DOCTORAL THESIS

Clinical Decision Support for screening, diagnosis and assessment of respiratory diseases: Chronic Obstructive Pulmonary Disease as a use case

Filip Velickovski

2016

Universitat de Girona

DOCTORAL THESIS

Clinical Decision Support for screening, diagnosis and assessment of respiratory diseases: Chronic Obstructive Pulmonary Disease as a use case

Filip Velickovski

2016

DOCTORAL PROGRAM IN TECHNOLOGY

Supervisors:
Ph.D. Robert Martí and Ph.D. Luigi Ceccaroni

This thesis is presented in fulfillment of the requirement for the conferral of the degree of Doctor of Philosophy by the University of Girona

Dr. Robert Marti of University of Girona, Girona, Spain

and

Dr. Luigi Ceccaroni[1] of 1000001 Labs, Barcelona, Spain

WE DECLARE:

That the thesis titles "*Clinical Decision Support for screening, diagnosis and assessment of respiratory diseases: Chronic Obstructive Pulmonary Disease as a use case"* presented by Filip Velickovski to obtain a doctoral degree, has been completed under my supervision

For all intents and purposes, we hereby sign this document.

Robert Marti
Girona, 24/06/2016

Luigi Ceccaroni
Barcelona, 24/06/2016

[1] Formerly affiliated with Eurecat, Barcelona, Spain

Science is a way of thinking
much more than it is a body of knowledge.

CARL SAGAN

Acknowledgements

I would like to acknowledge several individuals who have influenced, supported, and encouraged me in the completion of this thesis.

Firstly I would like to thank immensely my two thesis advisers Dr. Luigi Ceccaroni and Dr. Robert Martí who guided me through the difficult task of working at an organisation as well as working towards a PhD. Without their continuous support, availability for feedback, intellectual discussions, and guidance, this thesis would not have been possible.

Secondly I would like to thank all my colleagues in the E-Health unit at Eurecat, Barcelona who have provided a motivating, intellectually stimulating, and a very pleasant work environment. I give my gratitude to Dr. Felip Miralles, the manager of the E-Health unit, for his support, and in letting me combine a PhD. with work, and giving me access to research collaborations through the European FP7 Synergy-COPD project that partly funded the research in this thesis.

I am very grateful to Dr. Josep Roca and Dr. Felip Burgos from the University of Barcelona and IDIBAPS research group, who have always been there to provide me with the clinical perspective that has been important in my thesis, and have been key collaborators. Next I would like to thank Dr. Umberto Melia, Dr. Montserrat Vallverdú and Dr. Pere Caminal from the CREB research group at the Universitat Politècnica de Catalunya for our excellent collaboration in spirometry quality assurance, one of the principal elements to my thesis.

Many thanks to my parents Eva and Dragan for the values and qualities they have instilled in me, and motivating me to always move forward. Additionally, I would like to mention that my mum and G.P. at hand, Dr. Eva Velickovski, gave me the primary care perspective needed in this thesis and was relentless in supplying me with clinical material on COPD.

Finally and most importantly, I owe the deepest gratitude to my partner in life Mirella. She has been my pillar of support, inspiring me during moments of doubt, and has been the persistent force of enthusiasm and encouragement throughout the entire thesis.

Publications

The following are a list of publications resulting from the work completed during the thesis. They are categorized as: (i) *Core Research* are the three journal articles forming the main body of this thesis, which is presented as **a compendium of articles**; (ii) *Supporting Research* are journal articles and conference proceedings that have resulted from secondary aspects of the work presented in this thesis in the broader topics of *Health Informatics* and *Decision Support* in other non-clinical contexts.

Core research

[1] F. Velickovski, L. Ceccaroni, J. Roca, F. Burgos, J. B. Galdiz, N. Marina, and M. Lluch-Ariet. "Clinical Decision Support Systems (CDSS) for preventive management of COPD patients". In: *Journal of Translational Medicine* 12.2 (2014), pp. 1–10. DOI: 10.1186/1479-5876-12-S2-S9.

[2] U. Melia, F. Burgos, M. Vallverdu, F. Velickovski, M. Lluch-Ariet, J. Roca, and P. Caminal. "Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments". In: *PLoS ONE* 9.12 (Dec. 2015), pp. 1–14. DOI: 10.1371/journal.pone.0116238.

[3] F. Velickovski, L. Ceccaroni, R. Marti, F. Burgos, C. Gistau, X. Alsina-Restoy, and J. Roca. "Automated spirometry quality assurance: supervised learning from multiple experts". In: *Computer Methods and Programs in Biomedicine* (2016). Submitted.

Supporting research

Health informatics

[4] F. Velickovski, S. Orte, M. Sola, S. Tabozzi, and C. L. Lafortuna. "EAI International Conference on Wearables in Healthcare, Budapest, Hungary, June 14-15, 2016." In: LNCS. Springer International Publishing. Chap. Detection and

assessment of behaviours associated with the risk of obesity in adolescents. Accepted.

[5] F. Burgos, U. Melia, M. Vallverdú, F. Velickovski, M. Lluch-Ariet, P. Caminal, and J. Roca. “Clinical Decision Support System to Enhance Quality Control of Spirometry Using Information and Communication Technologies”. In: *JMIR Medical Informatics* 2.2 (Oct. 2014), e29. DOI: 10.2196/medinform.3179.

[6] C. Vargas, F. Velickovski, F. Burgos, A. Alonso, L. Ceccaroni, and J. Roca. “COPD case-finding application: SYNERGY project”. In: *European Respiratory Journal* 42.Suppl 57 (2014).

[7] A. A. Navarro, L. Ceccaroni, F. Velickovski, S. Torrellas, F. Miralles, B. Z. Allison, R. Scherer, and J. Faller. “Context-Awareness as an Enhancement of Brain-Computer Interfaces”. In: *Ambient Assisted Living: Third International Workshop, IWAAL 2011, Held at IWANN 2011, Torremolinos-Málaga, Spain, June 8-10, 2011. Proceedings*. Ed. by J. Bravo, R. Hervás, and V. Villarreal. Springer Berlin Heidelberg, 2011, pp. 216–223. ISBN: 978-3-642-21303-8. DOI: 10.1007/978-3-642-21303-8_30.

[8] A. Hoekstra, J. Domingue, L. Ceccaroni, F. Velickovski, and M. Viceconti. “Web of models”. In: *VPH-FET Research Roadmap-Advanced Technologies for the Future of the Virtual Physiological Human*. Ed. by M. Viceconti and G. Clapworthy. VPH-FET consortium, Sept. 2011, pp. 70–77.

Decision support in other application areas

[9] L. Ceccaroni, M. Blaas, M. R. Wernand, F. Velickovski, A. Blauw, and L. Subirats. *A decision support system for water quality in the Wadden Sea*. Presented at the 47th International Liege Colloquium on Ocean Dynamics Marine Environmental Monitoring, Modelling and Prediction. 4-8th May, 2015 Liège, Belgium. 2015.

List of Acronyms

ANN multilayer artificial neural network

ANSI American National Standards Institute

API application programming interface

ATS American Thorax Society

AUC area under the receiver operating characteristic curve

BEV back extrapolated volume

CAD computer aided diagnosis

CBR case based reasoning

CCD Continuity of Care Document

CDA Clinical Document Architecture

CDSS clinical decision support system

CF certainty factors

CG clinical guideline

CIG computer-interpretable guideline

CLIPS C Language Integrated Production System

COPD chronic obstructive pulmonary disease

CT computerised tomography

DRL Drools rules language

DSL domain specific language

EHR electronic health record

EOTV end of test volume

ERS European Respiratory Society‘

FET100 forced expiratory time

FEV1 forced expiratory volume in one second

FS forced spirometry

FT flow time

FV flow volume

FVC forced vital capacity

GEE guideline execution engine

GOLD Global Initiative for COPD

GP general practitioner

GUI graphical user interface

HIS health information system

HL7 Health Level Seven International

ICD International Classification of Diseases

ICT information communication technology

JESS Java expert System Shell

KBS knowledge-based system

kNN k-nearest neighbour

LLN lower limit of normal

LOINC Logical Observation Identifiers Names and Codes

MLM medical logic module

MRC Medical Research Council

MRI magnetic resonance imaging

NCD non-communicable disease

PEF peak expiratory flow

PHR personal health record

RIM Reference Information Model

ROC receiver operating characteristic

SNOMED-CT Systematized Nomenclature of Medicine Clinical Terms

SVM support vector machine

VMR Virtual Medical Record

VT volume time

List of Tables

List of Figures

Contents

Abstract

Clinical decision support systems (CDSS), which aim to provide healthcare staff with patient specific advice, can enhance the level of care delivered to citizens by offering access to advanced medical knowledge in non-specialist healthcare settings. The work presented in this thesis adapts, refines and contributes to methodologies in applied clinical decision support research, through a concrete use-case of Chronic Obstructive Pulmonary Disease (COPD) at the early stage, with the outlook of generalising these methods to a broader set of chronic respiratory diseases, and other non-communicable diseases.

Chronic obstructive pulmonary disease is a major cause of chronic morbidity and mortality worldwide, and along with other chronic repository diseases currently represents a high burden on global healthcare systems. The primary objective of this thesis is to facilitate through clinical decision support research and development, the clinical tasks related to early stage COPD detection.

In this thesis we propose a framework for designing, developing, a CDSS offering a suite of services for the early detection and assessment of COPD, and then demonstrate how these services can be integrated into the work-flow of healthcare providers. Furthermore, we focus on supporting spirometry, one of the main diagnostic tools in respiratory disease assessment. We present two methods to offer decision support in assuring the quality of a spirometry test that can be easily embedded into the CDSS framework. The first method is a novel algorithm that relies on a set of rules operating on 23 new parameters to define a high quality test. The second is a machine-learning approach, where we optimise the distinction between a good quality spirometry test and a poor one using a set of supervised-learning classifiers and hyper-parameters.

The application of the outcomes generated from this research has a credible potential to contribute to lowering the level of under-diagnosis, reducing the level of misdiagnosis, and improving the quality of lung function assessment performed in non-specialist settings for COPD as well as other chronic respiratory diseases.

Resumen

Los Sistemas de Soporte de Decisión Clínica (SSDC) tienen como objetivo proporcionar consejo e información específica del paciente al personal sanitario. Los SSDC pueden mejorar el nivel de atención médica aportado a los ciudadanos, gracias al acceso al conocimiento clínico avanzado en entornos de salud no-especialistas. El trabajo desarrollado en esta tesis adapta, mejora y contribuye al estado del arte en investigación aplicada a los SSDC mediante el estudio práctico de la Enfermedad Pulmonar Obstructiva Crónica (EPOC) en fase inicial, con la intención de aplicar este método a un conjunto más amplio de enfermedades pulmonares crónicas y otras enfermedades no transmisibles.

La EPOC genera un alto índice de morbididad crónica y mortalidad mundial, y junto con el resto de enfermedades pulmonares crónicas, supone una importante inversión de recursos para los sistemas sanitarios mundiales. El objetivo fundamental de esta tesis es facilitar mediante la investigación y el desarrollo de los SSDC, las tareas de detección de la EPOC en su fase inicial.

En esta tesis proponemos un marco para el diseño y desarrollo de un SSDC que ofrezca un conjunto de herramientas para el diagnóstico precoz y la evaluación de la EPOC. Al mismo tiempo demostramos como estos servicios se pueden integrar en el flujo de trabajo del personal sanitario. Además, nos centramos en la ayuda en espirometría, una de las herramientas de diagnóstico principales en la evaluación de enfermedades pulmonares. En base a lo expuesto anteriormente presentamos dos métodos de soporte de decisión que tienen como objetivo asegurar la calidad de las pruebas de espirometría, y que se pueden integrar fácilmente en el marco del SSDC. El primero de estos métodos es un nuevo algoritmo que se basa en un conjunto de reglas que definen lo que es considerado como una prueba de alta calidad, usando para ello 23 nuevos parámetros. El segundo método es un enfoque de aprendizaje automático donde se optimiza la distinción entre una prueba correcta de espirometría y una de mala calidad mediante el uso de un conjunto de clasificadores de aprendizaje supervisado y de híper-parámetros.

La aplicación de los resultados generados en esta investigación puede contribuir positivamente a la reducción del nivel de infra diagnóstico y del nivel de diagnóstico erróneo de la EPOC. Al mismo tiempo también puede mejorar la calidad de la evaluación

de la función pulmonar para EPOC y otras enfermedades pulmonares crónicas en entornos no-especialistas.

Resum

Els Sistemes de Suport de Decisió Clínica (SSDC) tenen com objectiu proporcionar consell i informació específica del pacient al sanitari personal. Els SSDC poden millorar el nivell d'atenció mèdica aportat als ciutadans, gràcies al accés al coneixement clínic avançat en entorns de salut no-especialistes. El treball desenvolupat en aquesta tesi adapta, millora i contribueix a l'estat de l'art en investigació aplicada als SSDC mitjançant l'estudi pràctic de la Malaltia Pulmonar Obstructiva Crònica (MPOC) en fase inicial, amb la intenció d'aplicar aquest mètode a un conjunt més ampli de malalties pulmonars cròniques i altres malalties no transmissibles.

La MPOC genera un alt índex de morbiditat crònica i mortalitat mundial, i juntament amb la resta de malalties pulmonars cròniques, suposa una important inversió de recursos per als sistemes sanitaris mundials. L'objectiu fonamental d'aquesta tesi és facilitar mitjançant la investigació i el desenvolupament dels SSDC, les tasques de detecció de la MPOC en la seva fase inicial.

En aquesta tesi proposem un marc per al disseny i desenvolupament d'un SSDC que ofereixi un conjunt d'eines per al diagnòstic precoç i l'avaluació de la MPOC. Alhora demostrem com aquests serveis es poden integrar en el flux de treball del personal sanitari. A més, ens centrem en l'ajuda a espirometria, una de les eines de diagnòstic principals en l'avaluació de malalties pulmonars. En base a l'exposat anteriorment presentem dos mètodes de suport de decisió que tenen com a objectiu assegurar la qualitat de les proves d'espirometria, i que es poden integrar fàcilment en el marc del SSDC. El primer d'aquests mètodes és un nou algoritme que es basa en un conjunt de regles que defineixen el que és considerat com una prova d'alta qualitat, utilitzant per a això 23 nous paràmetres. El segon mètode és un enfocament d'aprenentatge automàtic on s'optimitza la distinció entre una prova correcta d'espirometria i una de mala qualitat mitjançant l'ús d'un conjunt de classificadors d'aprenentatge supervisat i d'híper-paràmetres.

L'aplicació dels resultats generats en aquesta investigació pot contribuir positivament a la reducció del nivell d'infra diagnòstic i del nivell de diagnòstic erroni de la MPOC. Alhora també pot millorar la qualitat de l'avaluació de la funció pulmonar per MPOC i altres malalties pulmonars cròniques en entorns no-especialistes.

Thesis statement

This thesis is presented as a compendium of the following articles:

1. F. Velickovski, L. Ceccaroni, J. Roca, F. Burgos, J. B. Galdiz, N. Marina, and M. Lluch-Ariet. "Clinical Decision Support Systems (CDSS) for preventive management of COPD patients". In: *Journal of Translational Medicine* 12.2 (2014), pp. 1–10. DOI: 10.1186/1479-5876-12-S2-S9

    - Journal: Journal of translation medicine
    - Impact factor: 3.694
    - Quartile: 1st
    - Status: Published

2. U. Melia, F. Burgos, M. Vallverdu, F. Velickovski, M. Lluch-Ariet, J. Roca, and P. Caminal. "Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments". In: *PLoS ONE* 9.12 (Dec. 2015), pp. 1–14. DOI: 10.1371/journal.pone.0116238

    - Journal: PLOS ONE
    - Impact factor: 3.057
    - Quartile: 1st
    - Status: Published

3. F. Velickovski, L. Ceccaroni, R. Marti, F. Burgos, C. Gistau, X. Alsina-Restoy, and J. Roca. "Automated spirometry quality assurance: supervised learning from multiple experts". In: *Computer Methods and Programs in Biomedicine* (2016). Submitted

    - Journal: Computer Methods and Programs in Biomedicine
    - Impact factor: 1.862
    - Quartile: 1st
    - Status: Submitted

Chapter 1

Introduction

Modern medicine is a unique synthesis of a patient, healthcare professionals, and technology. For the past 10-15 years there has been a slow but steady increase in the use and storage of electronic machine readable formats known as electronic health records (EHRs). It is unlikely that there will be major improvements in the quality and cost of care, solely from the use of EHRs without the proper implementation and use of clinical decision support [4–6]. A clinical decision support system (CDSS) can be defined as "software that is designed to be of direct aid to clinical decision-making in which the characteristics of an individual patient are matched to a computerized clinical knowledge base, and patient-specific assessments or recommendations are then presented to the clinician and/or the patient for a decision" [7]. Figure 1.1 depicts the three principle elements generally required for a CDSS, which are:

- the *knowledge base* contains in a computer interpretable format the rules, associations, and clinical know-how for the task at hand (e.g. screening, diagnosis, treatment, prognosis);
- the *algorithms* determine how to combine the *knowledge base* to an instance of patient specific data, which is supplied to the system in order to generate an actionable recommendation or assessment of the patient;
- the *communication mechanism* is the manner in which the system inputs the patient specific data and outputs the recommendations or assessments to the clinician.

The algorithms employed by a CDSS can be characterised by the reasoning paradigm employed by the CDSS which is primarily of two types. The first type is the *knowledge-driven* paradigm, which employs an inference engine that applies the formalised clinical knowledge representations stored in the knowledge base to an instance of the patient's data. The knowledge base in this case, can contain rules for treatment

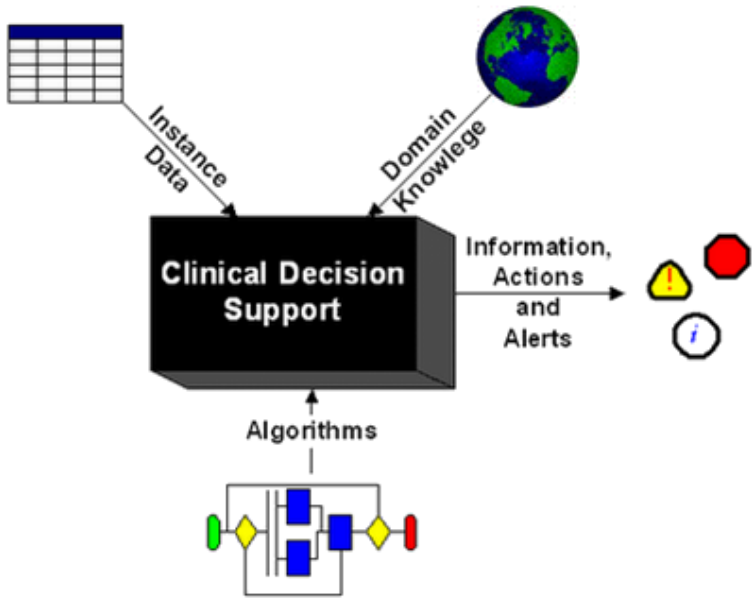

Figure 1.1: Principle elements of a clinical decision support system (CDSS)

or diagnosis, probabilistic associations of symptoms to diseases, drug-drug interactions, or clinical workflows with decision steps. The second type is the *data-driven* reasoning paradigm which employ methods from machine-learning such as neural networks, support vector machines, statistical methods such as regression models, and pattern recognition methods such as k-nearest neighbours (kNN) to detect patterns in clinical data. Rather than an inference engine, the *data-driven* approach uses a classifier, and the data contained in the knowledge-base may be thought of as the trained parameters or weights associated with the classifier.

Optimal use of CDSSs have the potential to improve healthcare processes and outcomes by ensuring compliance with the most up to date guidelines, reduce clinical errors, and reduce cost without compromising care [8–10]. A CDSS should be viewed as supportive tool available to the clinician to facilitate their task, and definitely not as her substitute.

1.1 Motivation and background

Non-communicable diseases (NCDs) are non-infectious or non-transmissible chronic diseases which last for long periods of time and progress slowly. They are the leading cause of death, causing 38 million deaths world wide [11]. Chronic obstructive pulmonary disease (COPD), is a preventable and treatable NCD characterized by airflow limitation that is not fully reversible and is mainly caused by smoking. The airflow limitation

is usually progressive and associated with an abnormal inflammatory response of the lung to noxious particles. Although primarily a pulmonary disease, COPD has some extra-pulmonary effects that may contribute to the severity in individual patients. Primary symptoms include cough, sputum production, and dyspnea (shortness of breath) on exertion. Episodes of acute worsening, known as exacerbations of these symptoms often occur. The presence of systemic inflammation in COPD has been linked to a variety of complications including: weight loss, cachexia (wasting syndrome), osteoporosis, and cardiovascular diseases.

Chronic obstructive pulmonary disease is a major cause of chronic morbidity and mortality worldwide [12] and along with other chronic repository diseases, currently represents a high burden on global healthcare systems [11, 13]. This trend is expected to increase as more people are concentrated in urban environments, and COPD has placed a great demand on primary care clinicians which do not have the specialist knowledge of a pulmonologist. Another important problem in healthcare is the significant gap between optimal evidence-based medical practice and the care actually applied. In a systematic review [14] of COPD guideline adherence, it was found that non-pharmacological treatment was infrequently explored for the in-hospital management of COPD exacerbations, and the assessment and therapy applied to the disease were suboptimal. This trend exists not only in COPD but across all chronic-disease care in general: in a multinational survey [15] of chronically ill adults, 14-23% of cases reported at least one medical error in the previous two years.

Clinical decision support systems, which aim to provide clinical staff, and patients with advice and personalized information, have the potential to enhance healthcare, and to help close the gap between optimal practice and actual clinical care. This is especially true for COPD due to the highly specialized knowledge that is needed to (i) recognise a potential case and assess the state of the subject's lung with a spirometer (a medical device to test lung function) (ii) confirm the diagnosis (iii) determine the severity and stratify the type of COPD patient (iv) assign the appropriate pharmacological and non-pharmacological treatment (v) monitor the patient over his or her lifetime. Due to the high prevalence of the disease, COPD must be detected and managed in non-specialists settings such as primary care, remote home care, and other allied health service providers.

This dissertation submitted as a compendium of articles, takes COPD as a NCD use case, and contributes to strategies, algorithms, and techniques for designing decision support tools (using both knowledge-driven and data-driven paradigms) primarily focused on early-stage COPD.

1.2 Context

Part of the work contained in this thesis was facilitated by the Synergy-COPD project (FP7 2010-2014) [16] whose general aim was to explore the potential of using a computer-based systems medicine approach to improve existing knowledge on the underlying mechanisms of COPD that should lead to better understanding of disease heterogeneity. The transfer of biomedical knowledge generated during the project into healthcare (aiming at enhancing chronic patient management) was an intrinsic component of Synergy-COPD.

The project addressed two main strategic areas. Firstly, contributing to foster the convergence between basic and clinical sciences. Secondly, the project applied a new approach not only to conventional care, but also to promoting the link between systems medicine and integrated care, aiming at personalized health for patients suffering from NCDs.

Synergy-COPD delivered and validated outcomes with outstanding impact both in clinical practise and research. In the first case, a CDSS (which is the subject of this thesis and the work presented is my own) to bring early diagnosis of COPD to primary and informal care, and support prognosis and therapy management of the chronic disease. In the second case, a Simulation Environment powered by a Knowledge Base, to allow researchers in biomedicine and bioinformatics to interact with the models and data, and furthermore to help improve their understanding about the underlying mechanisms of the disease.

1.3 Main objectives

The primary objective of this thesis is to facilitate through clinical decision support research and development, the tasks related to early stage COPD detection so that healthcare providers (primary care clinicians, nurses, and/or allied health service providers such as pharmacists) can obtain fast, reliable and directly applicable advice when facing an existing or potential COPD patient. Specifically these clinical tasks are (i) screening / case-finding, (ii) diagnosis, and (iii) patient stratification. Furthermore, the pulmonary function test of spirometry is necessary for all the aforementioned tasks. Hence, in this thesis there is a special emphasis on the goal of ensuring high quality spirometry assessment for non-expert users through novel algorithms and applications of artificial intelligence.

More specifically, the objectives of this work are:

- objective 1 - investigate, propose and implement the optimal CDSS architecture

and design in order to support primary care and allied healthcare providers with detecting and assessing early-stage COPD.

- objective 2 - investigate and develop CDSS algorithms (both knowledge-driven and/or data driven) for the clinical tasks involved in COPD management: (i) case-finding (ii) diagnosis and (iii) patient stratification.
- objective 3 - investigate and develop CDSS algorithms (both knowledge-driven and/or data driven) for the quality assurance of spirometry (the gold standard for lung function measurement).
- objective 4 - validate the CDSS algorithms against expert clinical professionals using existing datasets, and extending existing datasets where necessary.

Achieving the stated objectives will assist in early detection of COPD cases, and raise the diagnostic quality done in non-specialist settings for COPD as well as other chronic respiratory diseases. Furthermore these objectives are essential towards the deployment of remote care and monitoring technologies that have the potential to reduce hospital admissions.

1.4 Data protection

Data protection is an important consideration when it comes to information and communication technology for health. In the European Union (EU), data protection is explicitly enshrined in a legal framework consisting of the "Treaty of the functioning of the EU (Article 16)" [17], "EU charter of fundamental rights (Article 8)" [18], the "Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data" [19], as well as many individual national laws. The CDSS proposed in thesis involves the processing and transfer of sensitive health data, and if deployed in practice, therefore would be subject to rules governing data protection. Although it is not the primary objective, the CDSS framework proposed in this dissertation nevertheless employs a set of measures that enhance the privacy of the patient and protection of their data:

- Personally identifiable information (names, addresses, phone numbers, etc.) is not required and removed when transferring the patient data to the CDSS.
- Only the relevant clinical data (measurements, symptoms, comorbidities, etc.) that is needed to complete the decision support task is transferred to the CDSS and no other data.

- Strong encryption standards are used for the transfer of sensitive patient data
- The modular design allows for the CDSS to be deployed on-site on the intranet, not requiring any data to be transferred through the public internet
- Patient data is not permanently stored on the CDSS

1.5 Structure of this dissertation

This thesis is presented as a compendium of articles and is organized in seven chapters. Following this introduction, in Chapter 2 we review the state of the art in clinical decision support and research related to its application in respiratory diseases.

The next three chapters form the compendium of articles, and are presented in the same format as they are published or submitted. In Chapter 3, we propose a framework for designing, developing, a CDSS offering a suite of services for the early detection and assessment of COPD, and propose and demonstrate how they can be integrated into the workflow of healthcare providers. In Chapter 4 and Chapter 5, we focus on supporting spirometry (one of the main diagnostic tools in respiratory disease assessment) through two methods that could potentially be incorporated into a CDSS. The first, presented in Chapter 4, is a novel algorithm for automatic forced spirometry quality assessment relying on a set of rules or criteria operating on 23 new parameters to define a high quality test. The second, presented in Chapter 5, is an alternative new approach to same problem, however using a supervised-learning method that can automatically learn to distinguish between a good quality spirometry test and a poor one.

In Chapter 6, we discuss the main results, and contributions of the work, and finally, ending with Chapter 7, we state the conclusions and propose the future directions of this research.

Chapter 2

Literature review

In this chapter the features and aspects that differentiate clinical decision support systems (CDSSs) will be reviewed. Particularly the focus will be on the reasoning methods employed by CDSSs that are responsible for using personal clinical data of a patient to recommend to the clinician the best care path to follow. Finally we review the state of the art in CDSS for our particular use-case: chronic obstructive pulmonary disease (COPD).

2.1 Characterization of CDSSs

CDSSs come in a variety of forms, with various characteristics. In order to understand why some CDSSs are successful and others are not it is important to examine the critical features that describe a CDSS. One of the most common ways to classify a CDSS is by the type of knowledge and reasoning method it uses for arriving at the advice on the clinical decision.

Some CDSSs can be considered knowledge-based systems (KBSs), which arose

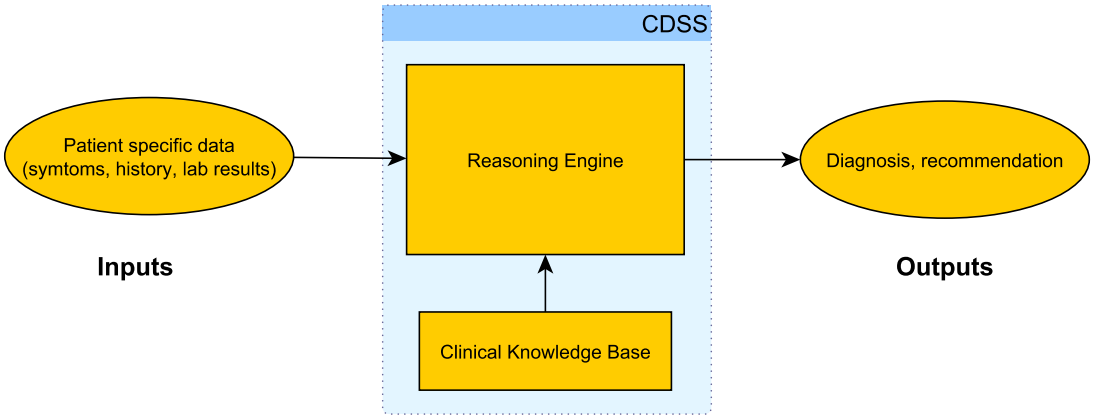

Figure 2.1: A general model of a knowledge-based CDSS

Table 2.1: Taxonomy of a CDSS: Context

Feature	Description	Examples
Clinical setting	Setting where the CDSS operates	inpatient, outpatient
Clinical task	Clinical task CDSS supports	screening, diagnosis, assessment, treatment
Outcomes optimized	Types of outcomes that are optimize	patient health outcomes, efficiency

from early expert systems research, in which the aim was to emulate human thinking in software. Medicine was considered a good domain for which this concept could be tested. The intent however in CDSSs is not to simulate an expert's decision making but to assist the clinician in her own decision making [4]. Figure 2.1 shows a general model of a KBS [20] consisting of three main parts: the knowledge-base, the reasoning or inference engine, and the communication environment to the user. Knowledge bases may include rules for treatment, probabilistic associations of symptoms to diseases, drug-drug interactions, or clinical workflows with decision steps. Furthermore, if the source of knowledge is from a reputable source such as clinical guidelines, knowledge-based systems can be used to promote evidence-based medical practice and ensure decisions made by the clinicians have a solid ground.

Other, more *data-driven* systems use a reasoning paradigm which employ methods from machine-learning to build an implicit model of the clinical data. The idea is for the system to learn from past experience and find patterns in the clinical data. Although in certain application (such as predicting pulmonary embolism [21]) they are more accurate than the clinicians, the disadvantage is that the reasoning behind them is not transparent, thus many clinicians might hesitate in adopting them.

In order to better analyse, evaluate and specify CDSSs, it is important to note the differences among these systems along other dimensions as well. There are many ways to categorize CDSSs; we have adopted an approach used by *Berlin et al.* [22] who identified a comprehensive list of descriptors or features that are grouped into five categories. Tables 2.1 through to 2.5 show the features along with descriptions that we deemed most important.

Kawamoto et al. performed a systemic review of CDSS publications that reported the performance of their systems and described the particular characteristics of the system [23]. The objective was to determine a correlation between successful CDSS and specific

Table 2.2: Taxonomy of a CDSS: Knowledge and data source

Feature	Description	Examples
Knowledge source	Source of the clinical knowledge used to generate recommendations	guidelines, experts' opinion
Patient data source	Source of the patient's data used to generate recommendations	paper chart, electronic health record (EHR), personal health record (PHR)
Data format	Format of data entered into the CDSS	free-text (requires natural language processing), standardized schema
Updatable knowledge	Mechanism for updating the clinical knowledge with new research	full replacement of CDSS, updating only knowledge base

Table 2.3: Taxonomy of a CDSS: Decision support

Feature	Description	Examples
Reasoning method	Method used by inference or reasoning engine to generate recommendation	rule-based, Bayesian network, neural network, case-based
Clinical urgency	Time criticality of recommendations suggested by CDSS	perform action within minutes, days, months

Table 2.4: Taxonomy of a CDSS: Information delivery

Feature	Description	Examples
Delivery format	Format the CDSS delivers its recommendation in	online screen, print out letter, email, xml message
Delivery mode	Whether the clinician requested the support (solicited, pull), or the recommendation came unsolicited (push)	pull or push
Delivery point	At what point in time and place the recommendation is provided	during patient consultation
Explanation	Whether an explanation is provided with the recommendation	explanation linked to clinical guideline
Type of advice	Whether the recommendations are actionable, or an assessment suggestion	therapy action, diagnosis class, severity class

Table 2.5: Taxonomy of a CDSS: Workflow

Feature	Description	Examples
User	End-users interfacing with the CDSS	specialist, primary care clinician, nurse, another health information system
Target decision maker	The person whose decisions or actions the CDSS is designed to influence	physician, GP
Workflow integration	The way the CDSS changes the current workflow.	integrated into existing EHR system, minimal change

features. They found successful CDSSs had the following three characteristics[a]:

- Decision support integrated into the workflow
- Decision support delivered at the time and place of decision making
- Actionable recommendations

This review does have limitations as CDSSs descriptions often omit some potentially explanatory features, such as system speed, use of intuitive user interfaces, and error rates in recommendations. Nevertheless, it provides a useful insight into what should be taken into account when designing a CDSS.

2.2 Reasoning methods

In this section we review the main reasoning methods used across CDSSs to provide their recommendations. We have grouped the reasoning methods into five categories:

1. *Workflow-driven*: Logical flows containing statements that reference and manipulate clinical data. These are usually executed in a serial manner with control structures that direct the flow of decision making through the procedure.
2. *Rule-based reasoning*: Medical knowledge is captured through a collection of IF-THEN expressions. Reasoning by *forward chaining* (the most common technique) links rules together until a conclusion is reached.
3. *Probabilistic reasoning (Bayesian networks)*: Bayesian networks are graphical representations that describes the causal relationships between diseases and symptoms with conditional probabilities.
4. *Pattern recognition and machine learning*: Machine learning and statistical techniques, such as regression by learning, used on existing, large datasets.
5. *Case based reasoning (CBR)*: Patients are treated or diagnosed by recalling past patient cases with a similar record.

