to detect obstructive sleep apnea (OSA) in a vulnerable urban population. J Am Board Fam Med 2015 Mar 06;28(2):257-264 [FREE Full text] [doi: 10.3122/jabfm.2015.02.140273] [Medline: 25748767]
14. Ferreira-Santos D, Monteiro-Soares M, Rodrigues PP. Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis. Stud Health Technol Inform 2018;247:126-130 [FREE Full text] [doi: 10.3233/978-1-61499-852-5-126] [Medline: 29677936]
15. Gallo C, Capozzi V. Clustering techniques for revealing gene expression patterns. In: Encyclopedia of Information Science and Technology, Third Edition. Hershey, PA: IGI Global; 2015:438-447.
16. Moore WC, Meyers D, Wenzel S, Teague WG, Li H, Li X, National Heart, Lung,Blood Institute's Severe Asthma Research Program. Identification of asthma phenotypes using cluster analysis in the Severe Asthma Research Program. Am J Respir Crit Care Med 2010 Feb 15;181(4):315-323 [FREE Full text] [doi: 10.1164/rccm.200906-0896OC] [Medline: 19892860]
17. Garcia-Aymerich J, Gomez FP, Benet M, Farrero E, Basagana X, Gayete A, PAC-COPD Study Group. Identification and prospective validation of clinically relevant chronic obstructive pulmonary disease (COPD) subtypes. Thorax 2011 May;66(5):430-437. [doi: 10.1136/thx.2010.154484] [Medline: 21177668]
18. Docampo E, Collado A, Escaramís G, Carbonell J, Rivera J, Vidal J, et al. Cluster analysis of clinical data identifies fibromyalgia subgroups. PLoS One 2013 Sep 30;8(9):e74873 [FREE Full text] [doi: 10.1371/journal.pone.0074873] [Medline: 24098674]
19. Erro R, Vitale C, Amboni M, Picillo M, Moccia M, Longo K, et al. The heterogeneity of early Parkinson's disease: a cluster analysis on newly diagnosed untreated patients. PLoS One 2013 Aug 1;8(8):e70244 [FREE Full text] [doi: 10.1371/journal.pone.0070244] [Medline: 23936396]
20. Topîrceanu A, Udrescu L, Udrescu M, Mihaicuta S. Gender phenotyping of patients with obstructive sleep apnea syndrome using a network science approach. J Clin Med 2020 Dec 12;9(12):4025 [FREE Full text] [doi: 10.3390/jcm9124025] [Medline: 33322816]
21. Bailly S, Grote L, Hedner J, Schiza S, McNicholas WT, Basoglu OK, ESADA Study Group. Clusters of sleep apnoea phenotypes: a large pan‐European study from the European Sleep Apnoea Database (ESADA). Respirology 2021 Apr;26(4):378-387. [doi: 10.1111/resp.13969] [Medline: 33140467]
22. Saaresranta T, Hedner J, Bonsignore MR, Riha RL, McNicholas WT, Penzel T, ESADA Study Group. Clinical phenotypes and comorbidity in European sleep apnoea patients. PLoS One 2016 Oct 4;11(10):e0163439 [FREE Full text] [doi: 10.1371/journal.pone.0163439] [Medline: 27701416]
23. Core R Team. R: a language and environment for statistical computing. In: R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
24. Weisberg S, Fox J. An R Companion to Applied Regression. Thousand Oaks, California: SAGE Publications; 2011:1-472.
25. Warnes GR, Bolker B, Lumley T. gmodels: various R programming tools for model fitting. In: R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
26. Aragon TJ. epitools: epidemiology tools. In: R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2017.
27. Comtois D. summarytools: tools to quickly and neatly summarize data. In: R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2019.

28. Torgo L. Data mining with R - learning with case studies. Minneapolis, USA: Chapman and Hall/CRC; Jun 20, 2020.
29. Scutari M. Learning Bayesian networks with the package. J Stat Soft 2010;35(3):1-22. [doi: 10.18637/jss.v035.i03]
30. Højsgaard S. Graphical independence networks with the grain package for R. J Stat Soft 2012 Feb 28;46(10):12031 [FREE Full text] [doi: 10.18637/jss.v046.i10]
31. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011 Mar 17;12:77 [FREE Full text] [doi: 10.1186/1471-2105-12-77] [Medline: 21414208]
32. Darwiche A. Modeling and Reasoning with Bayesian Networks. California, Los Angeles: Cambridge University Press; 2009.
33. Darwiche A. Bayesian networks. Commun ACM 2010 Dec;53(12):80-90 [FREE Full text] [doi: 10.1145/1859204.1859227]
34. Tsuchiya M, Lowe AA, Pae E, Fleetham JA. Obstructive sleep apnea subtypes by cluster analysis. Am J Orthod Dentofac Orthop 1992 Jun;101(6):533-542. [doi: 10.1016/0889-5406(92)70128-w] [Medline: 1598893]
35. Ye L, Pien GW, Ratcliffe SJ, Björnsdottir E, Arnardottir ES, Pack AI, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. Eur Respir J 2014 Sep 03;44(6):1600-1607. [doi: 10.1183/09031936.00032314] [Medline: 25186268]
36. Netzer NC, Stoohs RA, Netzer CM, Clark K, Strohl KP. Using the Berlin Questionnaire to identify patients at risk for the sleep apnea syndrome. Ann Intern Med 1999 Oct 05;131(7):485-491. [doi: 10.7326/0003-4819-131-7-199910050-00002] [Medline: 10507956]
37. Chung F, Yang Y, Brown R, Liao P. Alternative scoring models of STOP-bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. J Clin Sleep Med 2014 Sep 15;10(9):951-958 [FREE Full text] [doi: 10.5664/jcsm.4022] [Medline: 25142767]
38. Marti-Soler H, Hirotsu C, Marques-Vidal P, Vollenweider P, Waeber G, Preisig M, et al. The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. Lancet Respir Med 2016 Sep;4(9):742-748. [doi: 10.1016/s2213-2600(16)30075-3] [Medline: 27321086]

Abbreviations

AASM: American Academy of Sleep Medicine
AHI: apnea-hypopnea index
AUC: area under the curve
CPAP: continuous positive airway pressure
ESADA: European Sleep Apnea Database
NN: nearest neighbor
NoSAS: neck, obesity, snoring, age, sex
OSA: obstructive sleep apnea
PSG: polysomnography
ROC: receiver operating characteristic
STOP-BANG: snoring, tiredness, observed apnea, blood pressure, body mass index, age, neck circumference and gender

Edited by R Kukafka; submitted 20.10.20; peer-reviewed by T Penzel, B Sébastien; comments to author 23.12.20; revised version received 22.01.21; accepted 16.03.21; published 22.06.21

Please cite as:
Ferreira-Santos D, Rodrigues PP
Enhancing Obstructive Sleep Apnea Diagnosis With Screening Through Disease Phenotypes: Algorithm Development and Validation
JMIR Med Inform 2021;9(6):e25124
URL: https://medinform.jmir.org/2021/6/e25124
doi: 10.2196/25124
PMID:

4.7. Obstructive sleep apnea: a categorical cluster analysis and visualization.

Pulmonology. 16:53, 2021. doi:10.1016/j.pulmoe.2021.10.003

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

ARTICLE IN PRESS
JID: PULMOE [mSP6P;November 17, 2021;16:53]
Pulmonology 000 (xxxx) 1–7

PULMONOLOGY
www.journalpulmonology.org

ORIGINAL ARTICLE

Obstructive sleep apnea: A categorical cluster analysis and visualization

Daniela Ferreira-Santos[a,b,*], **Pedro Pereira Rodrigues**[a,b]

[a] *MEDCIDS-FMUP - Community Medicine, Information and Decision Sciences, Faculty of Medicine of the University of Porto, Porto, Portugal*
[b] *CINTESIS - Center for Health Technology and Services Research, Porto, Portugal*

Received 16 March 2021; accepted 24 October 2021
Available online xxx

KEYWORDS
Clinical presentations;
Cluster visualization;
Data mining;
Obstructive sleep apnea

Abstract
Introduction and Objectives: Obstructive sleep apnea (OSA) is a prevalent sleep condition which is very heterogeneous although not formally characterized as such, resulting in missed or delayed diagnosis. Cluster analysis has been used in different clinical domains, particularly within sleep disorders. We aim to understand OSA heterogeneity and provide a variety of cluster visualizations to communicate the information clearly and efficiently.
Materials and Methods: We applied an extension of k-means to be used in categorical variables: k-modes, to identify OSA patients' groups, based on demographic, physical examination, clinical history, and comorbidities characterization variables ($n = 40$) obtained from a derivation and validation cohorts (211 and 53, respectively) from the northern region of Portugal. Missing values were imputed with k-nearest neighbours (k-NN) and a chi-square test was held for feature selection.
Results: Thirteen variables were inserted in phenotypes, resulting in the following three clusters: Cluster 1, middle-aged males reporting witnessed apneas and high alcohol consumption before sleep; Cluster 2, middle-aged women with increased neck circumference (NC), non-repairing sleep and morning headaches; and Cluster 3, obese elderly males with increased NC, witnessed apneas and alcohol consumption. Patients from the validation cohort assigned to different clusters showed similar proportions when compared with the derivation cohort, for mild (C1: 56 vs 75%, $P = 0.230$; C2: 61 vs 75%, $P = 0.128$; C3: 45 vs 48%, $P = 0.831$), moderate (C1: 24 vs 25%; C2: 20 vs 25%; C3: 25 vs 19%) and severe (C1: 20 vs 0%; C2: 18 vs 0%; C3: 29 vs 33%) levels. Therefore, the allocation supported the validation of the obtained clusters.
Conclusions: Our findings suggest different OSA patients' groups, creating the need to rethink these patients' stereotypical baseline characteristics.

* Corresponding author at: MEDCIDS - Departamento de Medicina da Comunidade, Informação e Decisão em Saúde, Faculdade de Medicina da Universidade do Porto (CIM - FMUP), Rua Dr. Plácido da Costa, s/n; 4200-450 Porto; Portugal.
E-mail address: danielasantos@med.up.pt (D. Ferreira-Santos).

https://doi.org/10.1016/j.pulmoe.2021.10.003

Please cite this article in press as: D. Ferreira-Santos and P.P. Rodrigues, Obstructive sleep apnea: A categorical cluster analysis and visualization, Pulmonology (2021), https://doi.org/10.1016/j.pulmoe.2021.10.003

Introduction

Obstructive sleep apnea (OSA) is a common sleep disorder characterized by recurrent episodes of collapse of the upper airway during sleep, which is estimated to occur in nearly 1 billion adults aged 30–69 years worldwide and in more than 5 million in Portugal.[1] Although commonly observed in clinical practice, OSA which is a heterogeneous condition with various predisposing factors, pathophysiological mechanisms, clinical manifestations and consequences of respiratory events; has not been formally characterized, posing critical challenges to its clinical recognition and resulting in missed or delayed diagnosis.[2] A diagnosis is established when a patient has an apnea-hypopnea index (AHI) $\geq$5 with associated symptoms or an AHI $\geq$15 regardless of associated symptoms.[3] Nevertheless, it has long been recognized that AHI alone does not capture OSA's patients' heterogeneity, that one-size does not fit all.[4] The review of Zinchuk et al., presents numerous studies that used analytic approaches to gain advantage of this heterogeneity.[4] In addition, Allan Pack studies discuss P4 medicine and how it might be applicable to OSA, specifically if OSA diagnosis or treatment were personalized, providing several examples of this personalized, predictive, preventative and participatory medicine.[5] Cluster analysis, a hypothesis-generating strategy (or unsupervised learning), is being used to identify subtypes of patients with unique characteristics, classifying patients with OSA into smaller and more homogeneous disorder phenotypes. This is a statistical approach that studies the relationships present among groups of participants or variables in a population, aiming at the benefit of allowing more specific diagnosis and treatment strategies. This is accomplished by assessing similarity (or dissimilarity) between subjects using metrics such as correlation or distance based on the features used to characterize each individual. Ideally, each cluster member is as similar as possible to each other and as different as possible from those in other clusters. One approach in cluster analysis is the *k*-modes algorithm.[6] This algorithm extends the *k*-means paradigm to cluster categorical data by using a simple matching dissimilarity measure for categorical objects, modes instead of means for clustering, and a frequency-based method to update modes in the *k*-means fashion clustering process.[7] The dissimilarity measure can be defined by the total mismatches of the two objects' corresponding variable categories: the smaller the number of mismatches, the more similar the two objects are. To better understand OSA heterogeneity, this study applies categorical cluster analysis to various observable and measurable OSA characteristics, such as signs, symptoms, demographics, and comorbidities. It also provides a variety of cluster visualizations to communicate this information clearly and efficiently.

Material and methods

A literature review was previously conducted to identify the most relevant OSA variables to be collected from the medical records and a total of 51 variables were noted, such as: demographic variables (e.g., gender); physical examination (e.g., body mass index (BMI)); clinical history (e.g., snoring); and comorbidities (e.g., stroke).

Derivation cohort

All patients who underwent polysomnography (PSG) at the Vila Nova de Gaia and Espinho Hospital Center Sleep Laboratory were included in the study. All administrative records were collected retrospectively between January and May 2015; patients included were aged above 18 years old and were suspected of having OSA. Patients already diagnosed with OSA or with severe lung diseases or neurological conditions and pregnant women were excluded. In the case of duplicate exams, the one with better sleep efficiency was selected.[8] This study was approved by the Ethics Commission of Vila Nova de Gaia and Espinho Hospital Center, in accordance with the Declaration of Helsinki.

Validation cohort

We have prospectively included adult patients suspected of having OSA referred to perform PSG at the Sleep Laboratory of São João University Hospital, between December 2019 and March 2020, following the previously mentioned inclusion and exclusion criteria. This study was approved by the Ethics Commission of São João University Hospital, in accordance with the Declaration of Helsinki.

