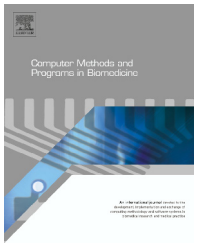
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 110 (2013) 361–368

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



Can computed tomography classifications of chronic obstructive pulmonary disease be identified using Bayesian networks and clinical data?

Lars P. Thomsen[a,*], *Ulla M. Weinreich*[b], *Dan S. Karbing*[a], *Vanja G. Helbo Jensen*[c], *Morten Vuust*[d], *Jens B. Frøkjær*[c], *Stephen E. Rees*[a]

[a] *Respiratory and Critical Care Group (RCARE), Center for Model-based Medical Decision Support, Department of Health Science and Technology, Aalborg University, Fredrik Bajers Vej 7E, DK-9220 Aalborg East, Denmark*
[b] *Department of Pulmonary Medicine, Aalborg University Hospital, Mølleparkvej 4, DK-9000 Aalborg, Denmark*
[c] *Department of Radiology, Aalborg University Hospital, Hobrovej 18-22, DK-9000 Aalborg, Denmark*
[d] *Department of Radiology, Sygehus Vendsyssel, Barfredsvej 83, DK-9000 Fredrikshavn, Denmark*

ARTICLE INFO

Article history:
Received 7 September 2012
Received in revised form
6 February 2013
Accepted 8 February 2013

Keywords:
Airflow limitation
Pulmonary gas exchange
Chronic obstructive pulmonary disease
Biomedical modeling

ABSTRACT

Diagnosis and classification of chronic obstructive pulmonary disease (COPD) may be seen as difficult. Causal reasoning can be used to relate clinical measurements with radiological representation of COPD phenotypes airways disease and emphysema. In this paper a causal probabilistic network was constructed that uses clinically available measurements to classify patients suffering from COPD into the main phenotypes airways disease and emphysema. The network grades the severity of disease and for emphysematous COPD, the type of bullae and its location central or peripheral. In four patient cases the network was shown to reach the same conclusion as was gained from the patients' High Resolution Computed Tomography (HRCT) scans. These were: airways disease, emphysema with central small bullae, emphysema with central large bullae, and emphysema with peripheral bullae. The approach may be promising in targeting HRCT in COPD patients, assessing phenotypes of the disease and monitoring its progression using clinical data.

1. Introduction

Chronic obstructive pulmonary disease (COPD) is a world leading cause of death with mortality rates predicted to increase [1]. Despite this, diagnosis of COPD can be seen as difficult, with COPD representing a heterogeneous group of disease categories including airways disease, emphysema, the type and severity of these; and the individual patient's clinical presentation. Only recently have recommendations been available which attempt to consider disease heterogeneity [2], with increasing effort being made to identify phenotypes of COPD that divide patients into subgroups expressing their measurements and outcomes [3]. As proposed by Friedlander et al. [4] the individual patient's phenotype might be described by combining data from one or more of three data sets: clinical data, physiological measurements, and radiological data.

Relating and combining data from these three sources is difficult. Physiologic data often describes the mechanical or gas exchange properties of the lung where radiology illustrates the lungs pathoanatomy involved in disease subtypes such as tissue degeneration in emphysema or bronchial wall

* *Corresponding author*. Tel.: +45 9940 9833.
E-mail address: lpt@hst.aau.dk (L.P. Thomsen).
0169-2607/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cmpb.2013.02.001