Workflow driven and rule-based systems require clinical knowledge such as diagnostic rules, procedures, treatment to be explicitly stored by the CDSS in its knowledge base. The knowledge of pattern recognition and machine learning is implicit and learnt from samples of clinical data. In probabilistic reasoning the knowledge may

[a] *Kawamoto et al.* reported actually a fourth feature "the decision support system was computer-based" which we consider implicit

be acquired explicitly and the network modelled by an expert. Alternatively, the network may be learnt through analysis of large datasets. Similarly, the probabilities may be either learnt from a dataset or explicitly entered from the literature. The CBR knowledge base is formed of many previous cases with solutions, which are used to solve a new case by comparison of features (analogy).

2.2.1 Workflow driven

Traditionally, medical practice has been based largely on individual clinical experiences. In the last two decades however the trend has shifted, and for good reasons there is a movement towards high-quality evidence. Evidence-based medicine aims to apply the best evidence available to clinical decision by employing the scientific method. Only when a gap in higher-quality evidence exists, is the opinion of the expert used. This gives rise to a notion of evidence quality, which can be from systemic reviews of double-blind, placebo-controlled trials at the top end of the scale, to conventional wisdom of the experts down at the lower end. In 2001 the committee of the Institute of Medicine in an influential report on the future of health [24] suggested authoring and dissemination of clinical guidelines (CGs) as a principle way to make evidence-based medicine more useful and accessible to clinicians and patients. CGs contains a set of directions to assist healthcare practitioner with patient care decisions about appropriate diagnostic, therapeutic or procedural steps to follow in specific clinical circumstances. These instructions are derived from systematically and periodically reviewing the literature, and updated or modified to best capture the most up-to-date evidence-based medical knowledge. Thus, their purpose ultimately is to standardize and improve healthcare.

CGs are often represented as flowchart or algorithms the clinician can follow with decision branches based on clinical criteria; an example of a differential diagnosis between COPD and asthma is shown in Figure 2.2.

This representation closely resembles computer algorithms. The early CDSSs reviewed by *Miller R. A* in the paper "Medical diagnostic decision support systemspast, present, and future: a threaded bibliography and brief commentary" [25] used computer programs written in conventional procedural programming languages. By intermixing clinical knowledge, inference, and control structure this representation made it very difficult to maintain the clinical knowledge, because the author must be familiar with not only the clinical domain but also the syntax and control features of the programming language. Moreover, edits to update the clinical knowledge may inadvertently modify the embedded control, thus affecting the accuracy of the CDSS. Furthermore, changes to the clinical knowledge would require recompilation, and redistribution of the software to the medical institutions.

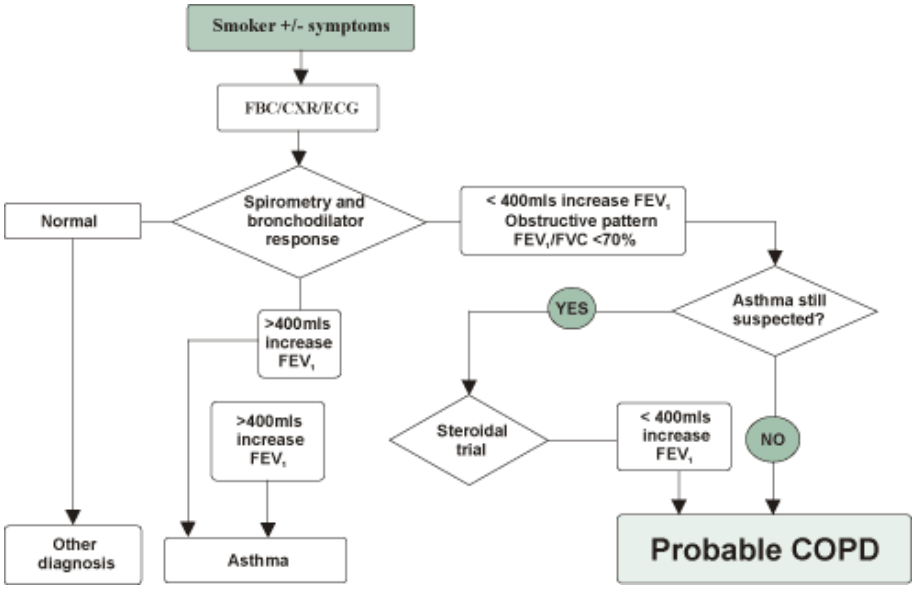

Figure 2.2: A flowchart showing a differential diagnosis of COPD and asthma from the New Zealand Best Practice Advocacy Centre.

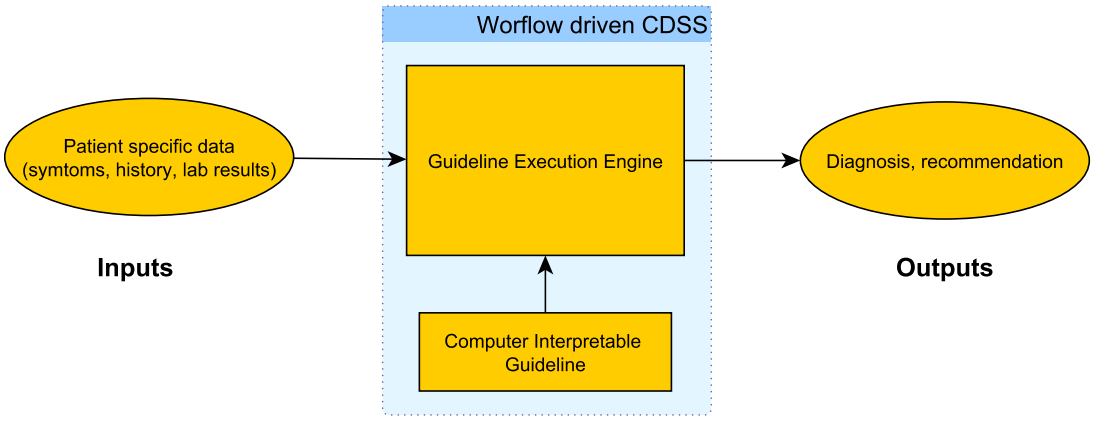

Figure 2.3: Generic model of a workflow-driven CDSS

The need to separate control structures and code from clinical knowledge, and the need to represent clinical guideline knowledge in a formal way that enables computer-based execution gave rise to one of the most common CDSS types, the workflow-driven computer-interpretable guideline (CIG).

Shown in Figure 2.3 is a generic representation of a system that uses the workflow-driven approach; it is in fact the same diagram as in Figure 2.2, except the clinical knowledge is encapsulated as a workflow-driven CIG and the reasoning-engine is a guideline execution engine (GEE). The CIG represents the CG as a workflow of component tasks that unfold over time. The tasks usually model medical actions, decisions and nested tasks. The CIG is usually authored with a composer GUI so that

the author does not need to be familiar with a computer programming language. The GEE is usually an interpreter which processes the CIG and combines it with patient data from an EHR or entry prompts. As the GEE steps through the CIG, the best path to follow will be recommended to the clinician but at the same time other possible paths can be offered. An alternative name to this CDSS design some authors prefer is also the task-network model [26].

Over the past 15 years we have seen numerous formats for CIG in the research community, each with their own motivations and features; also, several papers have reviewed these CIG specifications [27, 28]. For the more successful formats there has been development in various GEEs as well. The committee Health Level Seven International (HL7) [29] Clinical Guideline Special Interest Group (CGSIG) was established in 2001 to agree on a standard of a shareable representation of clinical guidelines, however it was soon apparent that agreeing on a fully comprehensive CIG format accepted by everyone was not achievable due to differences in opinion by the committee. What was possible, was standardizing components of CIG or CDSS models.

2.2.2 Rule-based reasoning

Another way to detach clinical knowledge from the internal programming language control code of the CDSS is through rule-based reasoning systems (also referred to as an expert system). Rules represent and manipulate knowledge in a declarative manner. Over the last several decades numerous representations and systems have spawned but essentially all of them are expressed as IF THEN statements containing two parts: the conditions and the actions. In the mathematical sense a rule is in the form $A \implies C$, where A is the set of conditions or the antecedent, and C is the set of actions or the consequent. Rules allow the declarative expression of first-order logic in an unambiguous, human readable form, at the same time retaining machine interpretability. Unlike the actions in workflow-driven CDSS, rules-based systems execute actions only in response to changes in the facts available to the rules-engine.

The rules of such a system typically reside in the “production memory” or “rules knowledge base” or simply “rule base” of a CDSS as illustrated in Figure 2.4. The rules are matched against “facts” (which represent patient data) stored in the “working memory” by the “inference engine”. If all conditions are matched, the rule is said to be “fired” and one or more actions may be performed which can assert (produce) new facts, retract them or modify the state of the “facts”.

In such a system facts are assertions that are known to be true, e.g. the input parameters of a specific patient. Using the rules in the system new facts can be deduced and finally possible diagnoses may be an output of the CDSS.

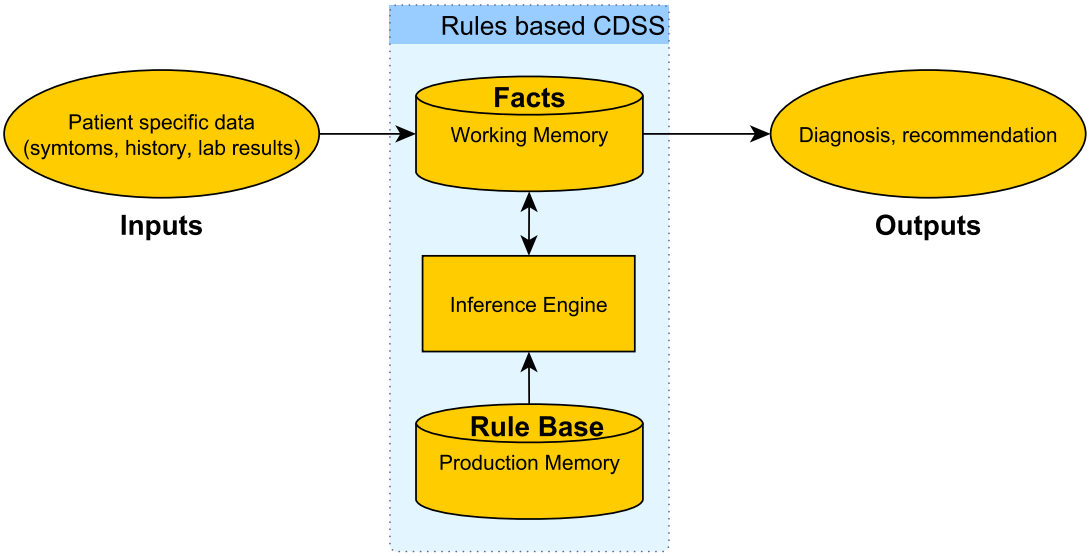

Figure 2.4: Generic model of a rule-based reasoning CDSS

There are two methods of processing rules for a rule-based system: Forward Chaining and Backward Chaining. In the Forward Chaining inference model, shown in Figure 2.5, facts are being asserted in the working memory, and rules begin to fire, thus updating the working memory with new facts. The process terminates when no new rules need to fire or a rule explicitly requires the process to terminate.

The Backward Chaining model, shown in Figure 2.6, is goal driven. The inference engine starts with a conclusion or goal and tries to satisfy it. If it can't, it searches for other conclusions or sub-goals that will help satisfy a part of the current goal. The process continues until either the initial conclusion is proven or there are no more sub-goals [30].

Originally rules engines for rules based systems used string comparison algorithms such as Boyer-Morre [31], Knuth-Morris-Prat [32], and Rabin-Karp [33] for rules matching. In 1974 Charles L. Forgy published the Rete algorithm [34] and current state-of-the-art engines all use slight variations of this. Rete made the rules to fact matching process significantly faster than its predecessors. When rules are added, Rete constructs a network of nodes that represent a pattern from the condition of the rules. The network resembles a tree with the leaves being the actions of the rules. If a path is traced from the root all the way to one of the leaves, a complete rule is described. As facts are added to the working memory, they are placed next to each node where the pattern matches the fact. Once a full path, from root to leaf is described, the rule fires and its actions are executed.

In the following subsection we will review some of the ways rules are represented and the popular rule-based systems used both commercially and by the research community.

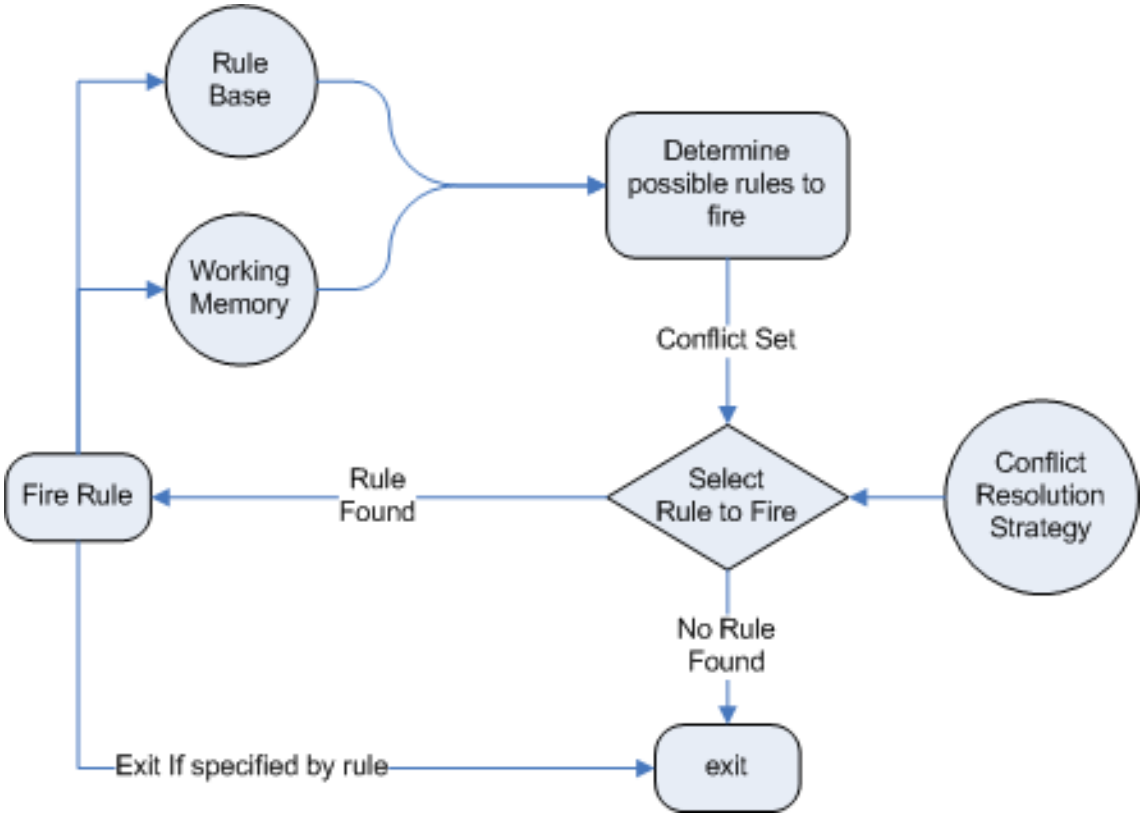

Figure 2.5: Forward chaining rules model [30]

With the exception of Arden Syntax [35], the tools (CLIPS [36], JESS [37], and Drools [38]) are general-purpose rules engines. Even though they are not specifically designed for handling medical data they are nevertheless important in CDSS applications as they have strong support from a wide range community making them very stable and accompanied by inference engines and authoring tools.

2.2.2.1 Arden Syntax

The Arden Syntax [35] is both a HL7 (1999) and American National Standards Institute (ANSI) (2002) standard for representing medical knowledge first introduced in 1989 and currently (2011) at version 2.7. The Arden Syntax encodes medical knowledge in medical logic modules (MLMs) which is a hybrid of a rules system and a procedural formulism described previously in section 2.2.1. Each MLM is invoked as a single-step "if-then" then executing serially as a sequence of instructions that include queries, calculations, logic statements, and EHR updates. Below is an example MLM from the paper "Fuzzy Arden Syntax: A fuzzy programming language for medicine" by *Vetterlein et al.* [39] based on the specification of nosocomial infections by the US-American Centers for Disease Control and Prevention [40]. The purpose of the MLM is determine whether a hospitalized patient has a symptomatic urinary tract infection (SUTI) maintenance:

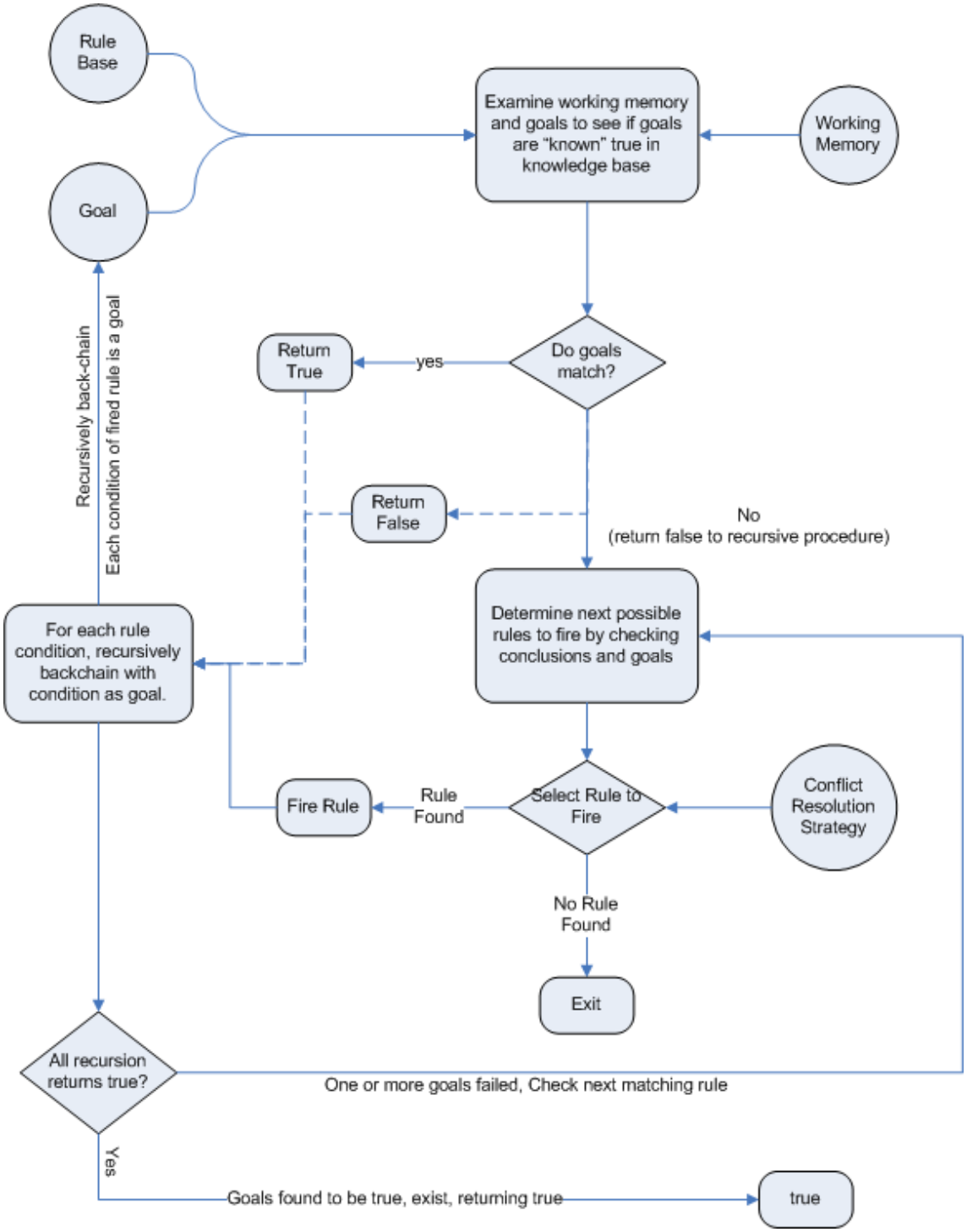

Figure 2.6: Backward chaining rules model [30]

```
mlmname: UTI_SUTI;;
   arden: version 2.5;;
   [...]
knowledge:
   [...]
   data:
      (Stay, Date) := argument;
      Temperature := read {temp (Stay, Date)};
      /* Body temperature. */
      if Temperature >= 38
         then Fever := true;
      endif;
      Urgency := read {urge_urinate (Stay, Date)};
      /* Urge to urinate? */
      Micturition := read {mict (Stay, Date)};
      /* Increased frequency of urination? */
      Dysuria := read {dys (Stay, Date)};
      /* Painful urination? */
      Suprapubic_tenderness := read {suprtend (Stay, Date)};
      /* Suprapubic tenderness? */
      Organ urine culture := read {org_urine_cult (Stay, Date)};
      /* Number of microorganisms of <= 2 species. */
      if Organ_urine_culture >= 1e5
      /* if number_of_species is     10^5/cm^3 */
         then Urine_culture := true;
      endif;
   ;;
   evoke:
   ;;
   logic:
      UTI SUTI := (Fever OR Urgency OR
                   Micturition OR Dysuria OR Suprapubic_tenderness)
                AND
                (Urine_culture);
      conclude true;
   ;;
   action:
      return UTI_SUTI;
   ;;
end
```

The result depends on real value and Boolean symptom parameters.

The aim of Arden Syntax was to encapsulate shareable clinical knowledge and was designed to support clinical decision making. The philosophy of an MLM was that

it contains sufficient logic to make a single medical decision. Although the original intention of the Arden Syntax is to make knowledge portable, in reality many MLMs developed for one hospital environment are not easily transferable to another. Most application which implement MLMs are for only compatible with individual vendors of hospital information systems. This probably in the research community is known as the curly braces problem and is common in many CIG formats.

Database schemas, clinical vocabulary and data access methods vary widely so any encoding of clinical knowledge must be adapted to the local institutions particularities. In Arden Syntax this is explicitly isolated in curly braces "..." in an MLM. Also to push out the output in the form of notifications and reminders, Arden Syntax relies on specific interaction with the local database through the curly braces. When an MLM is moved from one hospital system and adapted to another system, the query commands in the curly braces need to change.

2.2.2.2 CLIPS

C Language Integrated Production System (CLIPS) [36] is a general purpose expert system that provides a complete environment for both representing rules and a shell for performing inferences. It was created in 1985 by NASA and now is used widely in government, industry and academia. It is written in C and designed for portability, and efficiency. It can also be extendible and integrates within procedural code of many languages C, Java, Fortran, Python, etc. CLIPS is maintained as public domain software. Copies of CLIPS executables, documentation, and source code can be downloaded and used free of charge.

The CLIPS shell provides three basic elements:

1. Fact-list and instance-list stored in the global memory
2. Knowledge-base to store all the rules
3. Inference engine: control the execution of rules

Programs written with CLIPS consist of rules, facts and objects. The inference-engine decides which rules should be executed and when. In the event of multiple rules being simultaneously activated CLIPS uses a salience value to determine which rule is more important and should be executed first.

The example below is of a CLIPS rule for a CDSS that aids medical decision making for poising cases in the Philippines [41].

```
(defrule possible-organophosphate-mild
  (declare (salience 1000))
```

```
(or (symptom malaise yes)
    (symptom vomiting yes)
    (symptom nausea yes)
    (symptom diarrhea yes)
    (symptom sweating yes)
    (symptom abdominal-pain yes)
    (symptom salivation yes)
    (symptom miosis yes)) =>
(assert (poison-possible organophos-phate
 nil mild 0.0)))
```

The rule determines the necessary symptoms needed to suspect mild organophosphate poisoning. The CLIPS inference engine or shell processes rules such as the one above and reasons based on inputs from a patient's record what kind of poisoning the patient is most likely to have, calculating additionally confidence factors for each of the poison types with logic obtained from other rules. Additionally in order to provide a more user friendly interface for authoring tools, CDSS that use generic inference engine like CLIPS often create an additional GUI tool that produces CLIPS rule so that medical experts responsible for providing the rules do not need to be familiar with the CLIPS syntax. This is known as knowledge acquisition.

Another semi-automatic approach to knowledge-acquisition [42] is to map from workflow CIG to CLIPS rules. As rules based knowledge representation such as CLIPS are better supported for execution purposes than CIGs, and on the other hand CIGs are better suited to capture clinical knowledge, an option is to translate the workflows from CIGs such as GLIF [43], PROforma [44], and SAGE [45] into CLIPS rules that can be executed by the CLIPS shell.

CLIPS has also been extended into FuzzyCLIPS [46] by the National Research Council Canada with fuzzy concepts, uncertainty and reasoning and fully integrated with CLIPS facts and inference engine.

2.2.2.3 JESS

Java expert System Shell (JESS) [37] developed in 1995 is a rule engine and scripting environment entirely written in Java by Ernest Friedman-Hill at Sandia National Laboratories in Livermore, CA. It is based on CLIPS, initially intending to clone CLIPS for Java, but began to acquire a Java flavor of its own. The aim is to give Java applets and applications the ability to ”reason”. Jess rules are represented in both a CLIPS style format, and also a declarative XML rule language JessML which is designed to be easily transformed to other XML rule languages.

The following Jess rule below from the cancer Biomedical Informatics Grid [47] is

used for case-finding possible diabetes depending on the presence of three symptoms: Polydispia, Polyuria, and unexpected weight loss.

```
(data_element (name "Polydipsia") (avalue ?PolydipsiaValue))
 (test(eq ?PolydipsiaValue "true"))
(data_element (name "Polyuria") (avalue ?PolyuriaValue))
 (test(eq ?PolyuriaValue "true"))
(data_element (name "Unexplained Weight Loss") (avalue ?UnexplainedWeightLossValue))
 (test(eq ?UnexplainedWeightLossValue "true"))
         =>
        (store diabetes_suspicion "high")
```

Only when all 3 symptoms are present (have values true) should a patient be suspected of diabetes. JESS is available as a Java library for developers to integrate into their projects. The inference engine that processes the rules allows for direct interacting and reasoning with Java objects, the dynamic creation of new objects at runtime, and the ability to access internal variables from Java applications.

The same issues for knowledge-acquisition discussed for CLIPS apply for JESS as the JESS rules cannot be so easily interpreted by non technical medical staff. In the paper "Developing guideline-based decision support systems using protégé and jess" [48] *Chen et. al* describe a CDSS that automatically translate to JESS rules from workflow guideline representations that are more user friendly to clinicians.

2.2.2.4 Drools

Drools [38] is an open source effort started in 2001 by Bob McWhiter and in 2005 was absorbed by the larger JBoss organization, a division of Red Hat Inc. offering open source software. The Drools rule representation has a syntax easier to interpret than CLIPS and Jess, and is in particular targeted to Java users, however MVEL [49], Python [50], and Groovy [51] rule formats exist. Alternatively rules can be specified using a native XML format also. The rule engine essentially uses the Rete algorithm [34] however extended to support object oriented structures. The latest version of Drools is 5, and it is available under the Apache Software Foundation's open source license which is very liberal with the use of code.

The Drools rule syntax is much more user friendly than CLIPS/Jess. The general Drool rule structure is as a follows

```
rule <name>
<attribute> <value>
when
    LHS (condition)
then
    RHS (action)
```

```
end
```

Attributes impose further conditions on rule firing and the main mechanism is to control the order of rules firing. The following are the most common attributes:

- **no-loop** : When the Rule's consequence modifies a fact it may cause the Rule to activate again, causing recursion. Setting no-loop to true means the attempt to create the Activation will be ignored.
- **salience** : Each rule has a salience attribute that can be assigned an Integer number, defaults to zero, the Integer and can be negative or positive. Salience is a form of priority where rules with higher salience values are given higher priority when ordered in the Activation queue.
- **ruleflow-group** : Allows for workflows to control of the firing of rules. Only rules that are in the focus group are allowed to fire.

To demonstrate the syntax the following example is for assessing mild COPD severity according to the Australian and New Zealand COPD-X guidelines [52]

```
rule "COPD Mild Severity"
when
    $p : Patient( copd == true, bronchodilator == true,
                      fev1 > 60, fev1 < 80)
then
    $p.setCOPDSeverity(COPDSeverity.MILD);
end
```

The above rule defines a "mild COPD severity" of a patient confirmed diagnosis of COPD, and a forced expiratory volume in one second (FEV1) between 60 and 80 after the administration of bronchodilator medication.

Drools Flow An extra feature of Drools in comparison to other rule systems is the ability to use workflow representations as described in section 2.2.1. The workflow in Drools Flow describes the order in which a series of steps need to be executed, using a easy to interpret flowchart.

The idea is the process modeller can simply drag-and-drop blocks onto a canvas when constructing a workflow which include node types:

- Start and End
- Branching and synchronization nodes (split and join)
- Wait states

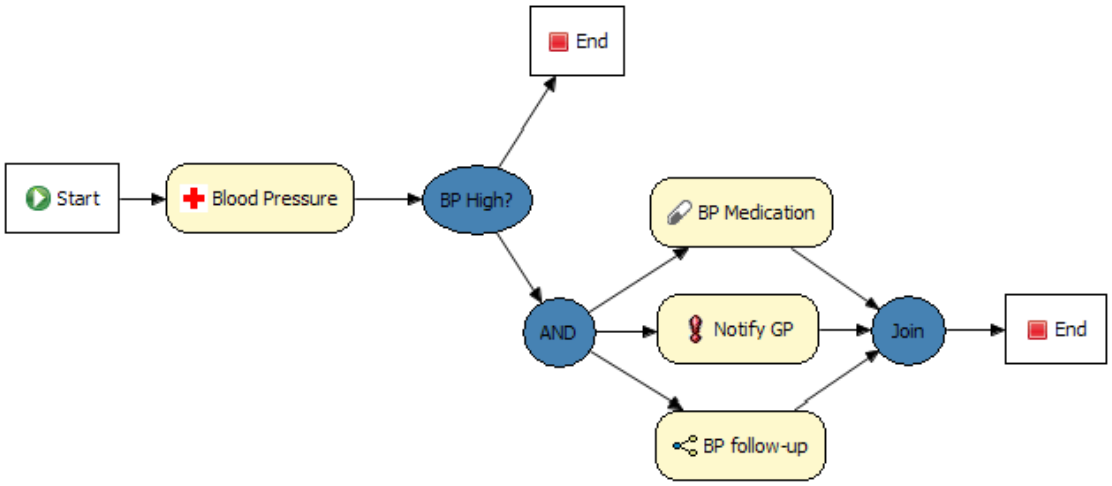

Figure 2.7: High blood pressure protocol [30].

- Timers
- Events
- Composition

These blocks may be domain specific work items defined using a domain specific language (DSL) (described next subsection). The process modeller therefore needs to have no knowledge of the underlying low level code that the work items may invoke. The following example Figure 2.7 [30] showing a model of a typical nursing task. The work items in Drools flow have been defined specifically for the clinical domain.

Normally, workflow-driven and rules-based systems are considered as two different paradigms for describing the reasoning engine for a CDSS. Drools however, has the flexibility for the combination of rules and workflows in a hybrid approach, and can offer the following powerful combinations:

- Rules can define which clinical processes (workflows) to invoke.
- High-level, domain specific rules can specify decisions in a clinical process
- Rules can specify exceptions to which a clinical process should deviate to
- Rules can be used to dynamically alter the behaviour of a clinical process

Domain Specific Languages One other key advantage of Drools over other rule representations is its support for domain specific language (DSL). These are high-level abstract programming languages that are used to increase productivity when programming for a very specific problem domain (e.g. Healthcare, COPD). In Drools a DSL allows

domain experts (such as medical specialists) interpret and create rules easily without understanding any programming code. The DSL definitions essentially transform the DSL "sentences" to the Drools rules language (DRL) constructs, which can be executed by the inference engine.

A DSL serves as a layer of separation between rule authoring and the technical intricacies resulting from the modelling of domain objects. The DSL rules can be easily read, interpreted and validated by the non technical expert as the technical details are obfuscated only revealing the logic. The following example is the same as the previous drools rule example expect written with a DSL.

```
rule "COPD Mild Severity"
when
    There is a person
        with disease COPD
        with fev1 between 60 and 80
        after administration of bronchodilators
then
    set the severity to mild
end
```

The rule above would be translated using a DSL template to a DRL rule interpretable by the Drools rules engine. In fact it is a clinical DSL representation of the previous example.

Although Drools was originally intended for business rules and process, it is general enough to be applied in the medical domain. Furthermore, the ability to extend the specification of rules and flows using DSL makes it very attractive for application in CDSSs.

2.2.3 Probabilistic Reasoning

In medical domains the CDSS's knowledge can at best provide a degree of belief in the relevant sentences. The main tool for dealing with degrees of belief is probability theory, which assigns to each sentence a numerical degree of belief between 0 and 1. Probability provides a way of summarizing the uncertainty that comes from the application of the knowledge being modelled on the data instance of the patient. Probability theory makes the same ontological commitment as logic (i.e. facts either do or do not hold in the world). The degree of truth, as opposed to degree of belief, is the subject of fuzzy logic. The beliefs and the relationships between parameters could be derived from:

- statistical data (e.g., the prevalence of COPD in current smokers is 12.1%)
- general rules

- some combination of evidence sources

Assigning a probability of 0 to a given statement or event corresponds to an unequivocal belief that the statement is false. Assigning a probability of 1 corresponds to an unequivocal belief that the statement is true. Probabilities between 0 and 1 correspond to intermediate degrees of belief in the truth of the statement. The statement itself is in fact either true or false. A degree of belief is different from a degree of truth. A probability of 0.8 does not mean "80% true", but rather an 80% degree of belief that something is true. In logic, a statement such as "The patient has COPD" is either true or false. In probability theory, a sentence such as "The probability that the patient has COPD is 0.8" is about the CDSS's belief. These beliefs depend on the percepts that the CDSS has received to date. These percepts constitute the evidence on which probability assertions are based. For example: a doctor studies a patient's general record; before looking at results of comprehensive specific exams, the doctor might assign a probability of 1 in 50 that the patient having a disease; after looking at results of a comprehensive specific exams (e.g. biopsy), an appropriate probability for the same proposition would be 0 or 1.

An assignment of probability to a proposition is analogous to saying whether a given logical sentence is entailed by the knowledge base, rather than whether or not it is true. All propositions have to indicate the evidence with respect to which probability is being calculated. When a CDSS receives new perceptions/evidence, its probability assessments are updated. Before the evidence is obtained, we refer to the prior or unconditional probability. After obtaining the evidence, we refer to the posterior or conditional probability.