Pre-processing phase

Although we had access to all the electronic medical records from the included patients, after screening all unstructured text reports, some predictive variables were not fully present or described, as physicians normally do not mention the absence of a disease, or it could only be noted in paper records (missing data proportions ranged from 0% for gender to 97% for bariatric surgery). Variables with more than 80% missing values were removed from the analysis (e.g., decreased libido). Also, daytime sleepiness exhibited contradictory results (higher percentage of patients in the normal group n = 77; 72%) with statistical significance. This was also described for the Epworth Sleepiness Scale (ESS), which presented a contradiction to the literature and the inherent meaning of the variables, and thus was not considered for statistical analysis. For more details, please refer to supplementary file (A). The outcome measure was OSA clinical diagnosis (AHI$\geq$5 plus excessive daytime sleepiness and at least two of the following three criteria: habitual loud snoring; witnessed apnea or gasping or choking; or diagnosed hypertension) obtained from PSG.[9]

We performed a pre-processing analysis, and continuous variables were categorized, plus *k*-nearest neighbours (*k*-NN) imputation was conducted to preserve all cases with missing data being replaced with a value obtained from related cases from the complete set of records.[10] After chi-square analysis, variables were selected if presenting a significant univariate association with the outcome (AHI), considering a 5% significance level. We used R software[11] to perform descriptive and associative analysis (packages gmodels[12] and epitools[13]), the *k*-modes categorical clustering

(package klaR[14]) and to create standard barplot (ggplot2[15]), heatmap (gplots[16]) and radar chart (fmsb[17]).

Results

Derivation cohort

In total, 211 patients were diagnosed with OSA (66%), of which 115 (55%) were categorized as mild, 50 (24%) as moderate and 46 (22%) as severe. Seventy percent were males with a mean age of 61 (53–68) years old and with OSA. The lower age category, 20–44 years old, had a lower percentage of patients with OSA, oppositely to categories 45–64 and 65–90 ($P<0.001$). Looking only at OSA patients, neck (NC) and abdominal circumference (AC) had a mean of 42 (39–44) cm and 107 (100- 113) cm, and a BMI median value of 30 (27–30) Kg/m^2 ($P = 0.008$). Several variables had no statistical significance but had a higher number of patients in the OSA group, namely craniofacial and upper airway abnormalities (CFA), snoring, nocturia, sleep fragmentation, insomnia, drivers, family history, myocardial infarction, arterial hypertension, pacemaker, stroke, renal failure, dyslipidaemia, and hypothyroidism. In contrast, a higher percentage of normal patients were seen in gasping/choking, behavioural changes, decreased libido, vehicle crashes, coffee, use of sedatives, respiratory alterations, and anxiety/depression.

A total of 13 variables were incorporated into the cluster analysis: gender, age, BMI, NC, modified Mallampati, witnessed apneas, nonrepairing sleep, morning headaches, driving sleepiness, alcohol consumption, congestive heart failure, arrhythmias, and pulmonary hypertension, resulting in three distinct clusters, presented in Table 1. Figure 1, 2 and 3 visually synthesizes the information obtained from Clusters 1 to 3.

Cluster 1 weighted the least and had non-increased neck circumference in middle-aged males. It had a higher percentage of driving sleepiness, and lower percentages in congestive heart failure, arrhythmias, pulmonary hypertension and modified Mallampati. In addition, it had the second higher percentage of witnessed apneas, nonrepairing sleep, morning headaches, and alcohol consumption. The mild severity level had a higher value. Cluster 2 was mainly middle-aged women with increased NC. It had the lowest percentage of witnessed apneas and alcohol consumption, and very low percentages of comorbidities. Finally, Cluster 3 presented older and obese men with increased NC. Witnessed apneas were reported in 75% of the patients, and alcohol consumption, congestive heart failure, and pulmonary hypertension had the highest percentages in all the clusters. In contrast, nonrepairing sleep, morning headaches, and driving sleepiness had the lowest rates. This cluster presented a lower value in the mild severity category and a higher value at the severe level.

Table 1 Clinical characteristics of the obstructive sleep apnea cohort by the defined clusters, % [95%CI].

	Cluster 1 ($n = 112$)	Cluster 2 ($n = 44$)	Cluster 3 ($n = 55$)	*P* value	P(C|F) (c_1, c_2, c_3)
Gender (male)	85 [77–91]	20 [10–36]	80 [67–89]	**<0.001**	(0.64, 0.06, 0.30)
Age (years)				**<0.001***	
20–44	6 [3–13]	11 [4–25]	11 [5–23]		(0.39, 0.28, 0.33)
45–64	65 [56–74]	57 [41–71]	27 [14–41]		(0.65, 0.22, 0.33)
65–90	29 [21–38]	32 [19–48]	62 [48–74]		(0.40, 0.18, 0.42)
Obesity	21 [14–29]	43 [29–59]	76 [63–86]	**<0.001**	(0.27, 0.23, 0.50)
Increased neck circumference	30 [22–40]	77 [62–88]	96 [86–99]	**<0.001**	(0.28, 0.28, 0.44)
Mallampati				**<0.001***	
Class I	22 [15–31]	11 [4–25]	36 [24–50]		(0.50, 0.10, 0.40)
Class II	48 [39–58]	61 [46–75]	5 [1–16]		(0.64, 0.32, 0.04)
Class III	26 [18–35]	23 [12–38]	47 [34–61]		(0.45, 0.15, 0.40)
Class IV	4 [1–9]	5 [1–17]	11 [5–23]		(0.33, 0.17, 0.50)
Witnessed apneas	64 [55–73]	30 [17–45]	75 [61–85]	**<0.001**	(0.57, 0.10, 0.33)
Nonrepairing sleep	46 [36–55]	70 [55–83]	29 [18–43]	**<0.001**	(0.52, 0.32, 0.16)
Morning headaches	45 [35–54]	84 [69–93]	18 [10–31]	**<0.001**	(0.52, 0.38, 0.10)
Driving sleepiness	13 [7–20]	5 [1–17]	2 [0–11]	**.05**	(0.82, 0.12, 0.06)
Alcohol consumption	85 [77–91]	20 [10–36]	91 [79–97]	**<0.001**	(0.62, 0.06, 0.32)
Congestive heart failure	10 [5–17]	14 [6–28]	18 [10–31]	.31	(0.41, 0.22, 0.37)
Arrhythmias	5 [2–12]	14 [6–28]	11 [5–23]	.15	(0.34, 0.33, 0.33)
Pulmonary hypertension	4 [1–9]	9 [3–23]	13 [6–25]	.07	(0.27, 0.27, 0.46)
Obstructive sleep apnea				.50	
Mild	56 [47–66]	61 [46–75]	45 [32–59]		
Moderate	24 [17–33]	20 [10–36]	25 [15–39]		
Severe	20 [13–28]	18 [9–33]	29 [18–43]		

* Fisher's exact test; $P<0.05$ are presented in bold; CI: confidence interval; P(C|F): probability of belonging to a cluster given the presence of a factor.

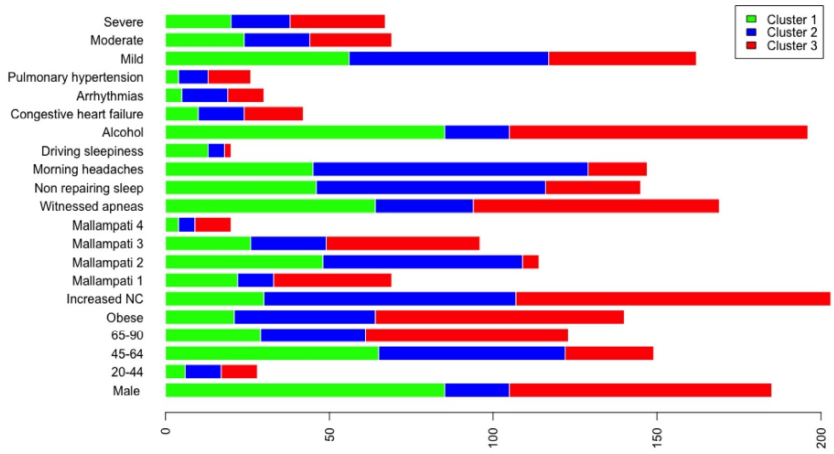


Fig. 1 Clinical characteristics of the obstructive sleep apnea cohort in Cluster 1, 2 and 3 visualized in a bar plot.

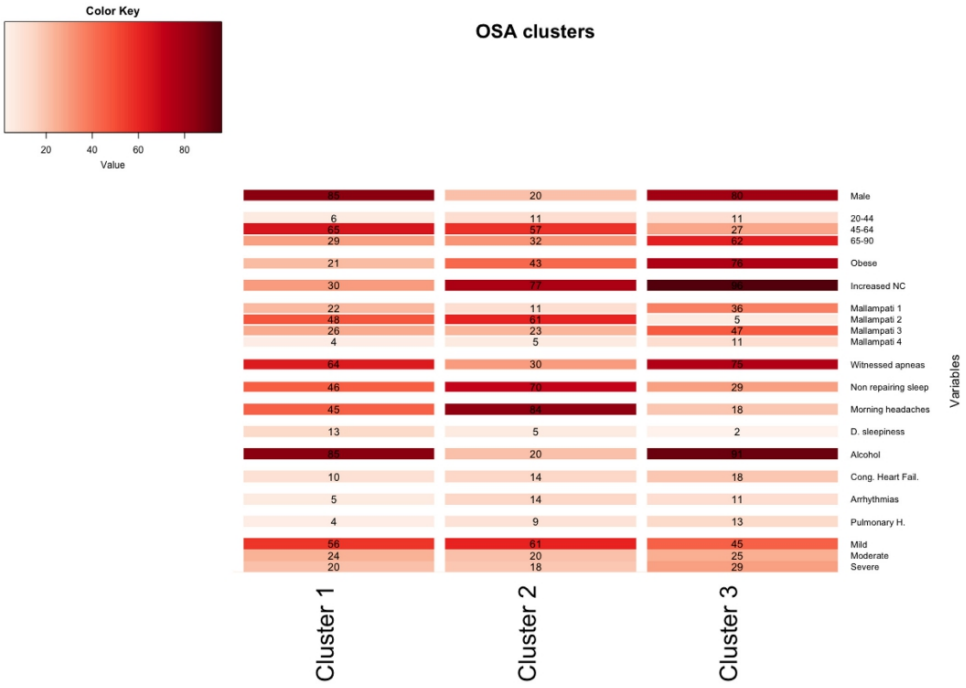


Fig. 2 Percentages of each clinical characteristics in obstructive sleep apnea patients' phenotypes visualized in a heatmap.

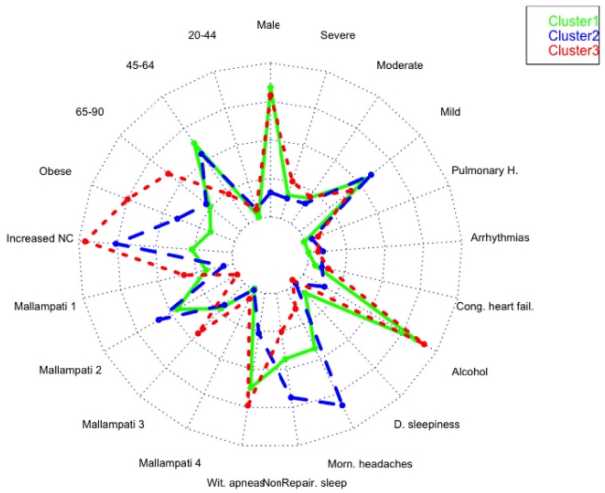


Fig. 3 Radar plot of obstructive sleep apnea patients' clinical characteristics distribution by cluster.

Validation cohort

Prospective data collection resulted in 53 OSA patients, 51% male. The median age was 59 (22–80) years old, with 8 (15%) in the 20–44 years old stratum, 30 (57%) in 45–64, and 15 (28%) in 65–90. Regarding NC and BMI, the median was 40 (31–48) cm and 31 (22–44) Kg/m^2, respectively. The highest proportion of patients (n = 35 (66%)) had been diagnosed with mild OSA, followed by moderate (11 (21%)), and severe (7 (13%)) levels. Concerning clinical history, witnessed apneas were reported in 64% of patients, followed by non-repairing sleep (26, 49%) and morning headaches (18, 34%). The most prevalent comorbidity was arrhythmias (4, 8%).

We calculated the dissimilarity between each new patient and the obtained clusters, based on the proposal of Huang,[6] as previously mentioned. Afterwards, we created a matrix and assigned the patients to their closest cluster (i.e., the cluster with the lowest dissimilarity), allocating 7 in Cluster 1, 17 in Cluster 2, and 18 in Cluster 3. Eleven patients had the same dissimilarity measure in two (n = 6) or three (5) clusters; we then assigned them randomly to any cluster, by generating 30 random assignments. This resulted in a mean (sd) allocation of 12 (1) in Cluster 1, 20 (1) in Cluster 2, and 21 (2) in Cluster 3. When comparing the proportion of patients in the derivation and validation cohorts in all clusters, we discovered no statistical differences for most variables (gender, age, witnessed apneas, non-repairing sleep, morning headaches, driving sleepiness, congestive heart failure, arrhythmias, pulmonary hypertension) and for the outcome, as showed in supplementary file (B). Regarding BMI and NC, statistical significance was found only in Cluster 1 (P = 0.001 and P = 0.024, respectively), while Mallampati and alcohol consumption was in Cluster 1 (P = 0.043 and P = 0.007) and 3 (P = 0.002 and P<0.001).

Discussion

In line with other studies, this research contributes to understanding OSA heterogeneity through exploring possible phenotypes by applying categorical cluster analysis (k-modes) in all severity levels. Our findings confirm that patients' phenotypes can be identified in the OSA population referred to the sleep laboratory (three in our case). Cluster 1 and 3 were middle-aged males or elderly, while Cluster 2 was middle-aged women. We verified that physical examination aspects, such as BMI and NC, were lower in Cluster 1 but extremely high in Cluster 3, especially NC. Regarding clinical history, the overall percentage of modified Mallampati was placed in the lower levels (Mallampati 1 and 2) and driving sleepiness percentages were low (higher value of 13% in Cluster 1). Alcohol consumption pre-sleep was widely described in Cluster 1 and 3. Reported witnessed apneas were higher in Cluster 3 and 1, while nonrepairing sleep and morning headaches in Cluster 2. All the chosen comorbidities (congestive heart failure, arrhythmias, and pulmonary hypertension) were low and without statistical significance in our analysis. The percentage of the outcome measure (AHI) was demonstrated in each cluster; Cluster 2 had 61% [46%–75%] of mild severity, followed by Cluster 1 (56% [47%–66%]) and Cluster 3, with a significantly smaller proportion (45% [32%–59%]), and high percentage in the severe level (29% [18%–43%]. When we observe the probability of belonging to a cluster given the presence of a factor, we can notice that factors such as male gender, lower levels of age

and modified Mallampati, witnessed apneas, nonrepairing sleep, morning headaches, driving sleepiness, alcohol consumption, congestive heart failure and arrhythmias put the patient in Cluster 1. On the other hand, elderly obese adults with increased NC, a level 4 modified Mallampati, and pulmonary hypertension, are allocated to Cluster 3. As a result, when applying our approach to the classification of a new case, physicians should choose the cluster with the greatest number of similar variables states, considering the ten statistically significant variables identified: gender, age, BMI, NC, modified Mallampati, witnessed apneas, nonrepairing sleep, morning headaches, driving sleepiness, and alcohol consumption. The validation cohort of OSA patients, confirmed the previous statistically significant variables as gender, age, witnessed apneas, nonrepairing sleep, morning headaches, and driving sleepiness. Additionally, BMI and NC can be considered in Cluster 2 ($P = 0.051$) and 3 ($P \geq 0.999$), while modified Mallampati and alcohol consumption only in Cluster 2 ($P = 0.417$ and $P = 0.085$), respectively.