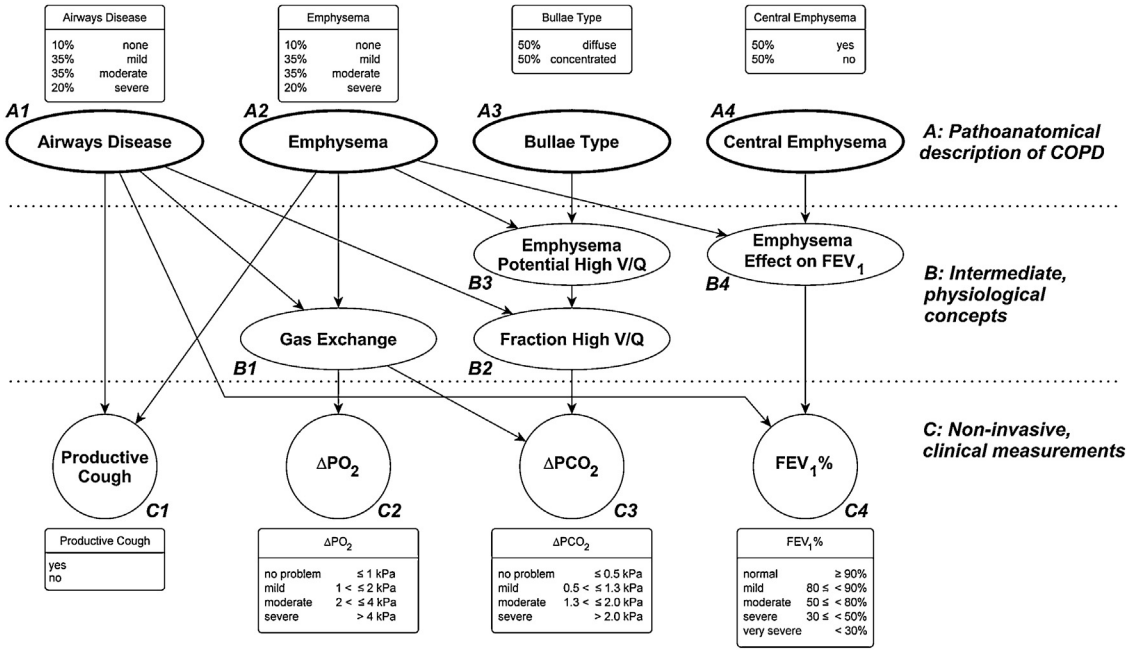


Fig. 1 – The causal probabilistic network divided into three horizontal layers, A–C, with directed arrows showing causal relation explained further in the text. The network includes three layers describing the pathoanatomy of COPD based on CT/HRCT (A), intermediate physiological concepts (B) and non-invasive clinical measurements (C), with details of the individual variables given in the text.

thickening in airways disease. Clinical data includes the level of dyspnea, number of exacerbations, and how the individual patient perceives the impact of disease. In the current clinical recommendations for diagnosing COPD [2], physiological and clinical data is combined into a risk assessment for the patient whereas the radiological data remains unrelated to this risk [5].

Combination of data from these three sources may be aided by a systematic application of causal reasoning, which has been shown applicable in other fields [6,7]. Several causal mechanisms can be postulated which link radiological, physiologic and clinical data. For instance, airways disease presenting on computed tomography (CT) as airway remodeling and bronchial wall thickening is more likely to cause sputum production than emphysema [8]. Similarly, using data from CT, emphysema can be described by grade, location and size of bullae, with each of these classifications perhaps related causally to profiles of pulmonary gas exchange, lung mechanics or clinical parameters.

Causal probabilistic networks (CPN) – or Bayesian networks – is a technique which allows causal reasoning to be integrated with measurements [9]. This paper investigates whether a CPN model can be built which integrates clinical, physiological and radiological knowledge into a network, and then by instantiating simple clinical measurements in the network whether it can adequately describe standard disease patterns obtained from CT data. This paper focusses on model formulation and proof of concept, showing the potential for application of the network in four well defined situations: small airways disease; emphysema with diffuse central bullae; emphysema with concentrated central bullae; and emphysema with diffuse peripheral bullae. Model descriptions are compared with radiological findings in these four cases.

2. Methods

In this section the CPN is presented along with the clinical measurements used as input for the network and the CT datasets used for comparison.

2.1. *Description of the CPN for COPD classification*

This section presents the CPN illustrated in Fig. 1. The aim of the CPN is to classify patients suffering from COPD into the two main disease categories: airways disease or emphysema, including their severity and type. These are selected as they represent the main radiologic subtypes of COPD found reliably from CT data [4,10], and as they make independent contributions to the cardinal sign of COPD – airflow obstruction [11]. The CPN consists of 12 variables and is divided into three layers (A–C). The variables in the top layer (A) describe the pathoanatomical subgroups airways disease and emphysema. The bottom layer (C) consists of the clinical measurements, and variables in the middle layer (B) describe intermediate physiological concepts necessary to establish the causal reasoning in the network. The following text describes the variables and causal links relating to Fig. 1.

In the top layer the two pathoanatomical subgroups *airways disease* and *emphysema* are described by variables A1 and A2, respectively. The severity of these are graded into none, mild, moderate, and severe where, a priori, it is assumed that 70%