The unconditional or prior probability associated with a proposition a is the degree of belief accorded to it in the absence of any other information. It is written as $P(a)$ for example:

$$P(\text{COPD} = \text{true}) = 0.1 \text{ or } P(\text{COPD}) = 0.1 \tag{2.1}$$

The above prior probability does include any conditional assertions, it could for example represent that the prevalence of COPD in the population is 10%. To model the probabilities of all the possible values of a discrete random variable, expressions such as $P(\text{smoker})$ are used to denoting a vector of values for the probabilities of each individual state of the patient's smoking. In this example the smoking status is domain is <non-smoker, ex-smoker, smoker>:

$$P(\text{smoker}) =< 0.5, 0.2, 0.3 > \tag{2.2}$$

The above equation may be used to express the distribution of smoking habits in a

population.

2.2.3.1 Bayes' rule

In probability theory and statistics, Bayes' rule [53] describes the probability of an event, based on conditions that might be related to the event. One interpretation of the rule is that it expresses how a subjective degree of belief should rationally change to account for the evidence. The Bayes rule can be expressed as:

$$P(A \mid B) = \frac{P(B \mid A)\,P(A)}{P(B)} \tag{2.3}$$

where A and B are the events and $P(B) \neq 0$, $P(A)$ and $P(B)$ are the probabilities of observing A and B without regard to each other. $P(A \mid B)$ is the conditional probability of observing event A given that B is true, and $P(B \mid A)$ is the conditional probability of observing event B given that A is true.

Thus utilising Bayes' rule, various probabilistic inferences can be calculated when new evidence is presented in order to update the degree of belief. For knowing age, smoking status, presence of symptoms changes the degree of belief that the obstructive lung disease is Asthma as opposed to COPD.

2.2.3.2 Bayesian networks

Bayesian networks [54] model a set of related random variables by representing their conditional dependencies via a directed acyclic graph (DAG). For example, the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases. They use the following syntax:

- a set of nodes, one per variable
- a directed links expressing conditional dependencies. Thus nodes that are not connected are conditionally independent of each other.
- a probability function $P(X_i \mid \text{Parents}(Xi))$ that takes as inputs the values of the nodes parent variables and as output gives the probability of the variables represented by the node

Through the use of efficient algorithms Bayesian networks may be used for either inference or learning the parameters (conditional probabilities) or structure (DAG) of the variables being modelled. This approach to diagnosing COPD based using Bayesian networks has been explored by *Himes et al.* [55].

2.2.3.3 Other representations of uncertainty

The certainty factor (CF) model is a method for managing uncertainty, and is primarily suited for rule-based systems. This model was introduced with the MYCIN [56] rules-based CDSS designed to diagnose and recommend treatment for meningitis and certain blood infections. The CF model came about to address some of the faulty assumptions in the Naive Bayes model that information science researchers were using in the medical domain. Heckerman and Shortliffe [57] however showed that the full Bayesian network overcomes many of the limitations of the CF model, and is grounded firmly in probabilistic theory.

Fuzzy logic [58] is another popular approach which makes computations on the degrees of truth rather than the probability of an event occurring or belief in a proposition. Its goal is different from probabilistic methods, as it is focused on dealing with partial truths and partial set memberships as in the case with fuzzy set theory which probability theory cannot capture.

2.2.4 Pattern recognition and machine learning

A very active area of research is the application of machine learning techniques to medicine. The knowledge of the CDSS is derived through inductive inference based on observing clinical data examples, in order to learn the statistical phenomena or pattern in the data so that prediction can be made about future data. The main application area of these techniques is in non invasive computer aided diagnosis (CAD) systems that classify clinical data obtained from medical imaging equipment (usually computerised tomography (CT), magnetic resonance imaging (MRI), or ultrasound). There are 4 main steps involved:

1. image preprocessing: Artefacts and signal noise is removed from the image, the image is normalized and contrast is often enhanced.

2. image segmentation: The image is segmented into structures corresponding to the human anatomy often aided by generic anatomic atlases. The organ of interest is identified and may be further partitioned into sub-segments.

3. feature extraction: The goal is to reduce the dimensionality of the image segments into a smaller feature vector that is easier to work with, and encodes less redundant information often through statistical techniques such as principal component analysis (PCA) in order to make the next step faster.

4. training, and classification: Prior to using the system on new images, the system's classifier is trained with a dataset of feature vectors that have already been labelled

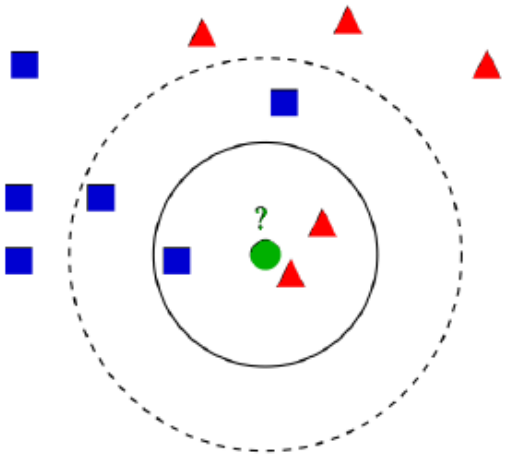

Figure 2.8: A kNN example.

with the correct classes (e.g. cancer tissue, normal tissue). When an unknown feature vector is presented the trained classifier must infer the true label using the knowledge gained from the training. Common classifiers applied in CAD include:

- k-nearest neighbour (kNN)
- Bayesian classifier
- multilayer artificial neural network (ANN)
- support vector machine (SVM)
- decision trees and random forests

2.2.4.1 k-nearest Neighbour

The kNN is amongst the simplest pattern recognition algorithms. It is an algorithm for classifying objects based on the closest training examples in the feature space. An object is classified by a majority vote of its closest k neighbours with the object being classified assigned to the class most common amongst its closest k neighbours. In a 2 class problem k is usually odd to avoid ties. When the feature space is continuous the Euclidean distance between the feature vector is used to find the closest neighbours. Figure 2.8 is an example of the kNN algorithm in simplified form with 2 classes and only 2 features. The green object which is being classified would be in labelled as a triangle class if k = 3 and a square if k = 5.

2.2.4.2 Artificial neural networks

The neural network classifier [59] applies biological concepts to recognize patterns. The outcome of this effort is the invention of artificial neural networks. Feed forward artificial

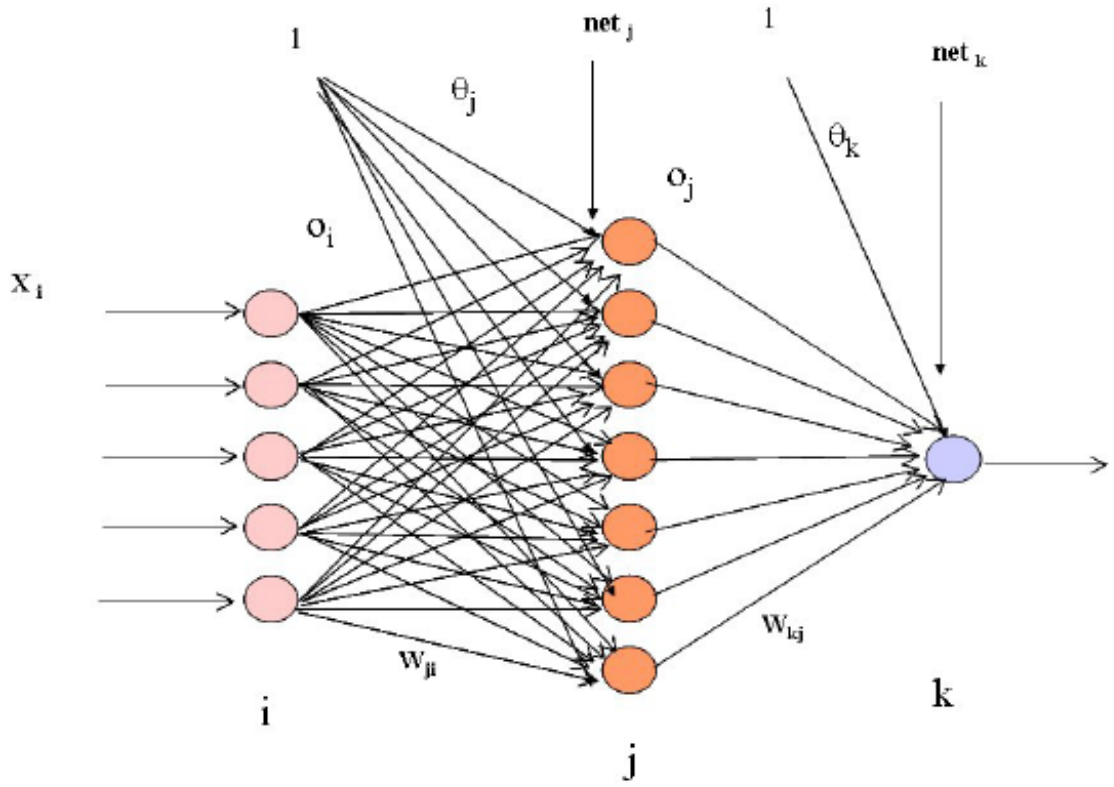

Figure 2.9: multilayer artificial neural network

neural networks ANN are well known for their pattern recognition ability hence the motivation for applying them to this problem. Figure 2.9 shows a diagram of a feed forward ANN model. It is composed of sets of neurons at 3 different layers: an input layer, a hidden layer and an output layer.

Applying the input values through the neural network is known as forward propagation and it starts at input layer which simply takes the value of the input features.

$$O_i = x_i \tag{2.4}$$

where O_i is the output of the input layer neuron i and x_i is the input feature value i. At each hidden layer neuron all the inputs are aggregated by a weighted sum:

$$\text{net}_j = \sum_j W_{ji} O_i + \theta_j \tag{2.5}$$

where W_{ji} is the adjustable weight between input neuron i and hidden layer neuron j, and θ_j is an adjustable bias. The output of the hidden layer neuron becomes:

$$O_j = \sigma(\text{net}_j) \tag{2.6}$$

where $\sigma(t)$ is a sigmoid function such that $\sigma(t) \to 1$ as $t \to \infty$, and $\sigma(t) \to 0$ as $t \to -\infty$. Usually we use:

$$\sigma(t) = \frac{1}{1+e^{-t}} \tag{2.7}$$

Similarly the output of the output layer neurons can be expressed as:

$$\text{net}_k = \sum_k W_{kj} O_j + \theta_k \tag{2.8}$$

where W_{kj} is the adjustable weight between hidden neuron j and output neuron k, and θ_k is an adjustable bias.

During training the weights and bias are tuned so that the input features produce an output in the ANN that corresponds to a desired target output T_k. The method used for training is known as back propagation which computes an error between T_k and O_k then adjusts first the weights W_{kj} then the weights W_{ji} so that this error is reduced. The weights are adjusted in the direction that reduces the error using a gradient descent approach.

Although neural networks have been around since the 1980s, advancement in cloud and parallel computer architectures has revitalised interest in them and have been re-branded as *deep learning*. Nevertheless many state-of-the-art algorithms in pattern-recognition for computer-vision, natural-language processing, speech recognition are of this type. *Larder et al.* have studied the state of the art of ANN in decision support [60].

2.2.4.3 Support vector machines

Support vector machines (SVM) [61] belongs to the classifier group of large margin classifiers. The main idea is to find an optimal hyperplane between the linearly separable classes, and extend this to non-linearly separable classes my mapping into the new space via kernel functions. The SVM are binary classifiers and the optimisation problem can be solved by optimization techniques by constructing a dual problem through the use of Lagrange multipliers. The formula to classify given test vector x is:

$$f(x) = \text{sign}(\sum_{i=1}^{k} \alpha_i y_i K(x, x_i) + b) \tag{2.9}$$

where α_i is the non-zero Lagrange multiplier for each support vector x_i (from the set of training set examples), k is the number of support vectors $y_i \in -1, +1$ is the class label of the support vector, b is the bias value for hyperplane, and $K(x, x_i)$ the kernel function. A polynomial kernel is often used and is defined as:

$$K(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^d \tag{2.10}$$

where $\langle ., . \rangle$ is the inner product, d is the polynomial degree, and when $d = 1$, becomes a linear SVM. Training can be formulated as finding the α_i values by maximizing the cost function:

$$Q(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j K(x_i, x_j) \tag{2.11}$$

subject to $\alpha_i \geq 0$ and $\sum_{i=1}^{n} \alpha_i y_i = 0$ and n being number of training examples. The x_i are input (training) vectors. This problem can be solved using quadratic programming (QP) methods, and the non-zero α_i will correspond to the support vectors used for classification in equation 2.9. Once α_i are computed, the bias b can be obtained as:

$$b = -\frac{\max_{i:y_i=-1} w^T x_i + min_{i:y=1} w^T x_i}{2} \tag{2.12}$$

with $w = \sum_{i=1}^{n} \alpha_i y_i x_i$

2.2.5 Decision trees

The goal of decision tree learning is to create a model that predicts the value or class of a target variable by learning the decision rules from the features of the training examples. Decision-tree learning algorithms "grow" the decision tree through a greedy iterative process by selecting for each node in the tree, the feature and threshold that will yield the largest information gain in the target variable.

The Random Forest is a learning method [62] grows many decision-trees by training with a new sample of training examples from the original set (usually with replacement). When a trained Random Forest is used to classify a new input sample, the predicted class is done by a weighted vote by the decision-trees in the forest, the weight is based on the their class probability estimates. The Random Forests learning method is used to improve the predictive accuracy and control over-fitting.

2.2.6 Case based reasoning

Case based reasoning (CBR) is both a paradigm for computer-based problem solvers and a model of human cognition. The central idea is that the problem solver reuses the solution from some past case to solve a current problem.

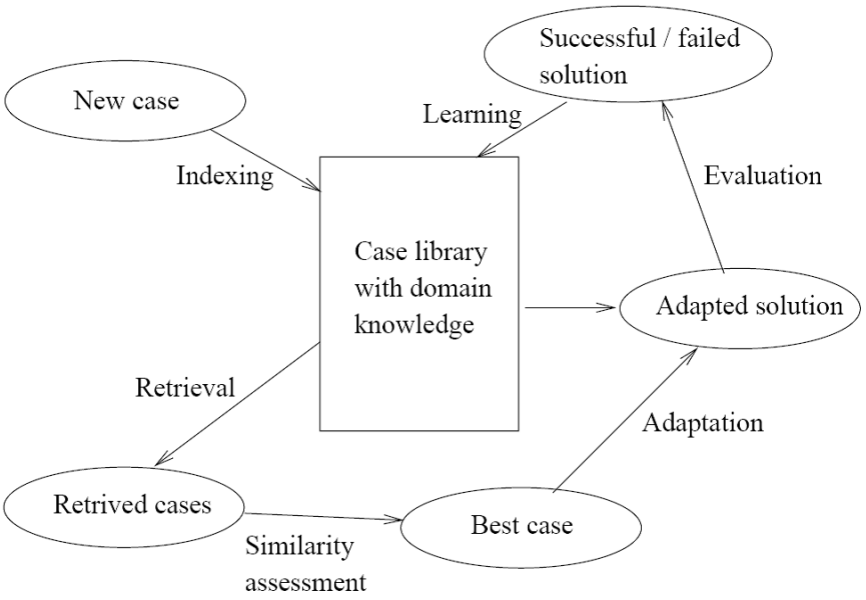

Figure 2.10: The case based reasoning paradigm

2.2.6.1 CBR as a computer program paradigm

As a paradigm for problem solvers, one of the advantages of CBR systems is that they improve their performance, becoming more efficient, by recalling old solutions given to similar problems and adapting them to fit the new problems. In this way they do not have to solve new problems from scratch. The memorization of past problems/episodes is integrated with the problem-solving process, which thus requires the access to past experience to improve the systems performance. Additionally, case-based reasoners become more competent during their functioning over time, so that they can derive better solutions when faced with equally or less familiar situations because they do not repeat the same mistakes (learning process). The basic steps in CBR are (see Figure 2.10):

1. Introducing a new problem (or situation) into the system.

2. Retrieving a past case (a problem and solution), whose problem part resembles the current problem. Past cases reside in case memory. The case memory is a library that contains rich descriptions of prior cases stored as units. Retrieving a past case involves determining what features of a problem should be considered when looking for similar cases and how to measure degrees of similarity. These are referred to as the indexing problem and the similarity assessment problem.

3. Adapting the past solution to the current situation. Although the past case is similar to the current one, it may not be identical. If not, the past solution may have to be adjusted slightly to account for differences between the two problems. This step is called case adaptation.

4. Applying the adapted solution and evaluating the results.

5. Updating the case memory (learning). If the adapted solution works, a new case (composed of the problem just solved and the solution used) can be formed (direct learning). If the solution at first fails, but can be repaired so the failure is avoided, the new case is composed of the problem just solved and the repaired solution. This new case is stored in case memory so that the new solution will be available for retrieval during future problem solving. In this way, the system becomes more competent as it gains experience. Updating case memory can include deleting cases (forgetting) too. This step is also part of the indexing problem.

Not all case-based problem solvers use all of the steps. In some, there is no adaptation step; the retrieved solution is already known to be good enough without adaptation. In others, there is no memory update step; the case memory is mature and provides adequate coverage for problems in the domain.

2.3 State of the art of CDSS in COPD

In order to gain an understanding of the state of the art in CDSS applications for the diagnosis and treatment of COPD we performed a literature search to find relevant publications. In this section we review analyse these systems in terms of their medical application area, and the techniques used to apply the decision support.

Using the search terms “COPD CDSS”, “COPD DSS” and “COPD decision support” we searched the PubMed publication resource database provide by the National Center for Biotechnology Information (NCBI). For the IEEE Xplore and ACM digital library we only used the search term “COPD” as these resources were already targeted at Computer Science and Engineering publications. Table 2.6 summarizes the resulting hits generated by the search.

Table 2.6: Literature search of CDSSs for COPD

Resource	**Search Terms**	**Hits**
PubMed	COPD CDSS	0
	COPD DSS	5
	COPD decision support	219
IEEEXplore	COPD	104
ACM Digital Library	COPD	203

Table 2.7: Relevant publications for clinical decision support applications in COPD

#	Publication	Clinical task	Methods
1	Er et al. [63]	diagnosis	neural network
2	Himes et al. [55]	case finding	Bayesian network
3, 4	Hosseini et al. [64, 65]	case finding assessment	Bayesian classifier statistical analysis image processing pattern recognition
5	Jafari et al. [66]	diagnosis	ANN
6	Kuilboer et al. [67]	management	rules based
7	Liang et al. [68]	diagnosis	image processing ANN
8	Mohktar et al. [69]	management	rules based
9	Rosso et al. [70]	management	workflow (CIG)
10	Song et al. [71]	management	rules based
11	Sahin et al. [72]	diagnosis	SVM
12	Uncu et al. [73]	diagnosis	rules based (fuzzy)

Additionally adhoc searches, meta-analysis, and review articles were used to obtained additional prospective papers. The resulting hits were filtered by title, abstract, and finally by content with any irrelevant publications excluded. The criteria was any publication that describes a computer system that is designed to be a direct aid to clinical decision-making in which the characteristics of an individual patient are processed by the system in order to provide a patient-specific assessment(s) or recommendation(s) that is useful to a clinician treating or diagnosing COPD.

After the initial filtering and selection, finally 12 papers published between 2001 - 2011 were identified as being relevant to the state of the art of CDSS in COPD cases. They are listed in Table 2.7.

2.3.1 Analysis

The methods presented in the 12 publications are spread across the 4 application areas of decision support in the treatment and diagnosis of COPD, these areas are:

- Case finding: Is a type of disease screening where individuals are identified in a population who are considered to be at risk of having the disease. These individuals may not exhibit strong manifestations of symptoms but have the presence of high risk factors associated with the disease. Case finding can lead to early detection of COPD once a diagnosis is confirmed, and can drastically improve the quality of life if a disease management plan is executed.

- Diagnosis: The criteria based that may be based on symptoms, patient characteristics, medical history and measurable parameters from tests or instruments used to confirm the absence or presence of the disease. When symptoms that are not unique to COPD resemble other diseases the clinician must perform a differential diagnosis to consider all candidate diseases.

- Assessment: Once the disease diagnosis has been confirmed a clinician must further characterize the disease with additional information that will capture its severity, progression, extra systemic effects, prognostic markers, etc. Assessment parameters also give the clinician an idea on how rapidly the disease is progressing and if the disease management plan is effective (outcome measures).

- Management: The selection of treatment, therapies, and creation of care plans for the specific individuals COPD case. This may include pharmacological therapies, therapies for non-respiratory manifestations, physical exercise therapies, and therapies for exacerbation stabilizing. There is no full cure for COPD, but the quality of life can drastically improve and disease progress can be slowed down with the application of an effective management plan.

2.3.1.1 Case finding

Himes et al. [55], and *Hosseini et al.* [64, 65] present two different case finding applications with two different methods for the discovery of COPD cases, *Himes et al.* uses Bayesian networks and the other two from *Hosseini et al.* uses image processing then classification (pattern recognition). *Himes et al.* attempt at finding future COPD cases in Asthma patients. Asthma is another similar complex obstructive respiratory disease, characterized by airway hyperresponsivenes but unlike COPD has reversible airflow limitation that often develops early in life and can completely disappear. According to *Pauwels et al.* [74] asthmatics have an increased risk of developing COPD as well.

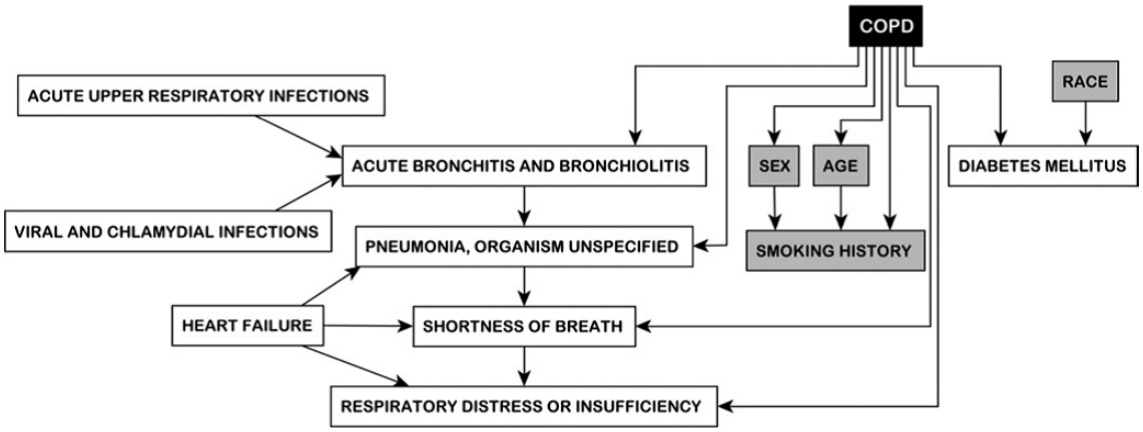

Figure 2.11: Predictive network of COPD [55]

The authors use a predictive modelling approach with data found in EHR records of 10,341 asthma patients (containing sex, age, smoking history, race, symptoms and co-morbidities) to create the Bayesian Network shown in Figure 2.11. The authors found that age was the most significant predictive factor producing an area under the receiver operating characteristic curve (AUC) value of 0.81, and performance improved with the addition of other factors only slightly. *Himes et al.* demonstrate how Bayesian Networks may be used in a CDSS for early prediction of chronic diseases.

Hosseini et al. [64, 65] present similar methods for detection of COPD, with only a limited validation on 24 subjects (half healthy, half COPD). In both cases inspiration and expiration sets of CT images are used.

The algorithm is outlined in Figure 2.12, a 5 step procedure with the first 4 steps consisting of image processing in order to separate the right and left lung and find a parenchyma variation parameter which characterize COPD airflow limitation. A Bayesian classifier is used as the pattern classifier that is applied to the resulting parameters after the image processing steps. The classifier separates a subjects into healthy or COPD. Although the authors report acceptable results, CT images come at a high cost and expose the subject to a high degree of radiation. Thus it would be very difficult to implement a screening protocol that uses such a technology. Furthermore, without a comparison to the relatively simple and inexpensive spirometry, it is difficult see the added value for a CDSS that uses CT for case finding. We find the parenchyma variation, and volume variation would be more appropriate for the further characterization of the disease once a COPD diagnosis is established.

2.3.1.2 Diagnosis

Computer aided diagnosis (CAD) has been a popular research line particular in the interpretation of medical images which combine areas of image processing and machine

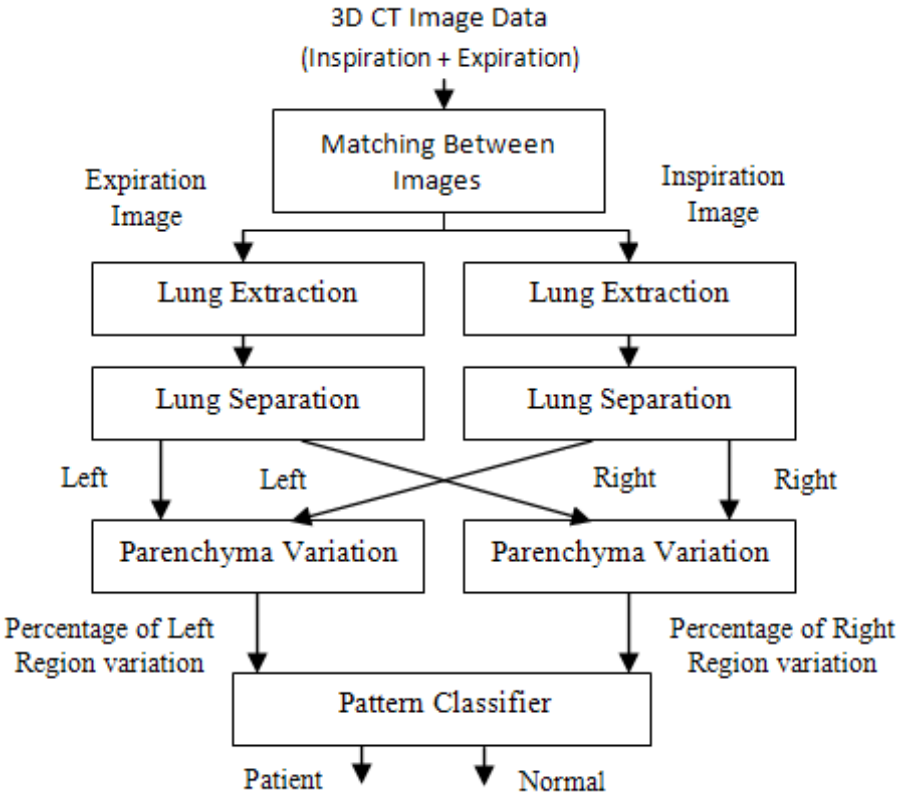

Figure 2.12: COPD identification algorithm [65]

learning. This is particularly true for the publications in Table 2.7, as 5 from 12 of them attempt to provide support in diagnosing COPD.

Popular amongst the diagnosis techniques is the use of pattern recognition and machine learning to train an ANN to learn the patterns of features which is consistent with a COPD diagnosis. *Er et al.* [63] uses such an approach with 38 ANN features mostly consisting of symptoms and blood test parameters: complaint of cough, body temperature, ache on chest, weakness, dyspnea on exertion, rattle in chest, pressure on chest, sputum, sound on respiratory tract, habit of cigarette, leucocyte (WBC), erythrocyte (RBC), trombosit (PLT), hematocrit (HCT), hemoglobin (HGB), albumin2, alkalen phosphatase 2 L, alanin aminotransferase (ALT), amylase, aspartat aminotransferase (AST), bilirubin (total+ direct), CK/creatine kinase total, CK-MB, iron (SERUM), gamma-glutamil transferase (GGT), glukoz, HDL cholesterol, calcium (CA), blood urea nitrogen (BUN), chlorine (CL), cholesterol, creatinin, lactic dehydrogenase (LDH), potassium (K), sodium (NA), total protein, triglesid, and uric acid. Although they claim high accuracy some key lung function features such as FEV1 and forced vital capacity (FVC) which are considered gold standards for COPD diagnosis in the medical literature are missing.

Liang et al. [68] also use ANNs except features are derived from CT images of lungs. A CT image is processed by undergoing image preprocessing steps of contrast

enhancement, vessel exclusion, and segmentation. Common texture features from each region such as entropy, variance, sum entropy, difference variance, and many others are used as inputs into the ANN.

It is important to observe that, although the authors claim to achieve almost perfect sensitivity and specificity the same criticism of this technique would apply here as it did to *Hosseini et. al*. That is, that the exposure to radiation and the high cost would deter its use as a general diagnosis method.

Spirometry is an essential tool for the functional diagnosis of COPD. *Jafari et. al* [66] and *Sahin et. al* [72] both provide decision support techniques using pattern recognition to provide differential diagnosis between restrictive diseases and obstructive diseases such as COPD based on spirometry curves which are the output signal of the spirometry device. From these curves, and patient demographic data such as weight, height, sex and race, lung function parameters such as FEV1, FVC, and FEV1 to FVC ratio can be computed. *Jafari et al.* used a parametric model of curve fitting, and used lung function parameters to train an ANN in classifying the patient in the 4 categories of respiratory disease (i) restrictive (ii) obstructive (iii) mixed (iv) normal. Similarly *Sahin et al.* did the same except by using SVM instead of ANNs and only on lung function parameters.

The last COPD diagnosis method presented in *Uncu et al.* [73] is a rules based model that supports fuzziness and operates on spirometric parameters. Essentially diagnosis of airflow obstruction can be made if the postbronchodilator FEV1 to FVC ratio < 0.7 and FEV1 < 0.8 predicted found in the UK COPD guidelines (NICE) [75]. The authors however represents these parameters with fuzzy values such as (FEV1 low, medium, high, very high). The advantage of representing the parameters using fuzzy sets is, it smoothens the sharp cut off criteria when the rules are applied (such as the 0.7 and 0.8 in the diagnosis rule) and it is also consistent with the way medical knowledge is expressed (often with uncertainty). The other advantage is rather than having a fixed Boolean results of COPD present or absent it allows as a COPD risk.

2.3.1.3 Assessment

As mentioned previously techniques that evaluate CT scans are better suited for COPD assessment purposes, than case finding and diagnosis. The methods in *Hosseini et al.* and *Liang et al.* could be used to complement the characterization of COPD by the parameters derived through the analysis of the CT scan.

2.3.1.4 Management

Kuilboer et al. [67] has a good example of a CDSS in use, used for supporting primary care clinicians in treating COPD and Asthma patients. The decision-support system

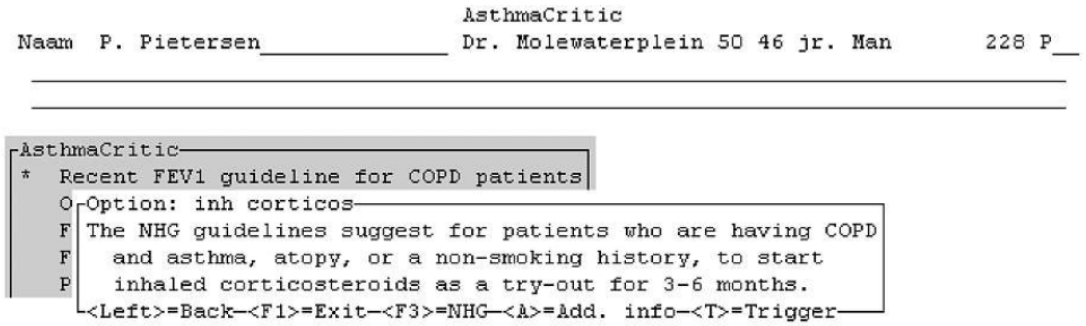

Figure 2.13: Feedback from AsthmaCritic [67]

Table 2.8: Parameters used in published Telehealth systems

Parameter	*Mohktar et al.* [69]	*Song et. al.* [71]	*Rosso et al.* [70]
Heart rate		X	X
Oxygen saturation	X	X	X
Respiration rate		X	
Temperature	X		
Lung function	X		
Blood pressure			X

AsthmaCritic provides the general practitioner with patient-specific feedback based on data solely obtained from the EHR. The model of use is: the clinician makes her decisions, the data of this decision is stored in the EHR and AsthmaCritic subsequently critiques these decisions. Figure 2.13 shows the feedback from the CDSS that is aware of the management guidelines for COPD.

The reasoning method is not disclosed by the authors, however be believe it is most likely it would be a rules based system as it responds only when additional facts are entered into the EHR.

Finally, *Mohktar et al.* [69], *Rosso et al.* [70] and *Song et. al.* [71] all presented telehealth systems that are designed to assist in the administration of a management plan to COPD patients at home. The main idea is to attach sensor devices to the patient that monitor the patient's vital parameters. Table 2.8 is a summary of the parameters in the publications.

Although the systems use similar parameters they address slightly different aspects to chronic disease management. *Mohktar et al.* describes a referral recommendation system that uses the patient parameters to decide whether a home patient should be treated at home or a carer be alerted to a worsening situation. The rules for referral are based on clinical guidelines. The decision logic is shown as a decision tree in Figure 2.14 where C is the number of referral criteria satisfied, F refers to a decrease in lung function, S a

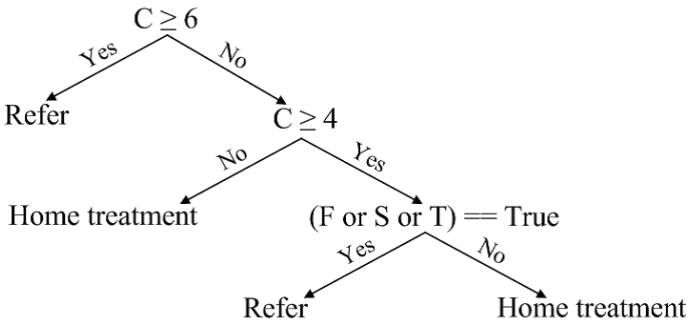

Figure 2.14: Decision criteria for patient referral [69]

decrease in oxygen saturation, and T an increase in temperature. *Song et. al* implemented rules on oxygen saturation, blood pressure and heart rate parameters to implement a exercise training regime in order to rehabilitate COPD patients. The knowledge is encoded in Drools described in section 2.2.2. *Rosso et al.* describes a European FP7 project CHRONIUS [70], a management platform for the telemonitoring of COPD and chronic kidney disease (CKD) patients via wearable sensor infrastructure (a t-shirt with sensors). There are two levels of decision support. The first is for immediate alerts to clinicians when certain vital parameters are abnormal done in realtime, and the second is offline and provides a deeper evaluation including history and laboratory data to propose possible actions. The method behind the reasoning engine is not disclosed.