Several authors have hypothesized the possible presence of OSA phenotypes, with the expected benefit of allowing a more precise diagnosis and treatment strategy, leading to improve patients outcomes, as described in the review of Zinchuk et al..[4] The referred studies were performed between 2012 and 2019 (a total of 17), with eight studies applying latent class analysis (LCA) as the clustering method, three applying hierarchical clustering (with multiple correspondences analysis for feature selection or principal component analysis (PCA)), one utilizing hierarchical clustering and *k*-means, two using *k*-means (with multiple correspondence for feature selection or PCA), one with *k*-means, one with time-series analysis and dynamic cluster analysis, and one exploiting *k*-modes (our previous work). Furthermore, the authors referred that these studies differ in terms of individuals included, sample size, patients features and outcomes, presented in the review. We also would like to point out that while our work chose PSG, as the standard diagnosis tool, other studies elected home studies or PSG mixed with home studies (seven studies). In the studies that selected PSG as the gold standard, different AHI cut-offs were applied, namely one study employed AHI ≥ 5[18] and two AHI ≥ 15,[19,20] while two other studies do not define the cut-off.[21,22] Analysing our results to the one that selected the same AHI cut-off, we noticed some differences. First, the study collected data from PSG, namely position, sleep state and arousals, while ours did not pre-select these variables as we focused our analysis only on pre-diagnostic data. Also, in our work we analysed the data as categorical while this study was numerical, reaching 6 clusters, all related to sleep position or sleep state. The only common aspect is that, like our study, it did not report outcomes. The two studies that reported a cut-off higher than 15 events per hour collected data from PSG in a supine position (not collected in our study) or symptoms and ESS. In this case, LCA was performed, and 4 clusters were found: 1) disturbed sleep; 2) minimally symptomatic; 3) excessively sleep, and 4) moderately sleepy. Although we could compare our results to this study, it is not our aim to assess the association with prevalent and incident cardiovascular disease. The main limitation of cluster analysis is that the phenotypes are always dependent on the number and the quality of the selected variables. Nevertheless, we believe that the process and the results are relevant. The inclusion of a comprehensive number of risk and diagnostic factors, as well as the observed robustness of the cluster-defined phenotypes in the validation cohort, enhances our understanding of OSA heterogeneity, emphasizing that the stereotypical OSA patient needs to be redefined.

Declarations of interest

None.

Acknowledgments

The work of DFS was supported by the Fundação para a Ciência e a Tecnologia (FCT) under Ph.D. grant number PD/BD/13553/2018 and the PhD Program in Clinical and Health Services Research (PD/00003/2013). We would like to acknowledge Pedro Amorim, for his valuable contribution regarding validation data collection.

Role of the funding source

The institutions who provided financial support had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report or in the decision to submit the article for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.pulmoe.2021.10.003.

References

1. Benjafield AV, Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med. 2019;7(8):687–98. https://doi.org/10.1016/S2213-2600(19)30198-5.
2. Ye L, Pien GW, Ratcliffe SJ, et al. The different clinical faces of obstructive sleep apnoea: a cluster analysis. Eur Respir J. 2014;44(6):1600–7. https://doi.org/10.1183/09031936.00032314.
3. Flemons WW, Buysse D, Redline S, et al. Sleep-related breathing disorders in adults: rrecommendations for syndrome definition and measurement techniques in clinical research. The report of an American Academy of Sleep Medicine Task Force. Sleep. 1999;22(5):667–89. https://doi.org/10.1093/sleep/22.5.667.
4. Zinchuk A, Yaggi HK. Phenotypic subtypes of OSA: aa challenge and opportunity for precision medicine. Chest. 2020;157(2):403–20. https://doi.org/10.1016/j.chest.2019.09.002.
5. Pack AI. Application of personalized, predictive, preventative, and participatory (P4) medicine to obstructive sleep apnea: a roadmap for improving care? Ann Am Thorac Soc. 2016;13(9):1456–67. https://doi.org/10.1513/AnnalsATS.201604-235PS.
6. Huang Z. A fast clustering algorithm to cluster very large categorical data sets in data mining. In: DMKD; 1997. p. 1–8.

7. Kaufman L., Rousseeuw P.J. Finding Groups in Data: An Introduction to Cluster Analysis. (Kaufman L, Rousseeuw PJ, eds.). John Wiley & Sons, Inc.; 1990. doi:10.1002/9780470316801
8. Ferreira-Santos D, Rodrigues PP. Improving diagnosis in obstructive sleep apnea with clinical data: aa bayesian network approach. In: IEEE 30th International Symposium on Computer-Based Medical Systems, Institute of Electrical and Electronics Engineers Inc.; 2017:612—7. https://doi.org/10.1109/CBMS.2017.19.
9. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: aan American academy of sleep medicine clinical practice guideline. J Clin Sleep Med. 2017;13(3):479—504. https://doi.org/10.5664/jcsm.6506.
10. Ferreira-Santos D, Monteiro-Soares M, Rodrigues PP. Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis. Studies in Health Technology and Informatics. Vol 247. Zagreb, Croatia: IOS Press; 2018. p. 126—30. https://doi.org/10.3233/978-1-61499-852-5-126.
11. Team R.C. R: A Language and Environment for Statistical Computing, Vienna, Austria. Published online 2021.
12. R. GW, Bolker B., Lumley T., C. RJ. gmodels: Various R Programming Tools for Model Fitting, Bethesda, Maryland. Published online 2018.
13. T.J. Aragon. Epitools: Epidemiology Tools. Published online 2020.
14. Weihs C, Ligges U, Luebke K, Raabe N. KlaR Analyzing German Business Cycles. Dortmund, Germany: Springer-Verlag; 2005.
15. Wickham H. Ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag; 2016.
16. Warnes G.R., Bolker B., Bonebakker L., et al. gplots: Various R Programming Tools for Plotting Data. Published online 2020.
17. Nakazawa M. fmsb: Functions for Medical Statistics Book with some Demographic Data. Published online 2021.
18. Joosten SA, Hamza K, Sands S, Turton A, Berger P, Hamilton G. Phenotypes of patients with mild to moderate obstructive sleep apnoea as confirmed by cluster analysis. Respirology. 2012;17(1):99—107. https://doi.org/10.1111/j.1440-1843.2011.02037.x.
19. Mazzotti DR, Keenan BT, Lim DC, Gottlieb DJ, Kim J, Pack AI. Symptom subtypes of obstructive sleep apnea predict incidence of cardiovascular outcomes. Am J Respir Crit Care Med. 2019;200(4):493—506. https://doi.org/10.1164/rccm.201808-1509OC.
20. Nakayama H, Kobayashi M, Tsuiki S, Yanagihara M, Inoue Y. Obstructive sleep apnea phenotypes in men based on characteristics of respiratory events during polysomnography. Sleep Breath. 2019;23(4):1087—94. https://doi.org/10.1007/s11325-019-01785-8.
21. Vavougios GD, George DG, Pastaka C, Zarogiannis SG, Gourgoulianis KI. Phenotypes of comorbidity in OSAS patients: combining categorical principal component analysis with cluster analysis. J Sleep Res. 2016;25(1):31—8. https://doi.org/10.1111/jsr.12344.
22. Zinchuk AV, Jeon S, Koo BB, et al. Polysomnographic phenotypes and their cardiovascular implications in obstructive sleep apnoea. Thorax. 2018;73(5):472—80. https://doi.org/10.1136/thoraxjnl-2017-210431.

Chapter 5: Implementation

Implementation

The third goal of this thesis is to build a tool that can improve the identification of OSA patients and see if we can transparently integrate another clinical decision tool in Portuguese care, by developing an online form and application to be used in daily consultation in adults with suspicion of OSA.

Two studies were conducted and corresponds to Implementation – Phase 6 in CRISP-DM.

5.1. Prospective validation of a Bayesian network model in the diagnosis of Obstructive Sleep Apnea
(submitted, Dez 2021)
Pedro Amorim, Daniela Ferreira-Santos, Marta Drummond, Pedro Pereira Rodrigues

5.2. Validation of OSABayes: a screening tool for obstructive sleep apnea
(finalizing details)
Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Both studies aimed at validating OSABayes, an auxiliary tool that can support the decision to undergo polysomnography. This tool was implemented an an online form and as an application for smartphones. It derives from article 4.5., which presented a tree-augmented Naïve Bayes network with six predictive variables.

Initially, the Bayesian model was developed with 194 patients (derivation cohort) and is now being externally validated with 216 and 619 patients in two cohorts from the same hospital. The main difference is that the first study establishes a new cut-off for the presence of OSA, and the second kept the previously defined (39%). Both studies have sensitivities above 95%, proving overall good results for OSABayes implementation.

Additionally, the student Pedro Amorim in his master's thesis, evaluated the app usability with a cognitive walkthrough method, analyzing 7 tasks performed by primary healthcare physicians and other specialities that refer more patients to the Sleep Laboratory. The results were that 82% of the tasks were easily and successfully performed, 14% were performed but with difficulty, and 4% were not concluded.

5.1. Prospective validation of a Bayesian network model in the diagnosis of Obstructive Sleep Apnea (finalizing details)

Pedro Amorim, Daniela Ferreira-Santos, Marta Drummond, Pedro Pereira Rodrigues

Abstract

Classification of obstructive sleep apnea (OSA) relies on results from polysomnography (PSG). Current guidelines recommend the development of clinical prediction algorithms in screening prior to PSG. A recent intuitive and user-friendly tool (OSABayes), based on a Bayesian network model using six clinical variables (gender, age, neck circumference, craniofacial abnormalities, witnessed apneas and nocturia), has been proposed to quantify the probability of OSA. Our aim is to prospectively validate OSABayes by comparing predicted probabilities with the apnea-hypopnea index (AHI) value from PSG.

For this purpose, we prospectively included adult patients with suspicion of OSA, without suspicion of other sleep disorders, who underwent level I or III diagnostic PSG at the sleep laboratory of São João University Hospital. AHI and OSABayes probabilities were obtained and compared using area under the ROC curve (AUC [95%CI]) for OSA diagnosis (AHI≥5/h) and for higher severity levels (AHI≥15/h) prediction.

Thus, a total of 216 subjects were included, with a mean (sd) age of 57 (13) years old, 55% males, performing PSG level I (34%) and level III (66%). Of the 177 (82%) subjects diagnosed with OSA, 78 (44%) were mild, 56 (32%) moderate and 43 (24%) had severe OSA. OSABayes presented an AUC of 83.6% [77.3-90.0%] for OSA diagnosis and 76.3% [69.9-82.7%] for moderate/severe OSA predicting, showing good response for both types of PSG.

These results show good discrimination power of OSABayes and validate its applicability in the identification of patients with high pre-test probability of OSA.

Keywords

Obstructive sleep apnea, Diagnosis, Polysomnography, Bayesian network, Artificial intelligence

Introduction

Obstructive sleep apnea (OSA) is a common sleep-related breathing disorder characterized by partial or complete collapse of the upper airway, repeated throughout sleep [1,2]. This may result in fragmented sleep, intermittent hypoxia and hypercapnia, intrathoracic pressure swings and increased sympathetic nerve activity [2,3]. Thus, it's possible that patients report not only witnessed apneas or snoring, but also nocturia, excessive sleepiness, non-restful sleep, morning headaches, decreased concentration, memory loss, or irritability [1,3–7].

OSA affects about 926 million adults worldwide [8], and untreated patients may be at increased risk of developing cardiovascular disease, metabolic dysregulation, diabetes, or neuropsychiatric conditions such as depression [1,4,5,9,10]. OSA symptoms and its association to other diseases reinforce the importance of early and accurate diagnosis [9]. However, despite being often associated with cardiovascular and neurocognitive deficits and symptoms, available data suggest that many cases of OSA remain undiagnosed and untreated [8].

The diagnosis of OSA requires either signs/symptoms or associated disorders coupled with predominantly obstructive apnea or hypopnea index (AHI) $\geq$ 5 events per hour (/h) in polysomnography (PSG). Alternatively, a frequency of obstructive respiratory events $\geq$ 15/h satisfies the criteria, even in the absence of associated symptoms or disorders [11]. Since a full PSG (PSG I) cannot be performed for all patients, due to high prevalence of OSA and the great demand for tests, portable recording devices (PSG III) have been developed to allow laboratory-equivalent diagnosis, although with fewer channels. These devices are particularly interesting in patients with a high probability of moderate to severe OSA pre-test [12,13].

The OSA diagnosis relies on PSG results, but its costs and waiting times are significant [3]. Therefore, tools that identify patients at high risk of OSA are important and, when integrated into a global approach, can play an important role in the diagnosis of these patients. Further, current American Academy of Sleep Medicine (AASM) guidelines recommend the development of clinical prediction algorithms in screening prior to PSG [1]. In this context, some studies have tried to show that, beyond questionnaires, machine learning methods also can be used in OSA [14,15]; they can prioritize OSA screening, facilitate clinical decision making and save costs [13,16].

One of these methods are Bayesian networks, a model of a joint probability distribution over a set of variables, that allows a general and versatile approach to capturing and reasoning with uncertainty [17]. In Bayesian networks each variable is represented by a node in the network and the edges represent the relationship between the connected nodes. Thus, the probability of one node is dependent on the set of variables represented by its ascendant nodes [17,18]. In OSA, Bayesian networks have already been used to classify respiratory events [19], to predict cardiovascular pathology [20] and to assess the effectiveness of continuous positive airway pressure treatment [21].