2.4 Infrastructural considerations

A CDSS rarely exists in a vacuum on its own. A CDSS needs to integrate well with existing health system infrastructure and be able to interface with these systems through standard interfaces. In this section we will review some of the common standards and infrastructural features that should be considered when implementing a CDSS. In summary we will consider the following:

- Knowledge acquisition tools
- Electronic health records (EHR)
- CDSS to EHR interfaces (HL7 Messages, Continuity of Care Document (CCD), Virtual Medical Record (VMR))
- Medical vocabularies: Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT), Logical Observation Identifiers Names and Codes (LOINC), International Classification of Diseases (ICD)

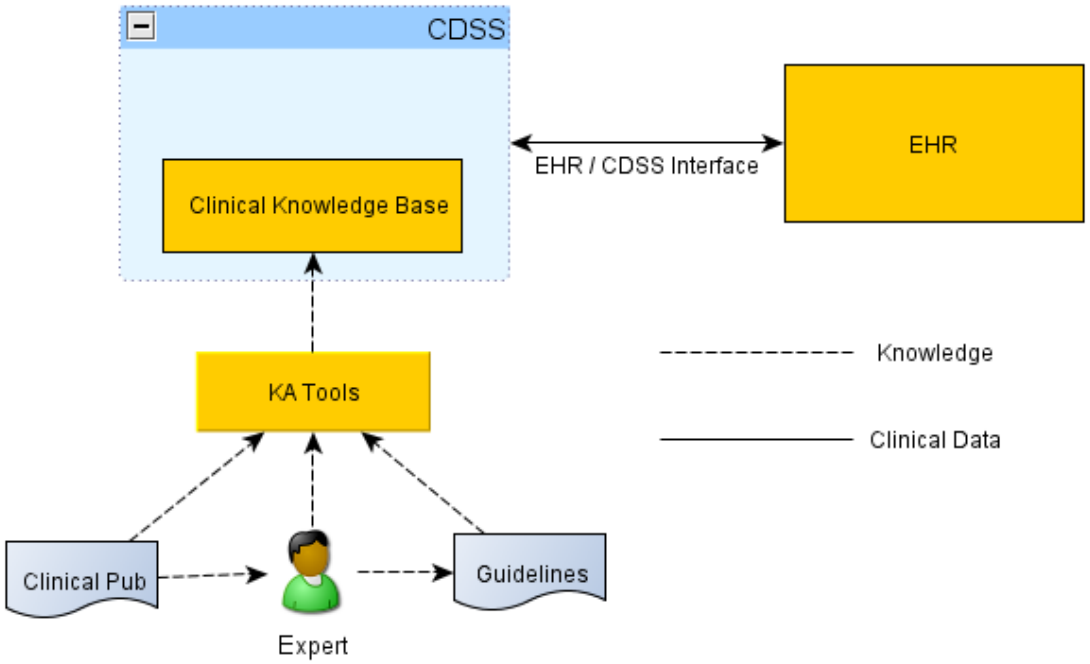

Figure 2.15: Knowledge and data in the CDSS environment

Their relationship is expressed in Figure 2.15 which shows the typical architecture of a CDSS. The connections in the diagram are:

- Clinical expert reviews new clinical research publications and based on the evidence updates clinical guidelines for the treatment of chronic disease.
- Knowledge acquisition tools (KA) may extract clinical knowledge from all three sources: (i) the clinical publication, (ii) directly from the clinical expert, (iii) the clinical guideline itself.
- The KA transforms the unstructured knowledge, or knowledge in the experts mind into a formalized format (e.g. rules or CIGs) stored in its knowledge base.
- The CDSS applies the personal clinical data sent with the formalized clinical knowledge in order to infer a recommendation. The patient data is received in a structured format often utilizing medical vocabulary standards through an application programming interface (API) interface.

2.4.1 Knowledge acquisition and authoring tools

One difficulty in developing a CDSS is the transfer of knowledge from medical literature, clinical guidelines, and clinical specialist in a computer interpretable format. Although

many systems can represent explicit clinical knowledge (either as a workflow, rules, or mixed), the support offered by KA tools is the conversion of these free text representation into a format that the CDSS reasoning engine operates with.

2.4.1.1 Natural language processing (NLP)

The ideal situation would be if a KA tool could process a new evidence based medical journal paper, extract the knowledge and using NLP algorithms update and modify the CDSS's knowledge base that contain the rules, or CIG with the new clinical knowledge from the journal paper. The reality is that NLP is still an evolving field of research and even though there have been some great success (e.g. IBM's Watson [76] with Deep Learning) at present only some parts of clinical knowledge can be obtained via NLP methods [77]. More realistic attempts are semi-automatic approaches such as the Stepper tool [78] that assist a knowledge engineer in developing a computable version from narrative guidelines via a step-by-step process.

2.4.1.2 CIG and rule authoring tools

Virtually every CIG needs software to support their creation, and in many circumstances large number of rules can be difficult task. Furthermore in the clinical domain the rule contrivers may not always be familiar with the rule language that is being used by the reasoning engine. This is why many rule based CDSS come with a KA mechanism for creating rules. For clinical systems, rules can be created in mainly three ways:

1. **Manually written** by knowledge engineers working with clinicians. The disadvantage is that it always requires both a technical and medical skilled staff, to create, or modify. The use of domain specific language (as described in section 2.2.2) makes it easier for non technical medical staff to author and modify.

2. Assistance with a **graphical user interface (GUI)**. Rule based systems often have a GUI editor so that non-technical staff can directly edit and manage rules.

3. Automatically extracted from another computer interpretable format, typically from a CIG [42, 48] or implicitly encapsulated in a classifier via machine learning algorithms.

2.4.2 Electronic health records

Over the last 20 years there has been a gradual replacement of paper based records by electronic medical records in the healthcare systems of the developed world. That said,

there has still not been the same level of penetration of computer systems in health as in finance, transport and manufacturing, and retail industries. Furthermore, different healthcare regulations and standards from country to country has varied the design and use of these systems. Even within a country or city itself, different primary care clinics, and hospitals use different types of systems. While there has been a push to standardize EHRs, this has not yet been achieved.

CDSS when connected to an EHR facilitate workflow integration within the medical center, and remove the need for redundant data entry. For this reason, the architecture of a CDSS often incorporates a mechanism to pull and push patient data directly to the EHR. Ideally the system can extract all the necessary parameters in order to make a decision from the patient's history all available in the EHR. Also in the ideal case the action decided by the clinician should be recorded back in the EHR record.

2.4.3 Communication interface between health information systems

Since it was difficult to standardize the way medical records are stored, research and industry communities such as HL7 and OpenEHR [79] have concentrated in creating standards in which pieces and blocks of clinical information are exchanged and represented between medical systems. This provides for a natural interface in which a CDSS can receive and exchanged patient data. There is still no consensus on the best format, and research is still ongoing. Listed below are the most popular formats relevant to CDSS.

- **HL7 V3 messages** An XML based format for interoperability between medical systems. These messages include concepts of message wrappers, sequential interactions, and model-based message payloads. It uses specific medical code vocabularies to represent clinical terms.

- **HL7 CCD** The HL7 Continuity of care document is an XML based specification containing mandatory textual part to ensure human interpretation and a structured part for software processing. The structured part uses an HL7 Reference Information Model (RIM) that provides a framework for refering to medical code vocabularies such as SNOMED-CT and LOINC.

- **HL7 VMR** The Virtual Medical Record (VMR) is a data model for representing clinical information relevant, specifically designed for clinical decision support, therefore excludes a lot of unnecessary administrative and financial information. The VMR encompasses data such as patient demographics, clinical history, as well as represents outputs from a CDSS such as recommendations, interventions,

and drug prescriptions. As with the other 2 HL7 models it is based on the same underlying RIM. The VMR standard is still under development and some medical informatics companies such as the Australian based Medical Objects have implement its own version of it.

- **HL7 FHIR** The Fast Healthcare Interoperability Resources has a strong focus on implementation fast and easy to implement. Strong foundation in current Web standards. Concise and easily understood specifications. Support service oriented architectures. Data may be represented in through either XML or JSON.

- **OpenEHR** Is a similar but alternative concept to HL7, and describes itself as a knowledge-oriented computing framework for representing high quality reusable clinical models of content and processes through what they have called archetype. OpenEHR modeling offers a multiple layer clinical data modeling approach, at the first level is a RIM similar to HL7 for representing basic data items which will form the building blocks for the next level. At the second layer are the archetypes which are formed of basic data types from the RIM layer and other high level archetypes with constraints on the data such as min/max values. The purpose of an archetype is to represent a reusable discrete clinical concept. For example the reusable concept of "blood pressure" is formed of the more basic concepts such as state (sitting, standing, lying), protocol (taken from leg, arm, left side, right side), multiple readings (baseline, 5 minute, 10 minute, etc), and finally the data (systolic, diastolic, comments).

2.4.4 Medical vocabularies

The RIM layer of clinical and patient data models often use coding terminology standards to label values of clinic data items such as symptoms, diseases, drugs, and laboratory measurements. There are several coding systems that overlap highly but with varying degree of generality and specificity in coding terms as shown in Table 2.9.

Table 2.9: Common medical vocabularies

Vocabulary	Description	Example
SNOMED-CT	Systematized Nomenclature of Medicine - Clinical Terms (SNOMED-CT) is a very diverse and comprehensive clinical terminology coding system that uses a controlled vocabulary. It is designed for capturing information about patient's history, illness, treatment and outcomes. It is mappable to LOINC, and ICD9	A COPD disease is represented by SNOMED-CT ID "13645005"
LOINC	Logical Observation Identifiers, Names and Codes (LOINC) is a public set of codes and names for storage and transmission of clinical laboratory results. The main objective is to ID laboratory tests, results and other clinical observations.	FVC after bronchodilation is represented by the LOINC ID "19874-7"
ICD	The International Statistical Classification of Diseases and Related Health Problems (ICD) is designed to support the collection, classification, processing, and presentation of mortality statistics. Main focus is to summarize the incidence of diseases and operations on national and international levels. Current version is ICD-10 it has 8000 categories and 12,500 codes in a hierarchy.	COPD is identified by "J44"

Chapter 3

Clinical Decision Support Systems (CDSS) for preventive management of COPD patients

F. Velickovski, L. Ceccaroni, J. Roca, F. Burgos, J. B. Galdiz, N. Marina, and M. Lluch-Ariet. "Clinical Decision Support Systems (CDSS) for preventive management of COPD patients". In: *Journal of Translational Medicine* 12.2 (2014), pp. 1–10. DOI: 10.1186/1479-5876-12-S2-S9

- Journal: Journal of translation medicine
- Impact factor: 3.694
- Quartile: 1st
- Status: Published

Velickovski et al. Journal of Translational Medicine 2014, 12(Suppl 2):S9
http://www.translational-medicine.com/content/12/S2/S9

JOURNAL OF TRANSLATIONAL MEDICINE

RESEARCH Open Access

Clinical Decision Support Systems (CDSS) for preventive management of COPD patients

Filip Velickovski[1,2*], Luigi Ceccaroni[3†], Josep Roca[4,6], Felip Burgos[4,6], Juan B Galdiz[5,6], Nuria Marina[5], Magí Lluch-Ariet[1,7]

Abstract

Background: The use of information and communication technologies to manage chronic diseases allows the application of integrated care pathways, and the optimization and standardization of care processes. Decision support tools can assist in the adherence to best-practice medicine in critical decision points during the execution of a care pathway.

Objectives: The objectives are to design, develop, and assess a clinical decision support system (CDSS) offering a suite of services for the early detection and assessment of chronic obstructive pulmonary disease (COPD), which can be easily integrated into a healthcare providers' work-flow.

Methods: The software architecture model for the CDSS, interoperable clinical-knowledge representation, and inference engine were designed and implemented to form a base CDSS framework. The CDSS functionalities were iteratively developed through requirement-adjustment/development/validation cycles using enterprise-grade software-engineering methodologies and technologies. Within each cycle, clinical-knowledge acquisition was performed by a health-informatics engineer and a clinical-expert team.

Results: A suite of decision-support web services for (i) COPD early detection and diagnosis, (ii) spirometry quality-control support, (iii) patient stratification, was deployed in a secured environment on-line. The CDSS diagnostic performance was assessed using a validation set of 323 cases with 90% specificity, and 96% sensitivity. Web services were integrated in existing health information system platforms.

Conclusions: Specialized decision support can be offered as a complementary service to existing policies of integrated care for chronic-disease management. The CDSS was able to issue recommendations that have a high degree of accuracy to support COPD case-finding. Integration into healthcare providers' work-flow can be achieved seamlessly through the use of a modular design and service-oriented architecture that connect to existing health information systems.

Introduction and background

An important problem in healthcare is the significant gap between optimal evidence-based medical practice and the care actually applied. A systematic review [1] of adherence to *chronic obstructive pulmonary disease* (COPD) guidelines by clinicians found that the assessment of the disease and the therapy applied to patients were suboptimal. This situation exists across all chronic-disease care in general: in a multinational survey [2] of chronically ill adults, 14-23% of cases reported at least one medical error in the previous two years.

Clinical decision support systems (CDSSs) can be defined as "software that is designed to be a direct aid to clinical decision-making in which the characteristics of an individual patient are matched to a computerized clinical knowledge base (KB), and patient-specific assessments or recommendations are then presented to the clinician and/or the patient for a decision" [3]. CDSSs have the potential to enhance healthcare and health, and to help close the gap between optimal practice and actual clinical care.

* Correspondence: fvelickovski@bdigital.org
† Contributed equally
[1]Barcelona Digital Technology Centre, 5th floor, 08018 Barcelona, Spain
Full list of author information is available at the end of the article

The primary objective of the work reviewed in this manuscript is to develop a set of decision-support services so that health professional staff (primary care clinicians and allied health professionals) can obtain fast, reliable and directly applicable advice when dealing with citizens at risk and early-stage patients with COPD, while minimising the impact in work-flows. Specifically, to tackle under-diagnoses, a suite of case-finding services has been developed in order to provide recommendations for both informal (e.g. pharmacy) and formal (e.g. primary care) clinical contexts at early stages of disease development. The case-finding services include a quality-control module to provide recommendations, and expert-quality classifications for forced spirometry tests performed by non-expert clinical providers (primary-care clinicians or allied healthcare providers, such as in a pharmacy) [4,5].

To support the management of disease heterogeneity, decision support services for patient stratification into treatment groups have been designed, relying on three main aspects: firstly, enhancing applicability of well-established rules recommended by the consensus report for the diagnosis, management, and prevention of COPD released by the Global Initiative for COPD (GOLD guidelines) [6]; secondly, using the latest consolidated knowledge on COPD management; and, thirdly, incorporating the knowledge generated by the Synergy-COPD European project, within which the research described in this paper is framed.

Related work

Ten of the most critical challenges facing the design, development, implementation and deployment of CDSS technology in healthcare were highlighted by a study Sittig et al., 2008 [7]. From these ten "grand challenges", reinforced subsequently by Fox et al., 2010 [8], and relevant to the context of this manuscript are (i) *disseminate best practices in CDS design, development, and implementation*; (ii) *create an architecture for sharing executable CDS modules and services*; (iii) *create internet-accessible clinical decision support repositories*. Furthermore, Kawamoto et al., 2005 [9], performed a systemic review of publications which reported performance of CDSS systems that included description of features. The objective was to determine a correlation between successful CDSS and specific features. They found successful CDSSs had the following three characteristics: (i) *Decision support integrated into the work-flow* ; (ii) *decision support delivered at the time and place of decision making* ; (iii) *actionable recommendations*.

Another systematic review of CDSSs was performed by Roshanov et al., 2011 [10], with the objective to determine if CDSSs improve the process of chronic care (in diagnosis, treatment, and monitoring) and associated patient outcomes. The authors identified 55 trials that measured and reported the impact of the CDSS on the process of care, and/or patient outcome. Out of the CDSSs that measured the impact on the process of care, 52% demonstrated a statistically significant improvement, and out of the trials that measured patient outcome 31% demonstrated benefits. Specifically, for chronic respiratory diseases (asthma and COPD), only one [11] of the nine reported a positive impact in the process of care: a CDSS for the management of drug therapy in severe asthma. From the five that measured impact on patient outcome, only two [12,13] reported a benefit.

Closely related to our work, Hoeksema et al., 2011 [14] performed a study to report the validity and accuracy of a CDSS designed for the assessment and management of asthma by leading medical institutions in the USA. The system used a similar approach to the Synergy-COPD CDSS by using rules extracted from the guidelines for the diagnosis and management of asthma (EPR-3) [15]. The system assesses the severity of asthma, by applying rules based on a set of inputs, from the patients symptoms, exacerbations, and spirometry (lung function) parameters. Furthermore it recommends the line of treatment, based on the severity level and other factors. The CDSS performed relatively accurately compared to clinicians for the asthma assessment task (pulmonologists agreed with the CDSS 67% of the time, and from the disagreements an expert panel determined that the CDSS was at error 68% of the time, making an overall accuracy level of 78% for the CDSS). The result for the CDSS was poor for the treatment recommendations (pulmonologists agreed with the CDSS 29% of the time, and from the disagreements an expert panel determined that the CDSS was at error 54% of the time, making an overall accuracy level of 62% for the CDSS).

Methods

Architecture of the CDSS

The design of the architecture of a CDSS has an important influence on its successful adoption [16]. Four principle architectural models were considered (see also Table 1):

(i) Standalone models: this architecture was used by early CDSSs. Since it has no integration to an external *health information system* (HIS) or *electronic health record* (EHR), it requires the user to enter all findings and clinical information, thus being time consuming. The advantage of such systems is that they are easily sharable and transferable to different centres (i.e. just by copying the software across).

(ii) Integrated models: this architecture is tightly coupled to the HIS or EHR. Such CDSSs may be proactive in issuing alerts, and make less input demands on users as the data are already available.

Table 1 Comparison of features in CDSS architectures.

Architectural model's feature	Stand-alone	Integrated	Standard-based	Service-oriented
Service transferable across clinical centres	Yes	No	Yes	Yes
Manual data-entry to CDSS minimized	No	No	Yes	Yes
Connected to EHR or HIS	No	Yes	Yes	Yes
Vendor independent EHR or HIS	N/A	No	Yes	No
Standardized clinical knowledge representation	No	No	Yes[†]	Sometimes
Standardized clinical data representation	No	No	Sometimes	Yes

[†] despite an on-going effort for the last two decades, there is still not a widespread adoption of standard-based systems, nor a widespread CIG format

The disadvantage of such a model is the difficulty to be shared, as it is dependent on vendor specific HIS or EHR;

(iii) Standard-based models: this architecture separates the CDSS from the HIS and the EHR. Interoperability is achieved through a standardization of the computerized representation of clinical knowledge through the use of *computer interpretable guidelines* (CIGs) [17-20].

(iv) Service-oriented models: this architecture (e.g., [21,22]) separates the CDSS from the HIS, but integrates them using standardized, service based interfaces. The interface encodes the clinical data and recommendations in a formal representation using ontologies and vocabularies. Thus, standardization is based on the data transferred between the HIS and CDSS instead of the the guidelines and clinical rules executed by the CDSS as in standard-based systems.

See [16] for an extensive review.

A service-oriented approach was selected for the CDSS as it covered the most critical features as summarized in Table 1. In this model the CDSS is interfaced through a web service protocol, with clinical data being exchanged through an interoperable format described later in the section on clinical data representation. The diagram in Figure 1 illustrates the architecture, showing the main interfaces between the external user systems and its internal components, which are described as follows.

Controller

The CDSS Controller is responsible for coordinating all communication between internal components and external systems during the execution of a decision support task. It manages user requests/responses that contain clinical data from the patient communicated in the HL7 *virtual medical record* (vMR) format, the running of the reasoning engine, the quality-control module, and the reference-value module.

Reasoning engine and clinical knowledge base

CDSSs may be classified by the reasoning or inference methods they use. These methods, along with demonstrated implementations for COPD management, are listed in Table 2. Approaches that explicitly model knowledge are preferred in the clinical domain because they facilitate the often-needed justification of the recommendation. The work-flow driven approach, by way of encapsulating clinical care protocols into computer interpretable guidelines, has been demonstrated by *J. Fox* and his team successfully through the PROforma language [23,24]. In Synergy-COPD, a rules-based reasoning paradigm was adopted for the CDSS. This approach was to complement existing HISs (Linkcare and Arezzo Pathways) that already implemented clinical work-flows, thus focusing on the critical clinical decision tasks in COPD management, modelled as production rules.

Rule-based programming has its foundations in symbolic production systems, and its basic approach is to decompose a computation into a set of elementary transformations, embodying, in the case of this research, clinical tasks. Each elementary transformation attempts to match its input against a set of templates, and, if some of these match, a rule corresponding to one of the templates is chosen, and the action associated with the rule is executed. Most rule-based inference engines use the Rete algorithm [25]. To represent rules, the CDSS uses the open-source Java-based *JBoss Drools* [26], which has an easier-to-interpret syntax than representations used by competing systems: CLIPS [27] and Jess [28].

Figure 2 is an illustration of the reasoning paradigm implemented by the CDSS. The rule-based engine operates on inserted facts about a patient that are transmitted to the CDSS by the external HIS requiring decision-support services. Facts may be particular clinical findings or measurements or demographic information about the patient (e.g. "forced vital capacity = 3.7 L"; "dyspnea's MRC severity grade = 4"; "gender = male"). Rules represent mathematical or logical knowledge that infers (produces) new facts from currently available facts. Clinical rules are a subclass of rules that represent clinical and medical knowledge that infers new facts or medical recommendations from currently available medical facts. Clinical rules operate within a modular context that allows, at any particular moment,

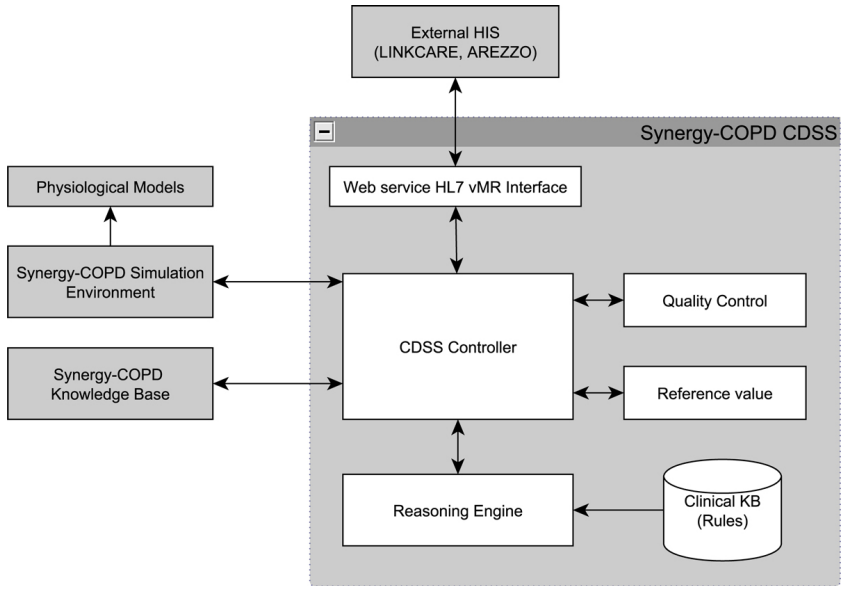

Figure 1 CDSS architecture depicting internal modules, external user HIS, and external supporting Synergy-COPD systems.

firing only the specific set of rules associated with the specific clinical task at hand (e.g. case-finding, diagnosis, assessment).

In this paradigm, the clinician has to ultimately take the final decision. The CDSS generates recommendations based on the patient's personal profile; each recommendation specifying a recommended course of action for the clinician (e.g. "Diagnose patient with COPD.") and the reason why this is the case (e.g. "Symptoms consistent with COPD according to GOLD guidelines' criterion: FEV1/FVC <0.7"). If a recommendation is accepted, it may either automatically create new facts into the system augmenting the patient's medical profile (e.g. COPD added to the list of diseases), or instruct the clinician to perform further actions (e.g. "Take a spirometry measurement after applying a bronchodilator and report back the results.").

Quality control module

The quality-control module implements an algorithm for assessing the acceptability of an individual forced spirometry manoeuvre. The automatic validation of the spirometry measurements consists of identifying wrong/flawed tests, or acceptable/valid tests. Hence an indication is provided regarding the quality of the measurements performed, and feedback or indication is provided regarding the reliability, or confidence level, of the manoeuvre or set of manoeuvres. This is used by an evaluator to assess the quality of a full spirometry test comprising more than one manoeuvre. No expert intervention is necessary and support can be provided in a clinical setting where the clinician or healthcare provider is not an expert in spirometry tests.

Reference value module

The reference-value module invokes continuous prediction equations and their *lower limit of normal* (LLN) for clinical parameters - specifically, it uses spirometric reference values specified by Hankinson et al., 1999 [29] and Quanjer et al., 2012 [30] for case-finding and diagnosis.

Table 2 Inference methods used in CDSS

Method	Description	Implementations
Work-flow driven [1]	Logical flows contain statements that reference and manipulate clinical data, usually executed in a serial manner, with control structures that direct the flow of decision making through the procedure.	[23,24]
Rules-based reasoning [1]	Medical knowledge is captured through a collection of IF-THEN expressions. Reasoning by forward chaining (the most common one) links rules together until a conclusion is reached.	[50-52]
Probabilistic reasoning [1,2]	Bayesian networks and graphical representation that describes the causal relationships between diseases and symptoms with conditional probabilities.	[53]
Machine learning (ML) [2]	Machine learning and statistical techniques, by learning or training, are used on existing, large datasets of clinical data.	[54,55]

[1]Clinical knowledge explicitly modelled

[2]Clinical knowledge derived or learnt from data of past cases

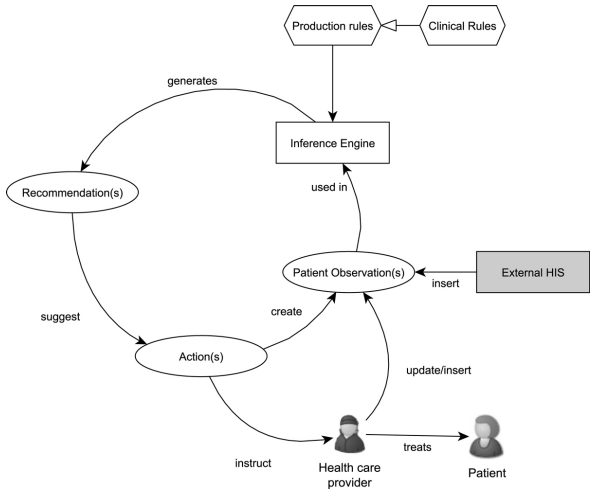

Figure 2 Reasoning paradigm.

External supporting Synergy-COPD systems

The interfaces to Simulation Environment [31,32] and Synergy-COPD Knowledge Base [33,34] (that host the predictive models [35-38] developed within the Synergy-COPD project) have been developed for prognostic extensions to the CDSS. Furthermore the Synergy-COPD Knowledge Base is accessed for drug-drug interaction data.

Clinical data representation

The service-oriented architecture allows the CDSS to deliver decision support capabilities to any external HIS that is able to provide the input clinical data of the patient and receive as output clinical recommendation through a well specified Simple Object Access Protocol (SOAP) interface defined in the web services description language (WSDL). The underlying format that was selected to contain the input clinical data was the HL7 virtual medical record (vMR) [39,40]. The vMR is a data model for representing clinical data specifically optimised for decision support tasks; it captures data about the patient's demographics, clinical history, and is also designed to capture CDSS-generated recommended actions such as suggested clinical interventions, therapies, procedures, and assessments. Data in the vMR are represented using user-defined vocabularies; to enhance interoperability, standardised vocabularies with clear semantics were used within the vMR messages to encode clinical concepts. The vocabularies are shown in Table 3, which includes an example concept and the associated code.

Development

The CDSS was constructed using an iterative and incremental development model adapted from [41]. Figure 3 shows the main development phases. After the initial requirements specification, and design phase, the framework containing the main CDSS components was developed. Three incremental cycles were completed to develop the CDSS web services, and within each cycle the following phases were executed:

(i) *requirements adjustment* - functionalities for subsequent clinical task refined;
(ii) *knowledge acquisition* - clinical guidelines interpreted by respiratory specialists and defined as rules or as algorithm;
(iii) *knowledge engineering* - translation of clinical rules into Drools rules representation and classification into specialized CDSS modules (quality control, reference value);
(iv) *validation and testing* - input test cases and expected output defined and tested against CDSS web services;
(v) *deployment* - secure web service interface exposed and integrated into an existing HIS platform.

Table 3 Standardised vocabulary used in clinical data exchange.

HL7 vMR item	Vocabulary	Example (Code)
Observation	SNOMED-CT [56]	forced vital capacity (50834005)
Procedure	SNOMED-CT	spirometry test (127783003)
Problem (Disease)	ICD-10 [57]	COPD (J44)
Ethnicity	Ethnicity - CDC [58]	white (2106-3)
Language	ISO 639 language code [59]	English (en)

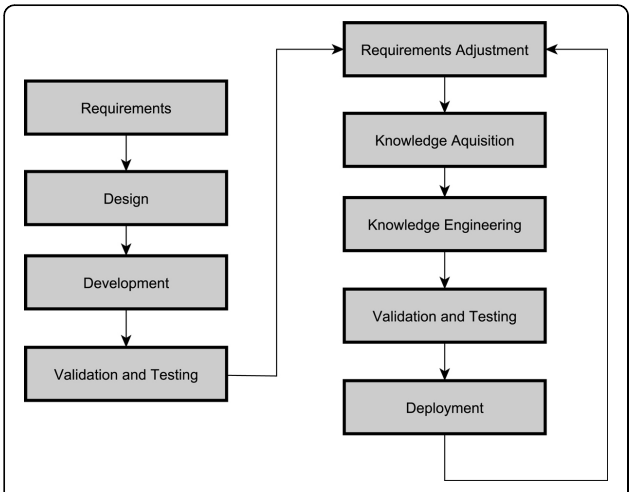

Figure 3 Adapted incremental software development model for the CDSS.

Results

Several decision-support web-services were deployed in a secured environment online for preventive management of COPD patients with performance benchmarked. The web services were incorporated into two existing HISs: all web services into the Linkcare platform [42] and the spirometry quality control web service into Arezzo Pathways [43].

Decision support web services

Spirometry quality control

This service, through spirometry-test results consisting of a set of raw signals from spirometry manoeuvres, determines: quality grade (A, B, C, D or F) of the spirometry test; best lung function parameters for the volume that has been exhaled at the end of the first second of forced expiration (FEV1), the vital capacity from a maximally forced expiratory effort (FVC), the highest forced expiratory flow (PEF), back extrapolated volume (BEV), and their associated manoeuvres; acceptability of each manoeuvre; ranking of each manoeuvre; for manoeuvres deemed to unacceptable, the reasons for their rejection.

case-finding: Eligibility for spirometry test

Through an inclusion/exclusion criteria, represented as Drools rules in the clinical knowledge-base, this service generates advice on subjects at risk of COPD. Subjects are selected for further investigation based on demographics, risk factors, and symptoms. The system produces patient-specific advice for: eligibility of the subject for a further spirometry test; recommendations for smokers based on their dependency.

Case finding: Preliminary evaluation

From the results of a pre-bronchodilation spirometry of an eligible subject, this service determines: requirement to refer the subject to primary care for further tests; preliminary evaluation of lung function; cessation advice for smokers based on their dependency.

Diagnosis: Primary care evaluation

From a patient's full exam consisting of pre-bronchodilation and post-bronchodilation spirometry, this service determines probable COPD cases, evaluates lung function, and issues cessation advice for smokers based on their dependency.

Assessment: Patient stratification

From a patient's post-bronchodilation spirometry result and index scores from standard questionnaires (COPD assessment test [44], modified Medical Research Council dyspnea scale [45]) the patient is stratified into the GOLD 2011 [6] categories: group A - low exacerbation risk, less symptoms; group B - low exacerbation risk, more symptoms; group C - high exacerbation risk, less symptoms; group D - high exacerbation risk, more symptoms. Each stratification group has an associated recommended set of pharmacological and non-pharmacological therapies.

Evaluation of the CDSS as a diagnosis service

Validation dataset

The performance of the CDSS diagnosis service was compared with an anonymised database of patients from Primary Care centres participating in forced-spirometry training in a web-based remote support program to enhance quality of forced spirometry done by non-expert professional in the Basque Country region of Spain. Forced-spirometry testing was done using a *Sibel 120 SIBELMED* spirometer. The spirometry quality and diagnosis evaluation was done by one respiratory specialist. Inclusion criteria to form the validation data set were:

(i) age of the patient greater than or equal to 40;
(ii) forced spirometry taken and recorded as an electronic record before and after the application of bronchodilators;
(iii) respiratory specialist used option menu to select the appropriate diagnosis (rather than entered through the free text field).

After applying the inclusion criteria, the validation set was formed containing 323 cases. The use of the dataset for validation purposes was approved by the Ethical Committee of the Hospital Clinic í Provincial de Barcelona.

Benchmarking the diagonisis service

The clinical data for each case in the validation set was fed into the CDSS diagnosis service, the result was compared against the respiratory specialist classification of the case. The mapping in Table 4 was used to compare the specialist classification to the CDSS for the purposes of validation.

Sensitivity and specificity of the CDSS were calculated for cases in the validation set classified as Likely COPD or Unlikely COPD.

$$sensitivity = \frac{TP}{TP + FN} \quad (1)$$

$$specificity = \frac{TN}{TP + FP} \quad (2)$$

Table 4 Mapping from respiratory specialist classification to CDSS diagnosis classification.

Specialist class	CDSS class
Normal, no obstruction pattern	Unlikely COPD
Mild, obstruction pattern	Likely COPD
Moderate obstruction pattern	Likely COPD
Severe obstruction pattern	Likely COPD

wherein TP (true positive) corresponds to cases classified as *Likely COPD* by both CDSS and the specialist; TN (true negative) corresponds to cases classified as *Unlikely COPD by the CDSS* and the by the specialist; FP (false positive) indicates cases classified as *Likely COPD* by the CDSS, but classified as class *Unlikely COPD* by the specialist; and, FN (false negative) corresponds to cases classified as *Unlikely COPD* by the CDSS, but as *Likely COPD* by the specialist. The CDSS produced 101 diagnosis recommendations as likely COPD, and 222 recommendations as unlikely COPD. 297 cases correctly matched the assessment of the specialists (92%). Sensitivity and specificity calculations were calculated to be 90% and 96%, respectively. Table 5 shows the details of these results as a confusion matrix.

Integration

The CDSS operates by receiving and sending standardized messages, and relies on an existing HIS to present its recommendations to the healthcare professional on screen or via the issuance of a report. Two such HISs have successfully implemented the CDSS web services. The CDSS response time for all decision support services was acceptable (within seconds) to the clinical task at hand, and thus allowed a seamless integration into the existing HIS.

Linkcare

Linkcare is an integrated-care open platform allowing healthcare professionals (specialists, general practitioners, case managers, nurses, etc.) to share clinical knowledge around a patient centric model. A Linkcare mobility module allows posting activities to be performed by patients, using their smart-phone, tablet or a web portal. Such activities include follow-up questionnaires and medical measurements, such as measurements by pulse-oximeters, glucometers and spirometers, and measurements of blood pressure. Healthcare professionals can exchange care protocols and clinical data around Integrated Practice Units or specific Clinical Research teams. Integrating the CDSS web services with the Linkcare platform allows healthcare professionals to be assisted in making clinical decision relating to case-finding, diagnosis, and stratification of COPD patient.

Arezzo Pathways

Arezzo Pathways combines best practice clinical guidelines with individual patient data to dynamically generate care pathways and provide decision recommendations specific to each patient at the point of care. This assists clinicians in managing patients with long-term conditions and in making timely and appropriate referrals. The CDSS web service offering spirometry quality-control and quality-assurance has been integrated into Arezzo Pathways.

Table 5 Confusion matrix of diagnosis

		Specialist diagnosis	
		Likely COPD	**Unlikely COPD**
CDSS Diagnosis	Likely COPD	78	23
	Unlikely COPD	3	219

Communication protocol

Figure 4 shows the use of the CDSS through a primary care scenario and the exchange of messages between Linkcare and the CDSS web services, with the objective for a clinician to confirm COPD in a patient. The clinician uses the Linkcare platform to enter the details of the patient in the system, or retrieve them from the EHR. The patient has already been assessed as being at risk of COPD, and the primary care clinician needs to confirm the COPD. For this, the primary care clinician needs to perform a full spirometry exam, i.e. two tests: one before the application of bronchodilators, and one after. To ensure the measurement taken with the spirometer satisfies criteria for an acceptable and reliable test, the full sampled signal, along with the lung function parameters are sent as two request messages, one for each test (pre and post-bronchodilation), to the Spirometry Quality Control web service. Linkcare uses the Diagnosis - Primary Care Evaluation web service to support the clinician in the decision of the diagnosis of the patient. The Linkcare platform sends the CDSS a request message with details of the lung function parameters obtained during the spirometry measurement. The CDSS replies with a response containing the evaluation to confirm the diagnosis and a recommendation to schedule an appointment for further evaluation and stratification.

Discussion and conclusion

A large epidemiological report on the prevalence and burden of respiratory disorders carried out in the general population of Catalonia [46] stresses two important facts in relation to COPD: (i) There is high prevalence in the population greater than 65 years of age (36% in men and 22% in women); (ii) there is a significant level of under-diagnoses (76%). Moreover, in the UK, over 25% of people with a diagnostic label of COPD have been wrongly diagnosed, usually because of poorly-performed spirometry [47]. This research addresses the above issues by targeting the identification of occult COPD cases aiming at a better delineation of the natural history of the disease. The CDSS services for detection and diagnosis provide this capability, and an initial validation of the diagnostic potential of the CDSS shows promising results (overall accuracy of 92%) in the ability to provide high quality recommendation service for the diagnosis of COPD.

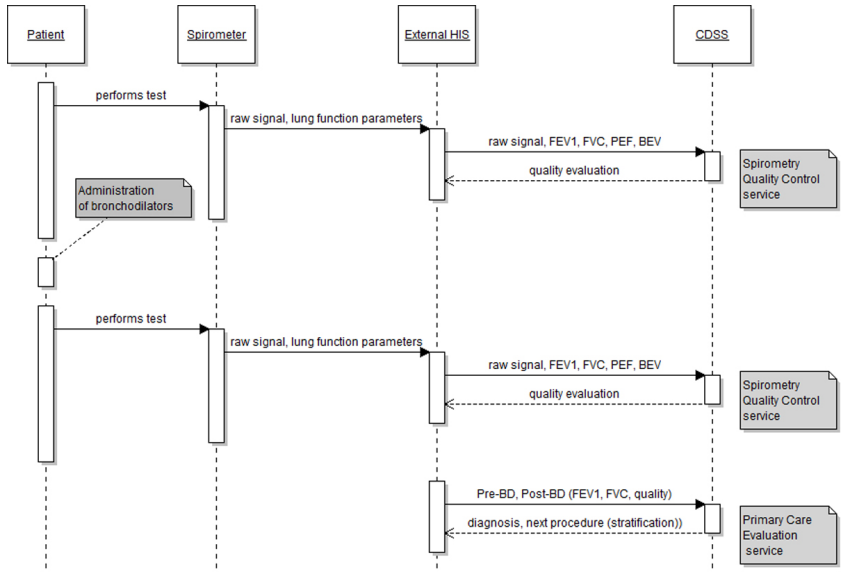

Figure 4 System interaction during confirmation of a COPD diagnosis in a primary care setting.

The *Global Initiative for Chronic Obstructive Lung Disease* (GOLD) consensus report released initially in 2011 *Global Strategy for the Diagnosis, Management, and Prevention of COPD* [6] recommended a major revision in the management strategy for COPD. An updated report released in January 2014 maintains the same treatment paradigm. Assessment of COPD is based on the patient's level of symptoms, future risk of exacerbations, the severity of the spirometric abnormality, and the identification of co-morbidities. This assessment has a limited practical applicability because of its complexity. To facilitate the adoption of the new GOLD classification, it has been incorporated as clinical rules into the clinical knowledge-base, and deployed as a CDSS service. And because an increasing number of reports indicate that the new GOLD classification is not providing added value in terms of clinical impact [48], future activities will be devoted to the development of richer stratification schemes that enrich the assessment capabilities of the CDSS using existing knowledge that is not incorporated in current schemes (i.e. information about general health status, disease severity, activity level, co-morbidities and use of healthcare resources), and by including new knowledge acquired in the Synergy-COPD European research project.

Another revision in the GOLD report was spirometry changed from being a supportive diagnostic tool, to be a requirement for the diagnosis of COPD. This has produced a strong need to support spirometry testing carried out by non-specialized professionals in primary care and allied health providers. This need is addressed through the spirometry quality control CDSS service capable of near expert level feedback on forced-spirometry manoeuvres. An article focusing on the module and performance of the quality control service is to appear in the Journal of Medical Internet Research [49].

Finally the research we present confronts the challenges and applies the characteristics that were originally highlighted in the related work. Firstly, it demonstrates through the modular design and service-oriented architecture of the CDSS framework, the capability of making available internet accessible decision-support modules and services shareable by multiple external HIS platforms. Furthermore the CDSS is able to be directly embedded into the user's work-flow by integration into existing HIS platforms with recommendations *generated at the time and place of decision making.*

Limitations

We acknowledge three principle limitations of the study. Firstly, only data from one respiratory expert was used as ground truth for comparison to the CDSS recommendation in the evaluation. Ideally further independent validation, involving a panel of experts would be more robust in evaluating CDSS performance. Secondly, although our design allows for multiple HIS distributed across the world to use the single CDSS specialised in COPD, we acknowledge guidelines in diagnosis, assessment, and treatment will differ across national borders to suit specific population. The CDSS's modular design allows for instances of the CDSS to be deployed that cater for the specific medical policy or protocol, only by modification of the rules. Thirdly, although a CDSS may achieve a high degree of accuracy and performance, the impact of when

it is deployed in an actual healthcare setting needs to be assessed separately before plans for large scale deployment are developed. As part of this deployment process, the current version of the CDSS is going through a qualitative evaluation using a focus group approach that includes: primary care physicians, nurses, pharmacists and respiratory specialists. A protocol to assess the clinical impact of the use of the CDSS is to be initiated.

Conclusion

Specialized decision support can be offered as a complementary service to existing policies of integrated care for chronic-disease management. The current research has generated a CDSS capable of addressing important issues facing COPD management in case-finding, diagnosis and stratification. The CDSS is able to issue recommendations that have a high degree of accuracy to support COPD case-finding. Moreover, integration into healthcare providers' work-flow has been demonstrated through the use of a modular design and service-oriented architecture that connect to existing health information systems already in use.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

FV and LC conceived and designed the initial CDSS. FV was the lead developer and research engineer of the CDSS. JR and FB provided clinical expertise for the design of the COPD rules in the case-finding, assessment, and diagnosis services. FV and JR supported the deployment of the CDSS into existing health information systems. MLA contributed to design and requirement refinements in the final stages and iterations of the CDSS development. FV, JBG and NM were involved in the evaluation of the diagnostic performance of the CDSS. FV and LC wrote the first draft of the manuscript. FV, LC, JR, FB, JBG, NM, and MLA contributed to the writing of the manuscript. FV, LC, JR, FB, JBG, NM, and MLA agree with manuscript results and conclusions. LC: part of this work was completed while at Barcelona Digital Technology Centre

Acknowledgements

We would like to thank all the members of the Synergy-COPD consortium for their support in the research described.

Declaration

Publication of this article has been funded by the Synergy-COPD European project (FP7-ICT-270086). The opinions expressed in this paper are those of the authors and are not necessarily those of Synergy-COPD project's partners or the European Commission.

This article has been published as part of *Journal of Translational Medicine* Volume 12 Supplement 2, 2014: Systems medicine in chronic diseases: COPD as a use case. The full contents of the supplement are available online at http://www.translational-medicine.com/supplements/12/S2.

Authors' details

[1]Barcelona Digital Technology Centre, 5th floor, 08018 Barcelona, Spain. [2]ViCOROB, Universitat de Girona, Campus Montilivi, 17071 Girona, Spain. [3]1000001 Labs, 08024, Barcelona, Spain. [4]Hospital Clínic, IDIBAPS, Universitat de Barcelona, 08036 Barcelona, Spain. [5]Servicio de Neumología, Hospital Universitario Cruces, 48903 Barakaldo, Bizkaia, Spain. [6]Centro de Investigacíon Biomédica en Red Enfermedades Respiratorias (CIBERES), 07110 Bunyola, Mallorca, Illes Balears, Spain. [7]Departament d'Enginyeria Telemática (ENTEL), Universitat Politécnica de Catalunya (UPC), 08034 Barcelona, Spain.

Published: 28 November 2014

References

1. Lodewijckx C, Sermeus W, Vanhaecht K, Panella M, Deneckere S, Leigheb F, Decramer M: **Inhospital management of copd exacerbations: a systematic review of the literature with regard to adherence to international guidelines.** *J Eval Clin Pract* 2009, **15(6)**:1101-1110.
2. Schoen C, Osborn R, How SKH, Doty MM, Peugh J: **In chronic condition: experiences of patients with complex health care needs, in eight countries, 2008.** *Health Aff (Millwood)* 2009, **28(1)**:1-16.
3. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, Tang PC: **Clinical decision support systems for the practice of evidence-based medicine.** *J Am Med Inform Assoc* 2001, **8(6)**:527-534.
4. Burgos F, Disdier C, de Santamaria EL, Galdiz B, Roger N, Rivera ML, Hervàs R, Durán-Tauleria E, Garcia-Aymerich J, Roca J: **Telemedicine enhances quality of forced spirometry in primary care.** *Eur Respir J* 2012, **39(6)**:1313-1318.
5. Castillo D, Guayta R, Giner J, Burgos F, Capdevila C, Soriano J, Barau M, Casan P: **Copd case finding by spirometry in high-risk customers of urban community pharmacies: a pilot study.** *Respir Med* 2009, **103(6)**:839-845.
6. Vestbo J, Hurd SS, Agusti AG, Jones PW, Vogelmeier C, Anzueto A, Barnes PJ, Fabbri LM, Martinez FJ, Nishimura M, *et al*: **Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: Gold executive summary.** *American journal of respiratory and critical care medicine* 2013, **187(4)**:347-365.
7. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, Campbell E, Bates DW: **Grand challenges in clinical decision support.** *Journal of biomedical informatics* 2008, **41(2)**:387-392.
8. Fox J, Glasspool D, Patkar V, Austin M, Black L, South M, Robertson D, Vincent C: **Delivering clinical decision support services: there is nothing as practical as a good theory.** *Journal of biomedical informatics* 2010, **43(5)**:831-843.
9. Kawamoto K, Houlihan CA, Balas EA, Lobach DF: **Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success.** *Bmj* 2005, **330(7494)**:765.
10. Roshanov PS, Misra S, Gerstein HC, Garg AX, Sebaldt RJ, Mackay JA, Weise-Kelly L, Navarro T, Wilczynski NL, Haynes RB, C.C.D.S.S.S.R.T: **Computerized clinical decision support systems for chronic disease management: a decision-maker-researcher partnership systematic review.** *Implement Sci* 2011, **6**:92.
11. Kattan M, Crain EF, Steinbach S, Visness CM, Walter M, Stout JW, Evans R, Smartt E, Gruchalla RS, Morgan WJ, *et al*: **A randomized clinical trial of clinician feedback to improve quality of care for inner-city children with asthma.** *Pediatrics* 2006, **117(6)**:1095-1103.
12. Plaza V, Cobos A, Ignacio-Garcia J, Molina J, Bergonon S, Garcia-Alonso F, Espinosa C: **Cost-effectiveness of an intervention based on the global initiative for asthma (gina) recommendations using a computerized clinical decision support system: a physicians randomized trial.** *Medicina clínica* 2005, **124(6)**:201-206.
13. McCowan , I.R.F.W.G.H.G.T.C., Neville RG: **Lessons from a randomized controlled trial designed to evaluate computer decision support software to improve the management of asthma.** *Informatics for Health and Social Care* 2001, **26(3)**:191-201.
14. Hoeksema LJ, Bazzy-Asaad A, Lomotan EA, Edmonds DE, Ramírez-Garnica G, Shiffman RN, Horwitz LI: **Accuracy of a computerized clinical decision-support system for asthma assessment and management.** *Journal of the American Medical Informatics Association* 2011, **18(3)**:243-250.
15. **National Asthma Education and Prevention Program: Expert panel report 3 (epr-3): Guidelines for the diagnosis and management of asthma-summary report 2007.** *J Allergy Clin Immunol* 2007, **120(5 Suppl)**:94.
16. Wright A, Sittig DF: **A framework and model for evaluating clinical decision support architectures.** *J Biomed Inform* 2008, **41(6)**:982-990.
17. Fox J, Thomson R: **Decision support and disease management: a logic engineering approach.** 1998, **2(4)**:217-228.
18. Peleg M, Tu S, Bury J, Ciccarese P, Fox J, Greenes RA, Hall R, Johnson PD, Jones N, Kumar A, Miksch S, Quaglini S, Seyfang A, Shortliffe EH, Stefanelli M: **Comparing computer-interpretable guideline models: a case-study approach.** *J Am Med Inform Assoc* 2003, **10(1)**:52-68.
19. Boxwala AA, Peleg M, Tu S, Ogunyemi O, Zeng QT, Wang D, Patel VL, Greenes RA, Shortliffe EH: **Glif3: a representation format for sharable**

computer-interpretable clinical practice guidelines. *J Biomed Inform* 2004, **37(3)**:147-161.
20. Isern D, Moreno A: **Computer-based execution of clinical guidelines: a review.** *Int J Med Inform* 2008, **77(12)**:787-808.
21. Tu SW, Hrabak KM, Campbell JR, Glasgow J, Nyman MA, McClure R, McClay J, Abarbanel R, Mansfield JG, Martins SM, Goldstein MK, Musen MA: **Use of Declarative Statements in Creating and Maintaining Computer-interpretable Knowledge Bases for Guideline-based Care.** 2006, 784-788.
22. Kawamoto K, Lobach DF: **Design, implementation, use, and preliminary evaluation of sebastian, a standards-based web service for clinical decision support.** In *AMIA Annual Symposium Proceedings. Volume 2005.* American Medical Informatics Association; 2005:380.
23. Fox J, Patkar V, Chronakis I, Begent R: **From practice guidelines to clinical decision support: closing the loop.** *Journal of the Royal Society of Medicine* 2009, **102(11)**:464-473.
24. Fox J, Glasspool D, Modgil S: **A canonical agent model for healthcare applications.** *Intelligent Systems, IEEE* 2006, **21(6)**:21-28.
25. Forgy CL: **Rete: A fast algorithm for the many pattern/many object pattern match problem.** *Artificial intelligence* 1982, **19(1)**:17-37.
26. Bali M: **Drools JBoss Rules 5.0: Developer's Guide.** Packt Publishing, Birmingham, United Kingdom; 2009.
27. CLIPS Website: 2013 [http://clipsrules.sourceforge.net/], [accessed August 24, 2014].
28. JESS Website: 2013 [http://herzberg.ca.sandia.gov/], [accessed August 24, 2014].
29. Hankinson JL, Odencrantz JR, Fedan KB: **Spirometric reference values from a sample of the general u.s. population.** *Am J Respir Crit Care Med* 1999, **159(1)**:179-187.
30. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, Enright PL, Hankinson JL, Ip MSM, Zheng J, Stocks J, E.R.S.G.L.F.I: **Multi-ethnic reference values for spirometry for the 3-95-yr age range: the global lung function 2012 equations.** *Eur Respir J* 2012, **40(6)**:1324-1343.
31. Huertas Miguelanez MdlM, Ceccaroni L: **A simulation and integration environment for heterogeneous physiology-models.** *e-Health Networking, Applications Services (Healthcom), 2013 IEEE 15th International Conference On* 2013, 228-232.
32. Huertas Miguelanez MdlM, Mora D, Cano I, Maier D, Gomez-Cabrero D, Lluch-Ariet M, Miralles F: **Simulation Environment and Graphical Visualization Environment: a COPD use-case.** *BMC Journal of Translational Medicine* 2014, **12(S2)**:S7.
33. Maier D, Kalus W, Wolff M, Kalko SG, Roca J, Marin de Mas I, Turan N, Cascante M, Falciani F, Hernandez M, Villa-Freixa J, Losko S: **Knowledge management for systems biology a general and visually driven framework applied to translational medicine.** *BMC Syst Biol* 2011, **5**:38.
34. Cano I, Tényi A, Schueller C, Wolff M, Huertas Miguelanez MdlM, Gomez-Cabrero D, Antczak P, Roca J, Cascante M, Falicani F, Maier D: **The COPD Knowledge Base: enabling data analysis and computational simulation in translational COPD research.** *BMC Journal of Translational Medicine* 2014, **12(S2)**:S6.
35. Wagner PD: **Determinants of maximal oxygen transport and utilization.** *Annu Rev Physiol* 1996, **58**:21-50.
36. Selivanov VA, Votyakova TV, Zeak JA, Trucco M, Roca J, Cascante M: **Bistability of mitochondrial respiration underlies paradoxical reactive oxygen species generation induced by anoxia.** *PLoS Comput Biol* 2009, **5(12)**:1000619.
37. Selivanov VA, Votyakova TV, Pivtoraiko VN, Zeak J, Sukhomlin T, Trucco M, Roca J, Cascante M: **Reactive oxygen species production by forward and reverse electron fluxes in the mitochondrial respiratory chain.** *PLoS Comput Biol* 2011, **7(3)**:1001115.
38. Cano I, Mickael M, Gomez-Cabrero D, Tegner J, Roca J, Wagner PD: **Importance of mitochondrial in maximal {O2} transport and utilization: A theoretical analysis.** *Respiratory Physiology & Neurobiology* 2013, **189(3)**:477-483.
39. Kawamoto K, Del Fiol G, Strasberg HR, Hulse N, Curtis C, Cimino JJ, Rocha BH, Maviglia S, Fry E, Scherpbier HJ, Huser V, Redington PK, Vawdrey DK, Dufour J-C, Price M, Weber JH, White T, Hughes KS, McClay JC, Wood C, Eckert K, Bolte S, Shields D, Tattam PR, Scott P, Liu Z, McIntyre AK: **Multi-national, multi-institutional analysis of clinical decision support data needs to inform development of the hl7 virtual medical record standard.** *AMIA Annu Symp Proc* 2010, **2010**:377-381.
40. Gomoi VS, Dragu D, Stoicu-Tivadar V: **Virtual medical record implementation for enhancing clinical decision support.** *Studies in health technology and informatics* 2012, **180**:118.
41. Fujii T, Dohi T, Fujiwara T: **Towards quantitative software reliability assessment in incremental development processes.** *Proceedings of the 33rd International Conference on Software Engineering* ACM; 2011, 41-50.
42. **Linkcare: The Integrated Health Care Shared Knowledge Community for Health Care Professionals.** Linkcare; 2014 [http://www.linkcare.es/], [accessed August 24, 2014].
43. **Infermed: Arezzo Pathways.** Infermed; 2014 [http://www.infermed.com/Clinical-Decision-Support/Overview.aspx], [accessed August 24, 2014].
44. Jones P, Harding G, Berry P, Wiklund I, Chen W, Leidy NK: **Development and first validation of the copd assessment test.** *European Respiratory Journal* 2009, **34(3)**:648-654.
45. Mahler D, Wells CK: **Evaluation of clinical methods for rating dyspnea.** *CHEST Journal* 1988, **93(3)**:580-586.
46. Duran-Tauleria E, Forastiere F, Janson C, Roca J, Viegi G, Weinmyer G, Anto J, the IMCA Group: **Report on the Respiratory Health of the Elderly Population Living in Catalunya.** Centre for Research in Environmental Epidemiology (CREAL), Barcelona; 2013, Centre for Research in Environmental Epidemiology (CREAL).
47. Holton K, Hill S, Kearney M, Winter R, Moger A: **Quality assured diagnostic spirometry (qads)-performance and competence.** *Am J Respir Crit Care Med* 2013, **187**:2836.
48. Soriano JB: **The gold rush.** *Thorax* 2013, **68(10)**:902-903.
49. Burgos F, Melia U, Vallverdú MA, Velickovski F, Lluch-Ariet M, Caminal P, Roca J: **Clinical Decision Support System to Enhance Quality Control of Forced Spirometry using Information and Communication Technologies.** *Journal of Medical Internet Research* 2014.
50. Kuilboer MM, Van Wijk MA, Mosseveld M, Van Der Does E, Ponsioen BP, De Jongste JC, Overbeek SE, Van Der Lei J: **Feasibility of asthmacritic, a decision-support system for asthma and copd which generates patient-specific feedback on routinely recorded data in general practice.** *Family practice* 2002, **19(5)**:442-447.
51. Song B, Wolf K-H, Gietzelt M, Al Scharaa O, Tegtbur U, Haux R, Marschollek M: **Decision support for teletraining of copd patients.** *Pervasive Computing Technologies for Healthcare, 2009* IEEE; 2009, 1-6, PervasiveHealth 2009. 3rd International Conference On.
52. Rosso R, Munaro G, Salvetti O, Colantonio S, Ciancitto F: **Chronious: an open, ubiquitous and adaptive chronic disease management platform for chronic obstructive pulmonary disease (copd), chronic kidney disease (ckd) and renal insufficiency.** *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* IEEE; 2010, 6850-6853.
53. Himes BE, Dai Y, Kohane IS, Weiss ST, Ramoni MF: **Prediction of chronic obstructive pulmonary disease (copd) in asthma patients using electronic medical records.** *Journal of the American Medical Informatics Association* 2009, **16(3)**:371-379.
54. Jafari S, Arabalibeik H, Agin K: **Classification of normal and abnormal respiration patterns using flow volume curve and neural network.** *Health Informatics and Bioinformatics (HIBIT), 2010 5th International Symposium On* IEEE; 2010, 110-113.
55. Sahin D, Übeyli ED, Ilbay G, Sahin M, Yasar AB: **Diagnosis of airway obstruction or restrictive spirometric patterns by multiclass support vector machines.** *Journal of medical systems* 2010, **34(5)**:967-973.
56. Stearns MQ, Price C, Spackman KA, Wang AY: **Snomed clinical terms: overview of the development process and project status.** *Proceedings of the AMIA Symposium* American Medical Informatics Association; 2001, 662.
57. World Health Organization: International Classification of Diseases (ICD-10): World Health Organization; 2014 [http://apps.who.int/classifications/icd], [accessed August 24, 2014].
58. **Centres for Disease Control and Prevention: Race and Ethnicity Code Set Version 1.0.** Centres for Disease Control and Prevention; 2000 [http://www.cdc.gov/nchs/data/dvs/Race_Ethnicity_CodeSet.pdf], [accessed August 24, 2014].
59. **Library of Congress: Codes for the Representation of Names of Languages.** Library of Congress; 2014 [http://www.loc.gov/standards/iso639-2/php/code_list.php], [accessed August 24, 2014].

doi:10.1186/1479-5876-12-S2-S9
Cite this article as: Velickovski *et al.*: Clinical Decision Support Systems (CDSS) for preventive management of COPD patients. *Journal of Translational Medicine* 2014 **12**(Suppl 2):S9.

Chapter 4

Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments

U. Melia, F. Burgos, M. Vallverdu, F. Velickovski, M. Lluch-Ariet, J. Roca, and P. Caminal. "Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments". In: *PLoS ONE* 9.12 (Dec. 2015), pp. 1–14. DOI: 10.1371/journal.pone.0116238

- Journal: PLOS ONE
- Impact factor: 3.057
- Quartile: 1st
- Status: Published

PLOS ONE

RESEARCH ARTICLE

Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments

Umberto Melia[1]*, Felip Burgos[2,3], Montserrat Vallverdú[1], Filip Velickovski[4,5,6], Magí Lluch-Ariet[4,5], Josep Roca[2,3], Pere Caminal[1]

1. Dept. d'Enginyeria de Sistemes, Automàtica i Informàtica Industrial (ESAII), Centre for Biomedical Engineering Research (CREB), Universitat Politècnica de Catalunya, CIBER of Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Barcelona, Spain, **2.** Department of Pulmonary Medicine. Hospital Clínic de Barcelona (ICT). IDIBAPS, Universitat de Barcelona, Barcelona, Spain, **3.** Centro de Investigación en Red de Enfermedades Respiratorias (CibeRes), Palma de Mallorca, Spain, **4.** Barcelona Digital Technology Centre, Barcelona, Spain, **5.** Dept. d'Enginyeria Telemàtica (ENTEL), Universitat Politècnica de Catalunya, Barcelona, Spain, **6.** ViCOROB, Universitat de Girona, Girona, Spain

*umberto.melia@upc.edu

OPEN ACCESS

Citation: Melia U, Burgos F, Vallverdú M, Velickovski F, Lluch-Ariet M, et al. (2014) Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments. PLoS ONE 9(12): e116238. doi:10.1371/journal.pone.0116238

Editor: Josué Sznitman, Technion - Israel Institute of Technology, Israel

Received: August 4, 2014

Accepted: December 7, 2014

Published: December 31, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by Inforegió (AGAUR) 2009; NEXES (Supporting Healthier and Independent Living for Chronic Patients and Elderly, CIP-ICT-PSP-2007-225025); FIS PI09/90634; Servicios Innovadores de Atencion Integrada para Pacientes Crónicos - PITES- ISCIII 2010-12; EC-FP7 Programme, Synergy-COPD, GA n° 270086; Catalan Master Plan of Respiratory Diseases (PDMAR); and TAMESIS (TEC2011-22746, Spanish Government). CIBER of Bioengineering, Biomaterials and Nanomedicine is an initiative of ISCIII. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

We hypothesized that the implementation of automatic real-time assessment of quality of forced spirometry (FS) may significantly enhance the potential for extensive deployment of a FS program in the community. Recent studies have demonstrated that the application of quality criteria defined by the ATS/ERS (American Thoracic Society/European Respiratory Society) in commercially available equipment with automatic quality assessment can be markedly improved. To this end, an algorithm for assessing quality of FS automatically was reported. The current research describes the mathematical developments of the algorithm. An innovative analysis of the shape of the spirometric curve, adding 23 new metrics to the traditional 4 recommended by ATS/ERS, was done. The algorithm was created through a two-step iterative process including: (1) an initial version using the standard FS curves recommended by the ATS; and, (2) a refined version using curves from patients. In each of these steps the results were assessed against one expert's opinion. Finally, an independent set of FS curves from 291 patients was used for validation purposes. The novel mathematical approach to characterize the FS curves led to appropriate FS classification with high specificity (95%) and sensitivity (96%). The results constitute the basis for a successful transfer of FS testing to non-specialized professionals in the community.

Introduction

Forced spirometry (FS) testing aims at a global assessment of lung and chest wall mechanics. Specifically, FS provides measurements of expiratory volume and flow during a maximal expiratory manoeuvre. It is a relevant test in the clinical setting useful to perform both diagnosis and assessment of functional reserve in various lung-related health disorders. The test is also used for pre-operative evaluation and assessment of disability/impairment. Moreover, there is evidence that key spirometric indices (FVC, forced vital capacity; and, FEV1, forced expiratory volume during the first second) predict survival in the general population. For all these reasons, it is forecasted that the role of FS testing will expand across healthcare tiers and beyond respiratory medicine.

As part of the FS testing procedure, the patient performs maximum expiratory maneuvers under the guidance of a healthcare professional who should: (1) aim for a proper cooperation of the patient; (2) assess the quality of different FS manoeuvres; and, (3) select the most suitable spirometric values using the ATS/ERS recommendations [1].

The equipment for FS measurements and the recommendations for testing are highly standardized [2, 3], as well as the quality assessment [1]. The current systems measure expired flow using different technologies [4] that generate two types of FS curves: (1) a volume-time curve (VT) representing volume (in liters, L) along the ordinate and time (in seconds, s) in the abscissa; and, (2) a flow-volume curve (FV) depicting expired flow (in liters per second) in the ordinate and expired volume (in liters) in the abscissa. Clinically useful spirometric indices (i.e FVC and FEV_1) are calculated from selected curves following the international recommendations [1–3]. FS testing requires a high degree of patient cooperation with the support of a health professional to ensure that the quality of the maneuvers follows the recommended standards [1]. The transfer of FS testing to non-specialized professionals in the community generates a challenge in terms of preserving the quality of the testing to preclude misdiagnosis due to poor quality of FS curves.

There is evidence that remote off-line support of quality testing shows both feasibility and cost-effectiveness, but requires supervision by a specialist [5]. Unfortunately, currently available equipment with functionalities for automatic assessment of quality of FS testing generates poor outcomes due to an inadequate application of the ATS/ERS recommendations on quality control [5, 6]. We recently reported the high potential of an automatic algorithm for real-time assessment of quality testing paving the way for the transfer of FS testing to the community [5, 7].

It is hypothesized that the regional deployment of a comprehensive program ensuring: *i)* reliable automatic quality assessment of forced spirometric testing; ii) off-line remote assistance to non-specialized health professionals; and, *iii)* accessibility to quality labeled forced spirometric information across healthcare tiers, may have a marked positive impact on quality of diagnosis, healthcare outcomes and may generate cost savings.

The current research reports the identification of new metrics based on a mathematical approach that describes the entire spirometric curve allowing a proper quality assessment of volume-time (VT); flow-time (FT); and, flow-volume (FV) curves. It is of note that the results are under review in the European Patent Office with the registration number (PCT/EP2013/068732).

Materials and Methods

Databases

Three databases were used for building and validating the algorithm: (1) 24 simulated curves recommended by the ATS [2, 3]; (2) 270 curves from 90 patients examined at the Hospital Clinic de Barcelona [5] (P1); and, (3) 778 curves from 291 patients (P2) from one of the Primary Care centers in Barcelona. Forced spirometry testing in P1 and P2 was performed with the same equipment (Sibel 120, SIBELMED, Barcelona Spain). The simulated curves permitted the elaboration of the initial version of the algorithm; whereas the two patient databases (P1 and P2) were considered for refinement and validation purposes, respectively. The study was approved by the Ethical Committee of the Hospital Clínic de Barcelona. All the participants signed informed consent.

Algorithm development

The 24 simulated ATS curves were used to perform a comprehensive characterization of the curve morphology to facilitate the application of the different quality criteria defined in the ATS/ERS recommendations [2, 3]. To this end, three different concepts were introduced as defined below:

Criterion

Specific feature of the spirometric testing that requires quality assessment (i.e. back extrapolation, end-of-curve, peak flow, etc…). The quality analysis of the different criteria considered by the algorithm will provide an overall quality assessment of the spirometric curve.

Metric

Mathematical description of a given criterion. Several criteria may require one or more metrics to be properly defined.

Threshold

Quantitative values of a given metric used to assess the quality of a criterion. It is of note that some metrics may have primary and secondary thresholds.

The ultimate aim of the algorithm was to integrate the new criteria to enhance current quality assessment [1] and to allow on-line quality control of testing.

Each criterion (C_n) defined with the 24 ATS curves used one (primary) or more (secondary) metrics (M_j) with the respective threshold(s). In each step of the

algorithm development, the results were compared with the criteria of one expert in the field of pulmonary function testing.

In order the perform a global assessment of each spirometric curve, five different zones were identified in the flow-volume graph, as indicated in Fig. 1 and described in detail in S1 File. The overall result of applying the new methodology was the identification of three different quality grades, namely: i) Grade 0 → curve to be rejected because of a bad morphology; ii) Grade 1 → curve with acceptable morphology; and, iii) Grade 2 → curve requiring specialized professional judgment for acceptability.

The refinement of the threshold values was performed in an iterative process to maximize the agreement between the human expert and the automatic classification in grades 0 and 1 and to minimize the number of curves automatically classified in grade 2. The first two categories, grades 0 and 1 allow proper real-time and automatic classification of FS; whereas grade 2 requires off-line expert assessment. The automatic grade assignment is made as described in the S1 File.

The algorithm issued from the analysis of the 24 ATS curves was subsequently evaluated using the P1 database following identical procedures.