In one of these studies, Ferreira-Santos *et al* [22] created an auxiliary diagnostic method – OSABayes - based on Bayesian networks, that could support the decision to perform PSG, based on risk and diagnostic factors. The authors studied 38 variables in a sample of individuals (derivation cohort) with suspected OSA, pre-PSG I, selecting 6 variables with a univariate significant association with the outcome. To build the model, a Tree Augmented Naïve Bayes classifier was used. This approach allows a more natural description of the relationships between variables, admitting up to two dependencies for each factor, which can be important, since, for example, not only OSA causes nocturia, but age as well [23]. So, an intuitive and user-friendly form was created with these 6 variables (gender, age, neck circumference (NC), craniofacial abnormalities (CFA), witnessed apneas and nocturia), that can be used in primary care setting to quantify the probability for having OSA and help referring only the suspected patients to sleep laboratories [24].

Given that OSABayes was validated following an internal cross-validation approach, but have not yet undergone external validation, our aims were to prospectively validate the existing form and analyse the model's response to a cutoff value that can be used in primary care setting.

Methods

Patients

We have prospectively included adult patients with suspicion of OSA referred to perform a PSG – level I or III - at the sleep laboratory of São João University Hospital, between December 2019 and March 2020. Patients with suspicion/confirmation of other sleep disorders, pregnant women, patients with nocturnal oxygen supplementation, or performing therapeutic studies with positive airway pressure therapy or oral appliances, were excluded.

Data collection and Preprocessing

Prior to PSG, all patients were observed by a sleep physiologist who collected OSABayes variables (gender, age, NC, CFA, witnessed apneas and nocturia) and other important variables to characterize the cohort, namely demographic variables (body mass index (BMI)), clinical history (Epworth sleepiness scale (ESS), snoring, morning headaches, non-restful sleep) and comorbidities (hypertension, diabetes, dyslipidemia, cardiac pathology, stroke and depression).

OSABayes probability was obtained by filling the form available online. The PSG review and analysis were performed by a sleep physiologist blind to the OSA probability resulting from OSABayes.

The continuous variables were categorized according to the following criteria:

- Age (<40, 40–54, 55-69, ≥70 years);
- NC (Female - ≤37cm: normal, >37cm: increased; Male - ≤42cm: normal, >42cm: increased) [25];
- BMI (< 30Kg/m^2: normal, ≥ 30Kg/m^2: obese) [26];
- ESS (0–10: normal, 11–24: daytime sleepiness) [27,28];
- AHI (<5/h: normal, 5-14: mild, 15-29: moderate, >30: severe) [1,29].

Polysomnography

All patients with suspected OSA initially undergone PSG III. PSG III was performed at home, with portable equipment that records respiratory flow through a nasal cannula, respiratory effort with inductance plethysmography belts, pulse oximetry and body position. PSG III was performed with Embletta® MPR [Embla Systems] equipment. An effective diagnosis was considered whenever the AASM criteria were met [11], with high-quality recording time of more than 4 hours [1]. Only patients who did not have a confirmed diagnosis with this exam underwent PSG I.

PSG I was performed in the sleep laboratory of São João University Hospital, with the presence of a sleep physiologist, and with the recording of electroencephalogram, electrooculogram, chin and legs

electromyography, electrocardiogram, respiratory flow through the nasal cannula and thermistor, respiratory effort with inductance plethysmography belts, pulse oximetry and body position. PSG I was performed either with Embla® S4500 [Embla Systems] or Alice® 5 [Respironics] equipment.

Sleep epochs were defined according to the criteria of AASM Scoring Manual version 2.5 [30]. For PSG respiratory analyses, an apnea was defined as the absence of airflow (≥ 90% reduction) for ≥ 10 seconds and hypopnea as an airflow reduction (≥ 30% and < 90%) of at least 10 seconds duration with a ≥ 3% drop in oxygen saturation (or final arousal on PSG I). Since all subjects in the study had suspected OSA - signs/symptoms or associated diseases - OSA diagnosis was confirmed with AHI values ≥ 5/h in PSG level I or III.

Statistical analysis

Categorical variables were described with absolute and relative frequencies and compared using Chi-square or Fisher's exact test. Continuous variables were described by means and standard deviations (unless otherwise specified). Comparisons of continuous variables were performed using Independent T test or Mann-Whitney. Statistical significance was defined at 5%, and the statistical analysis was all performed in SPSS®26 (SPSS Inc., Chicago, IL, USA).

The accuracy of the model was assessed by the area under the receiver operating characteristics curve (AUC). AHI and OSABayes probabilities were obtained and compared using AUC [95%CI] for OSA diagnosis (AHI ≥ 5/h) and for higher severity levels (AHI ≥ 15/h) of OSA prediction. The accuracy of the model was also assessed for the diagnosis of OSA performed by PSG I and PSG III. To assess the discriminative capacity of the model, a cutoff value was chosen. This value was obtained after assessing the ROC curve for OSA diagnosis (AHI ≥ 5/h), aiming at a sensitivity ≥ 95%.

Ethics

The study was approved by the Ethics Committee of São João University Hospital (CES 335-19), following the Declaration of Helsinki.

Results

Validation cohort

The validation cohort had 221 patients with OSA suspicion, being 5 excluded (2 with inconclusive PSG, 2 with suspicion of other sleep disorders and 1 therapeutic PSG), resulting in a total of 216 subjects included. The mean (sd) age was 57 (13) years old, with 55% of males. A total of 73 subjects performed PSG level I (34%) and 143 level III (66%). OSA diagnosis was present in 177 (82%), from which 78 (44%) were categorized as mild, 56 (32%) moderate, and 43 (24%) severe. Table 1 describes the validation cohort obtained from the questionnaire carried out to all subjects, comparing normal ones to OSA patients, while Table 2 shows a comparison between PSG I and PSG III, and by normal and OSA subjects in each PSG.

Insert table 1 (suggestion)

Among OSABayes variables, only the CFA did not differ between OSA patients and normal subjects. Male gender, age, NC, witnessed apneas and nocturia were significantly higher in patients with OSA. In addition, other variables such as snoring, hypertension, diabetes and dyslipidemia were also significantly higher in OSA patients, compared to normal subjects. The percentage of subjects with increased BMI (44 vs 56% in OSA group, $p = 0.151$) and ESS (57 vs 40% in OSA group, $p = 0.064$) did not differ significantly between groups with and without OSA. Non-restful sleep was superior in the group without OSA (74% vs 50% in OSA group, p value ≤ 0.05).

The subjects who performed PSG I or PSG III also exhibit some differences. Male gender, NC and snoring (83% vs 97% in PSG III, p value ≤ 0.001) were significantly higher in PSG III. In contrast, the percentage of patients with depression was significantly higher in PSG I. The percentage of patients with OSA was considerably higher in the PSG III group (66% vs 90% in PSG III, p value < 0.001).

Insert table 2 (suggestion)

Derivation and validation cohorts

Cohorts have a similar percentage of male patients, age distribution and nocturia presence, however, the percentage of patients with witnessed apneas, CFA and increased NC were significantly higher in the derivation cohort. Otherwise, the percentage of patients with OSA was considerably higher in the validation cohort (66% vs 82% in validation cohort, p value < 0.001). Table 3 describes the major characteristics of the patients included in the derivation and validation cohorts.

Insert table 3 (suggestion)

Model validation

The Bayesian network model - OSABayes - was validated following an external approach, with ROC analysis performed and the respective AUC, along with their 95% confidence intervals, illustrated in Figure 1. For OSA diagnosis (AHI ≥ 5/h), the AUC was 83.6% [77.3–90.0%], being slightly higher in PSG III (90.5% [83.0–98.0%]) than in PSG I (79.0% [68.7–89.3%]), although without statistical significance (p = 0.117). For moderate/severe OSA (AHI ≥ 15/h), the AUC was 76.3% [69.9–82.7%].

Cutoff definition

After an evaluation of the coordinates of the ROC curve for OSA diagnosis a cutoff value of 4.0% was defined, with a sensitivity > 95%. This cutoff can be used in primary care setting to quantify the probability of having OSA and help to refer patients to sleep laboratories. Table 4 shows the discriminatory capacity of the model in different settings: diagnosis of OSA with PSG I or PSG III, as well as in the prediction of moderate/severe cases. Several measures of diagnostic accuracy are depicted.

In general, it can be seen that a cutoff of 4.0% had a sensitivity of 96.6% and a specificity of 41.0%. The diagnostic odds ratio was 19.74 and the diagnostic accuracy of 86.6%, with an F-score of 0.91. For the diagnosis of OSA (AHI ≥ 5/h) in each PSG, specificity was slightly lower in PSG I. In PSG III, a positive post-test odds of 19.26 was found, with a diagnostic accuracy of 92.3% and a diagnostic odds ratio of 31.26. For moderate/severe OSA prediction, the cutoff of 4.0% had a general sensitivity value of 97.0%, and although having low specificity (27.4%), it had a negative predictive value of 91.5% and a negative post-test odds of 0.09.

Insert table 4 (suggestion)

Discussion

OSABayes is an intuitive and user-friendly tool, based on a Bayesian network, which allows clinicians to quantify the probability of having OSA with six clinical and demographic variables.

Our validation data, particularly the analysis of the various ROC curves, shows a good performance of OSABayes in the diagnosis of OSA (AHI ≥ 5/h), with a global AUC of 83.6% [77.3–90.0%]. Although the model was created with PSG I collected at a district hospital, it seems to have a high discriminatory power regardless the type of PSG used. These data show that the use of OSABayes is not confined to the place where it was derived and validate its applicability in identifying patients with high pre-test probability of the disease.

The fact that OSABayes respond well in patients who performed PSG III, means that it can be usefully used in a larger number of patients. Because, as we saw: 1) the model was created to improve the initial screening of OSA in primary health care and to help a more efficient hospital patients' referral, and 2) PSG III is an available test for the initial screening of these patients and, therefore, to make the diagnosis of OSA. In addition, our validation data shows that although in PSG I there are patients who, despite having OSA, have a milder severity of the disease, they do not have the most typical signs/symptoms or are less evident, and have other comorbidities, such as depression, which hinder the approach, the AUC for PSG I has an acceptable value of 77.0% [62.1–91.9%].

The relevance of this study is the validation of a model that allows only the most likely OSA cases to be referred for screening. Thus, for clinicians to be able to define these cases, it was important to define a cutoff: a probability above which a patient must be considered to perform PSG. Setting a cutoff value of 4.0%, we were able to have a sensitivity > 95% for the diagnosis of OSA (in general and in each PSG), with an overall diagnostic accuracy of 86.6% and an F-score of 0.91, ensuring the test quality. In the particular case of PSG III, if we apply the OSABayes form before the PSG, we can send 96.9% of patients with OSA for screening. With that, we would be reducing the number of unnecessary PSG to half (50.0%). In PSG I, the same cutoff presented high sensitivity, with a lower specificity (36.0%), but maintaining a good diagnostic odds ratio (12.52).

According to our validation data, we expect about 11% false positives for both PSGs, which despite being undesirable, are an improvement when compared to all patients at risk who are referred for sleep consultation and PSG. Nonetheless, we were able to rule-out 41% of healthy patients, which would alleviate health services, reducing the burden of unnecessary consultations and waiting lists for PSG, while identifying more than 95% of patients with OSA. In our laboratory, where around 280 diagnostic PSG I and 2450 diagnostic PSG III are performed annually, the reduction would be of 155 tests (35 PSG I and 120 PSG III). This decrease will certainly be associated with financial advantages, but also in workflow and waiting lists. Most importantly, the vacancies left by these patients may be filled by others, allowing a more efficient referral, and a faster diagnosis.

Despite the high sensitivity of OSABayes and its low proportion of false negatives (2.8%) in the OSA diagnosis, it has a negative predictive value of 72.7%. Thus, a negative result (<4.0%) should not be immediately interpreted as a rule-out, since 27% of individuals may have the disease. However, we are more confident to exclude the possibility of having a moderate to severe OSA, since the negative predictive value rises (91.5%) and the proportion of false negatives drops to 1%. In addition, OSABayes has a high sensitivity (97.0%) in prediction of moderate to severe OSA, with an AUC of 76.3% [69.9–82.7%], making it possible for the most severely ill patients to be tested.

Compared to other screening tools widely tested for OSA, OSABayes has a slightly higher AUC value (83.6% vs 79.6% and 82.6% in STOP-Bang [31,32] and NoSAS [33], respectively), although very similar; in OSABayes, the sensitivity value was higher, as was objective, although this has decreased specificity. So, we can say that OSABayes catches more patients with OSA than these tests, although also more non-patients. In addition to these good results, OSABayes has the advantage of being shorter than the STOP-Bang and the Berlin questionnaire [34,35].

There are few limitations in the present study. First, it does not have many patients who have performed PSG I (n=73). Second, the prevalence of OSA in our cohort is high; thus, the results may not be applicable to the general population, where the prevalence of OSA is lower. Also, we acknowledge that the specificity of the model is not as high as we would like. Even so, the specificity has values > 35% for the diagnosis of OSA, general and in each PSG. For the future, it is important to understand the behavior of OSABayes in other settings and to study the economic impact that the tool can have.

In summary, we think that OSAbayes can be a useful tool in daily clinical practice for identifying patients with high probability of the disease. The model is based on clinical and demographic variables, which are easy and quick to acquire, and works even in the absence of one or more variables [36]. The form is intuitive, user-friendly and can be applied quickly in a medical appointment to quantify the probability of having OSA, helping to refer patients to sleep laboratories. Our results showed the good performance of OSABayes in both types of PSG, reducing the number of unnecessary exams, which can help to reduce waiting times and health costs.

Acknowledgments

The authors would like to thank the sleep laboratory team of São João University Hospital, in particular the sleep physiologists Ana Sofia Pimentel, Andreia Neves, Joana Pipa, Patrícia Dantas, Ermelinda Moreira, Elisabete Santa-Clara and Paulo Viana. DFS acknowledges Fundacão para a Ciência e Tecnologia (FCT) under Ph.D grant number PD/BD/13553/2018.