The current algorithm incorporates the 4 traditional ATS/ERS criteria commonly used in commercially available equipment with automatic quality assessment and applies several other ATS/ERS criteria for quality control of FS as indicated below in the description of the corresponding zones and in Fig. 2. The metrics corresponding to the 4 traditionally used ATS/ERS recommendations in commercially available equipment are defined as follows:

- *BEV* refers to back extrapolated volume (BEV>0.15 L or BEV <5% of FVC), which is the volume value for t=T_{zer} (T_{zero} refers to the back extrapolated time, which is the time at which the volume curve tangent with maximum slope crosses the horizontal time axis).
- *EOTV* refers to the difference between maximum and minimum volume in the last 1 second of exhalation. (T_{ex} refers to the time from T_{zero} to the time in which the VT curve reaches *EOTV* <0.025 L or the end of exhalation, as depicted in Fig. 1A),
- *FET100* refers to the time from T_{zero} to the time in which the VT curve reaches *FVC*, as depicted in Fig. 1A (6 seconds in adult population).
- Repeatability criteria (three good maneuvers, two of them with differences in FVC and FEV_1 less than 0.15 L).

Five spirometric zones

Fig. 2 displays the rationale for the five spirometric zones considered in the current analysis. The first zone (Z1) encompasses the area from zero to peak flow; whereas Z2 relates to the profile or the peak expiratory flow (PEF). The decrease of flow rate after the peak is analyzed in Z3; the end-of-test area is examined in Z4 and, finally, Z5 considers the overall shape of the curve.

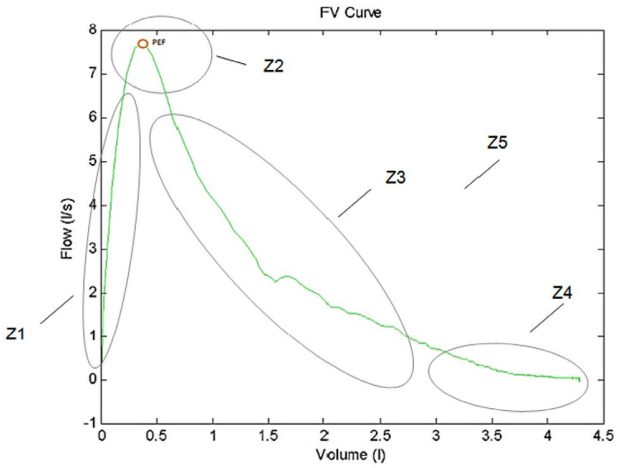

Fig. 1. Spirometer Zones. An example of FV curve with five zones described.

doi:10.1371/journal.pone.0116238.g001

The Z1 criteria ensure that the slope of the curve is regular and free from fluctuations. The calculations are based on the first and second derivatives of the FV curve in zone Z1. Fig. 3 depicts an example of an FV curve that presents a bad morphology in zone Z1. The criterion C_1 detects that irregular concavity or convexity exists. A second criterion C_{2a} detects that the profile has an irregular slope. The criterion C_{2b} detects that the profile has an irregular concavity or convexity.

The Z2 criteria ensure that the PEF occurs at an early point in the maneuver, it has an appropriate height to width ratio and there are no secondary peaks present. Fig. 4 depicts example curves in which the corresponding tests are performed in the zone Z2. Criterion C_3, detects that the PEF point has occurred too late. Criterion C_4 detects that the PEF point is too early. The criteria C_5 analyze the peak. C_{5a} and C_{5b} detect a flat peak (Fig. 4B). The criterion C_{5c} detects a situation of bimodal peaks as depicted in Fig. 4A (multiple peaks). The criterion C_6 detects if the V value in the position of the PEF is lower than a fixed threshold (Fig. 4C).

The Z3 criteria ensure that the slope of the curve is regular and free from fluctuations and are based on the first derivative, and definite integrals of the FV curve in zone Z3. Fig. 5 depicts an example curve in which the corresponding tests are performed in zone Z3. Fig. 5A depicts criterions C_{7a}, C_{7b} and C_{7d}. Fig. 5B depicts criterions C_{7c} and C_{7d}. The criterion C_{7a} detects a situation of high slopes during FV curve descent. The criterion C_{7b} detects an excessive variation in the slope of the FV curve in zone Z3. The criterion C_{7c} detects an excessive variation in the slope calculated in a V segment of the FV curve in the zone Z3. The criterion C_{7d} detects an irregular slope.

The Z4 criteria ensure that the flow at the end of the curve is regular and free from fluctuations and are based on the first order derivative of the FT curve in zone Z4, and the difference between the maximum and minimum volume in the last second of exhalation. Fig. 6 depicts an example curve in which the

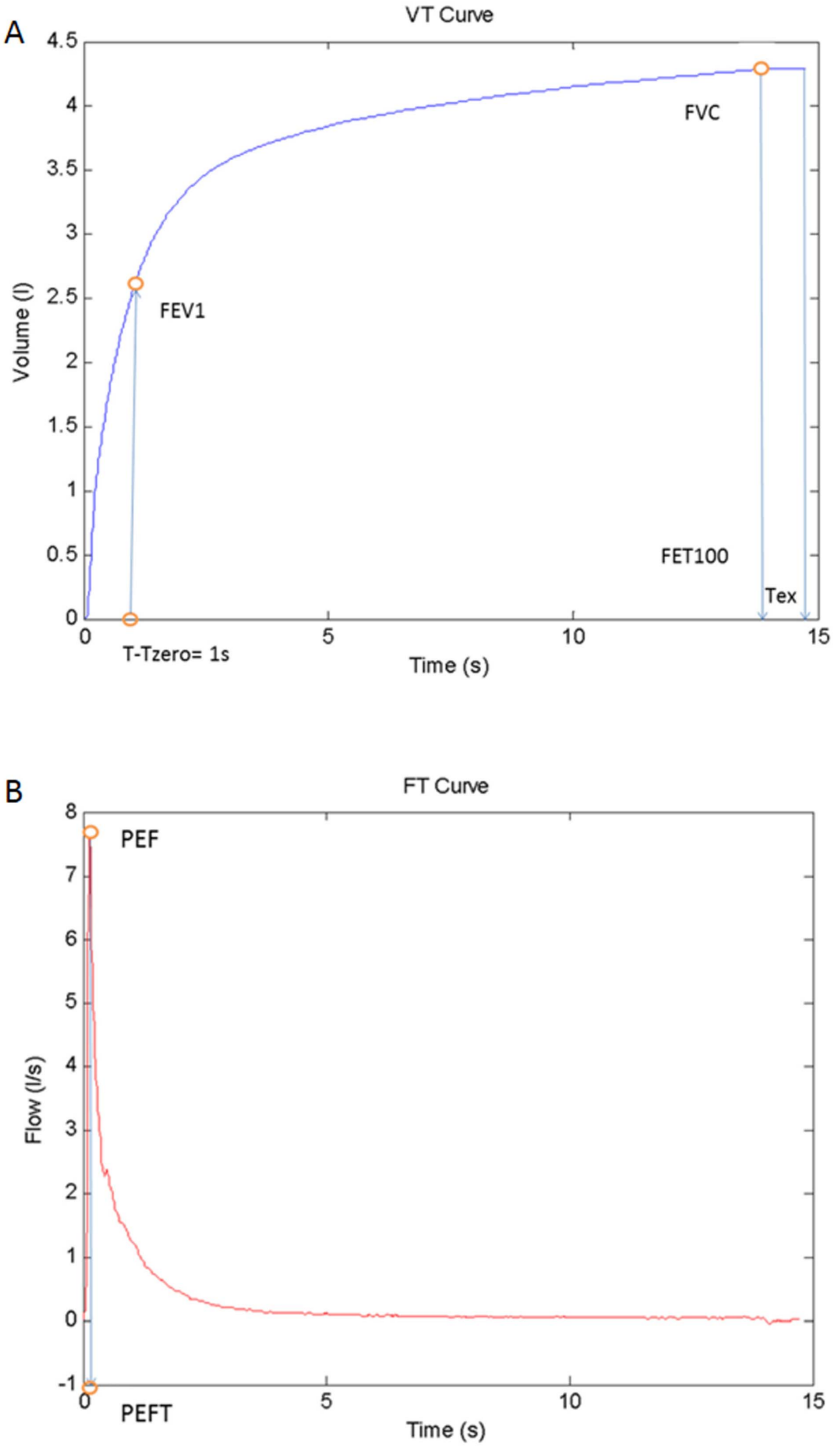

Fig. 2. Spirometer metrics. Metrics involved in the traditional criteria: (A) *FVC* and *FEV1*, (B) *PEFT*.

doi:10.1371/journal.pone.0116238.g002

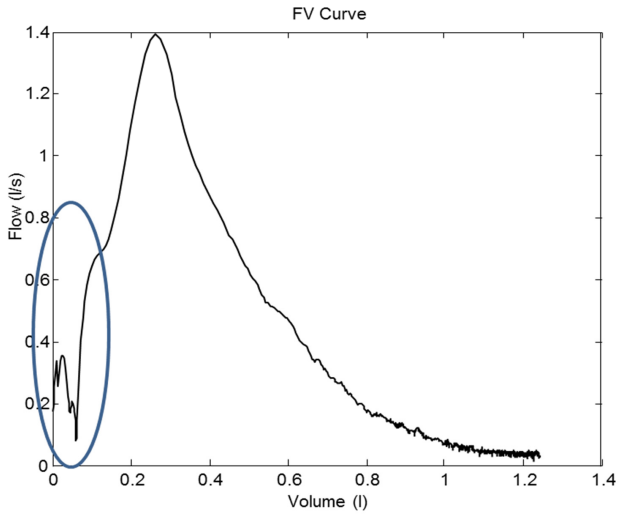

Fig. 3. Zone Z1 analysis. An example of a FV curve that presents irregularity on the ascent to the PEF.

doi:10.1371/journal.pone.0116238.g003

corresponding tests are performed in the zone Z4. Fig. 6A and 6B depict criterion C_{11}. Criteria C_8 is the traditional *BEV* criteria (*BEV*>0.15 L or *BEV* <5% of *FVC*). Criteria C_9 is the traditional *EOTV* criteria (V<0.025 L in $t \geq 1$ s). Criteria C_{10} is a combination of 5 criteria explained in the following lines. C_{10a} detects if the *EOTV* and *Tex* both does not satisfy their traditional criteria. C_{10b} defines a new period to calculate *EOTV* if the *Tex* traditional criteria is satisfied. C_{10c} defines a new threshold for *EOTV* if the *Tex* traditional criterion is satisfied. C_{10d} uses threshold to define the *EOTV(Tex)* and compare with the traditional threshold. C_{10e} defines *EOTV* calculated as a function of *Tex*. Criterion C_{11} detects irregularity or oscillation at the end part of FT curve.

The Z5 criteria ensure that there only exists one local maximum (the *PEF* point) and they are based on the derivative of the FV curve. Fig. 7 depicts criterions C_{12a} and C_{12b}. The criterion C_{12a} detects a situation of multiple peaks that typically occurs when the subject coughs. The criterion C_{12b} detects a situation of multiple peaks for values of V adjacent to FEV_1.

For more details, see S1 File.

Algorithm evaluation

The quality grades (0 to 2) generated by the algorithm using the P2 dataset were compared with those generated by the expert evaluator. Sensitivity and specificity of the algorithm were calculated for all curves classified as classes 0 or 1 in order to quantify the agreement between the algorithm and the evaluator and between the 4 traditional ATS criteria and the evaluator. Sensitivity is defined as the number of curves classified as class 0 by both the algorithm (or the 4 traditional ATS/ERS criteria) and the evaluator divided by the total of the curve evaluated in grade 0 by

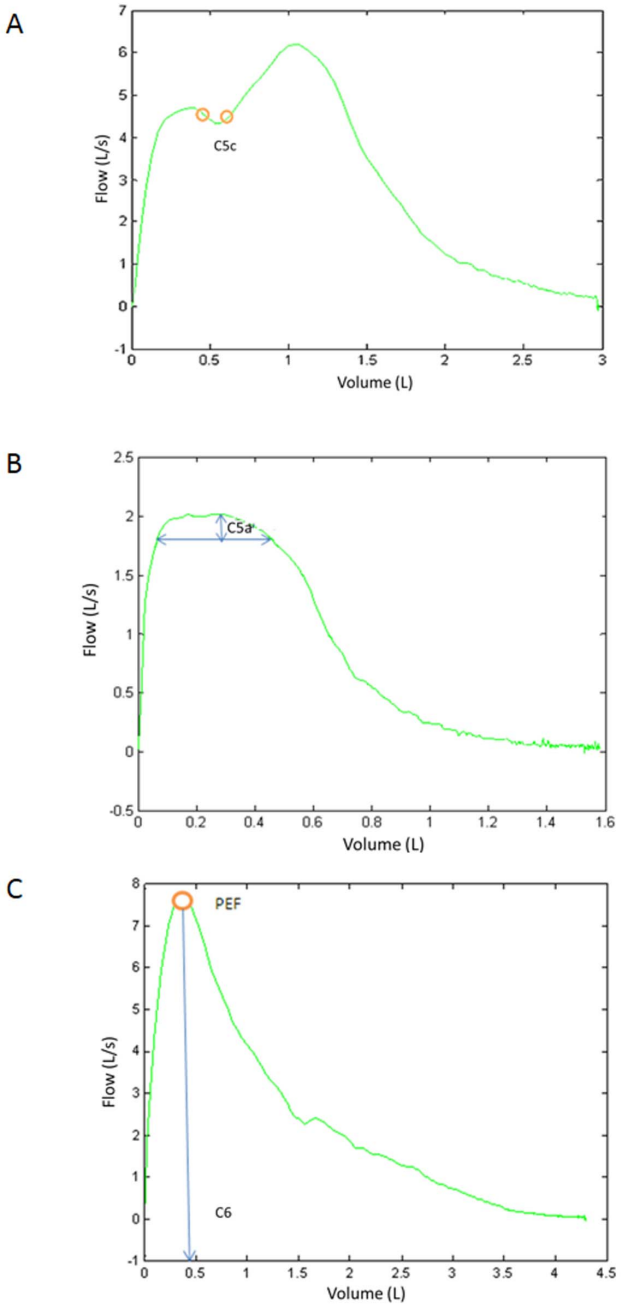

Fig. 4. Zone Z2 analysis. Examples of FV curves that present (A) bimodal peak; (B) flat peak and (C) slow peak.

doi:10.1371/journal.pone.0116238.g004

the evaluator, while specificity is defined as the number of curves classified as class 1 by both the algorithm (or the 4 traditional ATS/ERS criteria) and the evaluator divided by the total of the curve evaluated in grade 1 by the evaluator.

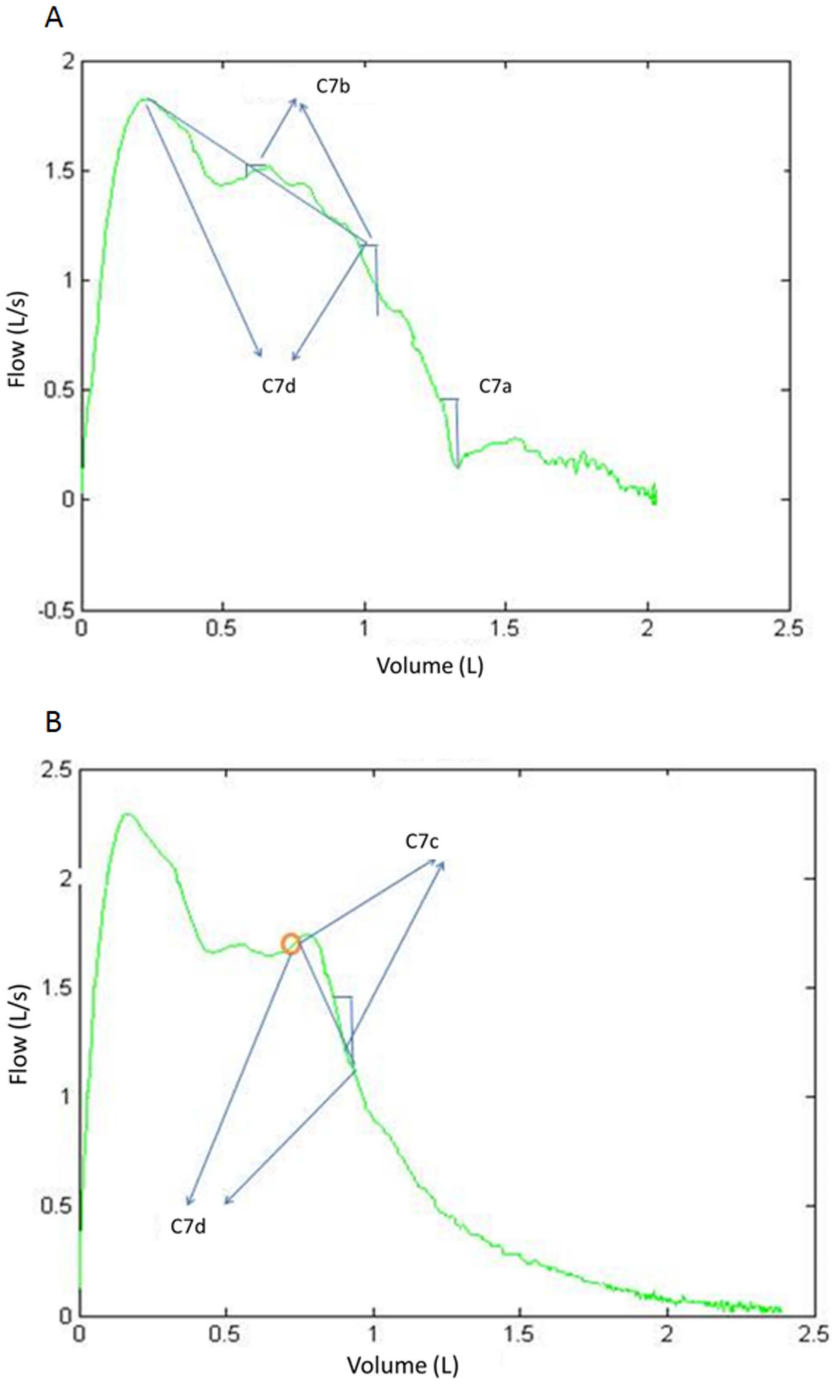

Fig. 5. Zone Z3 analysis. Examples of FV curves that present irregularity in the descent from PEF.

doi:10.1371/journal.pone.0116238.g005

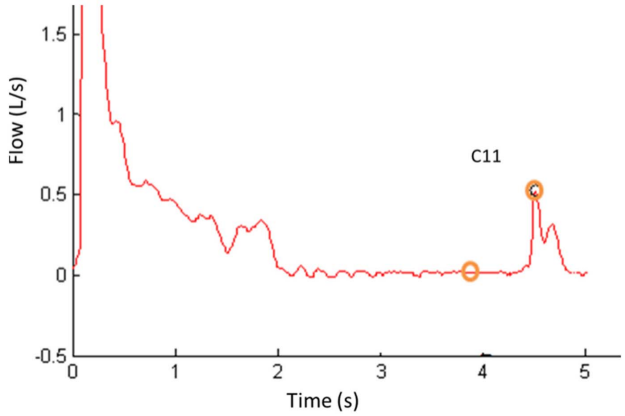

Fig. 6. Zone Z4 analysis. Example of FT curve that present irregularity in the final part.

doi:10.1371/journal.pone.0116238.g006

Results and Discussion

Table 1 summarizes the evaluation of the new algorithm showing that a reasonable percentage of FS curves (88%), could be automatically assessed as either acceptable (grade 1) or unacceptable (grade 0) in concordance to an expert evaluation. Twelve percent of the curves (n=93) were automatically classified as grade 2 requiring an expert opinion. It is of note that 43% of these grade 2 curves were evaluated as grade 0 and 57% as grade 1 by the expert evaluator. The table also compares the results of the current research against those obtained only using the four traditional ATS/ERS criteria for quality assessment. Think

Several alternative technological approaches [8–14] for automatic quality assessment of FS testing were considered during the current study design. But, we

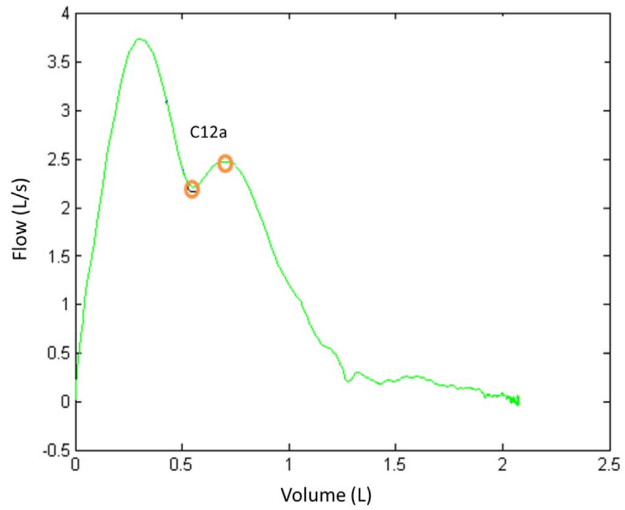

Fig. 7. Zone Z5 analysis. Examples of FV curve with peak and valley.

doi:10.1371/journal.pone.0116238.g007

Table 1. Computed sensitivity (Sen) and specificity (Spe) using the current automatic classification algorithm and using only the four traditional ATS/ERS quality criteria applied to P2.

Automatic Classification Algorithm	**Sen: 96.1%**
	Spe: 94.9%
Number of Curves Detected in each Grade	*Grade 0*: 266
	Grade 1: 419
	Grade 2: 93
Four traditional ATS/ERS Criteria	**Sen: 67.7%**
	Spe: 75.0%
Number of Curves Detected in each Grade	*Grade 0*: 320
	Grade 1: 458
TOTAL CURVES ANALYZED: 778	

doi:10.1371/journal.pone.0116238.t001

consider that the method reported allows a comprehensive and efficient application of the different quality control ATS/ERS criteria [1], it does not show limitations in terms of computer requirements, or unnecessary delays using regular computers used in the clinical setting, and it was well accepted by health professionals as on-line clinical decision support systems support during performance of FS testing.

We acknowledge as a limitation of the current study that the algorithm has been developed with the feedback of only one expert. Consequently, despite the positive results reported, there is a need for a formal assessment of the variability among various expert observers. Despite that internal interim data indicates that interobserver variability is not a relevant factor, we are planning its evaluation as part of a large prospective future trial analyzing both clinical and cost saving impact of the regional deployment of the high quality FS program, as described below.

The need for an external, likely centralized, quality control of FS testing [15–16] is widely accepted if based on well-established objective criteria. It is of note, however, that low specificity of any combination of the computer-based quality control criteria using only the four traditional ATS/ERS [3] has been reported [5, 6] such that automatic quality assessment using algorithms incorporated in commercially available equipment cannot replace the visual inspection by an expert. In contrast, our proposed algorithm shows two advantages: it enhances quality control of FS testing and allows on-line assessment of the testing.

Previous reports have indicated the potential of telemedicine to enhance both quality and diagnostic potential of FS testing carried out by non-expert professionals [5, 17–19], but the studies are based on off-line analyses by specialists [20–22]. The findings of the current research suggest that a vast majority of FS testing carried out by non-specialized professionals in primary care could be reliably assessed in real-time. Consequently, the results of the current study refine previous achievements [5] and open the way to explore extensive and efficient adoption of this type of high quality FS programs.

Several factors limiting regional deployment of a quality control program of FS using the current algorithm in the clinical practice are acknowledged, namely: (i) implementation of standardized raw spirometric data transfer through a clinical document architecture (CDA) [23]; (ii) an ICT architecture providing interoperability across healthcare tiers; (iii) design of an educational program for professionals; and, (iv) implementation of incentives fostering professional engagement. The region of Catalonia will be ready in 2015 for the regional deployment of a high quality FS program overcoming the barriers alluded to above. Such a comprehensive program: *i)* will likely have a positive clinical impact on the quality of diagnosis of patients with respiratory disorders, *ii)* should prevent unnecessary duplication of FS testing; *iii)* will likely enhance longitudinal follow-up of patients and support cost-effective preventive strategies aiming at modulating disease progress; *iv)* will pave the way to generate novel approaches to assess abnormal biological variability of FS testing; and, *v)* may likely produce cost savings. No doubt that such a program will require a proper evaluation on a longitudinal basis to assess its potential for generation of healthcare value.

Conclusion

The results of the current study provides a tool that makes operational a comprehensive application of the ATS/ERS recommendations for automatic quality control of FS testing. It constitutes a pivotal element facilitating the design and future deployment of a high quality FS program based on remote automatic evaluation of the testing.

Supporting Information

S1 File. Metrics and flow-chart. This file provides a detailed description of the metrics and the decision process used in each of the zone (Z1, Z2, Z3, Z4 and Z5) in order to automatically evaluate the FS curves.
doi:10.1371/journal.pone.0116238.s001 (PDF)

Aknowledgments

CIBER of Bioengineering, Biomaterials and Nanomedicine is an initiative of ISCIII.

Author Contributions

Conceived and designed the experiments: UM FB MV PC JR. Performed the experiments: UM FB. Analyzed the data: UM FB MV PC FV ML JR. Contributed reagents/materials/analysis tools: UM FB MV PC FV ML JR. Wrote the paper: UM FB MV PC FV ML JR.

References

1. **Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, et al.** (2005) Standardisation of spirometry. Eur Respir J. 26: 319–338.
2. **Gardner RM, Hankinson JL** (1988) Standardization of spirometry ATS update American Thoracic Society, J Occup Med March. 30 (3): 272–3.
3. Medical Section of the American Lung Association, Standardization of spirometry: 1994 update (1994), Am Rev Respir Dis. 152: 1107–1136.
4. **Burgos F, Roca J** (2013) Is Plug-and-Play Spirometry the Future? Buyers' Guide To Respiratory Care Products. 14: 11–22.
5. **Burgos F, Disdier C, de Santamaria EL, Galdiz B, Roger N, et al.** (2012) Telemedicine enhances quality of forced spirometry in primary care. Eur Respir J. 39 (6): 1313–8.
6. **Müller-Brandes U, Krämer M, Gappa G, Seitner-Sorge A, Hüls A, et al.** (2014) LUNOKID: can numeric ATS/ERS quality criteria replace visual inspection of spirometry? Eur Respir J. 43: 1347–1356.
7. **Burgos F** (2013) Impact of Information and Communication Technologies on Remote Testing: Forced Spirometry as a User Case. PhD Thesis, University of Barcelona 2013. http://hdl.handle.net/10803/284587.
8. **Sankar AB, Selvi JAV, Kumar D, Lakshmi KS** (2013) Effective enhancement of classification of respiratory states using feed forward back propagation neural networks. Sadhana, 38(3): 377–395.
9. **de Oliveira Camargo-Brunetto MA, Gonçalves AR** (2013) Diagnosing Chronic Obstructive Pulmonary Disease with Artificial Neural Networks using Health Expert Guidelines. In HEALTHINF 1: 207–214.
10. **Devi GM, Venkatesan P** (2013) Least Squares Support Vector Regression for Spirometric Forced Expiratory Volume (FEV1) Values. International Journal of Science and Technology 3: 1.
11. **Polak AG** (1998) A forward model for maximum expiration. Comp. Biol. Med. 28: 613–625.
12. **Pardaens J, van de Woestijne KP, Clément J** (1975) Simulation of regional lung emptying during slow and forced expirations. J. Appl. Physiol. 39: 191–198.
13. **Lambert RK** (1989) A new computational model for expiratory flow from nonhomogeneous human lungs. ASME Trans. J. Biomech. Eng. 111: 200–205.
14. **Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, et al.** (2012) Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. European Respiratory Journal 40(6): 1324–1343.
15. **Enright PL** (2003) How to make sure your spirometry tests are of good quality, Respir Care. 48(8): 773–776.
16. **Enright PL, Skloot GS, Cox-Ganser JM, Udasin IG, Herbert R** (2010) Quality of spirometry performed by 13,599 participants in the World Trade Center Worker and Volunteer Medical Screening Program. Respir Care 55(3): 303–309.
17. **Masa JF, Gonzalez MT, Pereira R, Mota M, Riesco JA, et al.** (2011) Validity of spirometry performed online. Eur Respir J. 37 (4): 911–918.
18. **Bellia V, Pistelli R, Catalano F, Antonelli I, Grassi V, et al.** (2000) Quality control of spirometry in the elderly, Respiratory Health in the Elderly. Am J Respir Crit Care Med. 161: 1094–100.
19. **Spirometry Monitoring Technology - SPIROLA.** NIOSH - Spirometry in Occupational Medicine. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Institute for Occupational Safety and Health. www.cdc.gov/niosh/spirometry/spirola.htlm.
20. **de Montes OM, Talamo C, Perez-Padilla R, Jardim JR, Muino A, et al.** (2008) Chronic obstructive pulmonary disease and body mass index in five Latin America cities: the PLATINO study, Respir Med. 102 (5): 642–50.
21. **Janssens W, Liu Y, Liu D, Kesten S, Tashkin DP, et al.** (2013) Quality and reproducibility of spirometry in COPD patients in a randomized trial (UPLIFT), Respir Med. 25.
22. **Perez-Padilla R, Vazquez-Garcia JC, Marquez MN, Menezes AM** (2008) Spirometry quality-control strategies in a multinational study of the prevalence of chronic obstructive pulmonary disease, Respir Care, 53 (8): 1019–26.

23. **Sala T, Rubies C, Gallego C, Muñoz P, Burgos F, et al.** (2011) Technical Requirements of Spirometers in the Strategy for Guaranteeing the Access to Quality Spirometry. Arch Bronconeumol. 47(9): 466–469.

Chapter 5

Automated spirometry quality assurance: supervised learning from multiple experts

F. Velickovski, L. Ceccaroni, R. Marti, F. Burgos, C. Gistau, X. Alsina-Restoy, and J. Roca. "Automated spirometry quality assurance: supervised learning from multiple experts". In: *Computer Methods and Programs in Biomedicine* (2016). Submitted

- Journal: Computer Methods and Programs in Biomedicine
- Impact factor: 1.862
- Quartile: 1st
- Status: Submitted

Automated spirometry quality assurance: supervised learning from multiple experts

Filip Velickovski[a,b], Luigi Ceccaroni[c], Robert Marti[b], Felip Burgos[d,e,f], Concepción Gistau[d,e,f], Xavier Alsina-Restoy[d,e,f], Josep Roca[d,e,f]

[a] *Eurecat, Barcelona, Spain*
[b] *Computer Vision and Robotics Institute, University of Girona, Girona, Spain*
[c] *1000001 Labs, Barcelona, Spain*
[d] *Respiratory Diagnostic Center, Hospital Clínic, Barcelona, Spain*
[e] *Institut d'Investigacions Biomédiques August Pi i Sunyer (IDIBAPS), Universitat de Barcelona, Barcelona, Spain*
[f] *Centro de Investigación Biomédica en Red de Enfermedades Respiratorias (CIBERES), Barcelona, Spain*

Abstract

Background and Objectives: Forced spirometry testing is gradually becoming available across different healthcare tiers including primary care. It has been demonstrated in earlier work that commercially available spirometers are not fully able to assure the quality of individual spirometry manoeuvres. Thus a need to expand the availability of high quality spirometry assessment beyond specialist pulmonary centres has arisen.

Method: In this work we propose a method to select and optimise a classifier using supervised learning techniques by learning from previously classified forced spirometry tests from a group of experts. Such a method is able to take into account the shape of the curve as an expert would during visual inspection.

Results: We evaluated the final classifier on a dataset set aside for evaluation yielding an area under the receiver operating characteristic curve of 0.88 and specificities of 0.91 and 0.86 for sensitivities of 0.60 and 0.82. Furthermore, other specificities and sensitivities along the ROC curve were close to the level of the experts when compared against each-other, and better than an earlier rules-based method assessed on the same dataset.

Email address: filip.velickovski@eurecat.org (Filip Velickovski)

Conclusion: We foresee key benefits (i) raising diagnostic quality, (ii) saving time, and (iii) reducing cost, and also (iv) improving remote care and monitoring services for patients with chronic respiratory diseases in the future if a CDSS with the encapsulated classifier is to be integrated into the work-flow of forced spirometry testing.

Keywords: spirometry, quality assurance, clinical decision support, supervised learning

---

1. Introduction

Spirometry is the measurement of airflow into and out of the lungs over a specified period of time. Generally, the forced spirometry (FS) test, for exhalation, involves the patient taking the deepest breath possible and exhaling into a spirometer device as hard as possible for as long as possible.

Forced spirometry is essential in the screening, diagnosis, monitoring and management of patients with respiratory diseases, especially chronic respiratory diseases which currently represent a high burden on healthcare systems worldwide [1, 2], this trend is expected to increase as more people are concentrated in urban environments. With this high prevalence of chronic respiratory diseases, and the creation of highly portable spirometry devices, the role of FS testing has begun shifting from exclusive specialist use to being available across different healthcare tiers, especially in primary care [3].

A common downside to FS is that it is highly dependent on patient cooperation and effort. On one hand patients need some initial training before they can complete a successful test which is reliable. They therefore normally repeat the manoeuvre at least three times to ensure repeatability and to follow international standardization [4]. On the other hand, a substantial number of patients do not provide their best performance, resulting in biased results affecting the physician's diagnosis. Poorly performed tests may have little value and may even provide misleading information. For example, in the UK a recent study has concluded that over 25% of people with a diagnostic label of COPD have

been wrongly diagnosed due to poorly performed FS [5]. Furthermore, diagnosing without the proper use of FS is unreliable, and it has been shown to likely miss up to 50% of cases [6]

We have demonstrated previously that it is possible to define expert rules, operating on well selected features of the spirogram signal to automatically assess the quality of the measurement in most spirometry manoeuvres at a level near to a lung function expert [7, 8]. In this paper we present a machine learning approach capable of learning quality assessment from previously classified FS tests from a group of experts, furthermore the principles in this approach can be generalised to other biosignals.

2. Background

2.1. Spirogram

A spirogram is the plot that is generated by a spirometer that measures the flow of air in continuous intervals produced by a subject expiring completely all the air in their lungs as fast as possible. Before commencing the procedure the subject is required to inspire as much air as possible. The spirogram is typically presented to the health professional in two views: volume-time, and flow-volume as shown in Figure 1. To assess lung function a clinician uses the pattern of the spirogram, and parameters extracted from the spirogram such as the forced vital capacity (FVC), the forced expiratory volume in one second (FEV_1), the peak expiratory flow (PEF) and the ratio of FEV_1 to FVC (FEV_1/FVC ratio).