References

1. Kapur, V. K. *et al.* Clinical Practice Guideline for Diagnostic Testing for Adult Obstructive Sleep Apnea : An American Academy of Sleep Medicine Clinical Practice Guideline. *J. Clin. Sleep Med.* **13**, 479–504 (2017).
2. Dempsey, J. A., Veasey, S. C., Morgan, B. J. & Donnell, C. P. O. Pathophysiology of Sleep Apnea. *Physiol. Rev.* **90**, 47–112 (2010).
3. Lam, J. C. M., Sharma, S. K. & Lam, B. Obstructive sleep apnoea: definitions, epidemiology & natural history. *Indian J. Med. Res.* **131**, 165–170 (2010).
4. Young, T., Skatrud, J. & Peppard, P. Risk Factors for Obstructive Sleep Apnea. *J. Am. Medican Assoc.* **291**, 2013–2016 (2004).
5. Al Lawati, N. M., Patel, S. R. & Ayas, N. T. Epidemiology, Risk Factors, and Consequences of Obstructive Sleep Apnea and Short Sleep Duration. *Prog. Cardiovasc. Dis.* **51**, 285–293 (2009).
6. Romero, E., Krakow, B., Haynes, P. & Ulibarri, V. Nocturia and snoring: predictive symptoms for obstructive sleep apnea. *Sleep Breath.* **14**, 337–343 (2010).
7. Bradley, T. D. & Floras, J. S. Obstructive sleep apnoea and its cardiovascular consequences. *Lancet* **373**, 82–93 (2009).
8. Benjafield, A. V *et al.* Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir. Med.* **7**, 687–698 (2019).
9. Epstein, L. J. *et al.* Clinical Guideline for the Evaluation, Management and Long-term Care of Obstructive Sleep Apnea in Adults. *J. Clin. Sleep Med.* **5**, 263–276 (2009).
10. Kapur, V. K. Obstructive Sleep Apnea: Diagnosis, Epidemiology, and Economics. *Respir. Care* **55**, 1155–1167 (2010).
11. American Academy of Sleep Medicine. International Classification of Sleep Disorders - Third Edition. (2014).
12. Masa, J. F. *et al.* Effectiveness of home respiratory polygraphy for the diagnosis of sleep apnoea and hypopnoea syndrome. *Thorax* **66**, 567–573 (2011).
13. Corral-Peñafiel, J., Pepin, J. & Barbe, F. Ambulatory monitoring in the diagnosis and management of obstructive sleep apnoea syndrome. *Eur. Respir. Rev.* **22**, 312–324 (2013).
14. Huang, W.-C., Lee, P.-L., Liu, Y.-T., Chiang, A. A. & Lai, F. Support vector machine prediction of obstructive sleep apnea in a large-scale Chinese clinical sample. *Sleep* 1–11 (2020) doi:10.1093/sleep/zsz295.
15. Ustun, B., Westover, M. B., Rudin, C. & Bianchi, M. T. Clinical prediction models for sleep apnea: The importance of medical history over symptoms. *J. Clin. Sleep Med.* **12**, 161–168 (2016).
16. Bozkurt, S., Bostanci, A. & Turhan, M. Can Statistical Machine Learning Algorithms Help for Classification of Obstructive Sleep Apnea Severity to Optimal Utilization of Polysomnography Resources? *Methods Inf. Med.* **56**, 308–318 (2017).
17. Lucas, P., Gaag, L. van der & Abu-Hanna, A. Bayesian networks in biomedicine and health-care. *Artif. Intell. Med.* **30**, 201–214 (2004).
18. Darwiche, A. *Modeling and Reasoning with Bayesian Networks*. (Cambridge University Press,

2009).

19. Fontenla-Romero, O., Guijarro-Berdiñas, B., Alonso-Betanzos, A., Fraga-Iglesias, A. del R. & Moret-Bonillo, V. A Bayesian Neural Network Approach for Sleep Apnea Classification. *Conf. Artif. Intell. Med. Eur.* 284–293 (2003).
20. Turhan, M., Bostanci, A. & Bozkurt, S. Estimation of cardiovascular disease from polysomnographic parameters in sleep-disordered breathing. *Eur. Arch. Oto-Rhino-Laryngology* **273**, 4585–4593 (2016).
21. Ryynänen, O., Leppänen, T., Kekolahti, P., Mervaala, E. & Toyras, J. Bayesian Network Model to Evaluate the Effectiveness of Continuous Positive Airway Pressure Treatment of Sleep Apnea. *Healthc. Inform. Res.* **24**, 346–358 (2018).
22. Ferreira-santos, D. & Rodrigues, P. P. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches. *Int. J. Data Sci. Anal.* **8**, 339–349 (2019).
23. Weiss, J. P. Nocturia: focus on etiology and consequences. *Rev. Urol.* **14**, 48–55 (2012).
24. OSABayes. http://servicosforms.gim.med.up.pt/form_test/osabayes.html.
25. Davidson, T. M. & Patel, M. R. Waist circumference and sleep disordered breathing. *Laryngoscope* **118**, 339–347 (2008).
26. Yumuk, V. *et al.* European Guidelines for Obesity Management in Adults. *Obes. Facts* **8**, 402–424 (2015).
27. Johns, M. W. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep* **14**, 540–545 (1991).
28. Pahwa, P. *et al.* Prevalence of high Epworth sleepiness scale scores in a rural population. *Can. Respir. J.* **19**, 10–14 (2012).
29. Flemons, W. W. *et al.* Sleep-related breathing disorders in adults: Recommendations for syndrome definition and measurement techniques in clinical research. *Sleep* **22**, 667–689 (1999).
30. Berry, R. B. *et al.* The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications - Version 2.5. (2018).
31. Chung, F. *et al.* STOP Questionnaire: A Tool to Screen Patients for Obstructive Sleep Apnea. *Anesthesiol. J. Am. Soc. Anesthesiol.* **108**, 812–821 (2008).
32. Chung, F., Yang, Y., Brown, R. & Liao, P. Alternative scoring models of STOP-Bang questionnaire improve specificity to detect undiagnosed obstructive sleep apnea. *J. Clin. Sleep Med.* **10**, 951–958 (2014).
33. Marti-Soler, H. *et al.* The NoSAS score for screening of sleep-disordered breathing: a derivation and validation study. *Lancet Respir. Med.* **4**, 742–748 (2016).
34. Ahmadi, N., Chung, S. A., Gibbs, A. & Shapiro, C. M. The Berlin questionnaire for sleep apnea in a sleep clinic population: Relationship to polysomnographic measurement of respiratory disturbance. *Sleep Breath.* **12**, 39–45 (2008).
35. Netzer, N. C., Stoohs, R. A., Netzer, C. M., Clark, K. & Strohl, K. P. Using the Berlin Questionnaire To Identify Patients at Risk for the Sleep Apnea Syndrome. *Ann. Intern. Med.* **131**, 485–491 (1999).
36. Ferreira-Santos, D., Monteiro-Soares, M. & Rodrigues, P. P. Impact of Imputing Missing Data in

Bayesian Network Structure Learning for Obstructive Sleep Apnea Diagnosis. *Stud. Heal. Technol. Informatics - IOS Press* **247**, 126–130 (2018).

Tables

Table 1 – Descriptive analysis of the patients included on the validation cohort, comparing OSA with normal subjects. Results shown are n (%) and p value result of Chi-square test, unless specified otherwise.

	Total (n=216)	Normal subjects (n=39)	OSA (n=177)	p value
Male gender	119 (55%)	12 (31%)	107 (61%)	**<0.001**
Age				
Mean (SD)	57 (13)	48 (11)	59 (12)	**<0.001**[#]
< 40	21 (10%)	9 (23%)	12 (7%)	
40-54	66 (30%)	20 (51%)	46 (26%)	**0.001**
55-69	92 (43%)	9 (23%)	83 (47%)	
>70	37 (17%)	1 (3%)	36 (20%)	
NC				
Median (IQR)	40 (37-43)	36 (33-39)	41 (38-44)	**<0.001**[+]
Increased	88 (41%)	5 (13%)	83 (48%)	**<0.001**
Witnessed apneas	120 (56%)	12 (31%)	108 (62%)	**<0.001**
Nocturia	123 (58%)	14 (36%)	109 (62%)	**0.003**
CFA	25 (12%)	2 (5%)	23 (13%)	0.264[*]
BMI				
Mean (SD)	31 (6)	28 (5)	31 (6)	**<0.001**[#]
Obese	116 (54%)	17 (44%)	99 (56%)	0.151
ESS				
Median (IQR)	9 (5-15)	11 (4-14)	9 (5-15)	0.729[+]
Increased	77 (43%)	20 (57%)	57 (40%)	0.064
Snoring	166 (92%)	26 (72%)	140 (97%)	**<0.001**
Morning headache	69 (39%)	20 (56%)	49 (35%)	**0.024**
Non-restful sleep	94 (55%)	26 (74%)	68 (50%)	**0.010**
Hypertension	95 (55%)	11 (31%)	84 (61%)	**0.002**
Diabetes	39 (23%)	1 (3%)	38 (28%)	**<0.001**[*]
Dyslipidemia	79 (46%)	10 (29%)	69 (50%)	**0.023**
Cardiac pathology	21 (12%)	2 (3%)	19 (14%)	0.054[*]
Stroke	11 (6%)	2 (6%)	9 (7%)	0.606[*]
Depression	48 (28%)	10 (29%)	38 (28%)	0.922

OSA: Obstructive sleep apnea; SD: Standard Deviation; IQR: Interquartile range; NC: Neck circumference; CFA: Craniofacial abnormalities; BMI: Body mass index; ESS: Epworth sleepiness scale. [#]*Independent T test,* [+]Mann-Whitney test; *Fisher's exact test. Bold text indicates $p \leq 0.05$.

Table 2 - Descriptive analysis of the patients included on the validation cohort, comparing PSG I and PSG III data. Results shown are n (%) and p value result of Chi-square test, unless specified otherwise.

	PSG I (n=73)	PSG III (n=143)	p value	PSG I (n=73)			PSG III (n=143)		
				Normal (n=25)	OSA (n=48)	p value	Normal (n=14)	OSA (n=129)	p value
Male gender	29 (40%)	90 (63%)	**0.002**	6 (24%)	23 (48%)	**0.038**	6 (43%)	84 (65%)	0.101
Age									
Mean (SD)	56 (12.4)	58 (13.2)	0.510[#]	51 (11.1)	59 (12.2)	**0.007**[#]	44 (9.8)	59 (12.7)	**<0.001**[#]
< 40	5 (7%)	16 (11%)		2 (8%)	3 (6%)		7 (50%)	9 (7%)	
40-54	22 (30%)	43 (30%)	0.691	13 (54%)	9 (19%)	**0.011**	6 (43%)	37 (29%)	**<0.001**
55-69	34 (47%)	58 (41%)		8 (33%)	26 (54%)		1 (7%)	57 (44%)	
>70	11 (16%)	26 (18%)		1 (4%)	10 (21%)		0 (0%)	26 (20%)	
NC									
Median (IQR)	38 (35-42)	40 (38-44)	**<0.001**[+]	36 (33-37)	40 (36-42)	**0.001**[+]	37 (36-39)	41 (38-44)	**0.010**[+]
Increased	24 (33%)	64 (46%)	**0.036**	3 (12%)	21 (44%)	**0.005**[*]	2 (15%)	62 (49%)	**0.018***
Witnessed apneas	39 (54%)	81 (57%)	0.704	8 (32%)	31 (66%)	**0.006**	4 (29%)	77 (60%)	**0.024***
Nocturia	38 (52%)	85 (60%)	0.184	10 (40%)	28 (58%)	0.137	4 (29%)	81 (64%)	**0.012**[*]
CFA	11 (15%)	14 (10%)	0.310	2 (8%)	9 (19%)	0.192[*]	0 (0%)	14 (11%)	0.363*
BMI									
Mean (SD)	30 (4.8)	31 (6.5)	0.055[#]	28 (4.7)	30 (4.7)	0.092[#]	27 (4.7)	32 (6.5)	**0.010**[#]
Increased	39 (53%)	77 (54%)	0.771	11 (44%)	28 (58%)	0.244	6 (43%)	71 (56%)	0.369
ESS									
Median (IQR)	8 (5-14)	10 (5-15)	0.254[+]	11 (5-15)	6 (3-12)	0.305[+]	12 (2-14)	9 (5-15)	0.965[+]
Increased	29 (40%)	48 (45%)	0.489	14 (56%)	15 (32%)	**0.047**	6 (60%)	42 (44%)	0.506[*]
Snoring	60 (83%)	106 (97%)	**<0.001**[+]	17 (68%)	43 (92%)	**0.015**[*]	9 (82%)	97 (99%)	**0.026**[+]
Morning headache	30 (42%)	39 (38%)	0.479	15 (60%)	15 (32%)	**0.021**	5 (46%)	34 (37%)	0.564
Non-restful sleep	43 (60%)	51 (52%)	0.364	20 (80%)	23 (49%)	**0.011**[*]	6 (60%)	45 (51%)	0.410*
Hypertension	35 (49%)	60 (59%)	0.296	9 (36%)	26 (57%)	0.099	2 (20%)	58 (63%)	**0.011***
Diabetes	10 (14%)	29 (28%)	**0.033**	1 (4%)	9 (20%)	0.068[*]	0 (0%)	29 (31%)	**0.029***
Dyslipidemia	28 (39%)	51 (50%)	0.225	9 (36%)	19 (41%)	0.662	1 (10%)	50 (54%)	**0.008***
Cardiac pathology	9 (13%)	11 (10%)	0.669[*]	0 (0%)	9 (20%)	**0.015**[*]	1 (10%)	10 (11%)	0.702*
Stroke	4 (6%)	7 (7%)	0.516[*]	2 (8%)	2 (4%)	0.441[*]	0 (0%)	7 (8%)	0.471*
Depression	25 (35%)	23 (23%)	**0.041**	9 (36%)	16 (35%)	0.918	1 (10%)	22 (24%)	0.284*
AHI									
AHI ≥ 5/h	48 (66%)	129 (90%)	**<0.001**						
Normal	25 (34%)	14 (10%)							
Mild	30 (41%)	48 (34%)	**<0.001**						
Moderate	11 (15%)	45 (31%)							
Severe	7 (10%)	36 (25%)							

NC: Neck circumference; CFA: Craniofacial abnormalities; BMI: Body mass index; ESS: Epworth sleepiness scale; AHI: Apnea-hypopnea index; PSG: Polysomnography; OSA: Obstructive sleep apnea. [#]Independent T test; [+]Mann-Whitney test; *Fisher's exact test. Bold text indicates $p \leq 0.05$.

Table 3 – Comparing major characteristics of patients included in the derivation and validation cohorts. In the derivation cohort, absolute and relative frequencies are presented in the variables with a high missing values. Results shown are n (%) and p value result of Chi-square test, unless specified otherwise.

	Derivation cohort (n=194)	Validation cohort (n=216)	p value
Male gender	123 (63%)	119 (55%)	0.109
Age			
Mean (SD)	58 (13.0)	57 (13.0)	0.696[#]
< 40	19 (10%)	21 (10%)	
40-54	56 (29%)	66 (30%)	0.860
55-69	79 (41%)	92 (43%)	
>70	40 (21%)	37 (17%)	
NC			
Median (IQR)	42 (39-45)	40 (37-43)	**<0.001**[+]
Increased	107 (55%)	89 (41%)	**0.004**
Witnessed apneas	104 (54%)-(70%)	120 (56%)	**0.012**
Nocturia	70 (36%)-(63%)	123 (57%)	0.299
CFA	43 (22%)-(92%)	25 (12%)	**<0.001**[*]
AHI			
AHI ≥ 5/h	128 (66%)	177 (82%)	**<0.001**
Normal	66 (34%)	39 (18%)	
Mild	63 (33%)	78 (36%)	**0.002**
Moderate	32 (17%)	56 (26%)	
Severe	33 (17%)	43 (20%)	

SD: Standard deviation; IQR: Interquartis range; NC: Neck circumference; CFA: Craniofacial abnormalities; AHI: Apnea-hypopnea index; IQR: Interquartile range. [#]Independent T test; [+]Mann-Whitney test; *Fisher's exact test. Bold text indicates $p \leq 0.05$.

Table 4 – Diagnostic accuracy measures of the model with a cutoff of 4.0%, in different settings: diagnosis of OSA with PSG I or PSG III, as well as in the prediction of moderate/severe *cases. Values shown in parentheses are 95% CI.*

	AHI ≥ 5/h			AHI ≥ 15/h		
	Total	PSG III	PSG I	Total	PSG III	PSG I
Sensitivity (%)	96.6 [92.8-98.4]	96.9 [92.3-99.2]	95.8 [85.9-99.5]	97.0 [91.4-99.4]	96.3 [89.6-99.2]	94.4 [72.7-99.9]
Specificity (%)	41.0 [25.6-57.9]	50.0 [23.0-77.0]	36.0 [17.8-57.5]	27.4 [19.5-36.4]	24.2 [14.2-36.7]	30.9 [19.1-44.8]
PPV (%)	88.1 [82.8-91.9]	94.7 [91.4-96.8]	74.2 [68.0-79.5]	53.1 [50.1-55.9]	62.4 [58.9-65.8]	30.9 [26.6-35.6]
NPV (%)	72.7 [52.8-86.5]	63.6 [36.9-84.0]	81.3 [51.0-94.7]	91.5 [77.1-97.1]	83.4 [60.2-94.3]	94.4 [70.9-99.2]
Dx accuracy (%)	86.6 [81.3-90.8]	92.3 [86.7-96.1]	75.3 [63.8-84.7]	59.3 [52.4-65.9]	65.0 [56.6-72.8]	46.6 [34.8-58.6]
Dx OR	19.74 [18.57-20.92]	31.26 [29.81-32.70]	12.52 [10.90-14.14]	12.20[10.98-13.43]	8.31 [7.02 -9.60]	7.54 [5.45-9.63]
LR +	1.64 [1.26-2.13]	2.09 [1.15-3.28]	1.50 [1.11-2.02]	1.34 [1.19-1.50]	1.27 [1.10-1.47]	1.37 [1.11-1.69]
LR -	0.08 [0.03-0.20]	0.06 [0.02-0.19]	0.12 [0.03-0.50]	0.11 [0.03-0.35]	0.15 [0.05-0.51]	0.18 [0.03-1.26]
Odds Post +	7.43 [6.51-8.35]	19.26 [16.18-22.33]	2.87 [2.34-3.40]	1.13 [1.08-1.18]	1.66 [1.49-1.83]	0.45 [0.33-0.56]
Odds Post -	0.38 [0.31-0.44]	0.57 [0.49-0.65]	0.23 [0.13-0.33]	0.09 [0.05-0.13]	0.20 [0.13-0.27]	0.06 [0.01-0.11]
F-score	0.91 [0.88-0.94]	0.98 [0.96-1.00]	0.77 [0.70-0.84]	0.55 [0.52-0.58]	0.65 [0.61-0.69]	0.33 [0.23-0.43]

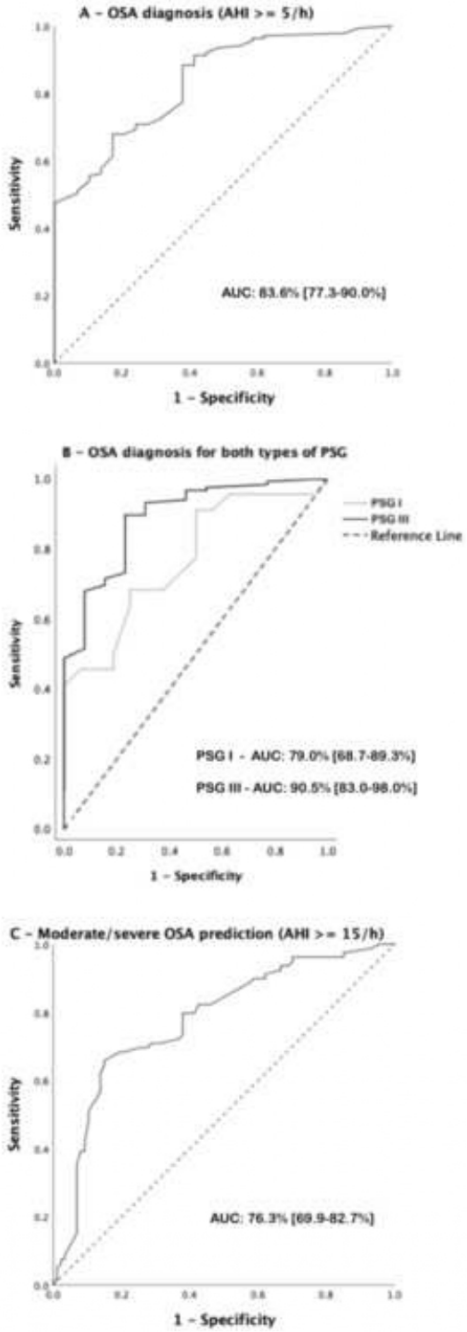
AHI: Apnea-Hypopnea index; PPV: Positive predictive value; NPV: Negative predictive value; Dx accuracy: Diagnostic accuracy; Dx OR: Diagnostic odds ratio; LR+: Positive likelihood ratio; LR-: Negative likelihood ratio; Odds Post +: Positive Post-test Odds; Odds Post -: Negative Post-test Odds.

Figure Caption List

Figure 1

Receiver operating characteristics analyses and area under the curve values for OSA diagnosis (A), for diagnosis in both types of PSG (B) and for moderate/severe OSA prediction (C). AUC values are shown with their 95% confidence intervals.

Figure1 Click here to access/download;Figure;Figure1.png



Disclosure Statement

Conflict of Interest: All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Financial Disclosure: none.

Non-financial Disclosure: none

Funding: No funding was received for this research.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the Ethics Committee of São João University Hospital (CES 335-19) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

5.2. Validation of OSABayes: a screening tool for obstructive sleep apnea

(finalizing details)

Daniela Ferreira-Santos and Pedro Pereira Rodrigues

Title page

Title:

Validation of OSABayes: a screening tool for obstructive sleep apnea

Authors:

Daniela Ferreira-Santos[1]

Pedro Pereira Rodrigues[1,2]

Authors affiliations:

[1]MEDCIDS-FMUP – Community Medicine, Information and Decision Sciences, Faculty of Medicine of the University of Porto, Portugal

[2]CINTESIS – Center for Health Technology and Services Research, Porto, Portugal

Corresponding author:

Daniela Ferreira-Santos

Rua Dr. Plácido da Costa, s/n 4200-450 Porto

danielasantos@med.up.pt

Institution where work was performed:

CINTESIS – Centre for Health Technology and Services Research

Abstract

Study objectives: OSABayes is an online auxiliary tool that can support the decision to undergo polysomnography in obstructive sleep apnea (OSA). We aim to validate this screening tool and establish a cutoff value for sleep experts and non-sleep-experts to refer suspected patients.

Methods: Six predictive variables were initially selected (gender, witnessed apneas, age, nocturia, craniofacial abnormalities, and neck circumference) from a previous study that employed the Bayesian network in the derivation cohort. The same variables were then collected in a new cohort (validation cohort) as well as the same network model was applied in this dataset, and a validity assessment was performed.

Results: A total of 194 patients were included in the derivation cohort, with an OSA prevalence of 66%, while 619 patients with a 77% OSA prevalence were found in the validation cohort. Looking at the previously defined cutoff value of 38.5%, aiming at a sensitivity higher than 95%, we can state that area under the curve for both cohorts overlapped. The sensitivity value in the validation cohort was almost 9% higher than before, with a diminishing in false negative results of 11% from the derivation to the validation cohorts.

Conclusions: OSABayes continues to prove overall good results, with 99% of OSA patients being elected to undergone polysomnography (rule-out approach). This tool is intuitive, simple, friendly, easy to access, and filled in. It has a defined clinical cutoff of 38.5%, meaning that if a patient has a probability in OSABayes lower than this, him/her should not be further referred.

Keywords: obstructive sleep apnea, screening, Bayesian network, external validation

Statement of significance

The American Academy of Sleep Medicine (AASM) believes that clinical prediction algorithms can be used to screen obstructive sleep apnea patients without replacing polysomnography, as they have the capability of identifying thus patients that most likely benefit from undergoing the good standard. OSABayes is a sensitive screening tool, based on Bayesian networks and consequently conditional probabilities, available online and with a pre-defined cutoff value to establish the need to be further referred to undergo the exam or a sleep consultation.

1. Introduction

In obstructive sleep apnea (OSA), the respiratory effort is maintained but ventilation decreases or disappears because of partial/total occlusion of the upper airway [1]. The standard method for assessing OSA diagnosis is in-laboratory polysomnography (in-Lab PSG), where the apnea-hypopnea index (AHI) is calculated, with the number of apneas and hypopneas per hour of sleep. An AHI≥ 5 per hour (/h) plus signs/symptoms or associated disorders indicates OSA presence; an apnea was defined as the absence of airflow (≥ 90%) for ≥10 seconds and hypopnea as an airflow reduction (≥30% and <90%) of at least 10 seconds duration with a >3% drop in oxygen saturation or final arousal in-Lab PSG) [1]. This exam is time-consuming, expensive, and it is neither a routine clinical practice nor an absolute suitable screening tool [2]. Additionally, almost all Portuguese suspected patients are referred by the primary care physician to a sleep appointment in urban areas, where the expert decide the need to perform PSG, which originates high waiting lists. As we know, primary care physicians do not have expert knowledge on sleep disorders, so they need support to decide who they should send to a sleep consultation. On the other hand, sleep experts have all the available knowledge on sleep disorders, but they also need help to decide on the ones that they immediately need to undergo in-Lab PSG.

The latest American Academy of Sleep Medicine (AASM) guidelines [1], suggested that clinical prediction algorithms can be used to screen OSA patients without replacing PSG. These models could reliably identify thus patients that most likely benefit from PSG depending on the purpose: rule-out (exclude OSA diagnosis when the probability is low) or rule-in (prioritizing patients in need of PSG when the probability is high).

Previous work [3] studied several factors and identified six associated with OSA screening (gender, witnessed apneas, age, nocturia, craniofacial abnormalities, and neck circumference). These factors were implemented in a Bayesian network, as they demonstrated the ability to explain decisions, transparency, and good performance even when handling data entry errors or missing information. The model was then implemented in a server to be used as an online form entitled OSABayes.

The aim of this work is to revisit OSABayes, an auxiliary tool that can support the decision, validating this screening form and confirming a cutoff value so that we can assist primary care physicians and sleep experts refer patients.

2. Methods

This study was designed according to the Standard for Reporting Diagnostic accuracy studies (STARD) list [4].

2.1. Patients

This study included patients referred to underwent PSG at the Sleep Laboratory of Vila Nova de Gaia/Espinho Hospital Center between January and May 2015 (derivation cohort) and the ones referred to underwent PSG at the Sleep Laboratory of São João University Hospital between December 2011 and December 2019 (validation cohort). All adults (>18 years old) with a suspicion of OSA were included, while patients with suspicion of another sleep disorder, patients already diagnosed (therapeutic studies), patients with severe lung diseases or neuromuscular diseases, and pregnant women were excluded.

This study was approved by the ethics committee of the Vila Nova de Gaia/Espinho Hospital Center (17/04/15) and São João University Hospital (CES 237/18), in accordance with the Declaration of Helsinki.

2.2. Predictive variables

Six predictive variables were selected in the previous study [3] including demographic data (gender and age), physical examination (neck circumference (NC)), and clinical history (witnessed apneas (WA), nocturia, and craniofacial abnormalities (CFA)). The same predictive variables were collected in the validation cohort, and variable categorization was preserved. Also, the previously developed Bayesian model, which was a Tree-augmented Naïve Bayes with the selected variables (TAN6), was maintained and directly applied to the validation cohort.

2.3. Data collection

Clinical information for the derivation cohort was retrospectively collected between 1st January 2015 and 31st May 2015. All 6 variables were extracted from the central electronic records along with sleep reports, making all clinical files available. NC was categorized for males and females, with males presenting an increased circumference if the value is higher than 42cm and females higher than 37cm. The outcome definition for OSA presence is an AHI≥ 5 per hour (/h) plus signs/symptoms or associated disorders as described in the Introduction section.

The same criteria were applied to the validation cohort, which was also retrospectively collected between 15th December 2011 and 31st December 2019.

2.4. OSABayes

A Bayesian network, the machine learning approach behind OSABayes, represents a joint distribution of one set of variables and the outcome. Each variable is represented by a node in the graph and is dependent on the remain set of variables, with this dependency being represented by a conditional probability [5]. The chosen Bayesian classifier was Tree-augmented Naïve Bayes (TAN), which allows for an optional dependence for each variable as the natural way of a clinical diagnosis given that variables/factors are not independent in clinical settings.