2.2. Quality Control

Spirometry quality control includes examination of parameter values and evaluation of both the volume-time and flow-volume curves of the spirogram for evidence of technical errors. When erroneous curves are produced and detected, additional manoeuvres are often needed. During testing, technicians should attempt to record a valid test, which is composed of at least 3 *acceptable* manoeuvres with consistent *repeatable* results ($< 0.15L$ in difference) for

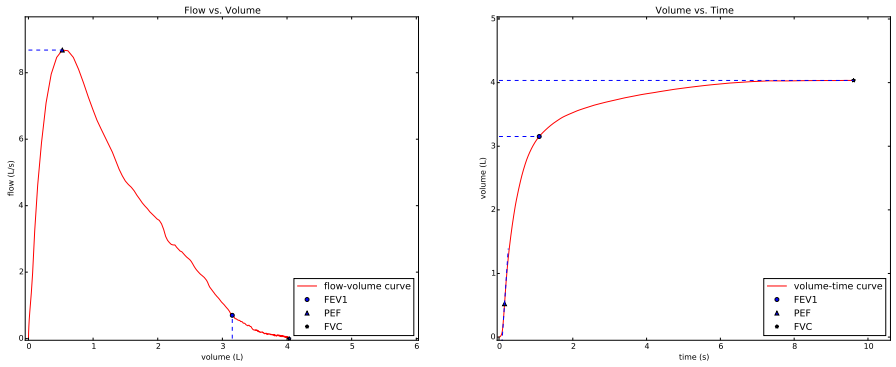

Figure 1: Flow-volume and volume-time curves of a spirogram, with key parameters FEV_1, FVC, and PEF marked in both curves.

both FVC and FEV_1 in at least 2 acceptable manoeuvres. To further illustrate this issue, Figure 2 depicts one normal flow-volume curve labelled (a) and four curves (b-e) illustrating some of the many possible errors that appear as artefacts in the spirogram from poorly executed spirometry tests. Curves similar to (b) suggest a cough during the execution of a manoeuvre, (c) an early ending, (d) a sub-maximal effort and (e) a hesitation at the start. The formal criteria for assessing the quality of forced spirometry test have been established by the American Thoracic Society / European Respiratory Society (ATS/ERS) [4].

2.3. Clinical Decision Support

Although a simple calculation suffices to check for repeatability and is available in most spirometers, the quality checks for an individual manoeuvre in currently available spirometers generates poor outcomes due to an inadequate application of the ATS/ERS recommendations on quality control [9]. Furthermore it has been demonstrated that numeric ATS/ERS criteria used in spirometers cannot replace visual inspections [10]. Thus, evaluating that a FS manoeuvre is acceptable requires highly specialist training that is often not available in non specialist settings such as primary care.

We propose embedding a clinical decision support system (CDSS) in the current work-flow of a FS test, that encapsulates a classifier trained with su-

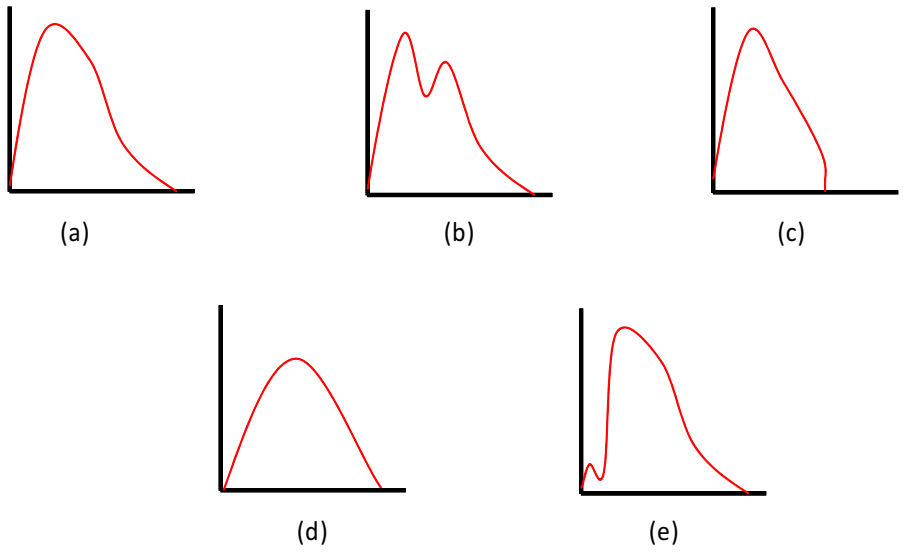

Figure 2: Depictions of normal (a) and problematic (b-e) flow-volume curves.

pervised learning methods. Technically the CDSS may be embedded into the device, or available as a centralised decision support web service to multiple institutions and patients at home as illustrated in Figure 3. Thus during FS testing, a CDSS may advise a health professional whether the subject has completed a successful manoeuvre or the test should be repeated. Furthermore with sufficient training, subjects may be able to perform spirometry testing in their own home in a remote care or monitoring health scenario [11].

3. Methods

3.1. Dataset preparation

A total of 900 spirograms representing individual manoeuvres were used from 300 spirometry tests taken from primary care centres participating in forced spirometry training in a web-based remote support program to enhance quality of forced spirometry done by non-expert professionals in the Basque Country region of Spain [9]. Forced spirometry testing was conducted using a *Sibel 120 (SIBELGroup, Barcelona, Spain)* device. The inclusion criteria were:

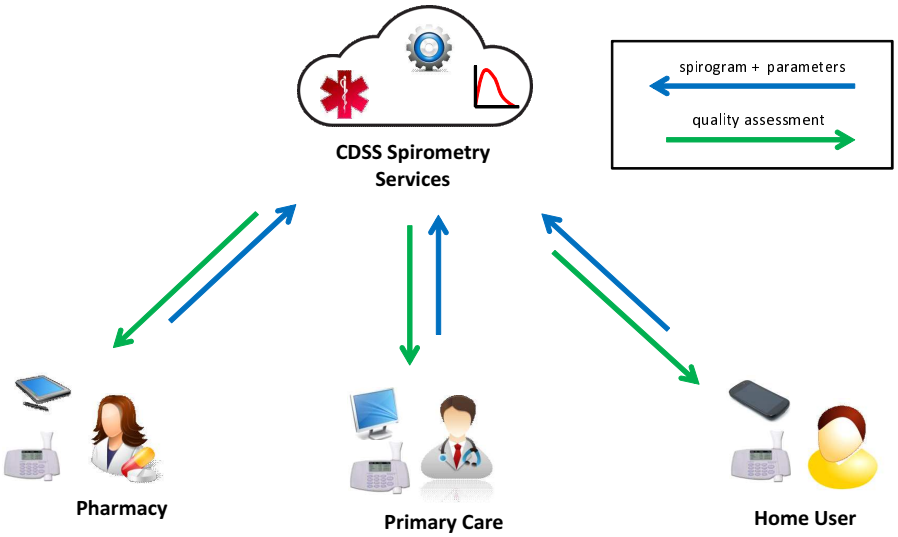

Figure 3: A clinical decision support system deployed in a secure cloud providing quality assessment services in multiple scenarios.

1. the subject was over 20
2. FS taken and recorded as an electronic record before the application of bronchodilators;
3. three manoeuvres were performed by each subject
4. the operating frequency of the spirometer was 100Hz

The mean age of the subjects were 51.83 ± 17.07 (years), the mean FEV_1, FVC, PEF, and FEV_1/FVC ratio were 2.59 ± 0.96 (litres), 3.58 ± 1.13 (litres), 6.15 ± 2.12 (litres/second), and 0.72 ± 0.11 respectively.

3.2. Expert evaluation

The spirograms were shuffled and divided into three equal sets (S1, S2, S3) containing 300 spirograms per set, and were evaluated by 3 clinical experts (E1, E2, E3) of the Lung Function Unit at the Hospital Clinic of Barcelona, Spain. The evaluation was performed using a specifically designed web application *Spirometry evaluation collector* shown in Figure 4. The application back-end

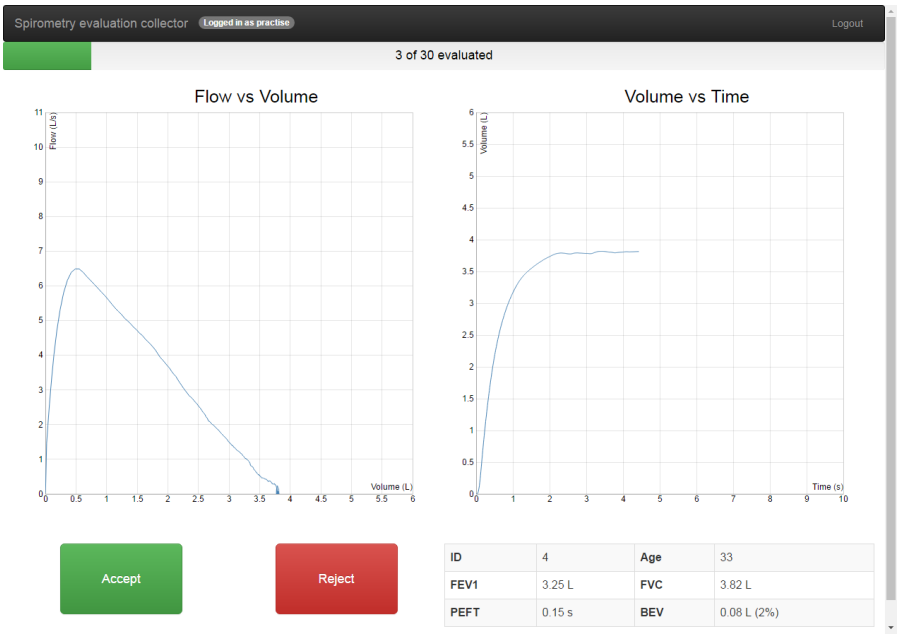

Figure 4: Spirometry evaluation collector: the spirogram quality acquisition tool.

was developed in Python, using the `Flask` web framework, and the front-end with `HTML5`, `javascript`, and the `Bootstrap` framework. The *Spirometry evaluation collector* visualises the spirograms in a high resolution, and allows the curves to be panned and zoomed easily. The evaluation strategy was designed as follows to ensure that each spirogram was evaluated by two experts, and each expert evaluated 600 spirograms:

- E1 evaluated sets S1 and S2 (600 spirograms)
- E2 evaluated sets S2 and S3 (600 spirograms)
- E3 evaluated sets S3 and S1 (600 spirograms)

The expert was asked to evaluate the quality of each spirogram as either *accepted* or *rejected*. The web application collected and stored the result along with the associated spirogram reference in a database.

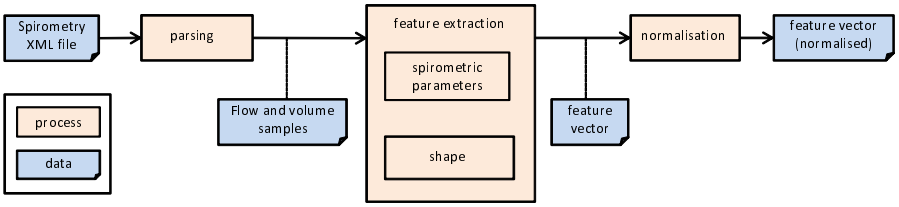

Figure 5: Processing pipeline showing the steps (i) parsing (ii) feature extraction consisting of (a) shape and (b) spirometric parameter extraction (iii) normalisation for converting an electronic XML spirometry file into a feature vector for suitable for supervised learning

3.3. Establishing the ground-truth

The spirograms were labelled by setting the target variable

- 1 (indicating acceptable) when two experts both evaluated the spirogram as acceptable
- 0 (indicating rejected) when two experts both evaluated the spirogram as rejected

The inter-observer agreement was analysed, and later used to benchmark performance of the classifiers. When the experts disagreed, the spirogram was not used for training the classifier nor validation. This resulted in a final dataset D_{full} consisting of 603 spirograms along with corresponding target variables.

3.4. Processing pipeline

A processing pipeline (parsing, feature extraction, normalisation) as shown in Figure 5 was implemented in Python utilising the scientific computing libraries `scipy` and `numpy` for converting the spirograms into a vector of normalised features suitable to be used in supervised learning techniques.

3.4.1. Parsing

The raw signal of the spirogram was represented as quantized flow-time samples is stored as a custom *SIBELGroup* `XML` format. The `XML` file was parsed and the flow samples were extracted and converted into litres per second. The

flow-time curve was converted to volume-time by integration, then both the flow-time and volume-time were combined to create the samples for the flow-volume curve. All samples representing flow were stored in litres per second, and volume in litres.

3.4.2. Feature extraction

The features used in the training of the classifier were formed from two main sources.

Spirometric parameters. Two distinct parameters already existing in spirometry quality control were extracted from the volume-time curve: (i) back extrapolated volume (BEV) which is the volume at the time at which the volume curve tangent with maximum slope crosses the horizontal time axis, (ii) forced expiratory time (FET) which measures the length of the expiration in seconds [4].

Shape extraction. Coefficients representing the shape of the spirogram were calculated as follows:

1. The spirometry signal (flow-time samples) were divided into two sections:
   - *ascent section*: from the beginning of expiration to the PEF point.
   - *descent section*: from the PEF point to the end of expiration.
2. Ascent coefficients were calculated by fitting a polynomial of degree a to the ascent section.
$$f(t) = c_0 + c_1 t + ... + c_a t^a \tag{1}$$
3. The descent section was re-sampled to 100 samples in order to discount variations in signal length.
4. Descent coefficients were calculated by fitting a polynomial of degree b to the 100 samples of the previous step.
$$f(t) = c_{a+1} + c_{a+2} t + ... + c_{a+b+1} t^b \tag{2}$$

5. Additionally the natural log transform of the descent samples was also evaluated to see if it would yield improved performance.

$$\ln f(t) = c_{a+1} + c_{a+2}t + ... + c_{a+b+1}t^b \tag{3}$$

The best degrees a and b were selected experimentally, by running 3-fold cross-validation with a logistic regression classifier on the cross-validation dataset D_{cv} that is defined in section 3.6. For the descent curve, performance using both equations 2 and 3 were compared.

It should be noted that flow-volume and volume-time curves are used for human interpretation, however both are derived from the raw signal (flow-time samples). We decided to characterise the shape from the flow-time samples directly as the other curves are implicitly encapsulated within it.

Feature vector. Finally the $a + 1$ ascent coefficients, $b + 1$ descent coefficients, and the 2 spirometric parameters (BEV, FET) were concatenated to form a feature vector of size $(a + 1) + (b + 1) + 2$.

3.4.3. Normalisation

Since the range of values of each feature had the potential to vary in relation to the other features, a standardization procedure was applied by subtracting the mean and dividing by the standard deviation of the values of each feature column.

3.5. Performance metrics

Due to the significant imbalance of classes (fraction of positive cases or acceptable spirograms to negative cases or rejected spirograms in D_{full}) we determined that accuracy (proportion of correctly classified cases) was not a suitable measure for assessing performance for neither the classifiers nor experts. To overcome this imbalance, the metrics sensitivity and specificity, separately measures the proportion of positive and negative cases that are correctly classified. Sensitivity and specificity are defined in equations 4 and 5 respectively where TP (true positives) is the number of correctly classified positive cases, TN (true

negatives) is the number of correctly classified negative cases, FP (false positive) is the number of cases incorrectly classified as positive, and FN (false negative) is the number of cases incorrectly classified as negative.

$$sensitivity = \frac{TN}{TN + FP} \tag{4}$$

$$specificity = \frac{TP}{TP + FN} \tag{5}$$

The probabilistic nature of the classifiers produced by supervised learning methods allows for manipulating the threshold at which the classifier determines the acceptance or rejection of a spirogram. Thus a receiver operating characteristic (ROC) curve can be defined by plotting the false positive rate (1 - specificity) against the sensitivity or true positive rate for all cut-off values or thresholds. The area under the ROC curve (AUC) of perfect classification is 1.0 and that of a useless or random classifier is 0.5. In medical tests the AUC can be thought of as a measure for the overall accuracy of a diagnostic procedure [12, 13]. Thus for comparing and determining the performance of the classifiers produced by the supervised learning methods, the AUC was used.

3.6. Model selection and evaluation

A random 30% (181) of spirograms in D_{full} were put aside forming the final evaluation set D_{test} to be used exclusively for reporting the final performance against the best classifier selected. The remaining 70% (422) forming set D_{cv} used for model selection and hyper-parameter optimization through the application of k-fold cross-validation where D_{cv} is split into a training and validation set.

Using the Python `scikit-learn` machine learning library, multiple classification algorithms were evaluated in order to search for the best classifier, using set D_{cv}, with the performance metric AUC. The following supervised learning methods and hyper-parameter settings were evaluated:

- Gaussian **Naive Bayes** [14] without any hyper-parameters

- **K Nearest Neighbours** (kNN) [15] with 12 hyper-parameter settings
  - the number of neighbours k : 3-8 (6 settings)
  - weighting strategy : uniform or by distance (2 settings)
- **Logistic Regression** [16] with 200 hyper-parameter settings
  - regularisation penalty : L1 or L2 (2 settings)
  - regularisation coefficient C : 10^{-4} to 10^{10} (100 settings)
- **Support Vector Machine** (SVM) [17] with 200 hyper-parameter settings
  - kernel type : linear or radial basis function (2 settings)
  - regularisation coefficient C : 10^{-2} to 10^{10} (100 settings)
- **Random Forest** [18] of decision trees with 14 hyper-parameter settings
  - number of decision trees : 50 to 10000 (7 settings)
  - decision tree splitting criterion : gini or entropy (2 settings)

Each binary classifier and hyper-parameter setting was trained and performance measured with 3-fold cross-validation of D_{cv} and the mean AUC was recorded. The best classifier model was chosen to be the one with the highest mean AUC. The best classifier and hyper-parameters setting was retrained with the full D_{cv} set, and performance tested on the dataset that was put aside for final evaluation D_{test}.

4. Results

4.1. Inter-observer agreement

Table 1, Table 2, and Table 3 correspond to confusion matrices of E1 vs. E2, E1 vs. E3, E2 vs. E3 quality classification of the 300 spirograms each expert had in common. Table 4 is an analysis of the inter-observer agreement, apart from calculating the agreement rate and kappa value, sensitivity and specificity was

Table 1: Expert 1 vs. Expert 2 confusion matrix

		Expert 2	
		Rejected	Accepted
Expert 1	Rejected	34	33
	Accepted	10	223

Table 2: Expert 1 vs. Expert 3 confusion matrix

		Expert 3	
		Rejected	Accepted
Expert 1	Rejected	60	4
	Accepted	140	96

also reported to be used as a means of comparison to the automatic classifier. The sensitivity, and specificity was calculated by taking the classifications of one expert along the row as true values, and other expert along the column as the predicted values. It should be noted that kappa and agreement rate are symmetric measures hence E1 vs. E2 is the same as E2 vs. E1, however sensitivity and specificity are not.

4.2. Curve fitting

The polynomial degrees that yielded the best performance were $a = 2$ for the ascent curve, and $b = 3$ fitted for the logarithm transformed descent curve. Thus the fitting equations

$$f(t) = c_0 + c_1 t + c_2 t^2 \tag{6}$$

Table 3: Expert 2 vs. Expert 3 confusion matrix

		Expert 3	
		Rejected	Accepted
Expert 2	Rejected	57	1
	Accepted	109	133

Table 4: Inter-observer agreement

	Expert 1	Expert 2	Expert 3
Expert 1		agreement: 0.86 kappa: 0.53 sensitivity: 0.96 specificity: 0.51	agreement: 0.52 kappa: 0.19 sensitivity: 0.41 specificity: 0.94
Expert 2	agreement: 0.86 kappa: 0.53 sensitivity: 0.87 specificity: 0.77		agreement: 0.63 kappa: 0.31 sensitivity: 0.55 specificity: 0.98
Expert 3	agreement: 0.52 kappa: 0.19 sensitivity: 0.96 specificity: 0.30	agreement:0.63 kappa: 0.31 sensitivity: 0.99 specificity: 0.34	

$$\ln f(t) = c_3 + c_4 t + c_5 t^2 + c_6 t^3 \tag{7}$$

yielded a feature vector of size 9 comprised of 7 fitting coefficients (c_0 to c_6), and 2 spirometric parameters (BEV and FET described earlier in section 3.4.2). The mean error of the fit between the samples and the reconstructed polynomial curve, was 5.25 ± 2.28 and 0.24 ± 0.14 for the descent curve.

4.3. Performance of the classifier models

Of the 427 classifiers trained and tested (one per type per hyper-parameter setting) with the cross validation set D_{cv} of 422 examples, the best of each supervised learning algorithm is reported in Table 5. The mean AUC scores ranged from 0.82 - 0.88 with the best classifier being the Random Forest method achieving a mean AUC of 0.88 $\pm$ 0.02, formed from 5000 decision trees, using the *entropy* metric to measure the quality of the splits. Figure 6 contains a plot the mean ROC curves (formed from the 3 folds in the cross-validation process) the best classifier of each learning method. The Random Forest ROC curve encompasses the others hence having the largest AUC.

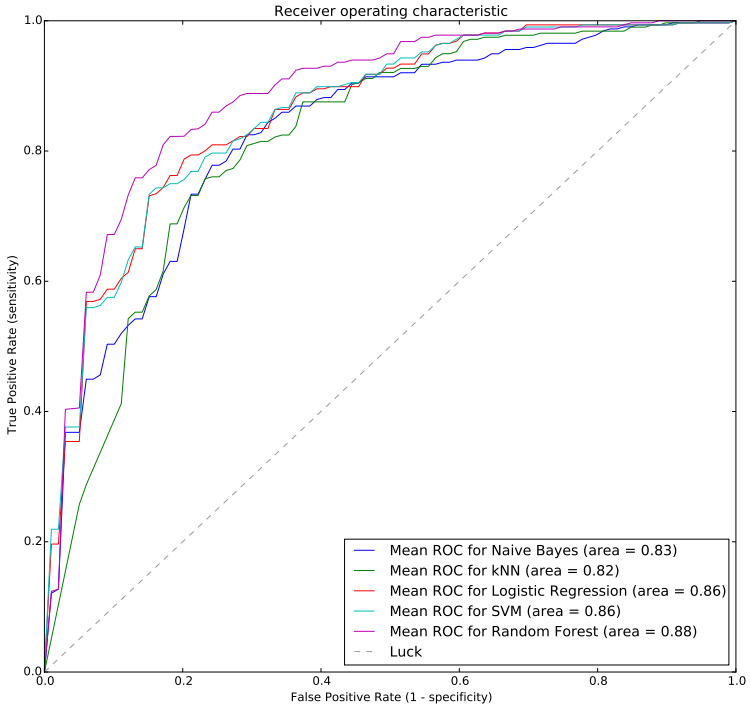

Figure 6: ROCs of the best classifier of each supervised learning method using the best hyper-parameter setting, trained and evaluated on D_{cv} using 3-folds.

Table 5: Performance of the all classifiers configured with best hyper-parameter setting

Classifier	**Hyper-parameter setting**	**Mean area under ROC**
Naive Bayes	N/A	0.83 ± 0.03
kNN	weights : distance no. of neighbours : 8	0.82 ± 0.02
Logistic Regression	C : 0.91 penalty : L2	0.86 ± 0.02
SVM	kernel : linear C : 0.04	0.86 ± 0.02
Random Forest	no. of trees : 5000 criterion : entropy	0.88 ± 0.02

4.4. Final evaluation

Figure 7 shows the plot of the ROC curve produced by retraining the best performing classifier (Random Forest: 5000 trees, entropy criterion) of the cross-validation testing on the full D_{cv} set and classifying (assigning probabilities of acceptance to) the 181 examples of the unseen dataset D_{test}. The AUC achieved was 0.88 matching the mean performance in the cross-validation testing. Furthermore, the specificity and sensitivity values of the expert evaluators of Table 4 are marked in Figure 7 as benchmarking points.

Additionally, the previous rules-based method for automated spirometry quality control published in 2014 [7, 8] (*SpiroQC 2014*) was evaluated on the same dataset D_{test}. It was not possible to generate a ROC curve, and only specificity and sensitivity could be obtained because the method could only classify rather than give a probability of acceptance. The specificity and sensitivity point of *SpiroQC 2014* (0.84 and 0.65 respectively) is marked also in Figure 7. Thus to compare the two methods the same false positive rate (1 - sensitivity = 1 - 0.84 = 0.16) must be used. For the false positive rate of 0.16, the new method presented in this work yielded a sensitivity of 0.85 (the sensitivity on the ROC curve at a false positive rate of 0.16) beating *SpiroQC*

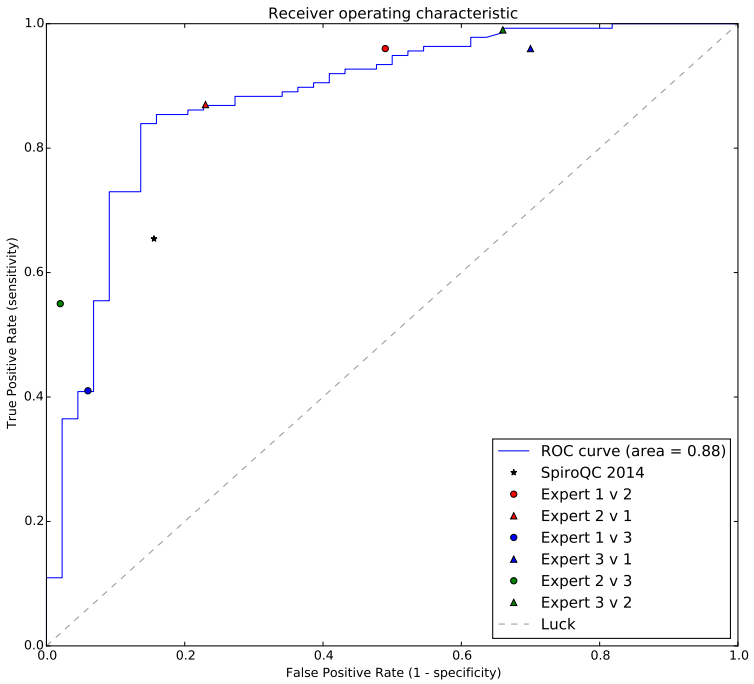

Figure 7: ROC of best classifier (Random Forest: 5000 trees, entropy criterion) evaluated on final test set D_{test}, sensitivity and sensitivity of experts, and previous quality control method *SpiroQC 2014*.

2014 by a statistically significant margin of 0.2 ($p < 0.0001$).

5. Discussion

In this work, we presented a technique to build a system for FS quality decision support using knowledge extracted from manual ratings of multiple clinical lung function experts by the application of supervised learning. Once the final classifier embedded in the system is trained, the processing time to classify a single spirogram is negligible ($< 1s$). Therefore, such a system would be suitable to be used in both on-line (providing instantaneous feedback) or

off-line settings.

For selection, training and evaluation of the supervised learning methods, three experts (E1, E2, E3) rated a total of 900 spirograms for acceptability, with each spirogram being rated by 2 different experts. As shown in Table 4, while E1 and E2 showed a good degree of inter-observer reliability ($agreement = 0.86$, $\kappa = 0.53$), E3 however, did not seem to coincide closely with the ratings of E1 ($agreement = 0.52$, $\kappa = 0.19$) nor E2 ($agreement = 0.63$, $\kappa = 0.31$). Upon closer analysis, however we observed that E3 is in fact a consistently stricter rater, and the disagreement are almost exclusively from positively (acceptable) classified spirograms of the other observers. With the negative cases there is a high level of agreement between E3 and the other experts. As shown in Table 2, out of the 64 negatively (rejected) classified spirograms by E1, E3 only disagreed in 4 cases, and similarly in Table 3, out of the 58 negatively classified cases by E2, E3 only disagreed with 1 case. The level of experience of E1 and E2 (40 years) was significantly higher than E3 (8 years) which further explained this discrepancy.

To establish the dataset D_{full} out of which the supervised learning would be conducted, spirograms where the two experts disagreed were excluded (removing 297 cases), this was done to be confident that the target variables of the ground truth data were labelled correctly. This could be seen as a limitation, because perhaps the more difficult cases to classify may fall in this group. However for these cases, it would be hard to establish the correct label without a rating from a third observer which was not available to us after the data had already been collected.

A part from the fitting of the piecewise polynomial functions used for shape extraction, fitting with other curve functions such as Rayleigh and Rice distribution was attempted, even though results are omitted due to poor performance.

As mentioned earlier in section 3.5 due to the imbalance of classes (452 positive examples and 151 negative examples in in D_{full}) accuracy was determined not to be a good metric for performance, and instead, AUC was used for comparison of hyper-parameter settings within a supervised learning method, and

between each method. As shown in Table 5 and Figure 6 after optimizing for the best hyper-parameter setting, the margin between the best learning method (Random Forest with AUC of 0.88) and worst learning method (kNN with AUC of 0.82) only varied by 0.06. Naturally the best method along with its hyper-paramter setting was selected as the final classifier. The reason to retrain this classifier on all the examples of D_{cv} before evaluation on the final test set D_{test}, was to take advantage of the extra 33% of examples that are set aside in each fold of the 3-fold cross-validation process. Alternative that could have also been considered was to choose the best classifier from the folds.

Although the results in the model selection process are reported for 3-folds, the experiment was re-run with 10-folds, taking longer, but did not yield major differences in the AUC scores, nor in the choice of the best classifier with the best hyper-parameter setting.

Finally the performance of the best classifier could be compared against the experts' inter-observer performance, and the previous algorithm *SpiroQC 2014* evaluated on D_{test} by plotting the ROC curve of the classifier, and marking the sensitivity / specificity points of the experts and previous method, as shown in Figure 7. Points below the curve can be interpreted as performance points that are worse than the classifier, and points above the curve, better than the classifier.

SpiroQC 2014 performed lower (sensitivity of 0.65 and specificity of 0.84) than the classifier, and considerably lower than in its evaluation in the earlier study [8] (sensitivity of 0.96 and specificity of 0.95). This may due to the fact that only one expert (E1) was used to establish the rules in *SpiroQC 2014* for evaluating FS, and the same expert produced the labels for the evaluation set, thus biasing the performance. A second reason may also be that the difficulty in the cases is higher in the datasets of this work.

Additionally, Figure 7 shows that the best classifier performance is close to the level of the multiple experts, having sensitivity slightly better in the *E3 vs. E1*, equal in the *E2 vs. E1* and *E1 vs. E3*, and slightly worse in the *E3 vs E1* (at the same specificity rate).

One advantage of this approach to quality classification is that the trade-off between sensitivity and specificity is adjustable, by being able to set the decision threshold probability so that the classifiers performs at a specific operating point on the ROC curve. For choosing the operating point, one needs to take into account that during spirometry testing, to achieve a Quality A, at **least 3 acceptable manoeuvres** need to be recorded during the session out of a maximum of $N = 8$ tries [4]. This maximum exists as each manoeuvre requires a considerable amount of exertion, and exceeding this may leave a subject quite light-headed. Furthermore, even if the subject performs perfectly, a classifier with low sensitivity may not classify enough manoeuvres as acceptable and the subject's results would be marked as unreliable. By solving for the sensitivity value p in:

$$Pr(X \geq 3) = 1 - Pr(X \leq 2) = 1 - \sum_{i=0}^{2} \binom{N}{i} p^i (1-p)^{n-i} \geq 0.95 \qquad (8)$$

it can be shown that setting the sensitivity on the ROC curve to $p = 0.60$ guarantees (with a probability of 95%) that a FS test (consisting of up to $N = 8$ manoeuvres) will have at least three classified as acceptable by the classifier (assuming the subject performs them perfectly). Furthermore, with frail subjects, it may be advantageous to impose a lower maximum number of manoeuvres (e.g. $N = 5$), and using the same method we can calculate the operating point to be 0.82 to ensure that perfect subjects will not have their test disqualified due to the false negatives. The corresponding specificity rates to sensitivities of 0.60, and 0.82 are 0.91 and 0.86 respectively.

We proposed earlier in section 2.3 encapsulating this method as a CDSS offering a centralised service for quality assessment in situations without the availability of lung function experts. Apart from this scenario, the CDSS could also be used as a training tool for patients and health professionals across all tiers (GPs, nurses, pharmacists) to improve the quality of FS results while at the same saving time, and reducing cost due to the reduced need of returning to repeat a erroneous test. Moreover, we would expect a reduction of misdiagnosis based on poorly performed FS tests.

Finally, the principles in this approach can be generalised to other biosignal applications beyond FS where there is a need to first ensure the biosignal is of an acceptable quality before interpretation can occur. Electrical biosignals where such a method may be useful include electrogardiograms, electromyographs, and galvanic skin response signals.

We acknowledge several limitation of this work. Firstly, the need for further feedback for failed manoeuvres will help patients and health care professionals in the next attempt. Thus in the future we plan to improve the *Spirometry evaluation collector* to capture the reasons for rejecting a spirogram. This will allow the formation of dataset for training of a learning method to label the rejected spirograms. Secondly, as mentioned previously, we removed 297 cases where the lung function experts did not agree in the quality assessment. We acknowledge that potentially difficult cases may fall within this set, and a re-evaluation of the method needs to be performed once a final label has been assigned to these in the future. Finally, only one type of spirometer model was used for this study, and more analysis needs to be done in the future on a broader range of equipment and its affect on assessing the quality of manoeuvres.

6. Conclusion

The incorporation of automated quality assurance through a CDSS integrated into the work-flow of primary care, remote care, and other healthcare tiers has the potential to raise the standard of FS tests thus reduce errors in their use in diagnosis and assessment. We have presented in this work an approach to generating a classifier of FS test quality derived from machine learning techniques, capable of being trained by data collected from a group of experts that have classified previously performed FS tests. The results (AUC = 0.88 and specificities of 0.91 and 0.86 for sensitivities of 0.60 and 0.82) indicate a credible potential to determine acceptable manoeuvres from ones with poor quality, with performance rates near the level of experts. Further investigation will be needed, and potentially an initial pilot trial before plans for large scale

deployment can commence.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgement

Part of the research leading to these results has received support from the following sources: Synergy-COPD (EU Grant FP7-ICT-270086); NEXES - Supporting Healthier and Independent Living for Chronic Patients and Elderly (EU Grant CIP-ICT-PSP-2007-225025); Inforegió (AGAUR 2008 FIS PI09/90634); Grant from the Spanish Government and, Catalan Master Plan of Respiratory Diseases (PDMAR); Grant from the University of Girona (MPCUdG 2016); Grant from the Spanish Ministry of Economics and Competitiveness (DPI2015-68442-R). The funders had no role in the design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] A. Alwan, et al., Global status report on noncommunicable diseases 2010., World Health Organization, 2011.

[2] T. Ferkol, D. Schraufnagel, The global burden of respiratory disease, Annals of the American Thoracic Society 11 (3) (2014) 404–406.