2.5. Evaluation methodology

The six selected variables were chosen as those with a significant univariate association with the outcome, considering a 5% significance level or a 10% significance level if at least 5 patients were observed in each outcome category (normal or OSA) and for which no quality problems were suspected.

Predictive variables in both cohorts were described with absolute and relative frequencies and compared using Chi-square, with statistical significance defined at 5% ($p<0.05$). We applied R software [6] to perform the analysis. OSABayes parameters were validated by comparing the area under the curve (AUC) and respective confidence intervals in the derivation and validation cohorts for an AHI≥ 5 per hour (/h) plus signs/symptoms or associated disorders. The previously obtained cutoff value (38.5%) was kept to assess the discriminative capacity of OSABayes.

3. Results

3.1. Cohorts

Regarding the derivation cohort, 194 patients were selected. More than half (63%) were males and 34% had a normal result – without OSA diagnosis after PSG. The validity assessment estimated from 10-times 2-fold cross-validation reached an AUC of 63.6 [60.9-66.3], a sensitivity of 90.2 [88.0-92.4], a specificity of 23.5 [20.0-27.0], and a positive and negative predictive value of 69.6 [68.8-70.4] and 56.6 [52.7-60.5], respectively.

The validation cohort included 619 patients, with a mean age of 52 years old and 77% with OSA. Table 1 describes the comparison between cohorts and the six predictive variables. Cohorts have a similar proportion of WA (70% vs. 63%) and nocturia (63% vs. 66%). However, the derivation cohort had a higher proportion of males, NC, and CFA. Also, while in the derivation cohort, the 55 to 69 age group had a higher proportion of patients (41%), the validation had 44 to 54 (40%). The validation cohort also demonstrated 10% fewer patients in the class higher than 70 years old.

The validation cohort presented an overlapping AUC (63.6 vs. 64.4). Regarding sensitivity, a higher value was obtained (98.7% [98.2-99.2]). This means that 99% of OSA patients would undergo PSG, rejecting only 1% of patients who could be referred to another follow-up consultation. Moreover, around 22% of false positives (less 8% than in the derivation cohort) is an improvement if we compare all patients at risk of being referred to a sleep laboratory and PSG. The overall results for comparison between AUCs and other metrics are stated in Table 2.

3.2. Inference

To give an example of how OSABayes could be implemented in primary care, the following figures present the interaction between the physician and the Bayesian model in the online form.

Assuming a new patient arrives at consultation with the following characteristics: a female 34 years old presenting a normal NC and without WA, CFA, and nocturia. OSABayes gives a 1% probability of having OSA. If now we swap this patient to be a male with the same characteristics, the probability changes to 2%, as seen in Figure 1.

Now, if we have an older male with the same characteristics, the probability is 30% as shown in Figure 2, while for a female, it is 13%. In both cases, as our clinical cutoff for defining a high probability of OSA is 39%, the physician should not send any patients to undergo PSG.

Supposing that in another consultation the patient is a male 40 years old, has an increased NC and reports WA and nocturia, and has no idea about craniofacial abnormalities, the probability is 79%, as described in Figure 3.

4. Discussion

The first step towards defining an auxiliary tool such as OSABayes in the clinical setting is assessing its need. In 2014, Leite et al. [7] described a proportion of normal results in OSA diagnosis of 48%, while in 2019, Ferreira-Santos et al. [3] showed 34% of patients not having OSA. Both results in Portuguese sleep

laboratories support the need for good clinical decision support, although OSA prevalence has decreased; in this study, the proportion of normal results is 23%.

OSABayes is an intuitive and user-friendly tool developed in a sample of 194 participants, resulting in six predictive variables. The present study performed its external validation in a sample of 619 participants, reaching an AUC value of 64.3% [63.4-65.3], showing fair discriminatory power. In addition, this approach performs well in the lack of information, which rises as a great advantage compared to other models and is based on conditional probabilities.

Since 1988, different studies aimed to predict whether an individual has OSA. The latest systematic review [8] tried to identify, gather, and analyze existing machine learning approaches intended to be used in OSA screening, namely in adults suspected of this disease. A total of 63 studies were included, divided into studies performing diagnostic models alone (n=23), adding internal validation (n=26), and applying the clinical prediction algorithm to an independent cohort (n=14). The manuscript demonstrated that the best model at rule-out (100% sensitivity in external validation results on 101 participants) included variables that are not available at primary care (respiratory conductance) for an OSA diagnosis definition of AHI≥15. Also, the created model was based on logistic regression with two variables in 168 participants, not presenting an AUC value and with a prevalence for OSA of 55%. In addition, this review also unveils a top5 for clinical factors in analyze (BMI, age, sex, NC, and snoring), revealing that a path for a consensus regarding clinical factors can be achieved.

Traditionally, OSA screening was based on questionnaires, such as the Berlin questionnaire (BQ) – one of the most recognized screening tools. In 2011, Vaz et al. [9] applied this questionnaire to screen patients with OSA in a sleep laboratory in Portugal, achieving 65% and 80% for sensitivity and specificity, respectively. Although these results revealed good discrimination, it also proved a poor performance in OSA identification.

Although OSABayes did not reach 100% sensitivity, as the logistic model referred to in the systematic review, the value was 99%, with the advantage of having 619 participants against 101 for external validation and predictive variables that are always available. Our model also included 3 out of 5 clinical factors more commonly used in prediction models.

A limitation of the previous study and this one is that both studies were conducted on patients referred to a sleep laboratory, which we presume results in higher OSA prevalence than of the general population (66% vs. 77%, respectively). However, Benjafield et al. [10] in 2019 aimed to estimate the global prevalence of

OSA. The conclusions were that most of the countries, including Portugal, do not provide data for OSA prevalence. Furthermore, they estimated that almost 1 billion people worldwide aged between 30-69 years have an AHI of five or more events per hour and that probably OSA prevalence exceeds 50% in some countries, which means that our assumption regarding OSA prevalence in the general population may be wrong. Also, we acknowledge the retrospective study design is a limitation, but all efforts were made to ensure that all information was within reach, especially in the validation cohort, as different clinical electronics records were consulted. Likewise, the low specificity (24% in the derivation cohort vs. 3% in the validation cohort) makes the model somewhat limited, nevertheless, we intended to follow a rule-out approach - always aiming at maximum sensitivity.

We conclude that this clinical prediction algorithm – OSABayes – can be applied in a non-expert or expert setting, with easy collectable and available variables, a standard OSA definition, a high sensitivity tool (rule-out approach – 99%), avoiding false negatives, i.e., a patient with OSA with no recommendation to perform PSG. Additionally, visually is very intuitive, simple, friendly, easy to access, and filled in.

Acknowledgments

Daniela Ferreira-Santos acknowledges Fundacão para a Ciência e Tecnologia (FCT) under PhD grant numbers PD/BD/13553/2018 and COVID/BD/152608/2022.

Disclosure statement

Non-financial or conflicts of interest that could be relevant in this context (Non-financial Disclosure: none)

Reference list

1. Kapur VK, Auckley DH, Chowdhuri S, et al. Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: an American academy of sleep medicine clinical practice guideline. *J Clin Sleep Med*. 2017;13(3):479-504. doi:10.5664/jcsm.6506
2. Sun LM, Chiu H-W, Chuang CY, Liu L. A prediction model based on an artificial intelligence system for moderate to severe obstructive sleep apnea. *Sleep Breath*. 2011;15:317-323. doi:10.1007/s11325-010-0384-x
3. Ferreira-Santos D, Rodrigues PP. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches. *Int J Data Sci Anal*. 2019;8:339-349. doi:10.1007/s41060-018-0118-x
4. Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *BMJ*. 2015;351. doi:10.1136/bmj.h5527
5. Lucas P. Bayesian analysis, pattern analysis, and data mining in health care. *Curr Opin Crit Care*. 2004;10(5):399-403. doi:10.1097/01.CCX.0000141546.74590.D6
6. Team RC. R: A language and environment for statistical computing. 2021.
7. Leite L, Costa-Santos C, Rodrigues PP. Can we avoid unnecessary polysomnographies in the diagnosis of Obstructive Sleep Apnea? A Bayesian network decision support tool. *Proc - IEEE Symp Comput Med Syst*. 2014:28-33. doi:10.1109/CBMS.2014.30
8. Ferreira-Santos D, Amorim P, Silva Martins T, Monteiro-Soares M, Pereira Rodrigues P. *Helping Early Obstructive Sleep Apnea Diagnosis with Machine Learning: A Systematic Review (Preprint)*. JMIR Publications Inc.; 2022. doi:10.2196/39452
9. Vaz AP, Drummond M, Caetano Mota P, Severo M, Almeida J, Winck JC. Translation of Berlin Questionnaire to Portuguese language and its application in OSA identification in a sleep disordered breathing clinic. *Rev Port Pneumol*. 2011;17(2):59-65. doi:10.1016/S0873-2159(11)70015-0
10. Benjafield A V., Ayas NT, Eastwood PR, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. *Lancet Respir Med*. 2019;7(8):687-698. doi:10.1016/S2213-2600(19)30198-5

List of figures

Tables

Table 1 - Comparing the six predictive variables in the derivation and validation cohorts. Results shown are n (%) and p-value result of Chi-square test.

Predictive variables	Derivation cohort (n=194)	Validation cohort (n=619)	p-value
Male gender	123 (63%)	283 (46%)	**<0.001**
Age			**<0.001**
<40	19 (10%)	102 (16%)	
44-54	56 (29%)	248 (40%)	
55-69	79 (41%)	199 (32%)	
>70	40 (21%)	70 (11%)	
Neck circumference increased	107 (55%)	187 (46%)	**0.040**
Witnessed apneas	104 (70%)	228 (63%)	0.142
Nocturia	70 (63%)	233 (66%)	0.521
Craniofacial abnormalities	43 (91%)	143 (52%)	**<0.001**
OSA presence	128 (66%)	479 (77%)	**0.001**

OSA: obstructive sleep apnea
Bold text indicates p≤0.05

Table 2 - Validity assessment for derivation and validation cohorts

Cohort	Cutoff	Sensitivity	Specificity	PPV	NPV	AUC
Derivation	38.5%	90.2	23.5	69.6	56.6	63.6
		[88.0-92.4]	[20.0-27.0]	[68.8-70.4]	[52.7-60.5]	[60.9-66.3]
Validation	38.5%	98.7	3.2	77.7	45.6	64.4
		[98.2-99.2]	[2.2-4.2]	[77.6-77.9]	[32.1-59.0]	[63.4-65.3]

PPV: positive predictive value; NPV: negative predictive value; AUC: area under the curve

Chapter 6: Discussion and recommendations

Discussion and recommendations

Screening and diagnosis of this chronic disease with high prevalence should eliminate the many cases that remain undiagnosed and untreated, as OSA presence can lead to decrease quality of life, increased risk of adverse advents, and even a high rate of morbidity and mortality. Furthermore, polysomnography as the gold standard for this disease diagnosis is time-consuming, labor-intensive, limited to urban areas, costly and with long waiting lists, and for all of these is not suitable as a screening tool. Given these aspects, it would be useful to develop clinical prediction models, in non-sleep settings, that could reliably identify patients that most likely benefit from PSG (high-risk patients).

To see if the development of clinical prediction models was performed in OSA, we establish Goal 1 of this thesis – Synthesis of evidence. This systematic review included 63 original studies, with most of them performing traditional frequentist statistical inference - logistic and linear regressions. Looking at the studies that performed external validation, one reached 100% of sensitivity and another 100% of specificity, meaning that the first is the best at rule-out (excluding healthy participants to undergo PSG), while the other is best at rule-in participants (identify high-risk probability OSA patients). Both studies performed logistic regression, which misses the graphical representation. Regarding predictive variables present in OSA screening, we need to continue paying attention to the ones referred in the AASM guideline, such as body mass index, age, sex, neck circumference, but renew attention needs to be given to oxygen saturation and blood pressure.

After understanding the problem, data collection was conducted. The acquired data was performed in two hospitals in the North Region of Portugal, one with 318 patients and the other with 619. In both datasets, after describing and exploring the data, we found a setback - data quality. Most of the predictive variables elected to be collected presented missing values. To deal with this, in the first of three steps in Goal 2 – Classification and Prediction, we utilize drugs to infer several active diseases and vice-versa. The study achieved a 5% to 13% error in deducing arterial hypertension, dyslipidemia, and diabetes, reaching up to 7% of missed diagnoses. The other developed study applied nearest neighbors in the obtained dataset and constructed four Bayesian network models to assess the impact of using this technique, resulting in overlapped AUCs. This first step, settling in Data Understanding and Data Preparation of CRISP-DM, shows good results that can be and were applied in future work. In the second and third steps of Goal 2, we group studies that applied various modeling techniques, namely clustering and Bayesian networks. In the first modeling technique, we developed two studies that applied categorical cluster analysis: study one with

healthy and OSA patients, and study two with only OSA patients. These studies allowed us to visualize clinical subtypes of OSA that can help us focus our attention on specific predictive variables. In the second modeling technique, modeling and evaluation were accomplished in three studies. The first two applied Naïve and Tree Augmented Naïve Bayes in significant variables, with one also creating a risk matrix and another utilizing the information obtained from clustering as a node in the network. Overall, the AUCs of all developed models were above 80%, showing good discriminating power. The third manuscript performed the external validation of previously developed OSA clusters and showed a new visualization, namely a radar plot.

Reaching the last Goal of this thesis (Goal 3 - Implementation), which corresponds to the last phase of CRISP-DM, we focused our attention on bringing Bayesian network models towards primary care facilities or other non-sleep settings. To do so, we created an online form based on six predictive variables (gender, age, neck circumference, witnessed apneas, nocturia, and craniofacial and upper-airway abnormalities), that posteriorly were transformed in a smartphone app. Results showed that this app is easy and quick to acquire data and even works in the absence of predictive variables. Its usability evaluation reveals that 82% of the tasks were easily performed by the target population.