[3] E. Derom, C. Van Weel, G. Liistro, J. Buffels, T. Schermer, E. Lammers, E. Wouters, M. Decramer, Primary care spirometry, European Respiratory Journal 31 (1) (2008) 197–203.

[4] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, J. Wanger, Standardisation of spirometry, European respiratory journal 26 (2) (2005) 319–338.

[5] K. Holton, S. Hill, M. Kearney, R. Winter, A. Moger, Quality assured diagnostic spirometry (qads) performance and competence, Am J Respir Crit Care Med 187 (2013) 2836.

[6] J. A. Walters, E. H. Walters, M. Nelson, A. Robinson, J. Scott, P. Turner, R. Wood-Baker, Factors associated with misdiagnosis of COPD in primary care, Prim Care Respir J 20 (4) (2011) 396–402.

[7] F. Burgos, U. Melia, M. Vallverdú, F. Velickovski, M. Lluch-Ariet, P. Caminal, J. Roca, Clinical decision support system to enhance quality control of spirometry using information and communication technologies, JMIR medical informatics 2 (2) (2014) e29.

[8] U. Melia, F. Burgos, M. Vallverdú, F. Velickovski, M. Lluch-Ariet, J. Roca, P. Caminal, Algorithm for automatic forced spirometry quality assessment: Technological developments, PloS one 9 (12) (2014) e116238.

[9] F. Burgos, C. Disdier, E. L. de Santamaria, B. Galdiz, N. Roger, M. L. Rivera, R. Hervàs, E. Durán-Tauleria, J. Garcia-Aymerich, J. Roca, Telemedicine enhances quality of forced spirometry in primary care, European Respiratory Journal 39 (6) (2012) 1313–1318.

[10] C. Müller-Brandes, U. Krämer, M. Gappa, G. Seitner-Sorge, A. Hüls, A. von Berg, B. Hoffmann, A. Schuster, S. Illi, M. Wisbauer, et al., LUNOKID: can numerical American Thoracic Society/European Respiratory Society quality criteria replace visual inspection of spirometry?, European Respiratory Journal 43 (5) (2014) 1347–1356.

[11] N. A. Jarad, Z. M. Sund, Telemonitoring in chronic obstructive airway disease and adult patients with cystic fibrosis, Journal of telemedicine and telecare 17 (3) (2011) 127–132.

[12] C. E. Metz, Basic principles of ROC analysis, in: Seminars in nuclear medicine, Vol. 8, Elsevier, 1978, pp. 283–298.

[13] J. A. Hanley, B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (ROC) curve, Radiology 143 (1) (1982) 29–36.

[14] G. H. John, P. Langley, Estimating continuous distributions in Bayesian classifiers, in: Proceedings of the Eleventh conference on Uncertainty in artificial intelligence, Morgan Kaufmann Publishers Inc., 1995, pp. 338–345.

[15] S. A. Dudani, The distance-weighted k-nearest-neighbor rule, Systems, Man and Cybernetics, IEEE Transactions on (4) (1976) 325–327.

[16] D. W. Hosmer Jr, S. Lemeshow, Applied logistic regression, John Wiley & Sons, 2004.

[17] V. Vapnik, The nature of statistical learning theory, Springer Science & Business Media, 2013.

[18] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

Chapter 6

Discussion

The work presented in this thesis has adapted, refined and contributed to the state-of-the art methodologies in applied clinical decision support research, through a concrete and practical use case of chronic obstructive pulmonary disease (COPD) at the early stage, with the outlook of generalising these methods to a broader set of chronic respiratory diseases, and other non-communicable diseases (NCDs). In this chapter we summarise the key results, contributions and limitations of the work.

6.1 Key results and contributions

The targeting of decision-support to early COPD is of particular importance for the following reasons. Firstly, we have learnt from a large epidemiological report on prevalence and burden of respiratory disorders carried out in the general population of Catalonia [80] three main facts: (i) prevalence of COPD in elderly people ($>$ 65 yrs.), 36% in men and 22% women (ii) elevated percentage of under-diagnoses (76%) (iii) the disease burden is explained by a relatively small proportion of severe COPD cases. Secondly, in the UK, a study found 10% of emergency COPD admissions are in people whose COPD has not previously been diagnosed [81]. Finally, there is evidence from recent studies [82, 83] that the rate of decline in lung function is faster in the earlier stages of COPD. The potential for altering the course of the disease and improving outcomes may therefore be greater in the earlier stages.

In Chapter 3 we outlined our framework for constructing a clinical decision support system (CDSS) for early stage COPD which directly addressed and thus contributed to three of the key *grand challenges* as initially outlined by *Sittig et al.* [84] and further reinforced by *Fox et al.* [85]

(i) disseminate best practices in CDSS design, development and implementation;

(ii) create an architecture for sharing executable CDSS modules and services;

(iii) create internet-accessible clinical decision support repositories.

The architecture presented this chapter was inspired by a service-oriented approach that achieved interoperability with external health information system (HIS) by modelling the patient specific data through an extension of the Health Level Seven International (HL7) Virtual Medical Record (VMR) [86] a model for representing the data that are analysed and/or produced by CDSS.

The VMR was extended by

(i) allowing the representation of bio-signals from medical devices such as spirometer;

(ii) extending therapy recommendations to allow for multiple option sets;

(iii) representation of disease risk.

We were able to demonstrate the flexibility of the adopted architectural model by integrating into two existing external HISs

(i) *Linkcare* an integrated-care open platform allowing healthcare professionals (specialists, general practitioners, case managers, nurses, etc.) to share clinical knowledge around a patient centric model. [87]

(ii) *Arezzo Pathways* combines best practice clinical guidelines with individual patient data to dynamically generate care pathways and provide decision recommendations specific to each patient at the point of care, utilised by a significant percentage of the primary care centres in the UK. [88]

By means of an inference engine using clinical knowledge formalised as JBoss Drools rules [89], the CDSS presented in Chapter 3 provided a suite of decision support services including diagnosis support which was validated against a dataset of 323 cases evaluated by a specialist respiratory clinician via a remote care application [90]. The results implied the CDSS is able to issue diagnosis recommendation with a high degree of accuracy, correctly classifying 92% of the cases, with sensitivity, and specificity of 90% and 96% respectively.

Next in Chapter 4 we focused on decision support algorithms for forced spirometry quality that can be embedded into the framework developed in the previous chapter. Forced spirometry is essential in the screening, diagnosis, monitoring and management of patients of COPD and other respiratory diseases. For example, in the UK a recent study has concluded that over 25% of people with a diagnostic label of COPD have been wrongly diagnosed due to poorly performed spirometry [81]. Furthermore, diagnosing

without the proper use of spirometry is unreliable, and it has been shown to likely miss up to 50% of cases [91].

We showed that just following criteria set [92] by the American Thorax Society (ATS) / European Respiratory Society‘ (ERS) applied from standard parameters or metrics derived from a spirometry curve can miss-classify the quality of over 28% of curves with specificity and sensitivities 0.75 and 0.68. This contributes further to the growing evidence that individual manoeuvre quality assessment in currently available spirometers generates poor outcomes due to an inadequate application of the ATS / ERS recommendations on quality control [90, 93]. We addressed this inadequacy, via an iterative process of manually acquiring specialised knowledge from one lung function assessment expert, and converting it to features and metrics to be extracted from the curve. This method generated 23 novel metrics developed by signal analysis techniques on the volume-time, flow-volume, and flow-time curves from a spirometry manoeuvre. Additionally we introduced 28 new rules that are applied to the new metrics to develop spirometry quality assessment method that can

(i) distinguish between acceptable and poor spirometry;

(ii) detect difficult cases to be deferred to a human expert;

The method demonstrated superior performance on quality assessment achieving a specificity and sensitivity of 95% and 96% (on the cases that are not-deferred) on a separate test set formed from the expert's evaluations. These results have demonstrated that automatic spirometry control can contribute to the successful transfer of high quality lung assessment in non-specialised clinical settings.

Finally, in Chapter 5 we addressed some limitations of the work that was presented in the previous chapter, namely the reliance on evaluation of the method only from a single expert. We noted that the evaluations of the two experts (E1 and E2) with a high level of experience (40 years) coincided well ($agreement = 0.86$, $\kappa = 0.53$). However the third expert (E3) with less experience (8 years) had a higher level of disagreement with the more experienced experts ($agreement = 0.52, 0.63$, $\kappa = 0.19, 0.31$). This was because E3 tended to have a much stricter criteria classifying many more negative cases than the other experts.

When evaluated against a dataset formed from multiple experts the algorithm from Chapter 4 showed reduction in performance comparing to the original evaluation. Thus in this chapter we presented an alternative technique to build a system for quality decision support using knowledge automatically extracted from the ratings of multiple clinical lung function experts, inspired by the novel application of supervised learning methods, and potentially generalisable to other bio-signals apart from spirometry manoeuvres. The

resulting method performed achieved an area under the receiver operating characteristic curve (AUC) of 0.88, and better than the previous method by a statistically significant margin of 0.2 in its sensitivity at the same specificity level. Additionally, it had an advantage over the previous algorithm in that the trade-off between sensitivity and specificity could be adjusted, by simply modifying one the decision threshold probability parameters. This allowed the classifier to perform at a specific operating point on the receiver operating characteristic (ROC) curve which we argue could be advantageous when the cost of a false-negative implies repeating a manoeuvre, which is particularly problematic in frail elderly patients (the typical COPD case).

6.2 Limitations

We acknowledge several principle limitations in the studies presented in the thesis.

The patient databases in Chapters 3, 4 and 5 were from persons that had visited primary care centers and were candidates for spirometry testing. This would have included patients with varying degrees of health thus capturing a diverse but realistic sample of the population. Healthy controls were unavailable but would have been excluded for further examination (through forced spirometry testing) at the case finding stage, as they would not exhibit any respiratory symptoms.

The evaluations of the diagnostic capability and quality assessment capabilities of the decision-support classification methods only used data evaluated by a limited number of experts (up to three) due to access constraints to specialist clinical staff. Ideally further independent validation is required, involving a panel of a significantly higher number of experts to have achieve a the necessary level of confidence in the system.

Next the outcomes of this research have produced a CDSS to support early COPD assessment that could be used by multiple HIS distributed across countries, however we acknowledge that clinical guidelines in diagnosis, assessment, and treatment will differ across national borders that suit the specific attributes to the population. This however could be addressed by having multiple instances of the CDSS deployed within regions catering for the specific medical policy, or screening protocol only by modifying the rules and not the design.

Finally, we recognise that evaluation of the classification accuracies and performance capabilities is not enough, the impact of a trial deployment at pilot sites in actual healthcare settings needs to be assessed separately before large scale deployment can commence.

Chapter 7

Conclusion and outlook

7.1 Conclusion

Due to high prevalence and under-diagnosis, non-specialist clinical settings such as primary care and other allied health services such as pharmacies will need to expand their capabilities in the early detection and assessment of chronic respiratory diseases. The incorporation of specialised decision-support can be offered as a complementary service to existing policies of integrated care for chronic-disease management. The work presented in this dissertation moves towards facilitating this through the following contributions:

- a clinical decision support system (CDSS) framework that includes an adapted incremental software development model and reasoning paradigm allowing the provision of a suite of decision support services for the early detection and assessment of chronic obstructive pulmonary disease (COPD);
- an extension and application of the Health Level Seven International (HL7) Virtual Medical Record (VMR), in order to represent COPD related concepts, and allowing for the representation of patient specific COPD data in an interoperable format;
- a software architecture model allowing COPD related decision support services to be integrated into existing health information systems (HISs) thus minimising the disruption to the healthcare providers' workflow;
- the formalisation of the medical guidelines for COPD screening, diagnosis, and assessment through a rule-based approach, their incorporation into the CDSS framework, and a benchmarking of the system against a respiratory clinician;
- decision support for spirometry quality assurance through 23 novel metrics

extracted from the analysis of the flow-time, volume-time and flow-volume curves of a spirometry test;

- 28 new rules or criteria applied to the 23 metrics to distinguish between acceptable and poor spirometry offering better performance than criteria found in international guidelines;
- spirometry quality assurance using knowledge automatically extracted from the ratings of multiple clinical lung function experts, inspired by the novel application of supervised learning methods and potentially generalisable to other bio-signals;
- evaluation of approaches of spirometry quality assurance against 3 lung function experts demonstrating near expert-level performance suggesting a credible potential to provide decision-support in spirometry testing in non-specialist settings;

In summary, we have demonstrated technically that decision support services for COPD can be incorporated directly into the workflow of health providers through existing health information systems using a framework proposed in this thesis. Several approaches have been presented in this thesis spanning from knowledge-driven methods to data-driven methods that can be embedded in the clinical decision support system framework. Evaluations of these methods show performance near the level of clinical experts, thus indicate a credible potential to assist non-specialist health providers in detecting early stage chronic obstructive pulmonary disease. Further investigation will be needed in the form of an initial pilot trial before plans for large scale deployment can commence.

7.2 Future research direction

Work beyond this thesis will be perused in a number of directions:

- expand CDSS capabilities into later-stages of COPD management with: (i) treatment recommendations based personalised to patient characteristics and within clinical guidelines (ii) exacerbation prognosis (COPD exacerbations are a major source of hospitalisation)
- issue recommendations for specific integrated care programs (e.g. pulmonary rehabilitation, promotion of physical activity) aiming at optimizing care. Allocation into a program will depend on two main factors: (i) health status and associated hospitalisation risk level; and, (ii) target health goals. A given patient can be simultaneously included in one or more programs.

- expand capabilities of spirometry quality assurance method to be able to classify failed manoeuvres into common reasons for rejection thus providing useful feedback to the patient and clinician administrating the test.

- Evaluate if in actual practise through a pilot trial the CDSS increases the quality and precision of COPD diagnosis and assessment of lung function in primary care. An initial protocol for regional implementation in Catalonia region of Spain has already been recently published [94].

- study the transferability of the CDSS framework into other non-communicable diseases (NCDs) prevalent in populations such as diabetes, and heart disease, and better address the co-morbid patient (patient with multiple conditions).

Bibliography

[1] F. Velickovski, L. Ceccaroni, J. Roca, F. Burgos, J. B. Galdiz, N. Marina, and M. Lluch-Ariet. "Clinical Decision Support Systems (CDSS) for preventive management of COPD patients". In: *Journal of Translational Medicine* 12.2 (2014), pp. 1–10. DOI: 10.1186/1479-5876-12-S2-S9.

[2] U. Melia, F. Burgos, M. Vallverdu, F. Velickovski, M. Lluch-Ariet, J. Roca, and P. Caminal. "Algorithm for Automatic Forced Spirometry Quality Assessment: Technological Developments". In: *PLoS ONE* 9.12 (Dec. 2015), pp. 1–14. DOI: 10.1371/journal.pone.0116238.

[3] F. Velickovski, L. Ceccaroni, R. Marti, F. Burgos, C. Gistau, X. Alsina-Restoy, and J. Roca. "Automated spirometry quality assurance: supervised learning from multiple experts". In: *Computer Methods and Programs in Biomedicine* (2016). Submitted.

[4] E. S. Berner. "Clinical decision support systems: state of the art". In: *AHRQ publication* 90069 (2009).

[5] H. S. Lin, W. W. Stead, et al. *Computational Technology for Effective Health Care:: Immediate Steps and Strategic Directions*. National Academies Press, 2009.

[6] J. R. Nebeker, J. M. Hoffman, C. R. Weir, C. L. Bennett, and J. F. Hurdle. "High rates of adverse drug events in a highly computerized hospital". In: *Archives of internal medicine* 165.10 (2005), pp. 1111–1116.

[7] I. Sim, P. Gorman, R. A. Greenes, R. B. Haynes, B. Kaplan, H. Lehmann, and P. C. Tang. "Clinical decision support systems for the practice of evidence-based medicine". In: *Journal of the American Medical Informatics Association* 8.6 (2001), pp. 527–534.

[8] S. Timmermans and A. Mauck. "The promises and pitfalls of evidence-based medicine". In: *Health Affairs* 24.1 (2005), pp. 18–28.

[9] H. J. Wong, M. W. Legnini, H. H. Whitmore, and R. S. Taylor. “The diffusion of decision support systems in healthcare: Are we there yet?/Practitioner application”. In: *Journal of Healthcare Management* 45.4 (2000), p. 240.

[10] V. A. Shaffer and L. S. Shaffer. “The use of recipe knowledge in medicine”. In: *International Journal of Technology* 2.7 (2007), pp. 73–80.

[11] A. Alwan et al. *Global status report on noncommunicable diseases 2010.* World Health Organization, 2011.

[12] B. R. Celli. “Management Guidelines for Chronic Obstructive Pulmonary Disease”. In: *COPD*. Springer, 2011, pp. 81–98.

[13] T. Ferkol and D. Schraufnagel. “The global burden of respiratory disease”. In: *Annals of the American Thoracic Society* 11.3 (2014), pp. 404–406.

[14] C. Lodewijckx, W. Sermeus, K. Vanhaecht, M. Panella, S. Deneckere, F. Leigheb, and M. Decramer. “Inhospital management of COPD exacerbations: a systematic review of the literature with regard to adherence to international guidelines”. In: *Journal of evaluation in clinical practice* 15.6 (2009), pp. 1101–1110.

[15] C. Schoen, R. Osborn, S. K. How, M. M. Doty, and J. Peugh. “In chronic condition: experiences of patients with complex health care needs, in eight countries, 2008”. In: *Health affairs* 28.1 (2009), w1–w16.

[16] D. Gomez-Cabrero, M. Lluch-Ariet, J. Tegnér, M. Cascante, F. Miralles, and J. Roca. “Synergy-COPD: a systems approach for understanding and managing chronic diseases”. In: *Journal of Translational Medicine* 12.2 (2014), pp. 1–10.

[17] *Treaty on European Union and the Treaty on the Functioning of the European Union*. European Union. 2012.

[18] *The Charter of Fundamental Rights of the European Union*. European Union. 2000.

[19] “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC”. In: *Official Journal of the European Union* (2016).

[20] J. K. Tan and S. B. Sheps. *Health decision support systems*. Jones & Bartlett Learning, 1998.

[21] H. Holst, K. Åström, A. Järund, J. Palmer, A. Heyden, F. Kahl, K. Trägil, E. Evander, G. Sparr, and L. Edenbrandt. “Automated interpretation of ventilation-perfusion lung scintigrams for the diagnosis of pulmonary embolism using artificial neural networks”. In: *European journal of nuclear medicine* 27.4 (2000), pp. 400–406.

[22] A. Berlin, M. Sorani, and I. Sim. “A taxonomic description of computer-based clinical decision support systems”. In: *Journal of biomedical informatics* 39.6 (2006), pp. 656–667.

[23] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach. “Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success”. In: *Bmj* 330.7494 (2005), p. 765.

[24] Institute of Medicine (US). Committee on Quality of Health Care in America. *Crossing the quality chasm: A new health system for the 21st century*. National Academy Press, 2001.

[25] R. A. Miller. “Medical diagnostic decision support systems–past, present, and future: a threaded bibliography and brief commentary.” In: *Journal of the American Medical Informatics Association* 1.1 (1994), p. 8.

[26] M. Peleg. *Guidelines and workflow models*. 2007, pp. 281–306.

[27] M. Peleg, S. Tu, J. Bury, P. Ciccarese, J. Fox, R. A. Greenes, R. Hall, P. D. Johnson, N. Jones, A. Kumar, et al. “Comparing computer-interpretable guideline models: a case-study approach”. In: *Journal of the American Medical Informatics Association* 10.1 (2003), pp. 52–68.

[28] F. Sonnenberg, C. Hagerty, et al. “Computer-interpretable clinical practice guidelines-Where are we and where are we going?” In: *IMIA Yearbook* (2006), pp. 145–158.

[29] *Health Level Seven web site*. HL7. 2016 [accessed August 28, 2016]. URL: `http://www.hl7.org/`.

[30] *Drools Expert Guide Version 5.3.0.CR1*. JBOSS Drools Team. 2011.

[31] R. S. Boyer and J. S. Moore. “A Fast String Searching Algorithm”. In: *Communications of the ACM* 20.10 (Oct. 1977), pp. 762–772.

[32] D. E. Knuth, J. H. J. Morris, and V. R. Pratt. “Fast Pattern Matching in Strings”. In: *SIAM Journal on Computing* 6.2 (1977), pp. 323–350.

[33] R. M. Karp and M. O. Rabin. “Efficient randomized pattern-matching algorithms”. In: *IBM Journal of Research and Development* 31.2 (Mar. 1987), pp. 249–260.

[34] C. L. Forgy. “Rete: A fast algorithm for the many pattern/many object pattern match problem”. In: *Artificial intelligence* 19.1 (1982), pp. 17–37.

[35] G. Hripcsak, P. Ludemann, T. A. Pryor, O. B. Wigertz, and P. D. Clayton. “Rationale for the Arden syntax”. In: *Computers and Biomedical Research* 27.4 (1994), pp. 291–324.

[36] J. C. Giarratano et al. *CLIPS User's guide*. 1993.

[37] E. Friedman-Hill. *JESS in Action*. Vol. 46. Manning Greenwich, CT, 2003.

[38] P. Browne. *JBoss Drools business rules*. Packt Publishing Ltd, 2009.

[39] T. Vetterlein, H. Mandl, and K.-P. Adlassnig. "Fuzzy Arden Syntax: A fuzzy programming language for medicine". In: *Artificial Intelligence in Medicine* 49.1 (2010), pp. 1–10.

[40] J. S. Gamer, W. R. Jarvis, T. G. Emori, T. C. Horan, and J. M. Hughes. "CDC definitions for nosocomial infections". In: *American Journal of Infection Control* 16.3 (1988), pp. 128–40.

[41] R. T. B. Batista-Navarro, D. A. Bandojo, M. J. K. Gatapia, R. N. C. Santos, A. B. Marcelo, L. C. R. Panganiban, and P. C. Naval. "ESP: An expert system for poisoning diagnosis and management". In: *Informatics for Health and Social Care* 35.2 (2010), pp. 53–63.

[42] B. Wu, X. Lu, and H. Duan. "An Automatic Knowledge Acquisition Mechanism for Independent Inference Engine Module of CDSS". In: *Bioinformatics and Biomedical Engineering, 2008. ICBBE 2008. The 2nd International Conference on*. IEEE. 2008, pp. 1293–1296.

[43] V. L. Patel, V. G. Allen, J. F. Arocha, and E. H. Shortliffe. "Representing clinical guidelines in GLIF". In: *Journal of the American Medical Informatics Association* 5.5 (1998), pp. 467–483.

[44] J. Fox, N. Johns, C. Lyons, A. Rahmanzadeh, R. Thomson, and P. Wilson. "PROforma: a general technology for clinical decision support systems". In: *Computer methods and programs in biomedicine* 54.1 (1997), pp. 59–67.

[45] S. W. Tu, J. R. Campbell, J. Glasgow, M. A. Nyman, R. McClure, J. McClay, C. Parker, K. M. Hrabak, D. Berg, T. Weida, J. G. Mansfield, M. A. Musen, and R. M. Abarbanel. "The SAGE Guideline Model: Achievements and Overview". In: *Journal of the American Medical Informatics Association* 14.5 (2007), pp. 589–598.

[46] R. A. Orchard. *FuzzyClips Version 6.04A User Guide*. National Research Council Canada. 1998.

[47] J. Saltz, J. Niland, P. Payne, H. Shah, and D. Stahl. *White Paper on Rules Engine Technologies Across caBIG Workspaces*. cancer Biomedical Informatics Grid (caBIG), 2007.

[48] C. C. Chen, K. Chen, C.-Y. Hsu, and Y.-C. J. Li. "Developing guideline-based decision support systems using protégé and jess". In: *computer methods and programs in biomedicine* 102.3 (2011), pp. 288–294.

[49] *MVEL (MVFLEX Expression Language) website*. MVEL. 2016 [accessed August 24, 2016]. URL: `https://github.com/mvel/mvel`.

[50] *Python website*. Python Software Foundation. 2016 [accessed August 24, 2016]. URL: `https://www.python.org`.

[51] *Groovy website*. Apache Software Foundation. 2016 [accessed August 24, 2016]. URL: `http://http://groovy-lang.org`.

[52] Lung Foundation. *COPD X Plan: Australian and New Zealand management guidelines for Chronic Obstructive Pulmonary Disease*. Lung Foundation, 2011.

[53] C. M. Bishop. "Probability Theory". In: *Pattern recognition amd Machine Learning*. Springer, 2006. Chap. 1, pp. 12–32.

[54] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[55] B. E. Himes, Y. Dai, I. S. Kohane, S. T. Weiss, and M. F. Ramoni. "Prediction of chronic obstructive pulmonary disease (COPD) in asthma patients using electronic medical records". In: *Journal of the American Medical Informatics Association* 16.3 (2009), pp. 371–379.

[56] E. Shortliffe. *Computer-based medical consultations: MYCIN*. Vol. 2. Elsevier, 2012.

[57] D. E. Heckerman and E. H. Shortliffe. "From certainty factors to belief networks". In: *Artificial Intelligence in Medicine* 4.1 (1992), pp. 35–52.

[58] G. Klir and B. Yuan. *Fuzzy sets and fuzzy logic*. Vol. 4. Prentice hall New Jersey, 1995.

[59] B. D. Ripley. *Pattern recognition and neural networks*. Cambridge university press, 2007.

[60] B. Larder, D. Wang, and A. Revell. "Application of artificial neural networks for decision support in medicine". In: *Artificial Neural Networks: Methods and Applications* (2009), pp. 119–132.

[61] V. Vapnik. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.

[62] L. Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[63] O. Er and F. Temurtas. “A study on chronic obstructive pulmonary disease diagnosis using multilayer neural networks”. In: *Journal of medical systems* 32.5 (2008), pp. 429–432.

[64] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor. “Assessing lung volumetric variation to detect and stage COPD”. In: *Biomedical Engineering (MECBME), 2011 1st Middle East Conference on*. IEEE. 2011, pp. 312–315.

[65] M. P. Hosseini, H. Soltanian-Zadeh, and S. Akhlaghpoor. “A novel method for identification of COPD in inspiratory and expiratory states of CT images”. In: *Biomedical Engineering (MECBME), 2011 1st Middle East Conference on*. IEEE. 2011, pp. 235–238.

[66] S. Jafari, H. Arabalibeik, and K. Agin. “Classification of normal and abnormal respiration patterns using flow volume curve and neural network”. In: *Health Informatics and Bioinformatics (HIBIT), 2010 5th International Symposium On*. IEEE. 2010, pp. 110–113.

[67] M. Kuilboer, M. Van Wijk, M. Mosseveld, E. Van Der Does, J. De Jongste, S. Overbeek, B. Ponsioen, J. van der Lei, et al. “Computed critiquing integrated into daily clinical practice affects physicians’ behavior - a randomized clinical trial with AsthmaCritic”. In: *Methods of information in medicine* 45.4 (2006), pp. 447–454.

[68] T. K. Liang, T. Tanaka, H. Nakamura, and A. Ishizaka. “A neural network based computer-aided diagnosis of emphysema using CT lung images”. In: *SICE, 2007 Annual Conference*. IEEE, 2007, pp. 703–709.

[69] M. S. Mohktar, J. Basilakis, S. J. Redmond, and N. H. Lovell. “A guideline-based decision support system for generating referral recommendations from routinely recorded home telehealth measurement data”. In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 6166–6169.

[70] R. Rosso, G. Munaro, O. Salvetti, S. Colantonio, and F. Ciancitto. “CHRONIOUS: an open, ubiquitous and adaptive chronic disease management platform for chronic obstructive pulmonary disease (COPD), chronic kidney disease (CKD) and renal insufficiency”. In: *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*. IEEE, 2010, pp. 6850–6853.

[71] B. Song, K.-H. Wolf, M. Gietzelt, O. A. Scharaa, U. Tegtbur, R. Haux, and M. Marschollek. “Decision support for teletraining of COPD patients”. In: *Pervasive Computing Technologies for Healthcare, 2009. PervasiveHealth 2009. 3rd International Conference on*. IEEE. 2009, pp. 1–6.

[72] D. Sahin, E. D. Übeyli, G. Ilbay, M. Sahin, and A. B. Yasar. "Diagnosis of airway obstruction or restrictive spirometric patterns by multiclass support vector machines". In: *Journal of medical systems* 34.5 (2010), pp. 967–973.

[73] U. Uncu, E. Koklukaya, A. Gencsoy, and O. Annadurdiyew. "A fuzzy rule-base model for classification of spirometric FVC graphs in chronical obstructive pulmonary diseases". In: *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*. Vol. 4. IEEE, 2001, pp. 3866–3869.

[74] R. A. Pauwels, A. S. Buist, P. M. Calverley, C. R. Jenkins, and S. S. Hurd. "Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease". In: *American journal of respiratory and critical care medicine* (2012).

[75] *Chronic obstructive pulmonary disease in over 16s: diagnosis and management*. National Institute for Health and Care Excellence. 2010 [accessed August 24, 2016]. URL: `https://www.nice.org.uk/guidance/cg101`.

[76] *IBM Watson*. IBM. 2016 [accessed August 24, 2016]. URL: `http://www.ibm.com/watson`.

[77] E. S. Chen, G. Hripcsak, H. Xu, M. Markatou, and C. Friedman. "Automated acquisition of disease–drug knowledge from biomedical and clinical documents: an initial study". In: *Journal of the American Medical Informatics Association* 15.1 (2008), pp. 87–98.

[78] M. Ruzicka and V. Svátek. "Mark-up based analysis of narrative guidelines with the Stepper tool." In: *Studies in health technology and informatics* 101 (2003), pp. 132–136.

[79] *openEHR*. The openEHR Foundation. 2016 [accessed August 24, 2016]. URL: `http://www.openehr.org`.

[80] E. Duran-Tauleria, F. Forastiere, C. Janson, J. Roca, G. Viegi, G. Weinmyer, J. Anto, and the IMCA Group. *Report on the respiratory health of the elderly population living in Catalunya.* Centre for Research in Environmental Epidemiology (CREAL). Barcelona, 2013.

[81] K. Holton, S. Hill, M. Kearney, R. Winter, and A. Moger. "Quality assured diagnostic spirometry (qads)-performance and competence". In: *Am J Respir Crit Care Med* 187 (2013), p. 2836.

[82] C. R. Jenkins, P. W. Jones, P. M. Calverley, B. Celli, J. A. Anderson, G. T. Ferguson, J. C. Yates, L. R. Willits, and J. Vestbo. "Efficacy of salmeterol/fluticasone propionate by GOLD stage of chronic obstructive pulmonary disease: analysis from the randomised, placebo-controlled TORCH study". In: *Respiratory Research* 10.1 (2009), pp. 1–9.

[83] M. Decramer, B. Celli, S. Kesten, T. Lystig, S. Mehra, D. P. Tashkin, E. Schiavi, J. C. Casas, E. Rhodius, R. Gene, C. B. Saenz, E. Giugno, C. Di Bartolo, M. C. De Salvo, E. Abbate, et al. "Effect of tiotropium on outcomes in patients with moderate chronic obstructive pulmonary disease (UPLIFT): a prespecified subgroup analysis of a randomised controlled trial". In: *Lancet* 374.9696 (Oct. 2009), pp. 1171–1178.

[84] D. F. Sittig, A. Wright, J. A. Osheroff, B. Middleton, J. M. Teich, J. S. Ash, E. Campbell, and D. W. Bates. "Grand challenges in clinical decision support". In: *Journal of biomedical informatics* 41.2 (2008), pp. 387–392.

[85] J. Fox, D. Glasspool, V. Patkar, M. Austin, L. Black, M. South, D. Robertson, and C. Vincent. "Delivering clinical decision support services: there is nothing as practical as a good theory". In: *Journal of biomedical informatics* 43.5 (2010), pp. 831–843.

[86] V. Gomoi, D. Dragu, and V. Stoicu-Tivadar. "Virtual Medical Record Implementation for Enhancing Clinical Decision Support". In: *Studies in health technology and informatics* 180 (2012), p. 118.

[87] I. Cano, A. Alonso, C. Hernandez, F. Burgos, A. Barberan-Garcia, J. Roldan, and J. Roca. "An adaptive case management system to support integrated care services: Lessons learned from the NEXES project". In: *Journal of biomedical informatics* 55 (2015), pp. 11–22.

[88] *Arezzo Pathways*. Infermed. 2016 [accessed June 1, 2016]. URL: `https://www.elsevier.com/solutions/infermed/clinical-decision-support`.

[89] M. Bali. *Drools JBoss Rules 5.0: Developer's Guide*. Birmingham, United Kingdom: Packt Publishing, 2009.

[90] F. Burgos, C. Disdier, E. L. de Santamaria, B. Galdiz, N. Roger, M. L. Rivera, R. Hervàs, E. Durán-Tauleria, J. Garcia-Aymerich, and J. Roca. "Telemedicine enhances quality of forced spirometry in primary care". In: *European Respiratory Journal* 39.6 (2012), pp. 1313–1318.

[91] J. A. Walters, E. H. Walters, M. Nelson, A. Robinson, J. Scott, P. Turner, and R. Wood-Baker. “Factors associated with misdiagnosis of COPD in primary care”. In: *Prim Care Respir J* 20.4 (2011), pp. 396–402.

[92] M. R. Miller, J. Hankinson, V. Brusasco, F. Burgos, R. Casaburi, A. Coates, R. Crapo, P. Enright, C. P. van der Grinten, P. Gustafsson, R. Jensen, D. C. Johnson, N. MacIntyre, R. McKay, D. Navajas, O. F. Pedersen, R. Pellegrino, G. Viegi, and J. Wanger. “Standardisation of spirometry”. In: *European respiratory journal* 26.2 (Aug. 2005), pp. 319–338.

[93] C. Müller-Brandes, U. Krämer, M. Gappa, G. Seitner-Sorge, A. Hüls, A. von Berg, B. Hoffmann, A. Schuster, S. Illi, M. Wisbauer, et al. “LUNOKID: can numerical American Thoracic Society/European Respiratory Society quality criteria replace visual inspection of spirometry?” In: *European Respiratory Journal* 43.5 (2014), pp. 1347–1356.

[94] C. Vargas, F. Burgos, I. Cano, I. Blanco, P. Caminal, J. Escarrabill, C. Gallego, M. A. Llauger, F. Miralles, O. Solans, M. Vallverdu, F. Velickovski, and J. Roca. “Protocol for regional implementation of collaborative lung function testing”. In: *NPJ primary care respiratory medicine* 26 (2016), p. 16024.