CRISP-DM proved to be good guidance throughout this thesis and allows us to see the interactions between the studies, most of the time creating new ideas to be developed, as it is an iterative model. Also, the thesis proves that OSA can have better screening and diagnosis, diminishing the number of unnecessary polysomnography being performed every day. According to the systematic review, there exists a good model, having been externally validated, with an overall low risk of bias and possible applicability [18]. The study used a logistic regression with age, waist circumference, Epworth Somnolence Scale, and minimum oxygen saturation for an AHI≥5. The AUC, sensitivity, and specificity were 98%, 94%, and 86%, respectively. The OSA prevalence was 87%, which could help explain an overestimated value for AUC, as the classes were unbalanced. Although our study described in 5.2. reached a higher value in sensitivity (99%), it only reached 3% for specificity. Nevertheless, we believe that all the taken steps, especially the implementation of OSABayes, demonstrated that this is a valid tool that can help any healthcare professional in the hard choice of sending or not a patient to a specialized appointment. We hope to have the opportunity to continue collecting data, as we need to continue re-assessing and training the Bayesian classifier as new patients are entered. We do not want this work to be another that stays on the shelf.

Bibliography

[1] V. K. Kapur *et al.*, "Clinical practice guideline for diagnostic testing for adult obstructive sleep apnea: An American academy of sleep medicine clinical practice guideline," *Journal of Clinical Sleep Medicine*, vol. 13, no. 3. American Academy of Sleep Medicine, pp. 479–504, 2017, doi: 10.5664/jcsm.6506.

[2] P. E. Peppard, T. Young, J. H. Barnet, M. Palta, E. W. Hagen, and K. M. Hla, "Increased prevalence of sleep-disordered breathing in adults.," *Am. J. Epidemiol.*, vol. 177, no. 9, pp. 1006–1014, Apr. 2013, doi: 10.1093/AJE/KWS342.

[3] J. A. Dempsey, S. C. Veasey, B. J. Morgan, and C. P. O'Donnell, "Pathophysiology of sleep apnea," *Physiol. Rev.*, vol. 90, no. 1, pp. 47–112, 2010, doi: 10.1152/PHYSREV.00043.2008/ASSET/IMAGES/LARGE/Z9J0011025260019.JPEG.

[4] V. Kapur *et al.*, "The Medical Cost of Undiagnosed Sleep Apnea," *Sleep*, vol. 22, no. 6, pp. 749–755, Sep. 1999, doi: 10.1093/SLEEP/22.6.749.

[5] A. V. Benjafield *et al.*, "Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis," *Lancet Respir. Med.*, vol. 7, no. 8, pp. 687–698, 2019, doi: 10.1016/S2213-2600(19)30198-5.

[6] T. Ononye, K. Nguyen, and E. Brewer, "Implementing protocol for obstructive sleep apnea screening in the primary care setting," *Appl. Nurs. Res.*, vol. 46, pp. 67–71, Apr. 2019, doi: 10.1016/J.APNR.2019.02.005.

[7] M. R. Bonsignore, P. Baiamonte, E. Mazzuca, A. Castrogiovanni, and O. Marrone, "Obstructive sleep apnea and comorbidities: a dangerous liaison," *Multidiscip. Respir. Med.*, vol. 14, no. 1, Feb. 2019, doi: 10.1186/S40248-019-0172-9.

[8] N. M. Al Lawati, S. R. Patel, and N. T. Ayas, "Epidemiology, Risk Factors, and Consequences of Obstructive Sleep Apnea and Short Sleep Duration," *Prog. Cardiovasc. Dis.*, vol. 51, no. 4, pp. 285–293, Jan. 2009, doi: 10.1016/J.PCAD.2008.08.001.

[9] R. J. O. Davies, N. J. Ali, and J. R. Stradling, "Neck circumference and other clinical features in the diagnosis of the obstructive sleep apnoea syndrome," *Thorax*, vol. 47, no. 2, pp. 101–105, 1992, doi: 10.1136/THX.47.2.101.

[10] J. M. Teich, J. A. Osheroff, E. A. Pifer, D. F. Sittig, and R. A. Jenders, "Clinical decision support in electronic prescribing: recommendations and an action plan: report of the joint clinical decision

support workgroup," *J. Am. Med. Inform. Assoc.*, vol. 12, no. 4, pp. 365–376, Jul. 2005, doi: 10.1197/JAMIA.M1822.

[11] M. J. Campbell, *Statistics at Square Two : Understanding Modern Statistical Applications in Medicine*. London: BMJ Books/Blackwell, 2006.

[12] P. Lucas, "Bayesian analysis, pattern analysis, and data mining in health care," *Curr. Opin. Crit. Care*, vol. 10, no. 5, pp. 399–403, Oct. 2004, doi: 10.1097/01.CCX.0000141546.74590.D6.

[13] T. Mitchell, *Machine learning*. Singapore: McGraw-Hill, 1997.

[14] A. Petrie and C. Sabin, *Medical Statistics at a Glance*. Wiley-Blackwell, 2009.

[15] C. Shearer, "The CRISP-DM Model: The New Blueprint for Data Mining," *J. Data Warehous.*, vol. 5, no. 4, pp. 13–22, 2000.

[16] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[17] D. Ferreira-Santos and P. P. Rodrigues, "A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches," *Int. J. Data Sci. Anal.*, vol. 8, pp. 339–349, 2019, doi: 10.1007/s41060-018-0118-x.

[18] J. Zou et al., "An effective model for screening obstructive sleep apnea: a large-scale diagnostic study," *PloS One,* vol. 8, no. 12, pp. e80704, Dec. 2013, doi: 10.1371/journal.pone.0080704.

Attachments

Attachment A:

Ethical permission

Autorização da Comissão de Ética do Centro Hospitalar de Vila Nova de Gaia/Espinho

Júlio Alberto Sampaio <jsampaio@chvng.min-saude.pt> sexta, 17/04/2015, 10:36

para mim

Exma. Sr.ª

Daniela Filipa Ferreira dos Santos

Informo que o pedido de autorização para a realização do Projecto de Investigação "***Bayesien network in obstrutive sleep apnea***", foi autorizado pela Direcção Clínica deste Centro Hospitalar.

Com os melhores cumprimentos,

CENTRO HOSPITALAR VILA NOVA DE GAIA|ESPINHO

Cuidamos de si.

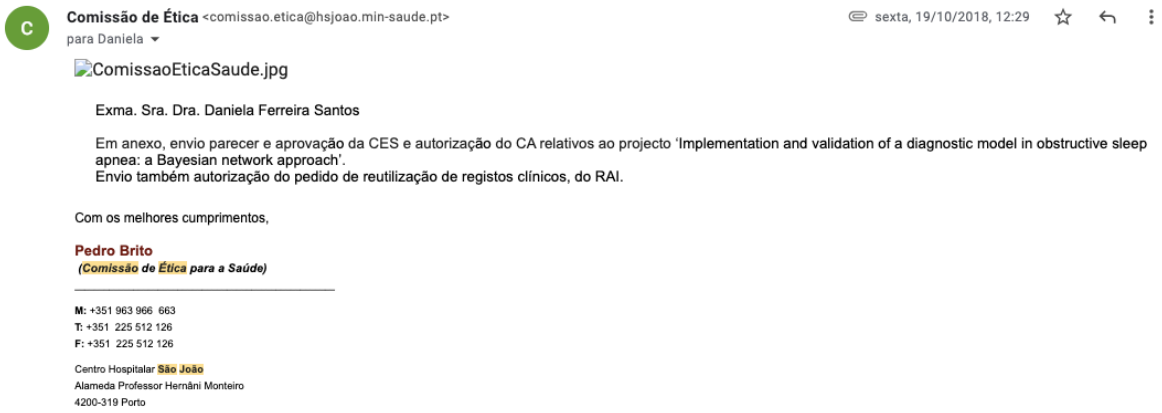
35 Aniversário Serviço Nacional Saúde

A cuidar dos portugueses

Júlio Alberto Sampaio
Responsável Serviço
Centro de Formação e Ensino
Serviço de Formação, Ensino e Investigação

Email jsampaio@chvng.min-saude.pt
Telefone: 227 865 127 Ext: 11452
Telemóvel: 962053725

Autorização da Comissão de Ética do Centro Hospitalar Universitário São João



Comissão de Ética <comissao.etica@hsjoao.min-saude.pt> sexta, 19/10/2018, 12:29

para Daniela

ComissaoEticaSaude.jpg

Exma. Sra. Dra. Daniela Ferreira Santos

Em anexo, envio parecer e aprovação da CES e autorização do CA relativos ao projecto 'Implementation and validation of a diagnostic model in obstructive sleep apnea: a Bayesian network approach'.
Envio também autorização do pedido de reutilização de registos clínicos, do RAI.

Com os melhores cumprimentos,

Pedro Brito
(Comissão de Ética para a Saúde)

M: +351 963 966 663
T: +351 225 512 126
F: +351 225 512 126

Centro Hospitalar São João
Alameda Professor Hernâni Monteiro
4200-319 Porto

Attachment B

Articles permissions to reuse

Impact of imputing missing data in Bayesian network structure learning for obstructive sleep apnea diagnosis

Impact of Imputing Missing Data in Bayesian Network Structure Learning for Obstructive Sleep Apnea Diagnosis

Authors	Daniela Ferreira-Santos, Matilde Monteiro-Soares, Pedro Pereira Rodrigues
Pages	126 - 130
DOI	10.3233/978-1-61499-852-5-126
Series	Studies in Health Technology and Informatics
Ebook	Volume 247: Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth

Abstract

Numerous diagnostic decisions are made every day by healthcare professionals. Bayesian networks can provide a useful aid to the process, but learning their structure from data generally requires the absence of missing data, a common problem in medical data. We have studied missing data imputation using a step-wise nearest neighbors' algorithm, which we recommended given its limited impact on the assessed validity of structure learning Bayesian network classifiers for Obstructive Sleep Apnea diagnosis.

Finding groups in obstructive sleep apnea patients: a categorical cluster analysis

31/12/21, 12:36 Rightslink® by Copyright Clearance Center

CCC RightsLink®

Home | Help | Email Support | Sign in | Create Account

IEEE Requesting permission to reuse content from an IEEE publication

Finding Groups in Obstructive Sleep Apnea Patients: A Categorical Cluster Analysis

Conference Proceedings:
2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)

Author: Daniela Ferreira-Santos

Publisher: IEEE

Date: Jun 2018

Copyright © 2018, IEEE

Thesis / Dissertation Reuse

The IEEE does not require individuals working on a thesis to obtain a formal reuse license, however, you may print out this statement to be used as a permission grant:

Requirements to be followed when using any portion (e.g., figure, graph, table, or textual material) of an IEEE copyrighted paper in a thesis:

1) In the case of textual material (e.g., using short quotes or referring to the work within these papers) users must give full credit to the original source (author, paper, publication) followed by the IEEE copyright line © 2011 IEEE.
2) In the case of illustrations or tabular material, we require that the copyright line © [Year of original publication] IEEE appear prominently with each reprinted figure and/or table.
3) If a substantial portion of the original paper is to be used, and if you are not the senior author, also obtain the senior author's approval.

Requirements to be followed when using an entire IEEE copyrighted paper in a thesis:

1) The following IEEE copyright/ credit notice should be placed prominently in the references: © [year of original publication] IEEE. Reprinted, with permission, from [author names, paper title, IEEE publication title, and month/year of publication]
2) Only the accepted version of an IEEE copyrighted paper can be used when posting the paper or your thesis on-line.
3) In placing the thesis on the author's university website, please display the following message in a prominent place on the website: In reference to IEEE copyrighted material which is used with permission in this thesis, the IEEE does not endorse any of [university/educational entity's name goes here]'s products or services. Internal or personal use of this material is permitted. If interested in reprinting/republishing IEEE copyrighted material for advertising or promotional purposes or for creating new collective works for resale or redistribution, please go to http://www.ieee.org/publications_standards/publications/rights/rights_link.html to learn how to obtain a License from RightsLink.

If applicable, University Microfilms and/or ProQuest Library, or the Archives of Canada may supply single copies of the dissertation.

BACK CLOSE WINDOW

© 2021 Copyright - All Rights Reserved | Copyright Clearance Center, Inc. | Privacy statement | Terms and Conditions
Comments? We would like to hear from you. E-mail us at customercare@copyright.com

https://s100.copyright.com/AppDispatchServlet#formTop 1/1

Obstructive sleep apnea: a categorical cluster analysis and visualization.

Pulmonology
Available online 17 November 2021
In Press, Corrected Proof

Original article

Obstructive sleep apnea: A categorical cluster analysis and visualization

Daniela Ferreira-Santos [a, b], Pedro Pereira Rodrigues [a, b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.pulmoe.2021.10.003

Phenotyping obstructive sleep apnea patients: a first approach to cluster visualization

Phenotyping Obstructive Sleep Apnea Patients: A First Approach to Cluster Visualization

Authors	Daniela Ferreira-Santos, Pedro Pereira Rodrigues

Pages	75 - 79
DOI	10.3233/978-1-61499-921-8-75
Series	Studies in Health Technology and Informatics
Ebook	Volume 255: Decision Support Systems and Education

Abstract

The varied phenotypes of obstructive sleep apnea (OSA) poses critical challenges, resulting in missed or delayed diagnosis. In this work, we applied k-modes, aiming to identify groups of OSA patients, based on demographic, physical examination, clinical history, and comorbidities characterization variables (n = 41) collected from 318 patients. Missing values were imputed with k-nearest neighbours (k-NN) and chi-square test was held. Thirteen variables were inserted in cluster analysis, resulting in three clusters. Cluster 1 were middle-aged men, while Cluster 3 were the oldest men and Cluster 2 mainly middle-aged women. Cluster 3 weighted the most, whereas Cluster 1 weighted the least. The same effect was described in increased neck circumference. The percentages of variables driving sleepiness, congestive heart failure, arrhythmias and pulmonary hypertension were very low (<20%) and OSA severity was more common in mild level. Our results suggest that it is possible to phenotype OSA patients in an objective way, as also, different (although not considered innovative) visualizations improve the recognition of this common sleep pathology.

A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches

About this article

Cite this article

Ferreira-Santos, D., Rodrigues, P.P. A clinical risk matrix for obstructive sleep apnea using Bayesian network approaches. *Int J Data Sci Anal* **8,** 339–349 (2019). https://doi.org/10.1007/s41060-018-0118-x

Download citation

Enhancing obstructive sleep apnea diagnosis with screening through disease phenotypes: algorithm development and validation

Enabling early obstructive sleep apnea diagnosis with machine learning: systematic review